Part I

Invited Papers

Gödel's program for new axioms: Why, where, how and what?

Solomon Feferman*

Department of Mathematics
Stanford University
Stanford, CA 94305 USA
email: sf@csli.stanford.edu

Summary. From 1931 until late in his life (at least 1970) Gödel called for the pursuit of new axioms for mathematics to settle both undecided number-theoretical propositions (of the form obtained in his incompleteness results) and undecided set-theoretical propositions (in particular CH). As to the nature of these, Gödel made a variety of suggestions, but most frequently he emphasized the route of introducing ever higher axioms of infinity. In particular, he speculated (in his 1946 Princeton remarks) that there might be a uniform (though non-decidable) rationale for the choice of the latter. Despite the intense exploration of the "higher infinite" in the last 30-odd years, no single rationale of that character has emerged. Moreover, CH still remains undecided by such axioms, though they have been demonstrated to have many other interesting set-theoretical consequences.

In this paper, I present a new very general notion of the "unfolding" closure of schematically axiomatized formal systems S which provides a uniform systematic means of expanding in an essential way both the language and axioms (and hence theorems) of such systems S. Reporting joint work with T. Strahm, a characterization is given in more familiar terms in the case that S is a basic system of non-finitist arithmetic. When reflective closure is applied to suitable systems of set theory, one is able to derive large cardinal axioms as theorems. It is an open question how these may be characterized in terms of current notions in that subject.

1. Why new axioms?

Gödel's published statements over the years (from 1931 to 1972) pointing to the need for new axioms to settle both undecided number-theoretic and set-theoretic propositions are rather well known. They are most easily cited by reference to the first two volumes of the edition of his *Collected Works.*[1] A number of less familiar statements of a similar character from his unpublished essays and lectures are now available in the third volume of that edition.[2]

* Invited opening lecture, Gödel '96 conference, Brno, 25-29 August 1996. This paper was prepared while the author was a fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, CA, whose facilities and support are greatly appreciated.

[1] Cf. in Gödel [1986] the items dated: *1931*(p.181, ftn.48a), *1934*(p.367), *1936*(p.397), and in Gödel [1990] those dated: *1940*(p.97, ftn.20[added 1965]), *1946*(p.151), *1947*(pp.181-183), *1964*(pp.260-261 and 268-270), and *1972a*, Note 2 (pp.305-306).

[2] Cf. in Gödel [1995] the items dated: **1931?*(p.35), **1993o* (p.48), **1951*(pp.306-307), **1961/?*(p.385) and **1970a,b,c*(pp.420-425).

Given the ready accessibility of these sources, there is n o need for extensive quotation, though several representative passages are singled out below for special attention.

With one possible exception (to be noted in the next section), the single constant that recurs throughout these statements is that the new axioms to be considered are in all cases of a set-theoretic nature. More specifically, to begin with, axioms of higher types, extended into the transfinite, are said to be needed even to settle undecided arithmetical propositions.[3] The first and most succinct statement of this is to be found in the singular footnote 48a of the 1931 incompleteness paper, in which Gödel states that "...the true reason for the incompleteness inherent in all formal systems of mathematics is that the formation of ever higher types can be continued into the transfinite...[since] the undecidable propositions constructed here become decidable whenever appropriate higher types are added". In an unpublished lecture from that same period Gödel says that analysis is higher in this sense than number theory and set theory is higher than analysis: "...there are number-theoretic problems that cannot be solved with number-theoretic, but only with analytic or, respectively, set-theoretic methods" (Gödel [1995], p.35). A couple of years later, in his (unpublished) 1933 lecture at a meeting of the Mathematical Association of America in Cambridge, Massachusetts, Gödel said that for the systems S to which his incompleteness theorems apply "...exactly the next higher type not contained in S is necessary to prove this arithmetical proposition...[and moreover] there are arithmetic propositions which cannot be proved even by analysis but only by methods involving extremely large infinite cardinals and similar things" (Gödel [1995], p.48). This assertion of the necessity of axioms of higher type — a.k.a. axioms of infinity in higher set theory — to settle undecided arithmetic (Π^0_1) propositions, is repeated all the way to the final of the references cited here in footnotes 1 and 2 (namely to 1972).

It is only with his famous 1947 article on Cantor's continuum problem that Gödel also pointed to the need for new set-theoretic axioms to settle specifically *set-theoretic* problems, in particular that of the Continuum Hypothesis CH. Of course at that time one only knew through his own work the (relative) consistency of AC and CH with ZF, though Gödel conjectured the falsity of CH and hence its independence from ZFC. Moreover, it was the question of determining the truth value of CH that was to preoccupy him almost exclusively among all set-theoretic problems — except for those which might be ancillary to its solution — for the rest of his life. And rightly so: the continuum problem — to locate $2^{\aleph_0}$ in the scale of the alephs whose existence is forced on us by the well-ordering theorem — is the very first chal-

[3] The kind of proposition in question is sometimes referred to by Gödel as being of "Goldbach type" i.e. in Π^0_1 form, and sometimes as one concerning solutions of Diophantine equations, of the form $(P)D = 0$, where P is a quantifier expression with variables ranging over the natural numbers; cf. more specifically, the lecture notes **193?* in Gödel [1995].

lenging problem of Cantorian set theory, and settling it might be considered to bolster its conceptual coherence. In his 1947 paper, for the decision of CH by new axioms, Gödel mentioned first of all, axioms of infinity:

> The simplest of these ... assert the existence of inaccessible numbers (and of numbers inaccessible in the stronger sense) $> \aleph_0$. The latter axiom, roughly speaking, means nothing else but that the totality of sets obtainable by exclusive use of the processes of formation of sets expressed in the other axioms forms again a set (and, therefore, a new basis for a further application of these processes). Other axioms of infinity have been formulated by P. Mahlo. [Very little is known about this section of set theory; but at any rate][4] these axioms show clearly, not only that the axiomatic system of set theory as known today is incomplete, but also that it can be supplemented without arbitrariness by new axioms which are only the natural continuation of those set up so far. (Gödel [1990], p.182)

However, Gödel goes on to say, quite presciently, that "[a]s for the continuum problem, there is little hope of solving it by means of those axioms of infinity which can be set up on the basis of principles known today...", because his proof of the consistency of CH via the constructible sets model goes through without change when such statements are adjoined as new axioms (indeed there is no hope in this direction if one expects to prove CH false):

> But probably [in the face of this] there exist other [axioms] based on hitherto unknown principles ... which a more profound understanding of the concepts underlying logic and mathematics would enable us to recognize as implied by these concepts. (*ibid.*)

Possible candidates for these were forthcoming through the work of Scott [1961] in which it was shown that the existence of measurable cardinals (MC) implies the negation of the axiom of constructibility, and the later work of Hanf [1964] and of Keisler and Tarski [1964] which showed that measurable cardinals and even weakly compact cardinals must be very much larger than anything obtained by closure conditions on cardinals of the sort leading to hierarchies of inaccessibles. But as we now know through the extensive subsequent work on large cardinals as well as other strong set-theoretic principles such as forms of determinacy, none of those considered at all plausible to date settles CH one way or the other (cf. Martin [1976], Kanamori [1994]). Gödel himself offered only one candidate besides these, in his unpublished 1970 notes containing his "square axioms" concerning so-called scales of functions on the $\aleph_n$'s. The first of these notes (**1970a* in Gödel [1995]) purports to prove that the cardinality of the continuum is $\aleph_2$ while the second (**1970b*, op.cit.) purports to prove that it is $\aleph_1$. However, there are essential gaps in

[4] The section enclosed in brackets was deleted from the 1964 reprinting of the 1947 article (cf. Gödel [1990], p. 260).

both proofs and in any case the axioms considered are far from evident (cf. the introductory note by R.M. Solovay to **1970a,b,c* in Gödel [1995], pp. 405-420).

Gödel's final fall-back position in his 1947 article is to look for axioms which are "so abundant in their verifiable consequences...that quite irrespective of their intrinsic necessity they would have to be assumed in the same sense as any well-established physical theory" (Gödel [1990], p.183). It would take us too far afield to look into the question whether there are any plausible candidates for these. Moreover, there is no space here to consider the arguments given by others in pursuit of the program for new axioms; especially worthy of attention are Maddy [1988, 1988a], Kanamori [1994] and Jensen [1995] among others.

My concern in the rest of this paper is to concentrate on the consideration of axioms which are supposed to be "exactly as evident" as those already accepted. On the face of it this excludes, among others, axioms for "very large" cardinals (compact, measurable, etc.), axioms of determinacy, axioms of randomness, and axioms whose only grounds for accepting them lies in their "fruitfulness" or in their simply having properties analogous to those of $\aleph_0$. Even with this restriction, as we shall see, there is much room for reconsideration of Gödel's program.

2. Where should one look for new axioms?

While the passage to higher types in successive stages, in one form or another, is sufficient to overcome incompleteness with respect to number-theoretic propositions because of the increase in consistency strength at each such stage, it by no means follows that this is the *only* way of adding new axioms in a principled way for that purpose. Indeed, here a quotation from Gödel's remarks in 1946 before the Princeton Bicentennial Conference is very apropos:

> Let us consider, e.g., the concept of demonstrability. It is well known that, in whichever way you make it precise by means of a formalism, the contemplation of this very formalism gives rise to new axioms which are exactly as evident and justified as those with which you started, and this process of extension can be iterated into the transfinite. So there cannot exist any formalism which would embrace all these steps; but this does not exclude that all these steps (or at least all of them which give something new for the domain of propositions in which you are interested) could be described and collected together in some non-constructive way. (Gödel [1990], p.151)

It is this passage that I had in mind above as the one possible exception to Gödel's reiterated call for new set-theoretic axioms to settle undecided number-theoretic propositions. It is true that he goes on immediately to say that "[i]n set theory, e.g., the successive extensions can most conveniently be

represented by stronger and stronger axioms of infinity". But note that here he is referring to set theory as an *example* of a formalism to which the general idea of expansion by "new axioms exactly as evident and justified as those with which you started" may be applied as a special case. That idea, in the case of formal systems S in the language of arithmetic comes down instead to one form or another of (proof-theoretic) reflection principle, that is a formal scheme to the effect that whatever is provable in S is correct. In its weakest form (assuming the syntax of S effectively and explicitly given), this is the collection of statements

$$(\mathrm{Rfn_S}) \qquad Prov_{\mathrm{S}}(\#(A)) \rightarrow A$$

for A a closed formula in the language of S, called the *local reflection principle*.[5] This is readily generalized to arbitrary formulas A uniformly in the free variables of A as parameters, in which case it is called the *uniform reflection principle* $\mathrm{RFN_S}$. The axioms $\mathrm{Rfn_S}$, and more generally, $\mathrm{RFN_S}$ may indeed be considered "exactly as evident and justified" as those with which one started. Moreover, as shown by Turing [1939], extension by such axioms may be effectively iterated into the transfinite, in the sense that one can associate with each constructive ordinal notation $a \in O$ a formal system S_a such that the step from any one such system to its successor is described by adjunction of the reflection principle in question, and where all previous adjunctions are simply accumulated at limit s by the formation of their union. These kinds of systematic extensions of a given formal system were called *ordinal logics* by Turing; when I took them up later in 1962, I rechristened them *(transfinite) recursive progressions of axiomatic theories* (cf. Feferman [1962, 1988]). While Turing obtained a completeness result for Π^0_1 statements via the transfinite iteration in this sense of the local reflection principle, and I obtained one for all true arithmetic statements via the iteration of the uniform reflection principle, both completeness results were problematic because they depended crucially on the judicious choice of notations in O, the selection of which was no more "evident and justified" in advance than the statements to be proved.

What was missing in this first attempt to spell out the general idea expressed by Gödel in the above quotation was an explanation of which ordinals — in the constructive sense — ought to be accepted in the iteration procedure. The first modification made to that end (Kreisel [1958], Feferman [1964]) was to restrict to *autonomous* progressions of theories, where one advances to a notation $a \in O$ only if it has been proved in a system S_b, for some b which precedes a, that the ordering specifying a is indeed a well-ordering. It was with this kind of procedure in mind that Kreisel called in his paper [1970] for the study of *all principles of proof and ordinals which are implicit in given concepts.* However, one may question whether it is appropriate at

[5] Note that the consistency statement for S is an immediate consequence of the local reflection principle for S.

all to speak of the concept of ordinal, in whatever way restricted, as being implicit in the concepts of, say, arithmetic. I thus began to pursue a modification of that program in Feferman [1979], where I pr oposed a characterization of that part of mathematical thought which is implicit in our conception of the natural numbers, without any prima-facie use of the notions of ordinal or well-ordering. This turned out to yield a system proof-theoretically equivalent to that proposed as a characterization of *predicativity* in Feferman [1964] and Schütte [1965]. Then in my paper [1991], I proposed more generally, a notion of *reflective closure* of arbitrary schematically axiomatized theories, which gave the same result (proof-theoretically) as the preceding when applied to Peano Arithmetic as initial system. That made use of a partial self-applicative notion of truth, treated axiomatically. The purpose of the present article is to report a new general notion of reflective closure of a quite different form, which I believe is more convincing as an explanation of *everything that one ought to accept if one has accepted given concepts and principles.* In order not to confuse it with the earlier proposal, I shall call this notion that of the *unfolding* of any given schematically formalized system. This will be illustrated here in the case of non-finitist arithmetic as well as the case of set theory. Exact characterizations in more familiar terms have been obtained for the case of non-finitist arithmetic in collaboration with Thomas Strahm; these will be described in Section 4 below. However, there is no space here to give any proofs.

3. How is the unfolding of a system defined?

As we shall see, it is of the essence of the notion of unfolding that we are dealing with schematically presented formal systems. In the usual conception, *formal schemata* for axioms and rules of inference employ *free predicate variables* $P, Q, \ldots$ of various numbers of arguments $n \geq 0$. An appropriate substitution for $P(x_1, \ldots, x_n)$ in such a scheme is a formula $A(x_1, \ldots x_n \ldots)$ which may have additional free variables. (Thus if P is 0-ary, any formula may be substituted for it.) Familiar examples of *axiom schemata* in the propositional and predicate calculi are

$$\neg P \to (P \to Q) \qquad \text{and} \qquad (\forall x)P(x) \to P(t)\ .$$

Further, in non-finitist arithmetic, we have the *Induction Axiom Scheme*

$$\text{(IA)} \qquad P(0) \wedge (\forall x)[P(x) \to P(x')\,] \to (\forall x)P(x)\ ,$$

while in set theory we have the *Separation* and *Replacement Schemes*

$$\text{(Sep)} \qquad (\exists b)(\forall x)[x \in b \leftrightarrow x \in a \wedge P(x)\,],\ \text{and}$$

$$\text{(Repl)} \qquad (\forall x \in a)(\exists! y)P(x,y) \to (\exists b)(\forall y)[y \in b \leftrightarrow (\exists x \in a)P(x,y)\,]\ .$$

Familiar examples of *schematic rules of inference* are, first of all, in the propositional and predicate calculi,

$$P, P \to Q \Rightarrow Q \quad \text{and} \quad [P \to Q(x)] \Rightarrow [P \to (\forall x)Q(x)] \text{ (for } x \text{ not free in } P),$$

while the scheme for the *Induction Rule* in finitist arithmetic is given by

$$\text{(IR)} \qquad P(0), P(x) \to P(x') \Rightarrow P(x) \ .$$

It is less usual to think of schemata for axioms and rules given by *free function variables* $f, g, \ldots$ But actually, it is more natural to formulate the Replacement Axiom Scheme in functional form as follows:

$$\text{(Repl)}' \quad (\forall x \in a)(\exists y)[f(x) = y] \to (\exists b)(\forall y)[y \in b \leftrightarrow (\exists x \in a) f(x) = y] \ .$$

Note that here, and for added compelling reasons below, our function variables are treated as ranging over *partial functions.*

The informal philosophy behind the use of schemata here is their *open-endedness.* That is, they are not conceived of as applying to a specific language whose stock of basic symbols is fixed in advance, but rather as applicable to *any* language which one comes to recognize as embodying meaningful basic notions. Put in other terms, *implicit in the acceptance of given schemata is the acceptance of any meaningful substitution instances.* But *which* these instances are need not be determined in advance. Thus, for example, if one accepts the axioms and rules of inference of the classical propositional calculus given in schematic form, one will accept all substitution instances of these schemata in any language which one comes to employ. The same holds for the schemata of the sort given above for arithmetic and set theory. In this spirit, we do not conceive of the function, resp. predicate variables as having a fixed intended range and it is for this reason that th ey are treated as *free* variables. Of course, if one takes it to be meaningful to talk about the totality of partial functions, resp. predicates, of a given domain of objects, then it would be reasonable to bind them too by quantification. In the examples of unfolding given here, it is only in set theory that the issue of whether and to what extent to allow quantification over function variables is unsettled.

Now our question is this: *given a schematic system* S, *which operations and predicates — and which principles concerning them — ought to be accepted if one has accepted* S? The answer for operations is straightforward: *any operation from and to individuals is accepted in the unfolding of* S *which is determined (in successive steps) explicitly or implicitly from the basic operations of* S. Moreover, the *principles* which are added concerning these operations are just those which are derived from the way they are introduced. Ordinarily, we would confine ourselves to the *total operations* obtained in this way, i.e. those which have been proved to be defined for all values of their arguments, but it should not be excluded that their introduction might depend in an essential way on prior *partial operations*, e.g. those introduced by recursive definitions of a general form.

We reformulate the question concerning predicates in operational terms as well, i.e.: *which operations on and to predicates — and which principles concerning them — ought to be accepted if one has accepted* S? For this, it is necessary to tell at the outset *which logical operations on predicates are taken for granted in* S. For example, in the case of non-finitist classical arithmetic these would be (say) the operations $\neg, \wedge$ and $\forall$, while in the case of finitist arithmetic, we would use just $\neg$ and $\wedge$. It proves simplest to treat predicates as propositional functions; thus $\neg$ and $\wedge$ are operations on propositions, while $\forall$ is an operation on functions from individuals to propositions. Now we can add to the operations from individuals to individuals in the unfolding of S also *all those operations from individuals and/or propositions to propositions which are determined explicitly or implicitly (in successive steps) fro m the basic logical operations of* S. Once more, the principles concerning these operations which are included in the expansive closure of S are just those which are derived from the way they are introduced. Finally, *we include in the expansive closure of* S *all the predicates which are generated from the basic predicates of* S *by these operations*; the principles which are taken concerning them are just those that fall out from the principles for the operations just indicated.

This notion of unfolding of a system is spelled out in completely precise terms in the next section for the case of non-finitist arithmetic. But the following two points ought to be noted concerning the general conception described here. First of all, one should not think of the unfolding of a system S as delimiting the range of applicability of the schemata embodied in S. For example, the principle of induction is applicable in every context in which the basic structure of the natural numbers is recognized to be present, even if that context involves concepts and principles not implicit in our basic system for that structure. In particular, it is applicable to impredicative reasoning with sets, even though (as will be shown in the next section) the unfolding closure of arithmetic is limited to predicative reasoning. Secondly, we may expect the language and theorems of the unfolding of (an effectively given system) S to be effectively enumerable, but we should not expect to be a ble to decide which operations introduced by implicit (e.g. recursive fixed-point) definitions are well defined for all arguments, even though it may be just those with which we wish to be concerned in the end. This echoes Gödel's picture of the process of obtaining new axioms which are "just as evident and justified" as those with which we started (quoted in Section 2 above), for which we cannot say in advance exactly what those will be, though we can describe fully the means by which they are to be obtained.

4. The expansive closure of non-finitist arithmetic: what's obtained

Here the starting schematic system NFA (Non-Finitist Arithmetic) has language given by the constant 0, individual variables $x, y, z, \ldots$, the operations

Sc and Pd for successor and predecessor, a free unary predicate variable P and the logical operations $\neg, \wedge$ and $\forall$.

Assuming classical logic, $\wedge, \rightarrow$ and $\exists$ are defined as usual.[6] We write t' for $Sc(t)$ in the following. The axioms of NFA are:

Ax 1. $x' \neq 0$
Ax 2. $Pd(x') = x$
Ax 3. $P(0) \wedge (\forall x)[P(x) \rightarrow P(x')] \rightarrow (\forall x)P(x)$.

Ax 3 is of course our scheme (IA) of induction. Before defining the full unfolding $\mathcal{U}(\text{NFA})$ of this system, it is helpful to explain a subsystem $\mathcal{U}_0(\text{NFA})$ which might be called the *operational unfolding* of NFA, i.e. where we do not consider which predicates are to be obtained. Basically, the idea is to introduce new operations via a form of generalized recursion theory (g.r.t.) considered axiomatically. The specific g.r.t. referred to is that developed in Moschovakis [1989] and in a different-appearing but equivalent form in Feferman [1991a] and [1996]; both feature *explicit definition* (ED) and *least fixed point recursion* (LFP) and are applicable to arbitrary structures with given functions or functionals of type level ≤ 2 over a given basic domain (or domains). The basic structure to consider in the case of arithmetic is $\langle \mathrm{N}, Sc, Pd, 0\rangle$, where N is the set of natural numbers. To treat this axiomatically, we simply have to enlarge our language to include the terms for the (in general) partial functions and functionals generated by closure under the schemata for this g.r.t., and add their defining equations as axioms. So we have terms of three types to consider: *individual terms, partial function terms* and *partial functional terms.* The types of these are described as follows, where, to allow for later extension to the case of $\mathcal{U}(\text{NFA})$, we posit a set Typ_0 of types of level 0; here we will only need it to contain the type ι of individuals, but below it will be expanded to include the type ι of propositions:

Typ 1. $\iota \in Typ_0$, where ι is the type of individuals. In the following κ, ν range over Typ_0 and $\bar{\iota}$, resp. $\bar{\kappa}$ range over types of finite sequences of individuals, resp. of objects of Typ_0.

Typ 2. τ, σ range over the types of partial functions of the form $\bar{\iota} \widetilde{\rightarrow} \nu$, and $\bar{\tau}$ ranges over the types of finite sequences of such.

Typ 3. $(\bar{\tau}, \bar{\kappa} \widetilde{\rightarrow} \nu)$ is used as types of partial functionals.

Note that objects of partial function type take only individuals as arguments; this is to insure that propositional functions, to be considered below, are just such functions. On the other hand, we may have partial functionals of type described under Typ 3 in which the sequence $\bar{\tau}$ is empty, and these reduce to partial functions of any objects of basic type in Typ_0.

[6] All our notions and results carry over directly to NFA treated in intuitionistic logic; the only difference in that case is that we take the full list of logical operations, $\neg, \wedge, \vee, \rightarrow, \forall$, and $\exists$ as basic.

The terms $r, s, t, u, \ldots$ of the various types under Typ 1 – Typ 3 are generated as follows, where we use $r : \rho$ to indicate that the term r is of type ρ.

Tm 1. For each $\kappa \in Typ_0$, we have infinitely many variables $x, y, z, \ldots$ of type κ.
Tm 2. $0 : \iota$.
Tm 3. $Sc(t) : \iota$ and $Pd(t) : \iota$ for $t : \iota$.
Tm 4. For each τ we have infinitely many partial function variables $f, g, h, \ldots$ of type τ.
Tm 5. $Cond(s, t, u, v) : (\bar{\tau}, \bar{\kappa}, \iota, \iota \widetilde{\to} \nu)$ for $s, t : (\bar{\tau}, \bar{\kappa} \widetilde{\to} \nu)$ and $u, v : \iota$.
Tm 6. $s(\bar{t}, \bar{u}) : \nu$ for $s : (\bar{\tau}, \bar{\kappa} \widetilde{\to} \nu)$, $\bar{t} : \bar{\tau}$, $\bar{u} : \bar{\kappa}$.
Tm 7. $\lambda \bar{f}, \bar{x}.t : (\bar{\tau}, \bar{\kappa} \widetilde{\to} \nu)$ for $\bar{f} : \bar{\tau}, \bar{x} : \bar{\kappa}, t : \nu$.
Tm 8. LFP $(\lambda f, \bar{x}.t) : (\bar{\iota} \widetilde{\to} \nu)$ for $f : \bar{\iota} \widetilde{\to} \nu, \bar{x} : \bar{\iota}$, $t : \nu$.

We now specialize this system of types and terms to just what is needed for $\mathcal{U}_0$(NFA), by taking $Typ_0 = \{\iota\}$. The formulas $A, B, C, \ldots$ of $\mathcal{U}_0$(NFA) are then generated as follows:

Fm 1. The atomic formulas are $s = t, s\downarrow$, and $P(s)$ for $s, t : \iota$.
Fm 2. If A, B are formulas then so also are $\neg A, A \wedge B$, and $\forall x A$.

As indicated above, formulas $A \vee B, A \to B$, and $\exists x A$ are defined as usual in classical logic. We write $s \simeq t$ for $[s \downarrow \vee t \downarrow \to s = t]$. Below we write $t\ [\bar{f}, \bar{x}]$, resp. $A\ [\bar{f}, \bar{x}]$ for a term, resp. formula, with designated sequences of free variables $\bar{f}, \bar{x}$; it is not excluded that t, resp. A may contain other free variables when using this notation. Since we are dealing with possibly undefined (individual) terms t, the underlying system of logic to be used is the *logic of partial terms* (LPT) introduced by Beeson [1985], pp. 97-99, where $t\downarrow$ is read as: t is defined. Briefly, the changes to be made from usual predicate logic are, first, that the axiom for $\forall$-instantiation is modified to

$$\forall x A(x) \wedge t\downarrow \to A(t)\ .$$

In addition, it is assumed that $\forall x(x\downarrow)$, i.e. only compound terms may fail to be defined (or put otherwise, non-existent individuals are not countenanced in LPT). It is further assumed that if a compound term is defined then all its subterms are defined ("strictness" axioms). Finally, one assumes that if $s = t$ holds then both s, t are defined and if $P(s)$ holds then s is defined. Note that $(s\downarrow) \leftrightarrow \exists x(s = x)$, so definedness need not be taken as a basic symbol.

The axioms of $\mathcal{U}_0$(NFA) follow the obvious intended meaning of the new compound terms introduced by the clauses Tm 5-8:

Ax 4. $(Cond(s, t, u, u))(\bar{f}, \bar{x}) \simeq s(\bar{f}, \bar{x}) \wedge [u \neq v \to (Cond(s, t, u, v))(\bar{f}, \bar{x}) \simeq t(\bar{f}, \bar{x})]$.
Ax 5. $(\lambda \bar{f}, \bar{x}.s[\bar{f}, \bar{x}])(\bar{t}, \bar{u}) \simeq s[\bar{t}, \bar{u}]$.
Ax 6. For $\varphi = \mathrm{LFP}(\lambda f, \bar{x}.t[f, \bar{x}])$, we have:
(i) $\varphi(\bar{x}) \simeq t[\varphi, \bar{x}]$

(ii) $\forall \bar{x}\{f(\bar{x}) \simeq t[f,\bar{x}]\} \to \forall \bar{x}\{\varphi(\bar{x}) \downarrow\!\to \varphi(\bar{x}) = f(\bar{x})\}$.

Finally, the predicate substitution rule for $\mathcal{U}_0$(NFA) is:

(Subst) $$A[P] \Rightarrow A[B/P]$$

where in the conclusion of this rule, B is any formula with a designated free variable x, $B[x]$, and we substitute $B[t]$ for each occurrence of $P(t)$ in A. This completes the description of $\mathcal{U}_0$(NFA).

In the following we shall write

$$\{\text{if } y = 0 \text{ then } s[\bar{f},\bar{x}] \text{ else } t[\bar{f},\bar{x}]\} \text{ for } (Cond(\lambda \bar{f},\bar{x}.s, \lambda \bar{f},\bar{x}.t, y, 0))(\bar{f},\bar{x}),$$

in order to meet the strictness axioms of LPT; this piece of notation has the property that the compound term is defined when $y = 0$ if s is defined, even if t is not defined, while it is defined when $y \neq 0$ and t is defined if s is not defined.

We shall use capital letters F for closed terms of function type such that NFA proves $\forall \bar{x}(F(\bar{x})\downarrow)$, i.e. for which F is proved to be total. Suppose given such terms G, H of arguments $(\bar{x})$ and $(\bar{x}, y, z)$, resp. Then we can obtain an F with

$$\begin{aligned} F(\bar{x}, 0) &= G(\bar{x}) \\ F(\bar{x}, y') &= H(\bar{x}, y, F(\bar{x}, y)) \end{aligned}$$

provable in $\mathcal{U}_0$(NFA). This is done by taking

$$\varphi = \text{LFP}[\lambda f, \bar{x}, y.\{\text{if } y = 0 \text{ then } G(\bar{x}) \text{ else } H(\bar{x}, Pd(y), f(\bar{x}, Pd(y)))\}] .$$

It is then proved by induction on y that $\varphi(y) \downarrow$; this is by an application of the substitution rule to the schematic induction axiom IA (Ax 3) together with part (i) of the LFP axiom (Ax 6). Then we can take F to be the term φ. It follows that $\mathcal{U}_0$(NFA) serves to define all primitive recursive functions, and so by IA and the substitution rule, we see that $\mathcal{U}_0$(NFA) contains the system of Peano Arithmetic PA in its usual first order (non-schematic) form. I believe this argument formalizes the informal argument (usually not even consciously expressed) which leads us to accept PA starting with the bare-bones system NFA.

Conversely, $\mathcal{U}_0$(NFA) is interpretable in PA, by interpreting the function variables as ranging over (indices of) partial recursive functions, and then the function(al) terms are interpreted as (indices of) partial recursive function(al)s. It follows that we have closure under the LFP scheme. Finally, one shows that if $A[P]$ is provable in $\mathcal{U}_0$(NFA) and B is any formula, and if A^*, B^* are their respective translations, then $A^*[B^*/P]$ is provable in PA. Thus we conclude:

Theorem 1. *$\mathcal{U}_0$(NFA) is proof theoretically equivalent to PA and conservatively extends PA.*

Now to explain the full expansive closure of NFA we treat (as already mentioned) *predicates as propositional functions*, more or less following Aczel with his notion of Frege structures (Aczel [1980]). For this purpose we add a new basic type π, the type of *propositions*, and explain propositional functions as *total* functions f of type $\bar{\iota}\widetilde{\to}\pi$. To fill out the language and axioms of $\mathcal{U}_0$(NFA) we thus begin by taking $Typ_0 = \{\iota, \pi\}$. As before, κ, ν range over Typ_0, τ, σ over types of the form $\bar{\iota}\widetilde{\to}\nu$ (and thus are either types of partial functions from individuals to individuals or partial functions from individuals to propositions), and $(\bar{\tau}, \bar{\kappa}\widetilde{\to}\nu)$ ranges over the types of partial functionals (of partial function, individual and propositional arguments, to individuals or propositions). Now the closure conditions on terms are expanded t o include the logical operations on and to propositions. These are given by the additional symbols $Eq, Pr, Neg, Conj$ and Un with the following clauses:

Tm 9. $Eq(s,t) : \pi$ for $s,t : \iota$.
Tm 10. $Pr(s) : \pi$ for $s : \iota$.
Tm 11. $Neg(s) : \pi$ for $s : \pi$.
Tm 12. $Conj(s,t) : \pi$ for $s,t : \pi$.
Tm 13. $Un(s) : \pi$ for $s : \iota\widetilde{\to}\pi$.

The intended meaning of these symbols is elucidated by Ax 7-11 below.

The formulas $A, B, C, \ldots$ of $\mathcal{U}$(NFA) are generated as follows, where $T(x)$ is an additional predicate which expresses that x is a true proposition:

Fm 1.
(a) $s = t, s\downarrow$, and $P(s)$ are atomic for $s,t : \iota$.
(b) $s = t, s\downarrow$, and $T(s)$ are atomic for $s,t : \pi$.

Fm 2. If A, B are formulas, so also are $\neg A, A \wedge B, \forall x A$.

The axioms of $\mathcal{U}$(NFA) are now as follows (in addition to Ax 1-6 above), where we reserve $x, y, \ldots$ as variables of type ι and $a, b, \ldots$ as variables of type π:

Ax 7. $Eq(x,y)\downarrow \wedge[T(Eq(x,y)) \leftrightarrow x = y]$.
Ax 8. $Pr(x)\downarrow \wedge[T(Pr(x)) \leftrightarrow P(x)]$.
Ax 9. $Neg(a)\downarrow \wedge[T(Neg(a)) \leftrightarrow \neg T(a)]$.
Ax 10. $Conj(a,b)\downarrow \wedge[T(Conj(a,b)) \leftrightarrow T(a) \wedge T(b)]$.
Ax 11. $(\forall x)(fx\downarrow) \to Un(f)\downarrow \wedge[T(Un(f)) \leftrightarrow (\forall x)T(f(x))]$, for $f : \iota\widetilde{\to}\pi$.

Because propositional terms in general implicitly depend on the predicate parameter P, we must restrict the rule $A[P] \Rightarrow A[B/P]$ to formulas A which do not contain any such terms. We write $Pred_n(t)$ for $(\forall \bar{x})(t(\bar{x})\downarrow)$ when $t : \bar{\iota}\widetilde{\to}\pi$ and $\bar{\iota}$ is of length n. Now the usual way of thinking of a *sequence* of n-ary predicates is as a function f of type $\iota\widetilde{\to}(\bar{\iota}\widetilde{\to}\pi)$ such that for each x, $f(x)\downarrow$ and $Pred_n(f(x))$. However, we do not have these types in our set-up (although that is easily modified to include them). Instead, a sequence of n-ary predicates is treated as being represented by a g of type $\iota, \bar{\iota}\widetilde{\to}\pi$ such that for each $x, \bar{y}$ we have $g(x,\bar{y})\downarrow$, in other words so that for each x, $Pred_n(\lambda \bar{y} \cdot g(x,\bar{y}))$. Such g can, at the same time, be considered as an

$(n+1)$-ary predicate and in that guise g is simply the *join* of the sequence it represents: $J(g) = g$.

Now the main result about proof-theoretic strength of $\mathcal{U}(\mathrm{NFA})$ is the following theorem, obtained in collaboration with Thomas Strahm.

Theorem 2. *$\mathcal{U}(NFA)$ is proof-theoretically equivalent to the system of ramified analysis up to but not including Γ_0, and conservatively extends that system.*

The system of ramified analysis up to and including level β is denoted RA_β, and the union of these for $\beta < \alpha$ is denoted $\mathrm{RA}_{<\alpha}$. For $\alpha = \omega \cdot \alpha$ this is proof-theoretically equivalent to the iteration of $(\Pi^0_1 - \mathrm{CA})$ through all levels $\beta < \alpha$. Using Kreisel's proposed characterization of predicative analysis in terms of the autonomous progression of ramified systems, the least impredicative ordinal was determined to be Γ_0 in Feferman [1964] and, independently, Schütte [1965]. Theorem 2 thus re-characterizes *predicativity* as *what ought to be accepted concerning operations and predicates if one has accepted the basic notions and principles of* NFA, *including the logical operations* $\neg, \wedge$ *and* $\forall$ *applied to variables for the natural numbers.* The proof of this theorem is rather involved and full details will be presented elsewhere; the following merely gives an indication of ho w to embed $\mathrm{RA}_{<\Gamma_o}$ in $\mathcal{U}(\mathrm{NFA})$, by means of the methods of Feferman [1979], sec.3.3. Basically, one shows for each initial segment $\prec_\alpha$ of the standard primitive recursive well-ordering of order type Γ_0 how to establish in $\mathcal{U}(\mathrm{NFA})$ the principle of transfinite induction up to α applied to arbitrary formulas A, in symbols, $\mathrm{TI}(\prec_\alpha, A)$. For this it suffices to prove $\mathrm{TI}(\prec_\alpha, P)$ and then apply the substitution rule. Now with the full scheme at hand, one can define the jump (Π^0_1) hierarchy relative to P along $\prec_\alpha$ by LFP recursion and prove that it defines a predicate by induction on this ordering. Note that the definition of this hierarchy makes use of arithmetical steps at successor stages, guaranteed by the axioms Ax 7-11, and of join at limit stages, guaranteed by the use of the J operator as explained above. As is shown in the reference loc.cit., by use of this hierarchy relative to P up to α , one can prove $\mathrm{TI}(\prec_\gamma, P)$ for $\gamma = \kappa^{(\alpha)}(0)$ in the Veblen hierarchy of critical functions. Define $\gamma_0 = 0, \gamma_{n+1} = \kappa^{(\gamma_n)}(0)$; then $\Gamma_0 = \lim_n \gamma_n$, so by this means we can embed RA_α in $\mathcal{U}(\mathrm{NFA})$ for each $\alpha < \Gamma_0$. The proof that $\mathcal{U}(\mathrm{NFA})$ is no stronger than $\mathrm{RA}_{<\Gamma_0}$ requires some interesting new arguments from infinitary proof theory. However, it is worth noting that in this proof, partial functions of type $\bar{\iota} \widetilde{\to} \iota$ are still interpreted as partial recursive functions. Indeed the same holds for functions of type $\bar{\iota} \widetilde{\to} \pi$ when propositions are treated intensionally.

Remarks

1. **Implicit definability of functions.** Another way of introducing partial functions given by implicit defining conditions is if we associate with each partial $f : \bar{\iota}, \iota \widetilde{\to} \iota$ *a* $g : \bar{\iota} \widetilde{\to} \iota$ with

(ID) $\quad \forall \bar{x}, y, z[f(\bar{x}, y) \simeq 0 \wedge f(\bar{x}, z) = 0 \rightarrow y = z]$
$\rightarrow \forall \bar{x}[(\exists y) f(\bar{x}, y) = 0 \rightarrow f(\bar{x}, g(x)) = 0].$

Adding (ID) as an axiom to $\mathcal{U}_0$(NFA) and $\mathcal{U}$(NFA) does not affect Theorems 1 and 2. It is plausible to include (ID) in the unfolding process applied to any system with a distinguished constant 0.

2. **Predicate types in place of the type of propositions.** We can treat predicates directly, instead of in terms of propositional functions, by introducing a basic type of n-ary predicates π_n for each $n \geq 1$. Then the atomic formulas to be used in the $\mathcal{U}$ process for this symbolism are of the form $s = t$ for $s, t : \pi_n, s \downarrow$ for $s : \pi_n$ and $(t_1, \ldots, t_n) \in s$ for $s : \pi_n$ and $t_j : \iota$ $(j = 1, \ldots, n)$. The axioms provide for suitable operations corresponding to atomic predicates and for the effect of *Neg* and *Conj* on each π_n and *Un* on π_{n+1} to π_n for each n. In addition, we include the *Join operator* J for each n, which when applied to a sequence of n-ary predicates, i.e. a total $f : \iota \rightarrow \pi_n$, produces the join predicate $J(f) : \pi_{n+1}$. In the language, so modified, the rule of substitution $A[P] \Rightarrow A[B/P]$ is restricted to A which do not contain terms of predicate type. Then Theorem 2 h olds as before. An advantage of the predicate type over the propositional type approach is that we can separate out the role of the Join operator from that of the logical operations while, as we saw, J is forced on us in the propositional type approach. Strahm has shown that if J is omitted, then the resulting system $\mathcal{U}^-$(NFA) is proof-theoretically equivalent to $RA_{<\omega}$.

3. **Quantifying function variables.** It was argued in Section 3 that for the general notion of unfolding, (partial) function variables in their schematic role ought not to be quantified. However, when we come to set theory and examine informal arguments that lead us to accept its basic principles and their immediate extensions, it is plausible to allow some degree or other of function quantification. Proof-theoretical strength there is sensitive to the decision as to whether to allow such quantification, and, if so, to what extent, as will be seen in the next section. Interestingly, it happens that in the case of NFA, even if we allow full function quantification in the language of $\mathcal{U}_0$(NFA), resp. $\mathcal{U}$(NFA), with suitable restrictions on the hypothesis $A[P]$ of the substitution rule as above, we do not alter proof-theoretic strength, i.e. Theorems 1 and 2 continue to hold as stated.

4. **The unfolding of finitist arithmetic.** Clearly the starting point for the study of this notion would be a *quantifier-free* system FA based on Axs 1 and 2 and, in place of Ax 3, the *induction rule*

$$P(0), P(x) \rightarrow P(x') \Rightarrow P(x).$$

Beyond this, there are various notions of unfolding to be considered, related to various informal and formal explanations of finitism in the

literature, due especially to Hilbert, Kreisel and Tait. Research on these notions is in progress.

5. The unfolding of set theory

This section is largely programmatic and, given the limitations of space, necessarily sketchy. On the face of it, set theory offers a prime candidate for the study of what is implicit in given notions and principles by means of the unfolding procedure, both for ZF as a schematic theory and for Gödel's program for new axioms. We begin with the former.

In the spirit of the functional formulation of the $\mathcal{U}_0$ and $\mathcal{U}$ procedures, we take the basic language of set theory to have individual variables $a, b, c, x, y, z, \ldots$, variables $f, g, h, \ldots$ for partial functions, the constants 0 and ω, the operation symbols $\{,\}, \bigcup, \wp$, and E (the characteristic function of the $\in$ relation) and the relation symbols $=$ and $\in$. In addition we have functionals $\mathbf{S}, \mathbf{R}$ and $\mathbf{A}$ whose meaning will be explained in a moment. The axioms of the system ST are, besides Extensionality, the expected ones for $0, \omega, \{,\}, \bigcup, \wp$, and E, and the following four function and predicate schemata:

$$(\mathbf{S}) \quad \forall x \in a[f(x)\downarrow] \to \mathbf{S}(f,a)\downarrow \wedge \forall x[x \in \mathbf{S}(f,a) \leftrightarrow x \in a \wedge f(x) = 0]$$

$$(\mathbf{R}) \quad \forall x \in a[f(x)\downarrow] \to \mathbf{R}(f,a)\downarrow \wedge \forall y[y \in \mathbf{R}(f,a) \leftrightarrow \exists x \in a(f(x) = y)]$$

$$I(\in) \qquad \forall x[(\forall y \in x)P(y) \to P(x)] \to \forall x(P(x))$$

$$(\mathbf{A}) \qquad \forall x[f(x)\downarrow] \to \mathbf{A}(f)\downarrow \wedge [\mathbf{A}(f) = 0 \leftrightarrow \forall x(f(x) = 0)].$$

Thus $\mathbf{S}$ gives Separation, $\mathbf{R}$ gives Replacement, $I(\in)$ is the positive (inductive) schematic form of the Axiom of Foundation, and $\mathbf{A}$ serves to represent every definable class by means of a characteristic function. This last allows $(\mathbf{S})$ and $(\mathbf{R})$ to take the place of the expected schemata:

$$\text{(Sep)} \qquad \exists b \forall x[x \in b \leftrightarrow x \in a \wedge P(x)], \text{ and}$$

$$\text{(Repl)} \qquad (\forall x \in a)\exists! y P(x,y) \to \exists b \forall y[y \in b \leftrightarrow (\exists x \in a)P(x,y)].$$

The point of doing it by the above function schemata instead is that we can treat a wide variety of set theories uniformly, with the only changes being the deletion or addition (with appropriate axioms) of various individual, function and functional constants. For example, if we omit $\omega, \wp$, and $\mathbf{A}$, we obtain a functional schematic form AST of *Admissible Set Theory.* To be more precise KP (taken with Δ_0-Replacement instead of Δ_0-Collection) is contained in $\mathcal{U}_0(\mathrm{AST})$, and the latter is interpretable in the constructible sets of the former by taking the function variables to range over the $\Sigma_1^{(L)}$ partial functions. It would be of interest to determine the strength of $\mathcal{U}(\mathrm{AST})$.

Quite a few useful general principles and functional constructions can be derived in $\mathcal{U}_0(\mathrm{AST})$ and $\mathcal{U}(\mathrm{AST})$, which then carry over to (the respective unfolding) of any set theory S extending AST. In particular, we can derive principles of induction for various classes C with an ordering $<_C$ in the form:

$$I(<_C) \qquad \forall x \in C[\forall y(y <_C x \to P(y)) \to P(x)] \to \forall x \in C[P(x)]\,.$$

Here $<_C$ might be much "longer" than the ordinals, for which we have $I(<)$ by the axiom $I(\in)$. Taking Ω as a symbol for the class of ordinals, we can define, for example, the lexicographic ordering $<_{\Omega^2}$ on pairs of ordinals by $\langle \xi, \eta\rangle <_{\Omega^2} \langle \alpha, \beta\rangle \leftrightarrow \xi < \alpha \vee \xi = \alpha \wedge \eta < \beta$, and prove $I(<_{\Omega^2})$ in $\mathcal{U}_0(\mathrm{AST})$. From this and the LFP construction we can derive a principle of recursion for hierarchies of functions $\lambda\alpha, \beta.\, f_\alpha(\beta)$ by means of any given functional G which determines each f_α in terms of $\langle f_\xi\rangle_{\xi<\alpha}$. More generally, I expect that we can establish $I(<_\rho)$ in $\mathcal{U}_0(\mathrm{AST})$ for each $\rho < \varepsilon_{\Omega+1}$ and similarly for each $\rho < \Gamma_{\Omega+1}$ in $\mathcal{U}(\mathrm{AST})$, where the ordering up to $\Gamma_{\Omega+1}$ is defined in AST on a suitable class of "notations" as in Feferman [1968]. We would then obtain related principles of recursion and construction of hierarchies as for $<_{\Omega^2}$ above. Note that the form of this ordering is independent of which set theory S we are in, but the interpretation in a standard model of S depends on what ordinal Ω turns out to be. What stronger S serve to do is supply a greater variety of functionals G for generating hierarchies associated with $<_\rho$ when $I(<_\rho)$ is provable.

Suppose S is an extension of our initial system ST to which we have added AC and the existence of arbitrarily large inaccessible cardinals. Then the preceding allows us to actually "name" specific large inaccessibles in the unfolding systems of S. In that sense, it already gives us some large cardinal axioms. But if we are to generate, e.g., hierarchies of Mahlo cardinals, we need to add to ST a new scheme which says in effect that *whatever holds in the universe of sets already holds in arbitrarily large transitive sets*, or what one would call a scheme of *Downwards Reflection*. This takes the following form:

$$\text{(D-Ref)} \qquad P \to \exists b[a \in b \wedge \mathit{Trans}(b) \wedge P^{(b)}]\,.$$

If this scheme is denoted $A[P]$ and B is a statement which involves both quantified individual variables and (possibly) quantified function variables, when forming $B^{(b)}$ in $A[B/P]$ we relativize the former variables to b as usual, and the latter variables to partial functions from b to b. Write $\mathit{Strans}(b)$ for $\forall x \in b \forall y[y \subseteq x \to y \in b]$. We can infer

$$\text{(D-Ref)}' \qquad P \to \exists b[a \in b \wedge \mathit{Strans}(b) \wedge P^{(b)}]$$

by substituting $P \wedge \forall x \exists y[\wp(x) = y]$ for P in (D-Ref). Thus with the substitution rule $A[P] \Rrightarrow A[B/P]$ taken to apply to any statement B in the unfolding language of ST in which function variables may be quantified unrestrictedly, we obtain a form of Bernays' downward second-order reflection principle (Bernays [1961], following on Levy [1960]). And as Bernays showed

op.cit., the existence of hierarchies of Mahlo cardinals then follows from this principle. Briefly, one begins by substituting for P in (D-Ref) the statement that expresses that the universe is closed under power set and replacement, i.e.

$$\forall x \exists y[\wp(x) = y] \wedge \forall u \forall g[\forall x \in u\, \exists y(g(x) = y)] \\ \rightarrow \exists v[\mathbf{R}(g, u) = v \wedge \forall y(y \in v \leftrightarrow \exists x \in u(g(x) = y)]\,.$$

Then the conclusion of this instance of D-Ref guarantees the existence of arbitrarily large inaccessible cardinals. It follows that any normal function on Ω has arbitrarily large inaccessible fixed-points. By substituting *that* statement for P in (D-Ref) we obtain the existence of arbitrarily large Mahlo cardinals — and so on.

Formulas involving (partial) function quantification are classified into the Π^1_n hierarchies as usual. The existence of Mahlo hierarchies follows from (D-Ref) by successively substituting suitable Π^1_1 statements for P. But if one is to obtain stronger large cardinal statements, e.g. the existence of weakly compact cardinals, it is necessary to make substitutions by more complicated formulas. For, as shown in the work of Hanf and Scott [1961], a cardinal κ is weakly compact iff it is Π^1_1 indescribable. The latter says that (D-Ref) holds in V_κ for all Π^1_1 statements, and saying *that* is Π^1_2. In general, we obtain the existence of arbitrarily large Π^1_n indescribables by suitably more complicated instances of (D-Ref). And that is *all* one can expect to follow from (D-Ref) in our languages using only function variables of type level 1 over the universe. And passing to higher types — however one were to argue for that — for substitution instances in (D-Ref), at most allows one to obtain the existence of Π^m_n indescribables for all m, n. But one certainly cannot obtain in this way the existence of measurable cardinals nor even some of its familiar consequences such as the existence of $0^\#$ (or even of some still weaker consequences from infinitary combinatorics, such as explained in Kanamori [1994] p.109).

However, as I see it, there is already a flat difference between the reasoning which leads us to the hierarchies of Mahlo cardinals, and that which leads, to begin with, to weakly compact cardinals. Here a quotation from Tarski is apropos:

> ... the belief in the existence of inaccessible cardinals $> \omega$ (and even of arbitrarily large cardinals of this kind) seems to be a natural consequence of basic intuitions underlying the "naïve" set theory and referring to what can be called "Cantor's absolute". On the contrary, we see at this moment no cogent intuitive reasons which could induce us to believe in the existence of cardinals $> \omega$ that are not strongly incompact, or which at least would make it very plausible that the hypothesis stating the existence of such cardinals is consistent with familiar axiom systems of set theory. As was pointed out at the end of Section 1, we do not know of any "constructively characterized"

> cardinal $> \omega$ of which we cannot prove that it is strongly incompact and for which therefore the problems discussed remain open. (Tarski [1962], p.134).

Gödel, commenting on this in a footnote (20) added in 1965 to his 1940 monograph (after referring to the work of Levy [1960] and Bernays [1961] leading to "all of Mahlo's axioms") said:

> Propositions which, if true, are extremely strong axioms of infinity of an entirely new kind have been formulated and investigated as to their consequences and mutual implications in Tarski [1962], Keisler and Tarski [1964] and the papers cited there. In contradistinction to Mahlo's axioms the truth (or consistency) of these axioms does not immediately follow from "the basic intuitions underlying abstract [sic] set theory" (Tarski [1962], p. 134), nor can it, as of now, be derived from them. However, the new axioms are supported by rather strong arguments from *analogy* ... (Gödel [1990] p. 97, italics mine).

What makes the separation of Mahlo from weakly compact cardinals reasonable is that when we substitute for P in (D-Ref) a Π^1_1 statement B, we may read B as asserting a *closure condition* in the ordinary sense on V under given function(al)s. But this reading is not plausibly extended to statements of higher function-quantifier complexity. From what Gödel says in the preceding quotation, it seems he would agree with this argument for demarcation.[7]

My personal attitude concerning the question of "actual" existence of various kinds of large cardinals, whether smaller or larger, is that *it is all pie in the sky.* This may make one wonder why I have even bothered with the present section. Well, the starting point was to see what one can say about which large cardinal statements are implicit in the basic notions and principles of set theory, *if* one accepts them, as Gödel and many other logicians certainly do, and to try to apply the unfolding procedure to begin to say something precise about that.[8] While that hypothetical acceptance does not apply to me, there are other potential values of great interest to me, which I hope will result from further pursuit of the present framework. The analogues to various large cardinal statements in admissible set theory are well-known.

[7] Tait [1990], p.76, ftn. 6, is puzzled by this view of Gödel's. But he says there that the existence of weakly compact cardinals follows from Π^1_1 reflection, which is mistaken, as we have seen.

[8] And if one is among the set theorists who believe there are reasons for accepting much larger cardinals than follow from $\mathcal{U}(\mathrm{ST})$, it should be of interest to make explicit what are the basic notions and principles that lead one to such conclusions, rather than depend on arguments from analogy or fruitfulness. In this respect, a suggestion of Gödel in his 1946 Princeton remarks is most provocative:"It is cert ainly impossible to give a combinational and decidable characterization of what an axiom of infinity is; but there might exist, e.g., a characterization of the following sort: An axiom of infinity is a proposition which has a certain (decidable) formal structure and which in addition is true." (Gödel [1990], p. 151)

The work earlier in this section with AST suggests to me that there should be a way of stating these as part of a *common generalization* via the unfolding of S+(D-Ref) for S$\supseteq$AST, and not merely an *analogue*. Still further, there has been a surprising use of recursive ordinal notation systems employing "names" for very large cardinals in current proof-theoretic ordinal an alyses of formal systems (cf. e.g. Rathjen [1995]). What I would really hope comes out of this is a generalization which encompasses these as well, and helps explain how it is that they come to be employed at all for these purposes.

References

P. Aczel [1980], Frege structures and the notions of proposition, truth and set, in (J. Barwise, et al, eds.) *The Kleene Symposium*, North-Holland, Amsterdam. 31-59.

M. Beeson [1985], *Foundations of Constructive Mathematics*, Springer-Verlag, Berlin.

P. Bernays [1961], Zur Frage der Unendlichkeitsschemata in der axiomatischen Mengenlehre, in (Y. Bar-Hillel, et al, eds.) *Essays on the Foundations of Mathematics*, Magnes Press, Jerusalem, 3-49. (See also Bernays [1976].)

P. Bernays [1976], On the problem of schema of infinity in axiomatic set theory, in (G. Müller, ed.) *Sets and Classes*, North-Holland, Amsterdam, 121-172. (English translation of Bernays [1961].)

S. Feferman [1962], Transfinite recursive progressions of axiomatic theories, *J. Symbolic Logic* 27, 259-316.

S. Feferman [1964], Systems of predicative analysis, *J. Symbolic Logic* 29, 1-30.

S. Feferman [1968], Systems of predicative analysis, II: Representations of ordinals, *J. Symbolic Logic* 33, 193-220.

S. Feferman [1979], A more perspicuous formal system for predicativity, in (K. Lorenz, ed.) *Konstruktionen vs. Positionen.* Vol. I, Walter de Gruyter, Berlin, 68-93.

S. Feferman [1988], Turing in the land of $O(z)$, in (R. Herken, ed.) *The Universal Turing Machine. A Half-century Survey*, Oxford Univ. Press, Oxford, 113-147.

S. Feferman [1991], Reflecting on incompleteness, *J. Symbolic Logic* 56, 1-49.

S. Feferman [1991a], A new approach to abstract data types, II. Computation on ADTs as ordinary computation, in (E. Börger, et al, eds.) *Computer Science Logic*, Lecture Notes in Computer Science 626, 79-95.

S. Feferman [1996], Computation on abstract data types. The extensional approach, with an application to streams, to appear in *Annals of Pure and Applied Logic.*

K. Gödel [1986], *Collected Works, Vol. I. Publications 1929-1936*, Oxford Univ. Press, New York.

K. Gödel [1990], *Collected Works, Vol. II. Publications 1938-1974*, Oxford Univ. Press, New York.

K. Gödel [1995], *Collected Works, Vol. III. Unpublished Essays and Lectures*, Oxford Univ. Press, New York.

W. P. Hanf [1964], Incompactness in languages with infinitely long expressions, *Fundamenta Mathematicae* 53, 309-324.

W. P. Hanf and D. Scott [1961], Classifying inaccessible cardinals (abstract), *Notices A.M.S.* 8, 445.

R. Jensen [1995], Inner models and large cardinals, *Bull. Symbolic Logic* 1, 393-407.

A. Kanamori [1994], *The Higher Infinite*, Springer-Verlag, Berlin.

H. J. Keisler and A. Tarski [1964], From accessible to inaccessible cardinals, *Fundamenta Mathematicae* 53, 225-308. Corrections ibid. 57 (1965) 119.

G. Kreisel [1958], Ordinal logics and the characterization of informal concepts of proof, *Proc. International Congress of Mathematicians (Edinburgh 1958)*, Cambridge Univ. Press, New York, 289-299.

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

A. Levy [1960], Axiom schemata of strong infinity in axiomatic set theory, *Pacific Journal of Mathematics* 10, 223-238.

P. Maddy [1988], Believing the axioms, I. *J. Symbolic Logic* 53, 481-511.

P. Maddy [1988a], Believing the axioms, II. *J. Symbolic Logic* 53, 736-764.

D. A. Martin [1976], Hilbert's first problem: The continuum hypothesis, in (F. Browder, ed.) *Mathematical Developments Arising from Hilbert's Problems*, Proc. Symposia in Pure Math. 28, A.M.S., Providence, 81-92.

Y. Moschovakis [1989], The formal language of recursion, *J. Symbolic Logic* 54, 1216-1252.

M. Rathjen [1995], Recent advances in ordinal analysis: Π^1_2-CA and beyond, *Bull. Symbolic Logic* 1, 468-485.

K. Schütte [1965], Eine Grenze für die Beweisbarkeit der transfiniten Induktion in der verzweigten Typenlogik, *Archiv für Math. Logik und Grundlagenforschung* 7, 45-60.

D. Scott [1961], Measurable cardinals and constructible sets, *Bull. de l'Acad. Polonaise des Sciences* 9, 521-524.

W. Tait [1990], The iterative hierarchy of sets, *Iyyun, A Jerusalem Philosophical Quarterly* 39, 65-79.

A. Tarski [1962], Some problems and results relevant to the foundations of set theory, in (E. Nagel, et al, eds.) *Logic, Methodology and the Philosophy of Science* (Proc. of the 1960 International Congress, Stanford), Stanford Univ. Press, Stanford, 125-135.

A. Turing [1939], Systems of logic based on ordinals, *Proc. London Math. Soc.*, ser. 2, 45, 161-228.