

PREDICATIVITY BEYOND Γ_0

NIK WEAVER

ABSTRACT. We reevaluate the claim that predicative reasoning (given the natural numbers) is limited by the Feferman-Schütte ordinal Γ_0 . First we comprehensively criticize the arguments that have been offered in support of this position. Then we analyze predicativism from first principles and develop a general method for accessing ordinals which is predicatively valid according to this analysis. We find that the Veblen ordinal $\phi_{\Omega^\omega}(0)$, and larger ordinals, are predicatively provable.

The precise delineation of the extent of predicative reasoning is possibly one of the most remarkable modern results in the foundations of mathematics. Building on ideas of Kreisel [27, 28], Feferman [10] and Schütte [40, 41] independently identified a countable ordinal Γ_0 and argued that it is the smallest predicatively non-provable ordinal. (Throughout, I take “predicative” to mean “predicative given the natural numbers”.) This conclusion has become the received view in the foundations community, with reference [10] in particular having been cited with approval in virtually every discussion of predicativism for the past forty years. Γ_0 is now commonly referred to as “the ordinal of predicativity”. Some recent publications which explicitly make this assertion are [1, 2, 3, 4, 6, 18, 19, 20, 23, 25, 26, 36, 37, 39, 44, 45].

This achievement is notable both for its technical sophistication and for the insight it provides into an important foundational stance. Although predicativism is out of favor now, at one time it was advocated by such luminaries as Poincaré, Russell, and Weyl. (Historical overviews are given in [18] and [35].) Its central principle — that sets have to be “built up from below” — is, on its face, reasonable and attractive. With its rejection of a metaphysical set concept, predicativism also provides a cogent resolution of the set-theoretic paradoxes and is more in line with the positivistic aspect of modern analytic philosophy than are the essentially platonic views which have become mathematically dominant.

Undoubtedly one of the main reasons predicativism was not accepted by the general mathematical public early on was its apparent failure to support large portions of mainstream mathematics. However, we now know that the bulk of core mathematics can in fact be developed in predicative systems [15, 43], and the limitation identified by Feferman and Schütte is probably now a primary reason, possibly *the* primary reason, for predicativism’s nearly universal unpopularity.¹ There do exist important mainstream theorems which are known to in various senses require provability of Γ_0 , and in any case Γ_0 is sufficiently tame that it is simply hard to take seriously any approach to foundations that prevents one from recognizing ordinals at least this large. Thus, it is of great foundational interest to examine carefully whether the Γ_0 limitation really is correct. If it is

Date: May 11, 2009.

not, predicativism could be more viable than previously thought and its current peripheral status in the philosophy of mathematics may need to be reconsidered.

I believe Γ_0 has nothing to do with predicativism. I will argue that the current understanding of predicativism is fundamentally flawed and that a more careful analysis shows the “small” Veblen ordinal $\phi_{\Omega^\omega}(0)$, and probably much larger ordinals, to be within the scope of predicative mathematics. It is my hope that this conclusion will open the way to a serious reappraisal of the significance and interest of predicativism. Elsewhere I introduce the term “mathematical conceptualism” for the brand of predicativism considered here and make a case that it is cogent, rigorous, attractive, and better suited to ordinary mathematical practice than all other foundational stances [46].

1. A CRITIQUE OF THE Γ_0 THESIS

At issue is the assertion that there are well-ordered sets of all order types less than Γ_0 and of no order types greater than or equal to Γ_0 which can be proven to be well-ordered using predicative methods (cf. [10], p. 13 or [42], p. 220). I call this *the Γ_0 thesis*.

As stated, this claim is imprecise because the classical concept of well-ordering has a variety of formulations which are not predicatively equivalent (see §1.4 and §2.4). In fact, previous discussions of predicativism have tended to ignore this distinction, and this will emerge as a crucial source of confusion (see §1.4). To fix ideas I will use the term “well-ordered” to mean of a set X that it is totally ordered and if $Y \subseteq X$ is progressive then $Y = X$. *Progressive* means that for every $a \in X$, if $\{b \in X : b \prec a\} \subseteq Y$ then $a \in Y$.

In principle, to falsify the Γ_0 thesis I need only produce (1) a well-ordering proof of an ordered set that is isomorphic to Γ_0 and (2) a convincing case that the proof is predicatively valid. However, no matter how convincing I could make that case, in light of the broad and sustained acceptance the thesis has enjoyed it would be unsatisfying to leave the matter there. The Γ_0 thesis has been repeatedly and forcefully defended by two major figures, Feferman and Kreisel. Many current authors simply assert it as a known fact. The only substantial published criticism of which I am aware appears in [24], but even that is somewhat ambivalent and seems to conclude in favor of the thesis. Therefore, I take it that I have a burden not only to positively demonstrate the power of predicative reasoning, but also to show where the more pessimistic previous assessments went wrong.

This is a somewhat lengthy task because a great deal has been written in support of the Γ_0 thesis from a variety of points of view. However, I believe that the entire body of argument is specious and can be decisively refuted. The goal of Section 1 is to do this in some detail.

One point before I begin. Predicativism is a philosophical position, and prior to the acceptance of a particular formalization there will be room for argument over its precise nature. Thus, a debate about the Γ_0 thesis could easily degenerate into a purely semantic dispute over the meaning of the term “predicative”. I therefore want to emphasize that the central claim of this section is that there is *no* coherent philosophical stance which would lead one to accept every ordinal less than Γ_0 but not Γ_0 itself. My polemical technique will be to examine various formal systems that have been alleged to model predicative reasoning and indicate how in each instance the informal principles that motivate the system actually justify a stronger system

which goes beyond Γ_0 . This is obviously independent of any special views one may have about predicativity. (A second major claim is that none of these supposedly predicative systems is actually predicatively legitimate. Evaluating the justice of this claim does require some understanding of predicativism, and I refer the reader who is not prepared to accept any assertion of this type at face value to Section 2, where I develop my views on predicativism in detail.)

1.1. Formal systems for predicativity. A variety of formal systems have been proposed as modelling predicative reasoning in some form of second order arithmetic. Among the main examples are Σ [27], H^+ , R^+ , H , R [10], RA^* [42], $P+\exists/P$ [13], $\text{Ref}^*(\text{PA}(P))$ [14], and $\mathcal{U}(\text{NFA})$ [19]. I give here a brief sketch of their most important features.

The systems Σ , H , and R are similar in broad outline and need not be distinguished in this discussion; likewise for the systems H^+ , R^+ , and RA^* . All six express a concept of “autonomy” which allows one to access larger ordinal notations from smaller ones and I will refer to them generally as “autonomous systems”. In the last three systems the idea is that once a predicativist has proven the well-foundedness of a set of order type α he is allowed to use infinite proof trees of height α to establish the well-foundedness of larger order types, the key infinitary feature being an “ ω -rule” which permits deduction of the formula $(\forall n)\mathcal{A}(n)$, where n is a number variable, from the family of formulas $\mathcal{A}(\bar{n})$ with \bar{n} ranging over all numerals. In the first three systems all proofs are finite and the key proof principle is a “formalized ω -rule” schema which, for each formula \mathcal{A} , concludes the formula $(\forall n)\mathcal{A}(n)$ from a premise which arithmetically expresses that for every number n there is a proof of $\mathcal{A}(\bar{n})$. This leads to a hierarchy of systems S_a where a is an ordinal notation and $S_{a\oplus 1}$ incorporates a formalized ω -rule schema referring to proofs in S_a . The predicativist is then permitted to execute a finite succession of proofs in various S_a ’s, subject to the requirement that passage to any S_a must be preceded by a proof that a is an ordinal notation.

The linked systems P and \exists/P are notable for their conception of predicativists as having a highly restricted yet not completely trivial ability to deal with second order quantification, in particular their being able to use only free or, to a limited extent, existentially quantified set variables. Second order existential quantification is actually permitted only in the “auxiliary” system \exists/P , but once a functional has been shown to exist uniquely one is allowed to introduce a symbol for it which can then be used in P . By passing back and forth between P and \exists/P one is able to produce functionals which provably enumerate larger and larger initial segments of the ramified hierarchy over an arbitrary set and use them to prove the well-ordering property for successively larger ordinal notations.

The system $\text{Ref}^*(\text{PA}(P))$ is obtained by applying a general construction Ref^* to a “schematic” form $\text{PA}(P)$ of Peano arithmetic. This construction involves extending the language of $\text{PA}(P)$ to allow assertions of truth and falsehood and adding axioms which govern the use of the truth and falsehood predicates. Paradoxes arising from a self-referential notion of truth are avoided by regarding the truth and falsehood predicates as partial and axiomatizing them in a way that expresses their ultimate groundedness in facts about $\text{PA}(P)$. The ability to reason about truth in effect implements the formalized ω -rule mentioned above, and this again enables one to prove the well-foundedness of successively larger ordinal notations. The general idea is that $\text{Ref}^*(S(P))$ embodies what one “ought to accept” given that one accepts

a schematic theory $S(P)$, and an argument can then be made that predicativism is fundamentally based on Peano arithmetic and therefore $\text{Ref}^*(\text{PA}(P))$ precisely captures what a predicativist ought to accept.

Like $\text{Ref}^*(\text{PA}(P))$, $\mathcal{U}(\text{NFA})$ is an instance of a general construction which applies to any schematic formal system and is supposed to embody what one ought to accept once one accepts that system. However, the exact claim is slightly different: here we are concerned with determining “which operations and predicates, and which principles concerning them, ought to be accepted” once one has accepted the initial system ([19], p. 75). This problem is approached from the point of view of generalized recursion theory and one is allowed to generate operations and predicates by using a least fixed point operator. It is easily seen that this recursive generation procedure rapidly recovers Peano arithmetic from a weaker theory NFA (“non-finitist arithmetic”); therefore, $\mathcal{U}(\text{NFA})$ is already supposed to capture predicative reasoning.

1.2. Outline of the critique. All of the proposed formalizations of predicative reasoning cited in §1.1 have the same provable ordinals, namely all ordinals less than Γ_0 . This by itself might be seen as good evidence in favor of the Γ_0 thesis simply because it seems unlikely that so many different approaches should all have settled on the same wrong answer.

Nonetheless, each of the formal systems in §1.1 is simultaneously too weak and too strong to faithfully model predicative reasoning and thereby verify the claim about Γ_0 . They are all too weak for a general reason I discuss in §1.3; in brief, anyone who accepts a given system ought to be able to grasp its global validity and then go beyond it. This is an old objection and there are several responses to it on record. However, these responses, which I review below, are not well-taken because they typically involve postulating (usually with little or no overt justification) a limitation on predicative reasoning which, if true, would actually have prevented a predicativist from working within the original system.

In addition, each system is manifestly impredicative in some way, and hence too strong. This fact does not seem to be widely appreciated, but it is hardly obscure. The autonomous systems impredicatively infer a transfinite iteration of reflection principles from a statement of transfinite induction. $P + \exists/P$ allows predicate substitution for Σ_1^1 formulas, so that for every Σ_1^1 formula \mathcal{A} it in effect lets one reason about $\{n : \mathcal{A}(n)\}$ as if this were a meaningful set, which in general is predicatively not the case. $\text{Ref}^*(\text{PA}(P))$ makes truth claims about schematic predicates which do not make sense unless one assumes an impredicative comprehension axiom. $\mathcal{U}(\text{NFA})$ employs a patently impredicative least fixed point operator and also treats schematic predicates in a way that again can only be justified by impredicative comprehension. I will elaborate on all of these points below.

The most striking impredicativity is the least fixed point operator of $\mathcal{U}(\text{NFA})$, but the other instances are actually more significant because they fit into a general pattern. The basic problem is that in each of these systems one proves the well-foundedness of successively larger ordinal notations by an inductive argument that at each step involves generating some kind of iterative hierarchy which is used to prove transfinite induction at the next level — but this does not justify the statement of transfinite *recursion* which is needed to generate the next hierarchy. In order to make this inference from induction to recursion one has to smuggle an

impredicative step somewhere into the proof, and this is the function of all the other examples of impredicativity noted above. I will return to this point in §1.9.

There are also more subtle instances of impredicativity which occur in the use of self-applicative schematic predicates in $\text{Ref}^*(\text{PA}(P))$ and $\mathcal{U}(\text{NFA})$ and in the use of self-applicative truth and falsehood predicates in $\text{Ref}^*(\text{PA}(P))$; see §2.5 and §2.6.

I next describe an objection that is generally applicable, and then go on to discuss the individual formal systems.

1.3. A general difficulty. Suppose A is a rational actor who has adopted some foundational stance. Any attempt to precisely characterize the limits of A 's reasoning must meet the following objection: if we could show that A would accept every member of some set of statements \mathcal{S} , then A should see this too and then be able to go beyond \mathcal{S} , e.g. by asserting its consistency. Thus, \mathcal{S} could not have been a complete collection of all the statements (in a given language) that A would accept. A similar argument can be made about attempts to characterize A 's provable ordinals.

There are a variety of ways in which this objection might be overcome. A may actually be unable to recognize \mathcal{S} as a legitimate set, for instance if \mathcal{S} is infinite and A is a finitist. Or the language in use may not be capable of expressing the consistency of \mathcal{S} . Or perhaps A can indeed see, as we do, that there exists a proof that he would accept for each statement in \mathcal{S} , but he cannot go from this to actually accepting every statement in \mathcal{S} (though it is difficult to imagine a plausible set of beliefs that would not allow him to take this step). Or it may be possible to identify some special limitation in A 's belief system which prevents him from grasping the validity of all of \mathcal{S} at once despite his ability to accept each statement in \mathcal{S} individually.²

Defenses of the Γ_0 thesis generally take the last approach. This is tricky for a slightly subtle reason. It is not hard to believe that A (or anyone) is unable to simultaneously identify exactly which statements are true from his perspective. But it is more difficult to reconcile this with the claim that we *do* know that he would accept each statement in \mathcal{S} . The most obvious way to establish this claim would be to explicitly show how A would prove each statement in \mathcal{S} , and this is actually the method used in the case at hand: each of the proposed formal systems for predicativism is accompanied by a recursive proof schema which is supposed to show how a predicativist could use the system to access every ordinal less than Γ_0 . What is confusing here is the suggestion that we can see that he would accept each proof in the schema but he cannot see this.

In fact this is highly implausible, for the following reason. Let $\gamma_0 = 1$ and $\gamma_{n+1} = \phi_{\gamma_n}(0)$, where $\{\phi_\alpha(\beta)\}$ is the Veblen hierarchy of critical functions, so that $\Gamma_0 = \sup_{n \in \omega} \gamma_n$. Now in general we are not merely given a recursive set of proofs which establish for each n that some notation a_n for γ_n is an ordinal notation; for each of the formal systems under discussion, at least at an intuitive level we have a *single* proof that for any n , a_n is an ordinal notation implies a_{n+1} is an ordinal notation. It therefore becomes hard to believe that someone who is presumed to grasp induction on ω (and even, allegedly, in “schematic” form [14, 16, 19]) would not be able to infer the single assertion that a_n is an ordinal notation for all n .

It is reasonable to expect that if a predicativist understands how to go from a_n to a_{n+1} for any single value of n , and if the passage from a_n to a_{n+1} is essentially the same for all n , then he *can* infer the statement that every a_n is an ordinal

notation. As this would enable him to immediately deduce the well-foundedness of an ordered set isomorphic to Γ_0 , advocates of the Γ_0 thesis have a crucial burden to explain why he cannot in fact do this. Yet the handful of attempts to establish this point that appear in the literature are brief, vague, and, I will argue, simply unpersuasive.

I now turn to the systems introduced in §1.1.

1.4. The finitary autonomous systems. The initial idea behind the finitary autonomous systems Σ , H , and R is that if a predicativist trusts some formal system for second order arithmetic, say ACA_0 (see [43]), then he should accept not only the theorems of the system itself, but also additional statements such as $\text{Con}(\text{ACA}_0)$ which reflect the fact that the axioms are true. Feferman [9] analyzed several such “reflection principles” and found the strongest of them to be the formalized ω -rule schema

$$(\forall n) [\text{Prov}(\ulcorner \mathcal{A}(\bar{n}) \urcorner) \rightarrow \mathcal{A}(n)],$$

where $\ulcorner \mathcal{A}(\bar{n}) \urcorner$ is the Gödel number of $\mathcal{A}(\bar{n})$ and Prov formalizes “is the Gödel number of a provable formula” (here, provable in ACA_0).

Having accepted this schema, the argument runs, the predicativist is then committed to a stronger system consisting of ACA_0 plus the ω -rule schema, and he should therefore now accept a version of the formalized ω -rule schema which refers to provability in this stronger system. This process can be transfinitely iterated, yielding a family of formal systems S_a indexed by Church-Kleene ordinal notations a . Kreisel’s idea [27] was that a predicativist should accept the system indexed by a when and only when he has a prior proof that a is an ordinal notation.³

Feferman [10] proved that when this procedure is carried out starting with a reasonable base system S_0 , Γ_0 is the smallest ordinal with the property that there is no finite sequence of ordinal notations a_1, \dots, a_n with a_1 a notation for 0, a_n a notation for Γ_0 , and such that S_{a_i} proves that a_{i+1} is an ordinal notation ($1 \leq i < n$). Thus, Γ_0 is the smallest predicatively non-provable ordinal.

There are two fundamental problems with this analysis. The first is that the plausibility of inferring soundness of S_a from the fact that a is an ordinal notation hinges on our conflating two versions of the concept “ordinal notation” — supports transfinite induction for arbitrary sets versus supports transfinite induction for arbitrary properties — which are *not* predicatively equivalent. What we actually prove about a is that, for a given partial order \prec on a subset of ω , if X is a set with the property that

$$(\forall b) [(\forall c \prec b)(c \in X) \rightarrow b \in X]$$

then every $b \prec a$ must belong to X . Classically this entails that for every formula \mathcal{A} the statement

$$(\forall b) [(\forall c \prec b) \mathcal{A}(c) \rightarrow \mathcal{A}(b)]$$

implies $\mathcal{A}(b)$ for all $b \prec a$ because we can use a comprehension axiom and reason about the set $X = \{b : \mathcal{A}(b)\}$. Predicatively this should still be possible if, for example, \mathcal{A} is arithmetical, but not in general. Now the statement $\mathcal{P}(b) \equiv$ “if $\text{Prov}_{S_b}(\ulcorner \mathcal{A} \urcorner)$ then \mathcal{A} , for every formula \mathcal{A} ” is not only not arithmetical, it cannot even be formalized in the language of second order arithmetic. So we should not expect there to be any obvious way to *predicatively* infer $\mathcal{P}(a)$ from what we have proven about a . Indeed, there are good reasons to suppose that this inference is not legitimate, for instance the fact that S_a proves the existence of arithmetical

jump hierarchies up to a , which is formally stronger than the fact that transfinite induction holds up to a for sets.

One may be tempted to dismiss this first objection as technical and to grant that predicativists can make the disputed inference, but that leads to a second basic problem: if a predicativist could somehow infer the soundness of S_a then he actually ought to be able to infer more. This point was made well by Howard [24]. I would put it this way: according to Kreisel, a predicativist is (somehow) always able to make the deduction

$$\text{from } I(\bar{a}) \text{ and } \text{Prov}_{S_a}(\ulcorner \mathcal{A}(\bar{n}) \urcorner), \quad \text{infer } \mathcal{A}(\bar{n}), \quad (*)$$

where $I(a)$ formalizes the assertion that a is an ordinal notation. Shouldn't he then accept the assertion

$$(\forall a)(\forall n) [I(a) \wedge \text{Prov}_{S_a}(\ulcorner \mathcal{A}(\bar{n}) \urcorner) \rightarrow \mathcal{A}(n)] \quad (**)$$

for any formula \mathcal{A} ?

As a straightforward consequence of [10], one can use $(**)$ to prove $I(\bar{a})$ with some standard notation for Γ_0 .⁴ The claim must therefore be that a predicativist can recognize each instance of $(*)$ to be valid but cannot recognize the validity of the general assertion $(**)$. In other words, whenever he has proven that a is an ordinal notation he can infer the statement that all theorems of S_a hold, but he does *not* accept the statement “if a is an ordinal notation then all theorems of S_a hold.” Why not?

(a) Kreisel's first answer. Kreisel addresses this point in [27]. He writes:

Here, too, though each extension is predicative provided $<$ has been recognized by predicative means to be a well-ordering, the general extension principle is not since [it is framed in terms of] the concept of predicative proof [which] has no place in predicative mathematics. ([27], p. 297; see also p. 290)

Although this comment sounds authoritative, it does not hold up under scrutiny because in whatever sense it could be said that $(**)$ presumes the concept of predicative proof, the same is true of any instance of $(*)$. If we had no concept of proof or validity then we ought not be able to make the inference $(*)$ in any instance. One can try to read something more subtle into Kreisel's comment, but I have not found any way to elaborate it into a convincing argument. Perhaps the best attempt appears in §1.4 (b) below.

Similar reasoning would actually support a more severe conclusion. Consider:

Although the inference of \mathcal{B} from \mathcal{A} is predicative provided \mathcal{A} has been recognized by predicative means to imply \mathcal{B} , the general principle of modus ponens is not since it is framed in terms of the concept of predicative truth, which has no place in predicative mathematics.

This is a parody, but not a gross one. In fact, I do not really see what could make one accept the first statement and not the second. (The rejoinder that modus ponens is not framed specifically in terms of *predicative* truth misses the point. To a predicativist, “truth” and “predicative truth” are the same thing, so it would not make sense to suggest that he can reason about truth but not about predicative truth.) If one did accept the second statement, of course, this would prevent any use of reflection principles since absent a general grasp of modus ponens the mere

acceptance of a set of axioms would not entitle one to globally infer the truth of all theorems provable from those axioms.

(b) *Kreisel's second answer.* A second argument in response to something like the objection raised above was made by Kreisel ([31], §3.631) and cited with approval by Feferman ([11], p. 134). Unfortunately, the cited passage is rather inscrutable, so it is hard to be sure what Kreisel had in mind. I think it is something like this. Predicativists are at any given moment only able to reason about those subsets of ω that have previously been shown to exist. The “basic step” of predicative reasoning is thus the passage from one level N_α of the ramified hierarchy over ω to the next ($N_{\alpha+1}$ = the subsets of ω definable by second order formulas relativized to N_α). Now the proof that (a notation for) γ_{n+1} is well-founded uses only sets in N_{γ_n} , so once N_{γ_n} is available this proof can be executed and one can pass to $N_{\gamma_{n+1}}$. However, we cannot go directly from N_{γ_n} to $N_{\gamma_{n+2}}$ since the proof that γ_{n+2} is well-founded uses sets in $N_{\gamma_{n+1}}$ which are not yet available. Thus, we cannot grasp the validity of the sequence of proofs as a whole since later proofs involve the use of sets that are not known to exist at earlier stages. Each individual proof is admissible, however, since there is a finite stage in the reasoning process at which the sets needed for that proof become available.

This neatly answers the question raised in §1.3 as to how each proof could be recognized as valid while the entire sequence of proofs cannot. But wait. Exactly how would one use the well-foundedness of γ_{n+1} proven using sets in N_{γ_n} to “pass to $N_{\gamma_{n+1}}$ ” and make sets at that stage available for future proofs? If we accept Kreisel's premise then it would seem that we cannot directly go even two levels up from N_{γ_n} to $N_{\gamma_{n+2}}$, let alone all the way to $N_{\gamma_{n+1}}$, because the construction of $N_{\gamma_{n+2}}$ uses sets in $N_{\gamma_{n+1}}$ which are not yet available. Thus, the argument that prevents us from getting up to Γ_0 should be equally effective at preventing us from getting from γ_n to γ_{n+1} .

This point may become clearer if we ask how a predicativist could establish the existence of N_ω . Starting with $N_0 = \emptyset$, he can use the basic step to directly pass to N_1 , then to N_2 , and so on, so that for each $n \in \omega$ he can give a finite proof of the existence of N_n . But in order to accept the existence of N_ω he has to somehow globally grasp that N_n exists for all n *without* sequentially proving their existence one at a time. Presumably he can accomplish this by recognizing the general principle that the existence of N_{n+1} follows from the existence of N_n and then making an induction argument. So evidently in this case he *can* accept the validity of the sequence of proofs as a whole despite the fact that later proofs involve the use of sets that are not known to exist at earlier stages. That is, *just getting up to N_ω already requires some ability to reason hypothetically about sets that are not yet available.* So Kreisel's argument (if this really is what he meant) appears to make little sense.

However, this entire discussion is speculative until we are told precisely why the proof that γ_{n+1} is well-founded is supposed to legitimate passage to $N_{\gamma_{n+1}}$. This takes us to Kreisel's final argument.

(c) *Kreisel's third answer.* Kreisel's most sophisticated analysis appears in [32]. Here he rightly addresses the central question of exactly how a predicativist would infer soundness of S_a once $I(a)$ has been proven. On my reading, the novel idea is that this inference (or something like it) would not be based on genuinely “understanding” the well-ordering property of a , which he now denies a predicativist

could do, but instead would be directly extracted from the structure of the proof of $I(a)$. If P is the property “the (formal) definitions at a level of the [ramified] hierarchy considered are understood if our basic concepts are understood” ([32], p. 498), then

Since we do not have an explicit definition for P . . . it seems reasonable to suppose that the *formal derivation of the well-foundedness of β is needed* . . . specifically, we expect to use the derivation as a (naturally, infinite) *schema* which need be applied only to instances of P whose meaning *is* determined at stage α . ([32], pp. 498-499; italics in original)

He adds in a footnote: “It seems likely that the work of Feferman and Schütte ‘contains’ all the formal details needed . . . the principal problem is conceptual: to formulate properly just what details are needed.”

It seems even more reasonable to suppose that if, ten years after his first attempt (in [27]) to refute the objection about (**), Kreisel is still not sure how to do this, then the objection is probably valid. Here he gives us not a fully realized refutation, but merely a speculation as to how one might be obtained. I do not think any attempt of this type is likely to succeed for the reasons discussed at the beginning of this section, in particular the fact that S_a proves the existence of arithmetical jump hierarchies up to a and this seems not to be predicatively entailed by $I(a)$ (cf. the end of § 2.4). Moreover, even if one could work out some way of converting formal derivations of well-ordering in the autonomous systems into informal verifications of soundness in some metatheory, then presumably the metatheory and the conversion process could be formalized, and then a predicativist should be able to apply a single instance of the formalized ω -rule to the metatheory and deduce (**) as a general principle. But again, this discussion is hypothetical.

(d) *Feferman’s position.* In [13] Feferman raises a version of the objection and notes that it “involve[s] the ordinal character of the proposal via progressions, hence [does] not apply to $P [+ \exists/P]$ ” ([13], p. 85). Similar comments appear in ([14], p. 3) and ([18], p. 24). It is certainly true that the systems $P + \exists/P$, $\text{Ref}^*(\text{PA}(P))$, and $\mathcal{U}(\text{NFA})$ do not presume any special ability to reason using well-ordered sets. However, Feferman nowhere openly repudiates the earlier systems, and I read his remark in [14] as implying that the later systems are merely more “perspicuous” than the earlier ones because they do not assume that predicativists have any understanding of ordinals. As far as I know he has never addressed the argument that a grasp of ordinals sufficient to justify (*) would also justify (**) and hence lead one beyond Γ_0 .⁵

1.5. The infinitary autonomous systems. In order to evaluate the infinitary (semiformal) autonomous systems we must first clarify in exactly what way these systems are supposed to model predicative reasoning. Surely they are not meant to be taken literally in this regard. Perhaps we can conceive of an idealized predicativist living in an imaginary world who is capable of actually executing proofs of transfinite length, but in this case the allowed proof lengths would merely depend on the nature of the imagined world, not on which well-ordering statements the predicativist is able to prove.

Presumably the infinitary autonomous systems are meant to be taken as modelling what an actual predicativist would consider a valid but idealized reasoning

process. In other words, the predicativist does not himself reason within any of these systems, but he believes that in principle they would prove true theorems if they could somehow be implemented (in some imaginary world). On this interpretation the fact that an ordinal α is autonomous within one of these systems could lead a predicativist to accept the well-foundedness of (some notation for) α *only if he knew this fact*. However, the only way he could know which ordinals are autonomous is via some kind of meta-argument about what is provable in the given system. This immediately suggests that he should be able to get beyond Γ_0 by performing a single act of reflection on the finitary (formal) system in which he actually reasons.

We can now see that just as in the case of the finitary autonomous systems one is faced with a dilemma: (1) why should a predicativist believe that the fact that some set of order type α is well-founded renders proof trees of height α valid, and (2) granting that he can draw this inference for any particular α , why does he fail to grasp that it is valid in general? The inference superficially seems reasonable because it is classically valid, but it is hard to imagine what its predicative justification could be. It is even harder to believe that a predicativist could recognize its validity in each instance but not as a general rule.

Working only in Peano arithmetic, a predicativist should be able to draw conclusions about what is provable in some infinitary system using proof trees of various heights. But in order to infer which proof trees are actually valid, he needs some new principle going beyond Peano arithmetic. Augmenting PA with an axiom schema which expresses the principle “if a well-founded proof tree proves \mathcal{A} , then \mathcal{A} ” in some form would yield a system which proves the well-foundedness of a notation for Γ_0 . Expressing the principle as a deduction rule schema rather than as an axiom schema would yield a system which proves the well-foundedness of notations for all ordinals less than Γ_0 but not Γ_0 itself, but we would need to explain why the deduction rules are valid while the corresponding axioms are not (and we would still be able to get beyond Γ_0 by a single act of reflection). One is virtually forced to assert that whenever a predicativist proves that a set is well-founded he is then able to infer the validity of proof trees of that height via an unformalizable leap of intuition, but I see no reasonable basis for such a claim.

1.6. The linked systems P and \exists/P . $P + \exists/P$ can be criticized in three different ways.

(a) *Obscure formulation.* The central feature of $P + \exists/P$, its division into two distinct but interacting formal systems, is so unusual that it would seem to call for an especially careful account of the underlying motivation. Although [13] contains a substantial amount of prefatory material, there is no explicit discussion of this seemingly crucial point. One gets a vague sense that part of the motivation is to allow use of second order quantifiers only during brief excursions into the “auxiliary” system \exists/P as a sort of next-best alternative to prohibiting them altogether, but nothing is said about why exactly this degree of usage is deemed acceptable. This makes it difficult to evaluate $P + \exists/P$, since one is left with the basic question of how we are supposed to regard the predicative meaning and reliability of statements proven in P as opposed to those proven in \exists/P .

There apparently is some basic distinction to be made between the conceptual content of the theorems of the two systems. I infer this from the requirement

both in the description of allowed formulas of \exists/P ([13], p. 76) and in the rules IV and V ([13], p. 78) that at least part of the premise specifically be proven in P . For instance, the functional defining axioms (IV) allow the introduction of a functional symbol provided existence of the functional has been proven in \exists/P and its uniqueness has been proven in P . Existence can only be proven in \exists/P since P lacks the necessary quantifiers, but no reason is given why uniqueness must be proven in P . Would a proof of uniqueness in \exists/P be unreliable in some way? If so, why should we trust other theorems of this system? *Why are we able to justify introducing a functional symbol when the functional's uniqueness has been proven in P , but not when its uniqueness has been proven in \exists/P ?*

The question is significant because an identical point can be made in the two other cases, and if they were all broadened to include premises proven in \exists/P then the system P would become superfluous: all reasoning could take place in \exists/P . Agreeing that P is indeed dispensable is not acceptable, since this would obviate the need for the functional defining axioms altogether and thereby void Feferman's justification for not allowing a predicativist to reflect on the validity of $P + \exists/P$ (see §1.6 (c)).

(b) *Too strong.* It is also unclear how to reconcile the proposed formalism surrounding second order existential quantification with the motivating idea that

we have partial understanding of 2nd order existential quantification, for example when a function or predicate satisfying an elementary condition is shown to exist by means of an explicit definition. Some reasoning involving this partial understanding may then be utilized, though 2nd order quantifiers are not to be admitted as logical operators in general. ([13], p. 71; italics in original)

For example, this intuition seems somewhat incompatible with the use of *negated* second order existential quantifiers, which are allowed in \exists/P . Even more problematic is the proof of transfinite recursion over well-ordered sets ([13], pp. 82-83), which conflicts rather severely with any understanding of second order existence in terms of “explicit definition”. The offending aspect of this proof is its use of predicate substitution with a Σ_1^1 formula, which is hard to reconcile with the idea that only “some” reasoning about “partially understood” second order quantifiers is available.⁶ General freedom to replace set variables with Σ_1^1 formulas seems to imply a *complete* ability to reason abstractly about second order existence.

Feferman mentions the prima facie impredicative nature of his predicate substitution rule (rule V) and responds that

By way of justification for the schema V it may be argued that the (predicative) provability of $\mathcal{B}(X)$ establishes its validity also for properties whose meaning is not understood, just as one may reason logically with expressions whose meaning is not fully known or which could even be meaningless. ([13], p. 92)

But this line of argument would equally well justify full comprehension. Indeed, for any formula $\mathcal{A}(n)$ the predicatively valid statement $(\exists Y)(n \in Y \leftrightarrow n \in X)$ yields $(\exists Y)(n \in Y \leftrightarrow \mathcal{A}(n))$ by predicate substitution. Even if we restrict ourselves to Σ_1^1 formulas \mathcal{A} , we could still infer Π_1^1 comprehension. So the idea that “the predicative provability of $\mathcal{B}(X)$ establishes its validity also for properties whose meaning is not understood” is clearly not acceptable as a general principle as it stands.

(c) *Too weak.* Now consider the general objection of §1.3. Feferman addresses it in the following way:

... this is not a good argument because the functional defining axioms are only given by a generation procedure and the predicative acceptability of these axioms is only supposed to be recognized at the stages of their generation. To talk globally about the correctness of P we have to understand globally the meaning of all functional symbols in P ; there is no stage in the generation process at which this is available ([13], p. 92).

The point here is that $P + \exists/P$ contains a rule which allows one to introduce a symbol for a functional $\vec{\alpha} \mapsto \beta$ once a unique β satisfying some formula $\mathcal{A}(\vec{\alpha}, \beta)$ (with $\vec{\alpha} = (\alpha_1, \dots, \alpha_n)$, and all free variables in \mathcal{A} shown) has been proven to exist for any $\vec{\alpha}$. The α_i and β are predicate variables.

I do not see how the fact that new symbols can be introduced could in itself prevent anyone from grasping the overall validity of the system. Surely, “to talk globally about the correctness of P ” we need only to accept the validity of the functional generating procedure, not necessarily “to understand globally the meaning of all functional symbols” beforehand.⁷

The more substantial question is whether the validity of the functional defining axioms of P might only be recognized in stages. Now it may be possible to imagine a set of beliefs which would lead one to accept the functional defining axioms only at the stages of their generation: perhaps someone could, by brute intuition, accept the validity of a specific functional definition after having grasped an explicit construction of the functional being defined, yet not be able to reason about functional existence in general terms. This seems something like the standpoint of “immediate predicativism” discussed on pp. 73 and 91 of [13]. The problem is that it tends to conflict with rule VII (relative explicit definition) and axiom VIII (unification) ([13], p. 78) of \exists/P , both of which do presume an ability to reason abstractly about second order existence (not to mention rule V, predicate substitution). Thus, Feferman’s argument belies his premise that a predicativist is capable of accepting rule VII and axiom VIII.⁸

1.7. The system $\text{Ref}^*(\text{PA}(P))$. The Ref^* construction applies to any schematic formal theory, but the case of interest for us is schematic Peano arithmetic $\text{PA}(P)$. This is formulated in the language L of first order arithmetic augmented by a single predicate symbol P . The axioms are the usual axioms of Peano arithmetic with the induction schema replaced by the single axiom

$$P(0) \wedge (\forall n)(P(n) \rightarrow P(n')) \rightarrow (\forall n) P(n),$$

and there is an additional deduction rule schema allowing substitution of arbitrary formulas for P . Now if $S(P)$ is any schematic theory then $\text{Ref}^*(S(P))$ is a theory in the language of $S(P)$ augmented by two predicate variables T and F whose axioms are the axioms of $S(P)$ together with “self-truth” axioms governing the partial truth and falsehood predicates T and F , and with a substitution rule which allows the substitution of formulas possibly involving T and F for P .

(a) *Too strong.* In §2.5 and §2.6 I will discuss the prima facie impredicativity both of self-applicative truth predicates and of schematic predicate variables. Leaving those issues aside for now, the first point to make here is that the key axiom which

distinguishes $\text{Ref}^*(\text{PA}(P))$ from a much weaker system $\text{Ref}(\text{PA}(P))$, axiom 3.2.1 (i)^(P) ([14], p. 19), has no obvious intuitive meaning. The reason for using schematic formulas, as opposed to ordinary second order formulas involving set variables, is that they are supposed to allow one to fully express principles such as induction without assuming any comprehension axioms ([14], p. 8). This means that we interpret a statement involving a schematic predicate symbol P not as making an assertion about a fixed arbitrary set, but rather as a sort of meta-assertion which makes an open-ended claim that the statement will be true in any intelligible substitution instance. However, the truth claim

$$T(\ulcorner P(\bar{n}) \urcorner) \leftrightarrow P(n),$$

a special case of 3.2.1 (i)^(P), cannot be given the latter interpretation since the number $\ulcorner P(\bar{n}) \urcorner$ does not change when a substitution is made for P in this formula. If we interpret P in a way that is compatible with 3.2.1 (i)^(P), i.e., as indicating membership in a fixed set, then the substitution rule $P - \text{Subst} : L(P)/L(P, T, F)$ ([14], Definition 3.3.2 (iii)) can only be justified by an impredicative comprehension principle (cf. [14], p. 8).

Feferman characterizes axiom 3.2.1 (i)^(P) as “relativizing T and F to P ” ([14], p. 19). I am not sure what this means, but the axiom clearly is not valid on arbitrary substitutions for P , yet one draws consequences from it to which one does apply a substitution rule (and this is crucial for the proof that $(\Pi_1^0 - \text{CA})_{<\Gamma_0} \leq \text{Ref}^*(\text{PA}(P))$). In fairness, I should point out that this problem is noted in ([14], §6.1.3 (i)), with the comment that “a fall-back line of defense could be that this substitution accords with ordinary informal reasoning. However, this seems to me to be the weakest point of the case for reflective closure having fundamental significance.”

I would argue that the above difficulty not only invalidates the idea that $\text{Ref}^*(\text{PA}(P))$ models predicative reasoning, it shows that the Ref^* construction indeed has no fundamental significance. There is no way to interpret the P symbol that simultaneously makes sense of the axiom 3.2.1 (i)^(P) and the substitution rule $P - \text{Subst} : L(P)/L(P, T, F)$.

(b) Too weak. The Ref^* construction is described in [14] as a “closure” operation and the question of its significance is discussed in terms of Kripke’s theory of grounded truth outlined in [33]. A casual reading of §6 of [14] might leave the impression that the statements \mathcal{A} such that $\text{Ref}^*(\text{PA}(P))$ proves $T(\ulcorner \mathcal{A} \urcorner)$ are supposed to be precisely the grounded true statements of the language $L(P, T, F)$. But this cannot be right because these statements are recursively enumerable, so that one can write a formula $(\forall n) T(\{\bar{n}\})(n)$ which asserts precisely their truth. This formula is grounded (in any reasonable sense) and true but the assertion of its truth is not a theorem of the system.

The more careful formulation that the self-truth axioms “correspond directly to the informal notion of grounded truth and falsity” ([14], p. 42) is well-taken, but we must not confuse this with the claim that the self-truth axioms *capture* the informal notion. Their failure to do so can be traced to axiom (vi) ([14], Definition 3.2.1). The first conjunct of this axiom, for example, asserts that if $\mathcal{A}(\bar{n})$ is true for every n then $(\forall n) \mathcal{A}(n)$ is true. But this does not fully capture the informal idea that “the truth of $\mathcal{A}(n)$ for all n implies the truth of $(\forall n) \mathcal{A}(n)$ ” in the sense that there exist formulas $\mathcal{A}(n)$ which can be proven true for each numerical value of n but such

that there is no proof in $\text{Ref}^*(\text{PA}(P))$ that for all n , $\mathcal{A}(\bar{n})$ is true — in particular, $\mathcal{A}(n) \equiv T(\{\bar{r}\}(n))$ where $\{r\}$ enumerates all a such that $\text{Ref}^*(\text{PA}(P))$ proves $T(\bar{a})$.

Now consider the claim that in general $\text{Ref}^*(S(P))$ encapsulates what one “ought to accept” given that one has accepted $S(P)$ ([14], p. 2). This has an air of paradox since one has to ask whether the claim itself is something that anyone ought to accept. However, that point is not crucial to the question of what a predicativist can prove since it need not attach to the specific assertion that $\text{Ref}^*(\text{PA}(P))$ encapsulates what one ought to accept given that one has accepted $\text{PA}(P)$. We may suppose that the predicativist does not realize (and indeed, *ought not accept*) that his commitment to Peano arithmetic obliges him to accept every theorem of $\text{Ref}^*(\text{PA}(P))$, although this is in fact the case. This leads us back to the question posed in §1.3. Evidently we are dealing with a claim that predicativists can affirm each theorem of $\text{Ref}^*(\text{PA}(P))$ individually but cannot accept this system globally.

This point is not explicitly addressed in [14], but the informal notion of “partial truth” has the flavor of a forever incomplete process and might seem like it could support such a claim. For example, it is suggested in §6.1.1 of [14] that the passage from $S(P)$ to $\text{Ref}(S(P))$ should not be iterated because this would “vitate the informal idea behind the use of partial predicates of truth and falsity.” A possibly more straightforward question which avoids the issue of using multiple partial truth predicates is whether one could justify augmenting $\text{Ref}^*(\text{PA}(P))$ by the single statement $(\forall n)T(\{\bar{r}\}(n))$ described above.

Surely a predicativist *can* justify adding this statement if he is able to generally recognize that every statement proven true by $\text{Ref}^*(\text{PA}(P))$ is indeed true. Given that $\text{Ref}^*(\text{PA}(P))$ is finitely axiomatized and that the predicativist is presumed to accept each theorem of $\text{Ref}^*(\text{PA}(P))$ individually, it is unclear how this could be plausibly denied. Indeed, axiom (vi) clearly affirms that the predicativist *is* able to reason about the collective truth of an infinite set of statements themselves involving assertions of truth and falsehood.

In §6.1.3 (ii) of [14] Feferman considers the question “have we accepted too little?” in terms of logically provable statements, e.g. of the form $\mathcal{A} \vee \neg \mathcal{A}$, whose truth is not provable because \mathcal{A} is not grounded. This leads into a brief discussion of the relative merits of Kripke’s minimal fixed point approach versus van Fraassen’s more liberal “supervaluation” approach to self-applicative truth. But this discussion is misleading because $\text{Ref}^*(\text{PA}(P))$ does not even prove the truth of every statement in Kripke’s minimal fixed point; in particular, if this needs repeating, it does not prove the statement $(\forall n)T(\{\bar{r}\}(n))$. This formula is not logically provable but it is grounded, and it seems a rather stronger candidate for a statement that “ought” to be accepted as true.

1.8. The system $\mathcal{U}(\text{NFA})$. Distinct, not obviously equivalent, versions of $\mathcal{U}(\text{NFA})$ are presented in [16] and [19]. I give priority to the later version in [19].

(a) *Too weak.* Like Ref^* , \mathcal{U} is presented in [19] as a general construction (“unfolding”) which can be applied to any schematic formal system $S(P)$. As usual, granting that acceptance of $S(P)$ justifies acceptance of every theorem of $\mathcal{U}(S(P))$, we can ask why it fails to justify accepting a formalized ω -rule schema referring to theorems of $\mathcal{U}(S(P))$. This question is not addressed in either [16] or [19]; the closest I can find to an answer is the following passage in [16]:

[W]e 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 able 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 . . . 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. ([16], p. 10)

Here reference is made to the fact that \mathcal{U} uses partial operations, which is seen as having fundamental significance. It is true that the question of which partial operations of $\mathcal{U}(\text{NFA})$ are total is (unsurprisingly) not decidable, though this in itself seems a questionable basis for forbidding us from proceeding beyond $\mathcal{U}(\text{NFA})$ when it did not prevent us from formulating this system in the first place or from working within it.

If $S(P)$ involves no basic objects of type 2 (as is the case for NFA) then an argument could be made that applying the \mathcal{U} construction twice is conceptually different from applying it once in that $\mathcal{U}(S(P))$ does employ higher type objects and thus the original system $S(P)$ can possibly be seen as being “concrete” in a way that $\mathcal{U}(S(P))$ is not. However, this should not prevent one from accepting a formalized ω -rule schema applied to $\mathcal{U}(\text{NFA})$, which would seem to require only that one accept $\mathcal{U}(\text{NFA})$ is sound.⁹

(b) *Way too strong.* $\mathcal{U}(\text{NFA})$ is actually flatly impredicative in two distinct ways. First, the \mathcal{U} construction suffers from the same nonsensical treatment of schematic predicates as Ref^* . Here the offending axiom is Ax 7 ([19], p. 82), which does not make sense if P is understood as a schematic predicate. It is valid if we regard P as indicating membership in a fixed set, but then, just as for $\text{Ref}^*(\text{PA}(P))$, use of the substitution rule (Subst) ([19], p. 82) would have to presume impredicative comprehension.

The really striking impredicativity of $\mathcal{U}(\text{NFA})$, however, is its use of a least fixed point operator, which apparently informally assumes the legitimacy of generalized inductive definitions in the sense of [7]. This not only vitiates any claim of $\mathcal{U}(\text{NFA})$ to model predicative reasoning, it more broadly undermines the idea that $\mathcal{U}(\text{NFA})$ has any fundamental philosophical significance, since it would seem that anyone who accepts the \mathcal{U} construction and Peano arithmetic ought to at least accept ID_1 [7], which is far stronger than $\mathcal{U}(\text{NFA})$.¹⁰

1.9. Summary of the critique. At the beginning of this section I made strong claims about the weakness of the case for the Γ_0 thesis. Were they borne out?

First, I stated that each of the formal systems of §1.1 is motivated by informal principles which actually justify a stronger system that proves the well-foundedness of an ordered set that is isomorphic to Γ_0 . In the case of Σ , H , and R , the informal principle is “ a is an ordinal notation implies S_a is sound”, which is needed to justify (*) but in fact justifies (**) (see §1.4). In H^+ , R^+ , and RA^* the principle is “ a is an ordinal notation implies proof trees of height a are sound”. In $P + \exists/P$ we accept that it is legitimate to substitute arbitrary predicates for set variables, which justifies full comprehension. $\text{Ref}^*(\text{PA}(P))$ assumes an informal grasp of a

self-applicative concept of truth, which justifies the inference of a statement that asserts the truth of every theorem proven true by $\text{Ref}^*(\text{PA}(P))$. $\mathcal{U}(\text{NFA})$ informally assumes the legitimacy of generalized inductive definitions, which actually justifies ID_1 .

Second, I stated that the responses on record to the objection in §1.3 are brief, vague, and unpersuasive. The only such responses of which I am aware are Kreisel's answers in [27], [31], and [32] (see §1.4 (a), (b), (c)) and Feferman's answer about $P + \exists/P$ in [13] (see §1.6 (c)). In [27] and [13] the response is barely more than a flat assertion with no real explanation given; in [31] it is a cryptic passage whose most reasonable interpretation is clearly self-defeating; and in [32] it is merely an implausible speculation. With regard to $\text{Ref}^*(\text{PA}(P))$ and $\mathcal{U}(\text{NFA})$, as far as I am aware the objection has not even been discussed in the literature, except tangentially by an argument in [14] that the Ref^* construction should not be iterated.

Finally, I said that each of the formal systems of §1.1 is manifestly impredicative in some way. The most blatant example of this is the least fixed point operator in $\mathcal{U}(\text{NFA})$, but in all three of $P + \exists/P$, $\text{Ref}^*(\text{PA}(P))$, and $\mathcal{U}(\text{NFA})$ there is a basic impredicativity involving the ability to substitute possibly meaningless formulas for free set variables. In $\text{Ref}^*(\text{PA}(P))$ and $\mathcal{U}(\text{NFA})$ this is hidden by employing a substitution rule involving a "schematic" predicate symbol, but elsewhere treating this predicate symbol in a way that only makes sense if it is thought of as a classical predicate indicating membership in a fixed set.¹¹

As I mentioned in §1.2, the reason one needs a substitution rule is because one wants to convert statements of transfinite induction into statements of transfinite recursion so that one can construct iterative hierarchies. In the autonomous systems this is accomplished by simply postulating that a statement of transfinite induction legitimates a transfinite application of reflection principles, which allows one to pass to a stronger system that proves the existence of the next hierarchy. Thus, *every* system uses an impredicative step to get from transfinite induction to transfinite recursion. This is not surprising, as there is a predicatively essential difference between induction and recursion (see the end of §2.4).

There is still a question as to why so many different attempted formalizations of predicative reasoning happen to have proof-theoretic ordinal Γ_0 . One answer is that they all employ essentially the same well-ordering proof and to a substantial extent appear to have been *built around* different versions of this proof. (This is especially evident in the case of $P + \exists/P$.) Another possible answer is that the earlier systems all access the same ordinals because they embody the same fallacies revolving around the idea of autonomous generation of ordinals, while the later systems were formulated against the background of the earlier systems which had already seemed to attain the correct answer. This could have made it difficult to free oneself from a conclusion that had already been formed and seemed well-justified. However, a properly functioning scientific community should be expected to debate and criticize major ideas, not to passively accept them, regardless of the stature of their author and the complexity of the argument. That this apparently was not done in a serious way in the present case suggests that the community as a whole did not function as it should have.

2. AN ANALYSIS OF PREDICATIVE PROVABILITY

As I discussed in §1.2 and §1.9, all of the formal systems of §1.1 employ impredicative methods in order to pass from transfinite induction to transfinite recursion. This presents a basic obstacle to obtaining predicative ordinals by means of the general technique employed by those systems. Our goal in this section is to develop new methods of producing predicative well-ordering proofs.

Considering the variety of impredicative features that have appeared in previous attempts to model predicative reasoning (and there are two other major ones besides those I discussed in Section 1; see §2.5 and §2.6), it seems fair to say that not enough attention has been paid to the basic conceptual content of predicativism. Therefore our discussion will incorporate a general conceptual analysis of predicativist principles.

As many commentators have noted, the vicious-circle principle — generally taken as the defining principle of predicativism — does not in itself constitute a well-defined foundational stance, as it is compatible with a variety of attitudes about which principles of set construction should be accepted as basic. Indeed, the version of predicativism under consideration here is nowadays often referred to as “predicativism given the natural numbers”, a phrasing which I find unfortunate, as it gives no indication as to why one should take the natural numbers as basic, as opposed to any other set. I take the essential basis for this view to be a conception of mathematical reality according to which sets have no independent prior existence but must be constructed, together with the idea that infinite constructions are legitimate, but only if one is able to conceive them in a completely precise and explicit way. The special role that the natural numbers play in this account arises from the fact that “we have a complete and clear mental survey of all the objects being considered, together with the basic [order] interrelationships between them” ([13], p. 70). Elsewhere I term this stance *mathematical conceptualism* and argue in support of it [46]. For our purposes here we may encapsulate it in three basic principles:

- (i) the mathematical universe is a variable entity that can always be enlarged
- (ii) every set must be constructible from logically prior sets
- (iii) constructions of length ω are legitimate.

I take it as granted that these principles constitute a coherent foundational stance and I am not concerned here with trying to justify them. However, a brief explanation is in order. The idea of (i) is that there is no well-defined complete universe of sets because any particular collection of sets can itself be identified as a set that does not belong to the collection. Any currently available partial universe can always be extended, and this extension process can be iterated; since we accept constructions of length ω it can even be iterated transfinitely. But the general concept of extension cannot be fully formalized since one can always go one step beyond any given partial formalization.

Assertion (ii) is an informal version of the vicious circle principle. Although that principle is notoriously difficult to formulate precisely (in particular, we do not attempt to define “logically prior”), the underlying intuition seems clear enough to be used in most cases in evaluating whether a proposed formal system is predicatively acceptable.

The idea behind (iii) is that we have an intuitively clear conception of what it would be like to carry out a process (construction, computation, proof) of length ω and therefore the results of such procedures are legitimate objects of study. By iteration we can accept processes of length ω^2 , ω^ω , etc., but for now we leave open just what the “etc.” entails. As with (i) we do not expect to be able to fully formalize exactly how far we can go.

2.1. The power set of ω . Unlike naive set theory, predicativism obviously does not support the principle “for any property \mathcal{A} of sets, $\{X : \mathcal{A}(X)\}$ exists”. Indeed, the intuitive appeal of this general comprehension principle seems to rest on an implicit belief in the existence of a well-defined complete universe of sets. If there were such a universe V , then for any definite property \mathcal{A} one might have some idea of forming $\{X : \mathcal{A}(X)\}$ by extracting from V just those sets satisfying \mathcal{A} . But if we reject the existence of such a universe then this idea fails, and in fact for many properties \mathcal{A} (e.g., “ $0 \in X$ ”) we clearly *cannot* imagine any way to form the set of all X ’s which satisfy \mathcal{A} . It is important to understand that this can be so even if \mathcal{A} is *definite* in the sense that $\mathcal{A}(X)$ is recognized to have a well-defined truth value for any conceivable X .

Thus, we do not accept a set as legitimate if it can only be defined “from above” in the form $\{X : \mathcal{A}(X)\}$. We do accept sets which can be defined by restricted comprehension (i.e., are of the form $\{X \in Y : \mathcal{A}(X)\}$) relative to a set which has already been accepted, provided the property \mathcal{A} is definite in the sense just indicated,¹² and we also accept sets which can be constructed “from below”. Principle (iii) gives us a powerful ability to build up countable sets from below, but we would not expect that we can predicatively reach any classical uncountable set in this manner. Indeed, on the basis of the principles enumerated above it is reasonable to assume that every predicatively acceptable infinite set should be not only classically but predicatively countable (i.e., predicatively known to be in bijection with ω).

For our purposes here we need not insist on such an “axiom of countability”, but we have to accept that, at the very least, we cannot assume that the power set of ω is a predicatively legitimate set. As explained above, this does *not* contradict the definiteness of the property “ X is a subset of ω ”, and indeed the latter could be justified by an appeal to (iii). Given any set X , an informal “computation” of length $\omega + 1$ could verify or falsify the claim that $X \subseteq \omega$: for each $n \in \omega$ check whether n belongs to X ; if so, remove it; at step ω check whether any elements remain. Thus, the property of being a subset of ω is predicatively definite.¹³

2.2. Predicatively valid logic. Classical logic is unsuited to reasoning about a variable universe. Since the general extension process by which new sets are recognized cannot be completely formalized, the universe is inherently ill-defined and so we do not expect every assertion about sets to have a well-defined truth value. Rather, we should regard the family of true statements as a variable entity which is always capable of enlargement, much like the mathematical universe itself. *This makes intuitionistic logic the appropriate tool for general predicative reasoning.*

Of course, this is not to say that the predicative notion of truth can be identified with intuitionistic truth. Predicatively there is no reason to believe (and undoubtedly good reason not to believe) that every true statement can be proven by a finite argument. Conversely, for reasons I do not understand, it seems that most

intuitionists accept impredicative constructions. Nevertheless, I maintain that the logical apparatus of intuitionism is exactly suitable for predicativism. To say that the law of the excluded middle always holds is just to say that *every* formula is definite in the sense of §2.1. Predicatively the definiteness of any statement that quantifies over subsets of ω is initially suspect, so it is highly implausible that a predicativist could be led to accept even that all formulas of second order arithmetic are definite.

For a specific example of the presumable failure of the law of the excluded middle, notice that well-ordering assertions can apparently fail to have a well-defined truth value because the inherent ambiguity of the mathematical universe could lead to uncertainty about whether or not a given totally ordered set has a proper progressive subset. If no such subset is currently available, indefiniteness about whether such a subset will appear in some future enriched universe could be a reasonable consequence of the fact that *we do not know how new sets might arise*. An even sharper example is given by the set $S = \{n : \mathcal{A}_n \text{ is true}\}$ where (\mathcal{A}_n) is some recursive enumeration of the sentences of second order arithmetic. The set S is obviously impredicative since it is defined in terms of quantification over the power set of ω , but if we accepted $(\forall n)(\mathcal{A}_n \vee \neg \mathcal{A}_n)$ then the restricted comprehension principle mentioned in §2.1 would allow us to form S . This shows that $\mathcal{A} \vee \neg \mathcal{A}$ must not be assumed to hold in every case.

On the other hand, we do regard statements relativized to any well-defined partial universe as definite, so that any such statement should have a well-defined truth value. For example, in the setting of second order arithmetic, principle (iii) should at least assure us that any arithmetical statement is definitely true or false since we can imagine checking it mechanically. This is so even if the statement contains set variables as parameters, since for any particular $n \in \omega$ and $X \subseteq \omega$ the atomic formula “ $n \in X$ ” has a definite truth value. Thus, at the level of arithmetical statements our logic is classical.

Similar considerations were discussed in [8], leading to the suggestion that predicativists can adopt the *numerical omniscience schema*

$$(\forall n) (\mathcal{A}(n) \vee \neg \mathcal{A}(n)) \rightarrow [(\forall n) \mathcal{A}(n) \vee (\exists n) \neg \mathcal{A}(n)]$$

(where here \mathcal{A} is any formula of second order arithmetic and n is a number variable). Together with the assumption $\mathcal{A} \vee \neg \mathcal{A}$ for every atomic formula \mathcal{A} , this implies the law of the excluded middle for every arithmetical formula.

A word about terminology. If we do not assume the law of the excluded middle then we may have to consider assertions whose sense is understood but which are not known to have a definite truth value. To keep this distinction clear I will say an assertion is *meaningful* if it has a definite truth value and *intelligible* if its sense is understood. Thus, every meaningful assertion must be intelligible and every intelligible assertion is potentially meaningful.

2.3. Second order quantification. First order (numerical) quantification is unproblematic by principle (iii). The legitimacy of second order quantification is less clear since we do not regard the power set of ω as a well-defined entity over which set variables could be imagined ranging. This has been a recurrent concern in the literature on predicativity. For instance, it was cited as motivation for the strong restrictions on second order quantification in [13].

To what extent, if any, are second order quantifiers acceptable? First, because the concept “set of numbers” is definite (§2.1), we should at least be able to make some limited constructive sense of existential quantification. There are situations in which we can recognize that we are (in principle) able to construct a set of numbers with some property, and this should license some use of second order existential quantifiers. This was also the position taken in [13].

In addition, we do seem to be able to predicatively accept some statements as being true of any set of numbers. Despite the unfixed nature of the mathematical universe, we can still affirm general assertions like $0 \in X \vee 0 \notin X$ as holding for any conceivable $X \subseteq \omega$. Not only is this statement true for all currently available sets, it must remain true in any future universe. We can be sure that we will never come across a set for which the assertion fails because its truth is inherent in the concept “set of numbers”. As another example, given any $X \subseteq \omega$, principle (iii) should justify asserting the (constructive) existence of its complement. Thus, we ought to be able to somehow express that for every X there is a Y such that $n \in Y \leftrightarrow n \notin X$. Finally, the principle of induction in the form $0 \in X \wedge (\forall n)(n \in X \rightarrow n' \in X) \rightarrow (\forall n)(n \in X)$ is recognizably true for any $X \subseteq \omega$. Given that we accept processes of length ω , we can be certain that any set which satisfies the induction premise must contain every number since we can imagine verifying this conclusion mechanically. Again, this must hold not only for all currently available sets but for all sets in any conceivable future universe.

In [18], following Russell, a distinction is drawn between the concepts “for all” (ranging over a well-defined collection) and “for any” (ranging over a “potential totality”). I find this distinction helpful, but in the present setting I do not accept Russell’s suggestion, adopted in [13], that the “for any” intuition is captured by using free set variables. Consider the following example. We have already agreed that predicativists can acknowledge that any subset X of ω has a complement Y . But they should then also agree that Y has properties like: for any Z , $Z \subseteq Y$ if and only if $Z \cap X = \emptyset$. Indeed, given any X and Z we can imagine constructing Y (using (iii)) and then verifying the relation between X , Y , and Z (again using (iii), specifically a version of the numerical omniscience schema, together with definiteness of the assertions $Z \subseteq Y$ and $Z \cap X = \emptyset$). Since the construction of Y did not depend on Z this means that we can affirm the statement

$$(\forall X)(\exists Y)(\forall Z)(Z \subseteq Y \leftrightarrow Z \cap X = \emptyset)$$

under the interpretation $\forall =$ “for any” and $\exists =$ “there can be constructed a”. This shows that alternating second order quantifiers can make predicative sense. Moreover, the idea *cannot* be expressed without using at least one universal quantifier, which shows that Russell’s free variable suggestion is inadequate here.

An alternative possibility is to allow use of both universal and existential second order quantifiers and to reason using an intuitionistic predicate calculus. Given the conception of predicativism developed above and the interpretation of second order quantifiers just indicated, this logical apparatus appears perfectly acceptable. *Intuitionistic logic legitimates the predicative use of set quantifiers.*

I have been careful to restrict this discussion to sets of numbers. Using quantifiers to range over *all sets*, with no qualification, is less tenable because the general concept of a set may be predicatively unclear. To a predicativist, the assertion that some object is a set may not only be indefinite (in the sense of §2.1) but unintelligible (in the sense of §2.2). In fact, I believe that quantification over all

sets is not predicatively valid for just this reason. However, I would still reject Russell's free variable convention here; rather, I would conclude that predicativists simply cannot make general statements about all sets.¹⁴

(There might be ways around this difficulty. For instance, if one interprets sets in terms of well-founded trees coded by subsets of ω , cf. §VII.3 of [43], then the concept may become predicatively intelligible.)

2.4. Predicative well-ordering. In the previous section I justified a second order induction statement using principle (iii). For which formulas \mathcal{A} of second order arithmetic would a similar argument lead us to accept $\mathcal{A}(0) \wedge (\forall n)(\mathcal{A}(n) \rightarrow \mathcal{A}(n')) \rightarrow (\forall n)\mathcal{A}(n)$?

If it contains set variables, the formula $\mathcal{A}(n)$ might not have a definite truth value (§2.2). However, once we have proven $\mathcal{A}(0)$ we must at least agree that this instance is definitely true. If, moreover, we have proven $(\forall n)(\mathcal{A}(n) \rightarrow \mathcal{A}(n'))$ then we can be successively brought to the same conclusion about $\mathcal{A}(1)$, $\mathcal{A}(2)$, etc., and recognizing this, we should therefore accept $(\forall n)\mathcal{A}(n)$ as true. Regarding the family of true statements as a variable entity always capable of enlargement, this shows that predicativists should accept induction for *every* formula $\mathcal{A}(n)$.

This may need further explanation in light of my insistence in §1.6 (b) that it is generally not valid to substitute arbitrary, possibly meaningless, formulas for set variables. I stand on this assertion: for example, $(\forall n)(n \in X \vee n \notin X)$ is predicatively true but $(\forall n)(\mathcal{A}(n) \vee \neg\mathcal{A}(n))$ is presumably not if, e.g., $\mathcal{A}(n)$ asserts that n is a Church-Kleene ordinal notation. However, this does not entail that possibly meaningless formulas can never appear in true statements. A predicativist should accept complete induction (provided he is using intuitionistic logic) since he can generally recognize that the truth of the premise of any induction statement would entail the truth of its conclusion even if the latter was not initially known to be meaningful.

Next let us consider the extent to which predicativists can understand the general concept of a well-ordered set. It is sometimes said that the well-ordering concept is not available to predicativists because it involves quantification over power sets. On the other hand, it seems to be generally accepted that predicativists are able to assert relatively strong versions of the statement that ω is well-ordered. If we agree with the conclusions of §2.3 then statements of transfinite induction of the form $(\forall X)\text{TI}(X, a)$ (transfinite induction up to a on a totally ordered subset of ω) are predicatively intelligible.¹⁵ Here I use the abbreviations

$$\begin{aligned} \text{TI}(X, a) &\equiv \text{Prog}(X) \rightarrow (\forall b \prec a)(b \in X) \\ \text{Prog}(X) &\equiv (\forall b)[(\forall c \prec b)(c \in X) \rightarrow b \in X]. \end{aligned}$$

I argued above that complete induction on ω is predicatively valid. Note, however, that if we know $\{b : b \prec a\}$ is well-ordered, i.e., we have verified $(\forall X)\text{TI}(X, a)$, we cannot in general infer $\text{TI}(\mathcal{A}, a) (\equiv \text{Prog}(\mathcal{A}) \rightarrow (\forall b \prec a)\mathcal{A}(b))$ where $\text{Prog}(\mathcal{A}) \equiv (\forall b)[(\forall c \prec b)\mathcal{A}(c) \rightarrow \mathcal{A}(b)]$ for arbitrary formulas \mathcal{A} . The latter schema is genuinely stronger because $(\forall X)\text{TI}(X, a)$ only asserts induction for *sets* that are by assumption well-defined, whereas $\text{TI}(\mathcal{A}, a)$ can hold if \mathcal{A} is not meaningful, and it can even be used to prove that $\mathcal{A}(b)$ *is* meaningful for all $b \prec a$. It may in fact be the case that whenever there is a predicatively valid proof of $(\forall X)\text{TI}(X, a)$ there is also a proof of $\text{TI}(\mathcal{A}, a)$ for any intelligible formula \mathcal{A} . However, inferring the latter statement from the former seems to me clearly predicatively illegitimate.

2.5. Schematic assertions. Even the complete induction schema does not entirely capture a predicative understanding of induction on ω since it only covers formulas that can be written in the language that is currently in use. If the expressive power of the language were strengthened in a predicatively intelligible way, then a predicativist should accept the induction schema for all formulas of the new language too.

This issue is addressed in [14] and [16] by a proposal to use a “schematic” predicate symbol P and to express the principle of induction in a single schematic formula. Together with an informal commitment to continue to accept all substitution instances of this statement if the language is enriched in any intelligible way, this does seem to fully capture a predicative understanding of induction on ω . However, it seems unlikely that a predicativist could agree to accept such a schematic formulation because of the circularity involved in having a formula $\mathcal{A}(P)$ which contains a schematic predicate symbol P that ranges over a class of formulas that includes $\mathcal{A}(P)$. In fact, without some special justification this usage seems clearly impredicative.

There should be no problem in using a schematic predicate symbol to range over all formulas of a previously accepted language, or even a previously accepted set of languages, as this would present no possibility of circularity. However, because of the inherently impredicative quality of a self-applicative predicate variable it seems to me that the general concept “intelligible predicate” is itself not intelligible and that it is therefore not possible for a predicativist to legitimately make assertions about all intelligible predicates (cf. §2.3). This leads to the conclusion that *predicativists have an open-ended ability to affirm induction statements on ω but are not capable of formally expressing this fact.*

The difficulties involved with schematic predicates shed light on the predicative unacceptability of some formal systems which superficially have a strong predicative flavor. For example, in [28] and [30] the possibility is raised that under intuitionistic logic theories of generalized inductive definitions might be predicatively valid, and this idea does have superficial appeal. However, on close examination there is a clear circularity even in the intuitionistic case. This is seen as follows.

Suppose we want to introduce a predicate symbol for the class defined by some inductive definition. Classically we could define this class “from above” as the intersection of all classes satisfying the relevant closure condition, but this is clearly impredicative. In the intuitionistic setting we instead conceive of the class as an incomplete entity that can always be enlarged by repeatedly applying the closure condition, which seems to be a predicatively legitimate idea. The problem is in verifying the minimality property of this class. Let \mathcal{A} be any formula in the language of first order arithmetic enriched by a predicate symbol I_X which is to represent the class X being defined; assuming \mathcal{A} satisfies the same closure condition as X , we must affirm $(\forall n)(I_X(n) \rightarrow \mathcal{A}(n))$. Now what is immediately clear from our conception of X is that this statement is progressive in the sense that if it holds at all previous stages in the construction of X then it will still hold at the immediately following stage since \mathcal{A} satisfies the same closure condition as X . This suggests that the statement should be verified by a transfinite induction and we must therefore imagine the stages in the construction of X as corresponding to elements of a well-ordered set. The difficulty then lies in specifying what we mean by “well-ordered”. If we had the ability to make assertions like $\text{TI}(P, a)$ where P is a

schematic predicate variable, then we could take “well-ordered” to mean “supports transfinite induction for a schematic predicate”, and we should then be able to carry out the transfinite induction needed to prove minimality. But if the most we can say of any totally ordered set is that it supports transfinite induction for all formulas of a given previously accepted language, then X cannot be conceived as being built up along sets that support transfinite induction for formulas of a language that includes I_X . This would be circular because the well-ordering assertion would refer to the class X which it is being used to define. But proving the minimality statement requires that we be able to carry out transfinite induction for formulas of this language. Hence there is no (or at least no obvious) way to predicatively verify minimality.

Kripke-Platek set theory also has a superficial predicative plausibility, assuming that the point raised at the end of §2.3 does not preclude any predicative treatment of arbitrary sets, but it fails for a similar reason. Namely, the KP foundation schema is impredicative for essentially the same reason that inductive definitions are. For a statement of the KP axioms that is intuitionistically suitable, see, e.g., [5]. Their intuitionistic justification involves a conception of an incomplete universe of sets which is built up in stages. In order to verify any instance of the foundation schema we would therefore need to carry out a transfinite induction with respect to the well-ordered sets along which this universe is being constructed. But the formula being proven by induction is a formula of the language of KP and would implicitly make reference to the universe being defined. Thus, in order to verify the foundation schema we would need to build up the KP universe along sets that are known to be well-ordered with respect to formulas which refer to that universe. Again, this is circular and hence impredicative.¹⁶

2.6. Truth theories. Without using some kind of reflection principle I doubt that predicativists can get beyond ordinals in the neighborhood of γ_2 or γ_3 , at most. In order to progress significantly further we need a systematic way of iterating the process of reflecting on the truth of a given theory to get a slightly stronger theory. One might hope to do this using a self-applicative truth predicate, as in [14]. On its face, the predicative legitimacy of a self-applicative truth theory is problematic — indeed, this seems just the sort of thing that predicativist principles tend to forbid. We can try to get around the prima facie circularity of such a theory by regarding the truth predicate as partial and built up in stages, giving it the flavor of a generalized inductive definition. Now I argued in §2.5 that theories of generalized inductive definitions are impredicative, but the difficulty with such theories is their assertion of minimality axioms, which we do not require of a truth predicate. On the contrary, the concept of belonging to an inductively defined class does not seem predicatively objectionable on its own; for example, according to §2.4 the assertion “ n is a Church-Kleene ordinal notation” is predicatively intelligible. A parallel could also be drawn with the predicative conception of the power set of ω as a necessarily incomplete entity that can always be enlarged. Therefore, it seems that *provided intuitionistic logic is used* self-applicative truth theories could be predicatively justifiable. The systems of [14] are firmly embedded in classical logic, but I suppose it is likely that there is an intuitionistic version of, say, the Ref(PA) construction of [14] that could be accepted as predicatively legitimate. However, such a theory would presumably have proof-theoretic ordinal only in the

neighborhood of γ_2 or γ_3 . So self-applicative truth theories do not seem a promising route to obtaining strong predicative well-ordering proofs.

Perhaps surprisingly, I find that it is possible to predicatively prove relatively strong well-ordering assertions using hierarchies of Tarskian (i.e., non self-applicative) truth predicates. The remainder of this paper will develop this approach.

For the sake of readability I will begin by defining a single-step Tarskian truth theory. Let Z_1^i be the theory in the language of second order arithmetic with (1) the axioms and rules of a two-sorted intuitionistic predicate calculus and (2) the Peano axioms including induction for all formulas of Z_1^i . We do not assume any comprehension axioms. *For the remainder of the paper all theories will be assumed to have the axioms and rules of a two-sorted intuitionistic predicate calculus and to be expressed in the language of second order arithmetic possibly extended by a countable family of unary relation symbols.* For each n fix a recursive bijection $\langle \cdot, \dots, \cdot \rangle$ from ω^n to ω with corresponding recursive projections $\pi_i = \pi_i^n$, so that $\pi_i(\langle a_1, \dots, a_n \rangle) = a_i$. I will write (a_1, \dots, a_n) for $(\langle a_1, \dots, a_n \rangle)$ below.

Definition 2.1. Let S be a formal theory which extends Z_1^i . Fix a Gödel numbering of its formulas. Assume there exist recursive functions ax and ded such that ax enumerates the Gödel numbers of the axioms of S and ded enumerates all triples $\langle \ulcorner \mathcal{A} \urcorner, \ulcorner \mathcal{B} \urcorner, \ulcorner \mathcal{C} \urcorner \rangle$ such that S has a deduction rule that infers \mathcal{C} from \mathcal{A} and \mathcal{B} (perhaps with $\mathcal{A} = \mathcal{B}$). Also fix recursive functions f and g such that if $n = \ulcorner \mathcal{A}(v_i) \urcorner$ then $f(n, i, k) = \ulcorner \mathcal{A}(\bar{k}) \urcorner$ (i.e., all free occurrences of v_i are replaced by \bar{k}) and $g(n, i) = \ulcorner (\forall v_i) \mathcal{A}(v_i) \urcorner$, where v_i is the i th number variable symbol. If n is not the Gödel number of a formula, assume $f(n, i, k) = g(n, i) = 0$.

We define the *Tarskian truth theory of S* , $\text{Tarski}(S)$, to be the theory whose language is the language of S together with one additional unary relation symbol T and whose non-logical axioms are those of S , with the induction schema extended to the language of $\text{Tarski}(S)$, together with the three axioms

$$\begin{aligned} & T(\text{ax}(n)) \\ & T(\pi_1(\text{ded}(n))) \wedge T(\pi_2(\text{ded}(n))) \rightarrow T(\pi_3(\text{ded}(n))) \\ & (\forall k) T(f(n, i, k)) \leftrightarrow T(g(n, i)) \end{aligned}$$

and the axiom schema

$$\mathcal{A}(v_1, \dots, v_j) \leftrightarrow T(\ulcorner \mathcal{A}(\bar{v}_1, \dots, \bar{v}_j) \urcorner)$$

for all formulas \mathcal{A} in the language of S with no free set variables and with all free number variables among v_1, \dots, v_j .

Less rigorously (but perhaps more readably), the three extra axioms of $\text{Tarski}(S)$ assert $T(\ulcorner \mathcal{A} \urcorner)$ for every axiom \mathcal{A} of S ; $T(\ulcorner \mathcal{A} \urcorner) \wedge T(\ulcorner \mathcal{B} \urcorner) \rightarrow T(\ulcorner \mathcal{C} \urcorner)$ whenever there is a deduction rule of S that infers \mathcal{C} from \mathcal{A} and \mathcal{B} ; and the ω -rule $(\forall k) T(\ulcorner \mathcal{A}(\bar{k}) \urcorner) \leftrightarrow T(\ulcorner (\forall v) \mathcal{A}(v) \urcorner)$. The extra axiom schema is of course just Tarski's truth condition.

The above definition is imprecise in that one really needs to specify not merely the functions ax , ded , f , and g but their codes (and similarly for the unnamed function used in the final axiom schema to substitute numerals for variables in Gödel numbers, though this could be defined in terms of f). Moreover, these codes must be chosen in a natural way in order to allow the formalization of proofs that are given below only informally. The most straightforward way to handle this rigorously would be to make the definitions explicit as primitive recursive functions

for the systems with which we will be concerned; however, this would be somewhat tedious and I leave the reader to convince himself that it is possible. This comment will also apply to similar definitions that appear later.

Note that it is easy to define a provability predicate Prov in terms of ax and ded and to show that $\text{Tarski}(S)$ proves every instance of the schema $(\forall n)[\text{Prov}(\ulcorner \mathcal{A}(\bar{n}) \urcorner) \rightarrow \mathcal{A}(n)]$.

The general concept of truth may or may not be philosophically problematic, but here we are only using a limited non self-referential form which I do not think should be controversial, even for use by a predicativist. Indeed, one could argue that $\text{Tarski}(S)$ minus its induction and ω -rule schemas merely formalizes the assertion that one accepts S , and that these schemas are clearly predicatively legitimate.¹⁷

2.7. Iterated truth theories. We now describe a way to construct iterated families of truth theories.

Definition 2.2. Let S be a theory which extends Z_1^i and let \prec be a recursive total order on ω . We define the *iterated Tarskian truth theory of S along \prec* , $\text{Tarski}_\prec(S)$, as follows. Its language is the language of S together with additional unary relation symbols Acc and T_a (for each $a \in \omega$). Its non-logical axioms are the axioms of S , with the induction schema extended to the language of $\text{Tarski}_\prec(S)$, together with the axiom

$$\text{Prog}(\text{Acc})$$

(stating progressivity of Acc with respect to \prec). It also has an additional set of deduction rules whose statement requires some preparation.

Say that a formula is *readable by T_a* if it is a formula of the language of S enriched by the unary relation symbols T_b for $b \prec a$. Fix a Gödel numbering of the formulas of $\text{Tarski}_\prec(S)$ such that both the function $a \mapsto \ulcorner T_a \urcorner$ and the relation $\text{Rd}(a, n)$ indicating that n is the Gödel number of a formula readable by T_a are recursive. We also assume there is a recursive function ax such that $\text{ax}(a, \cdot)$ enumerates the Gödel numbers of the axioms of S together with all logical axioms and the induction schema extended to all formulas readable by T_a , and a recursive function ded such that $\text{ded}(a, \cdot)$ enumerates the logical deduction rules extended to all formulas readable by T_a (via triples, in the same manner as in Definition 2.1). We also require recursive functions f and g satisfying similar modifications of the corresponding conditions in Definition 2.1; a recursive function h such that

$$h(a, \ulcorner \mathcal{A} \urcorner) = \ulcorner \mathcal{A}(v_1, \dots, v_j) \leftrightarrow T_a(\ulcorner \mathcal{A}(\bar{v}_1, \dots, \bar{v}_j) \urcorner) \urcorner$$

for every formula \mathcal{A} in the language of $\text{Tarski}_\prec(S)$ with no free set variables and with all free number variables among v_1, \dots, v_j ; and a recursive relation Bd of S such that $\text{Bd}(\ulcorner \mathcal{A} \urcorner)$ holds if and only if \mathcal{A} has no free set variables, for every formula \mathcal{A} of $\text{Tarski}_\prec(S)$. The extra deduction rules of $\text{Tarski}_\prec(S)$ then state, for each $a \in \omega$, that one can infer from $\text{Acc}(\bar{a})$ the assertions

$$\begin{aligned} & T_a(\text{ax}(\bar{a}, n)) \\ & T_a(\pi_1(\text{ded}(\bar{a}, n))) \wedge T_a(\pi_2(\text{ded}(\bar{a}, n))) \rightarrow T_a(\pi_3(\text{ded}(\bar{a}, n))) \\ & \text{Rd}(\bar{a}, n) \rightarrow [(\forall k)T_a(f(\bar{a}, n, i, k)) \leftrightarrow T_a(g(\bar{a}, n, i))] \\ & [b \prec \bar{a} \wedge \text{Rd}(b, n) \wedge \text{Bd}(n)] \rightarrow T_a(h(b, n)) \end{aligned}$$

and the assertions

$$\mathcal{A}(v_1, \dots, v_j) \leftrightarrow T_a(\ulcorner \mathcal{A}(\bar{v}_1, \dots, \bar{v}_j) \urcorner)$$

for all formulas \mathcal{A} readable by T_a with no free set variables and with all free number variables among v_1, \dots, v_j . This completes the definition of $\text{Tarski}_{\prec}(S)$.

The statement $\text{Acc}(a)$ is supposed to signify that one accepts the truth predicate T_a . Thus, this set-up allows the presence of truth predicates which are not initially known to be intelligible and can be reasoned with only after some criterion that convinces us of their legitimacy is satisfied. Specifically, the criterion for accepting T_a is that we should be assured of the intelligibility of all formulas to which it applies, which just means that we should have accepted all prior truth predicates. Thus, the one extra axiom of $\text{Tarski}_{\prec}(S)$ states that Acc is progressive. The extra deduction rules allow a predicativist, once he has accepted the truth predicate T_a , to invoke all of the axioms appropriate to that predicate. So if a predicativist accepts S he should also accept $\text{Tarski}_{\prec}(S)$, for any recursive total order \prec on ω .

The system $\text{Tarski}_{\prec}(S)$ is open to the objection that to a limited extent it allows one to reason with predicates T_a that are not, or not yet, known to be intelligible. A more stringent and perhaps preferable formulation of the theory would ban any use of formulas involving T_a until $\text{Acc}(\bar{a})$ has been proven; this could be accomplished by deleting all axioms and all logical deduction rules which involve any truth predicates, and adding new deduction rules which allow their implementation after an appropriate acceptability statement has been proven. Similarly, if there are concerns about interpreting $T_a(n)$ when n is not the Gödel number of a formula readable by T_a , it is possible to set up a system of distinct Gödel numberings, one for each $a \in \omega$, such that every $n \in \omega$ is the Gödel number of a formula readable by T_a in the appropriate numbering, for each a . This approach would involve heavy use of a recursive translation function which relates distinct Gödel numberings. In any case the net result would be a slightly more complicated theory with precisely the same deductive power. In particular, the results presented below would still hold.

2.8. $\text{Tarski}_{\Gamma_0}^{\omega}(Z_1^i)$ and Γ_0 . I will now present a predicatively valid system $\text{Tarski}_{\Gamma_0}^{\omega}(Z_1^i)$ which proves well-ordering in a strong sense for notations for every ordinal less than Γ_0 . The interest of this system is that it shows that every ordinal less than Γ_0 is predicatively provable. This claim has been made many times before, but as I pointed out in Section 1, all previous efforts have crucially involved impredicative reasoning. Moreover, it will be evident that the present construction can easily be pushed further to obtain predicative well-ordering proofs for notations for even larger ordinals. I will describe stronger systems that go significantly further in §2.9 and §2.10.

Definition 2.3. Let \prec be a standard recursive ordering of ω of order type Γ_0 whose least elements are 0 and 1. We write Tarski_{Γ_0} for Tarski_{\prec} . Define $\text{Tarski}_{\Gamma_0}^0(Z_1^i) = Z_1^i$ and inductively set $\text{Tarski}_{\Gamma_0}^{n+1}(Z_1^i) = \text{Tarski}_{\Gamma_0}(\text{Tarski}_{\Gamma_0}^n(Z_1^i))$. Observe that $\text{Tarski}_{\Gamma_0}^{n+1}(Z_1^i)$ extends $\text{Tarski}_{\Gamma_0}^n(Z_1^i)$. Let $\text{Tarski}_{\Gamma_0}^{\omega}(Z_1^i)$ be the union of the theories $\text{Tarski}_{\Gamma_0}^n(Z_1^i)$.

I argued in §2.7 that if a predicativist accepts a theory S then he should also accept $\text{Tarski}_{\prec}(S)$ for any recursive total order \prec . In particular, if he accepts S he should also accept $\text{Tarski}_{\Gamma_0}(S)$. Granting that he accepts Z_1^i , by iteration he should accept each $\text{Tarski}_{\Gamma_0}^n(Z_1^i)$, and recognizing this he should also accept $\text{Tarski}_{\Gamma_0}^{\omega}(Z_1^i)$. Now there is no reason to stop at ω , and by going further we can obtain predicative

well-ordering proofs of larger ordinals. In particular, Tarski($\text{Tarski}_{\Gamma_0}^\omega(Z_1^i)$) proves that a notation for Γ_0 is well-ordered, yielding the falsification of the Γ_0 thesis promised in the title and at the beginning of Section 1.

The well-ordering proof is based on the following lemma. Let a_n be the notation for γ_n according to \prec . If a is the notation for α and b is the notation for β then let $a \dot{+} b$ be the notation for $\alpha + \beta$, $\dot{\omega}^a$ the notation for ω^α , etc. (Let $a \cdot b$ be the notation for $\alpha\beta$.) We may assume that these are all recursive functions of a and b .

Lemma 2.4. *Let S be a formal theory that extends Z_1^i and satisfies all assumptions needed in Definition 2.2. Then for any $n \in \omega$, $\text{Tarski}_{\Gamma_0}(S)$ plus the transfinite induction schema $\text{TI}(\mathcal{A}, a_n)$ for every formula \mathcal{A} in its language proves $\text{TI}(\mathcal{A}, a_{n+1})$ for every formula \mathcal{A} in the language of S .*

Proof. For $n = 0$ one simply carries out the proof of Lemma 1 on page 180 of [42] within $\text{Tarski}(S)$. Thus, fix $n \geq 1$. Let $\mathcal{B}_{a,b}(m)$ be the formula

$$\mathcal{B}_{a,b}(m) \equiv (\forall z) \left[\text{Rd}(\dot{\omega}^a \cdot b, z) \wedge \text{Bd}(z) \rightarrow T_{\dot{\omega}^a \cdot b}(\ulcorner \text{Prog}[[z]] \rightarrow \mathcal{J}(\ulcorner [z], \dot{\phi}_a(\overline{m}) \urcorner) \urcorner) \right]$$

where

$$\mathcal{J}(\mathcal{A}, a) \equiv (\forall y) \left[(\forall x \prec y) \mathcal{A}(x) \rightarrow (\forall x \prec y \dot{+} a) \mathcal{A}(x) \right]$$

and the $[[z]]$ notation indicates that the formula with Gödel number z is to be inserted at that point. Note that $\mathcal{B}_{a,b}(m)$ is not a single formula in the language of $\text{Tarski}_{\Gamma_0}(S)$ because of the presence of the varying unary relation symbols $T_{\dot{\omega}^a \cdot b}$, but

$$\mathcal{C}(a) \equiv (\forall b) [0 \prec b \prec a_n \rightarrow T_{a_n}(\ulcorner \text{Prog}_m \mathcal{B}_{a,b}(m) \urcorner)]$$

is a single formula with parameter a .

Since Acc is progressive, the hypothesis about transfinite induction yields $\text{Acc}(a_n)$, so that the axioms for T_{a_n} are available. We will use them to prove in $\text{Tarski}_{\Gamma_0}(S)$ that $\mathcal{C}(a)$ is progressive over $a \prec a_n$. First, $\mathcal{C}(0)$ can be proven by carrying out, within T_{a_n} , the proof of Lemma 1 on page 180 of [42]. $\mathcal{C}(a)$ holds at limit values of a if it holds at all smaller values by a straightforward verification using the facts that $\phi_a(0) = \sup_{\tilde{a} \prec a} \phi_{\tilde{a}}(0)$, $\phi_a(m) = \phi_{\tilde{a}}(\phi_a(m))$ for all $\tilde{a} \prec a$, and $\phi_a(m+1) = \sup_{\tilde{a} \prec a} \phi_{\tilde{a}}(\phi_a(m) + 1)$, all of which are provable in Z_1^i . At successor stages, we assume $\mathcal{C}(a)$ and prove $\mathcal{C}(a \dot{+} 1)$ as follows. Fix $0 \prec b \prec a_n$ and $m \in \omega$ and, working within T_{a_n} , suppose $\mathcal{B}_{a \dot{+} 1, b}(m)$ holds; we must prove $\mathcal{B}_{a \dot{+} 1, b}(m \dot{+} 1)$. This will verify progressivity at successor stages. $\mathcal{B}_{a \dot{+} 1, b}(m)$ is proven similarly for $m = 0$, and it is trivial at limit stages assuming it holds at all previous stages.

The following argument is carried out within T_{a_n} . To prove $\mathcal{B}_{a \dot{+} 1, b}(m \dot{+} 1)$, observe that (by $\mathcal{C}(a)$) $\mathcal{B}_{a, r \cdot b}(s)$ is progressive in s for all $r \prec \dot{\omega}$. Since $\mathcal{B}_{a, r \cdot b}(s)$ is readable by $T_{\dot{\omega}^{a+1} \cdot b}$ and has no free set variables, the hypothesis that $\mathcal{B}_{a \dot{+} 1, b}(m)$ holds then implies

$$\mathcal{J}(\mathcal{B}_{a, r \cdot b}(s), \dot{\phi}_{a+1}(m))$$

for all r . Fixing $j, r \prec \dot{\omega}$ we successively infer

$$\mathcal{J}(\mathcal{B}_{a, (r \dot{+} j) \cdot b}(s), \dot{\phi}_{a+1}(m))$$

and hence

$$\mathcal{B}_{a, (r \dot{+} j) \cdot b}(\dot{\phi}_{a+1}(m) \dot{+} 1);$$

then if $j \geq 1$, since $\mathcal{B}_{a, (r \dot{+} j \dot{-} 1) \cdot b}(s)$ is readable by $T_{\dot{\omega}^a \cdot (r \dot{+} j) \cdot b}$,

$$\mathcal{J}(\mathcal{B}_{a, (r \dot{+} j \dot{-} 1) \cdot b}(s), \dot{\phi}_a(\dot{\phi}_{a+1}(m) \dot{+} 1))$$

and hence

$$\mathcal{B}_{a,(r+j\dot{-}1)\cdot b}(\dot{\phi}_a(\dot{\phi}_{a\dot{+}1}(m\dot{+}1)));$$

and so on, down to

$$\mathcal{B}_{a,r\cdot b}(\dot{\phi}_a^j(\dot{\phi}_{a\dot{+}1}(m\dot{+}1))).$$

Since $j \prec \dot{\omega}$ is arbitrary, this implies

$$\mathcal{B}_{a,r\cdot b}(\dot{\phi}_{a\dot{+}1}(m\dot{+}1)),$$

and since this is true for all $r \prec \omega$ (and every formula readable by $T_{\dot{\omega}^{a+1}\cdot b}$ is readable by $T_{\dot{\omega}^a\cdot r\cdot b}$ for some r), we infer $\mathcal{B}_{a\dot{+}1,b}(m\dot{+}1)$, as desired.

We conclude that $\mathcal{C}(a)$ is progressive, so our hypothesis about transfinite induction yields $\mathcal{C}(a)$ for all $a \prec a_n$. Taking $b = 1$ and $m = 0$, we infer

$$(\forall a \prec a_n)T_{a_n}(\ulcorner \mathcal{B}_{a,1}(0) \urcorner).$$

In particular, for every formula \mathcal{A} in the language of S with no free set variables and one free number variable we have

$$(\forall a \prec a_n)T_{a_n}(\ulcorner \text{Prog } \mathcal{A} \rightarrow \mathcal{J}(\mathcal{A}, \dot{\phi}_{\bar{a}}(0)) \urcorner)$$

and therefore

$$\text{Prog } \mathcal{A} \rightarrow (\forall a \prec a_n)\mathcal{J}(\mathcal{A}, \dot{\phi}_{\bar{a}}(0))$$

and finally (since $a_{n+1} = \phi_{a_n}(0) = \sup_{a \prec a_n} \phi_a(0)$)

$$\text{Prog } \mathcal{A} \rightarrow (\forall a \prec a_{n+1})\mathcal{A}(a).$$

This proves $\text{TI}(\mathcal{A}, a_{n+1})$ for every formula \mathcal{A} in the language of S with no free set variables and one free number variable. We can reduce to this case by replacing an arbitrary formula \mathcal{A} with the formula $\text{Prog}(\mathcal{A}) \rightarrow \mathcal{A}$ and then universally quantifying all parameters. This yields a formula which is automatically progressive and we conclude by the above that it holds for all $a \prec a_{n+1}$; interchanging the order of universal quantifiers then yields $\text{Prog}(\mathcal{A}) \rightarrow (\forall a \prec a_{n+1})\mathcal{A}(a)$. \square

This lemma could also be proven by adapting the proof of Theorem 3 in [19]. The point is that one can model the jump hierarchy used there by using the recursion theorem to find $r \in \omega$ such that

$$\{r\}(0, m) = \ulcorner \mathcal{A}(\bar{m}) \urcorner$$

and for all $a \succ 0$

$$\{r\}(a, m) = \ulcorner (\forall z)[z \prec \bar{a} \rightarrow \mathcal{J}(T_a(\{\bar{r}\}(z, x)), \dot{\phi}(e(\bar{a}), \bar{m})) \urcorner,$$

both provably in Z_1^i . Setting $\mathcal{B}(a, m) \equiv T_{a_{n+1}}(\{\bar{r}\}(a, m))$, we can then prove in $\text{Tarski}_{\Gamma_0}(S)$ that for all $a \leq a_n$

$$\mathcal{B}(a, m) \leftrightarrow (\forall z)[z \prec a \rightarrow \mathcal{J}(\mathcal{B}(z, x), \dot{\phi}(e(a), m))].$$

The argument used to prove Theorem 3 of [19] can then be used to complete the proof of the lemma. (It can even be slightly simplified because $\omega^{\gamma_n} = \gamma_n$ for $n \geq 1$ and the condition $h(a) \preceq z$ is not really necessary, there or here.) However, the proof given above generalizes better to the situation in §2.9 and §2.10.

Theorem 2.5. $\text{Tarski}_{\Gamma_0}^{\omega}(Z_1^i)$ proves transfinite induction up to any ordinal less than Γ_0 for all formulas in its language.

Proof. Inductive application of the lemma shows that $\text{Tarski}_{\Gamma_0}^n(S)$ proves transfinite induction up to γ_n for all formulas of S . Applying this result with $S = \text{Tarski}_{\Gamma_0}^k(Z_1^i)$ for arbitrary n and k , and observing that every formula of $\text{Tarski}_{\Gamma_0}^\omega(Z_1^i)$ is a formula of $\text{Tarski}_{\Gamma_0}^k(Z_1^i)$ for some k , we obtain the desired result. \square

The main point of this proof is that we finesse the induction versus recursion issue encountered in Section 1 by proving a stronger result, namely that transfinite induction holds for arbitrary formulas rather than just for sets.

I believe this is the first predicatively valid demonstration of the well-foundedness of notations for all ordinals less than Γ_0 . Now its predicative validity requires the validity of three stages of abstraction: reasoning using Tarskian truth predicates (going from S to $\text{Tarski}(S)$), reasoning about which of a sequence of truth predicates are acceptable (going from S to $\text{Tarski}_{\prec}(S)$), and iterating the preceding step (going from S to $\text{Tarski}_{\prec}^\omega(S)$).

If we grant that a predicativist can always pass from S to $\text{Tarski}_{\prec}(S)$, then he should be able to consider the sequence of theories $\text{Tarski}_{\prec}^n(S)$ and reason that the validity of each one implies the next, hence they are all valid and therefore so is $\text{Tarski}_{\prec}^\omega(S)$. Thus, if we accept the second stage then we should also accept the third stage. Conceivably one could try to make a case that a predicativist can pass to $\text{Tarski}_{\prec}^{n+1}(S)$ once he has actually accepted $\text{Tarski}_{\prec}^n(S)$ but he cannot recognize that this passage is valid in general. This would enable him to accept each $\text{Tarski}_{\prec}^n(S)$ but not $\text{Tarski}_{\prec}^\omega(S)$. Presumably the idea would be that predicativists are capable of making individual judgements about the acceptability of particular theories but cannot reason about the acceptability of theories in general. However, it would seem that one could with equal justice replace “theories” with “truth predicates” in this assertion, yielding the claim that predicativists can make individual judgements about whether to accept particular truth predicates but cannot reason about such questions abstractly, which would render $\text{Tarski}_{\prec}(S)$ illegitimate. In other words, we have just as much reason to go from the second stage to the third as we do to go from the first stage to the second.

On the other hand, if we reject passage from the first stage (accepting $\text{Tarski}(S)$) to the second (accepting $\text{Tarski}_{\prec}(S)$) I think it would have to be on the grounds just suggested, i.e., that predicativists are capable of making individual judgements about the acceptability of particular truth predicates but cannot reason abstractly about their acceptability in general. But then we could just as well replace “truth predicates” with “statements” and argue that predicativists can make individual judgements about whether to accept particular statements but cannot reason about such questions abstractly. This would render $\text{Tarski}(S)$ illegitimate. But not only that; as I pointed out in §1.4 (a), this line of argument, if accepted, would actually prevent a predicativist from recognizing modus ponens, or really any deduction rule, as a general principle. In fact it ought to forbid any use of statements involving variables of any kind since these already imply an ability to reason hypothetically about the truth of such a statement on all possible substitutions of values for the variables, which evidently requires some abstract sense of truth or acceptability. Ultimately we would be left only with the ability to make concrete numerical assertions. The point is that *there is a smooth progression in reasoning at successively higher levels of abstraction that takes one from Peano arithmetic up to $\text{Tarski}_{\Gamma_0}^\omega(Z_1^i)$ and beyond*, so that any attempt to cut this progression off at some point is bound

to appear arbitrary. At any rate there is no better reason to cut it off at Γ_0 than anywhere else.

The element of truth to the objection is that the general concept of statements “being predicatively acceptable” is impredicative when there is no limitation on the domain of discussion, because it would be circular to talk about the predicative acceptability of statements which themselves involve the concept of predicative acceptability. Similarly, the ideas of truth predicates being predicatively acceptable or of theories being predicatively acceptable are impredicative as unrestricted general concepts. But these are just versions of the fact noted in §2.5 that the general concept of “being intelligible” is not itself intelligible. In each of these cases, if attention is restricted to a well-defined previously grasped domain, I see no impredicativity.

2.9. Tarski $_{\kappa^\kappa}^\omega(\mathbf{Z}_1^i)$ and $\phi_{\Omega^2}(0)$. In the last section we considered a formal system, $\text{Tarski}_{\Gamma_0}(\mathbf{Z}_1^i)$, in which we were able to reason about the (predicative) acceptability of a hierarchy of truth predicates by means of an additional predicate Acc . The Tarski_{Γ_0} construction was then iterated ω times. By systematizing the process of iterating constructions involving acceptability predicates we can access ordinals well beyond Γ_0 . I will next illustrate this claim by describing a predicatively valid formal system that proves well-ordering statements for notations for all ordinals less than the Ackermann ordinal $\phi_{\Omega^2}(0)$. In §2.10 I will sketch a way to carry the construction further and get a predicative well-ordering proof for the “small” Veblen ordinal $\phi_{\Omega^\omega}(0)$.

The development is similar to that in §2.8 and will be presented here in slightly less detail. Let $\kappa = \phi_{\Omega^\omega}(0)$ and fix a notation system for κ^κ (e.g., see the introduction to [34]). In the following I will identify ordinals with their notations and I will use α, β, γ to range over ordinals $< \kappa$ and a, b, c to range over ordinals $< \kappa^\kappa$. Every nonzero a can be uniquely written in the form $a = \kappa^{\alpha_1}\beta_1 + \dots + \kappa^{\alpha_n}\beta_n$ such that $\alpha_1 \succeq \dots \succeq \alpha_n$; $\alpha_i = \alpha_{i+1}$ implies $\beta_i \succeq \beta_{i+1}$; and each β_i is either 1 or a limit ordinal. Let $h(a) = \kappa^{\alpha_1}\beta_1 + \dots + \kappa^{\alpha_{n-1}}\beta_{n-1}$. We define the *canonical sequence associated to a* to be $\{h(a) + \kappa^{\alpha_n}\gamma : \gamma < \beta_n\}$ if β_n is a limit ordinal; if $\beta_n = 1$ and $\alpha_n = 0$ then it is the single element $\{h(a)\}$; if $\beta_n = 1$ and α_n is a limit ordinal it is $\{h(a) + \kappa^\gamma : \gamma < \alpha_n\}$; and if $\beta_n = 1$ and $\alpha_n = \tilde{\alpha}_n + 1$ then it is $\{h(a) + \kappa^{\tilde{\alpha}_n}\gamma : \gamma < \kappa\}$. In the last case ($\beta_n = 1$ and α_n a successor) we say that a is of *type 1*, and otherwise it is of *type 0*. (Cf. Definition 1 of [22].) We consider 0 to be of type 0 and we let its canonical sequence be empty. Let Typ_0 be a formula such that $\text{Typ}_0(a)$ holds if and only if a is of type 0, and let Seq be a formula such that $\text{Seq}(x, a)$ holds if and only if x belongs to the canonical sequence associated to a .

Definition 2.6. Let S be a theory which extends \mathbf{Z}_1^i . We define $\text{Tarski}_{\kappa^\kappa}(S)$ as follows. Its language is the language of S together with a unary relation symbol Acc and two families of unary relation symbols T_a and Acc_a . In this setting a formula is *readable by T_a* if it is a formula of the language of S enriched by the unary relation symbols T_b and Acc_b for all $b < a$. The non-logical axioms are the axioms of S , with the induction schema extended to the larger language, together with an axiom which states that for any a , if a is of type 0 with associated canonical sequence (x_γ) then $(\forall\gamma)\text{Acc}(x_\gamma) \rightarrow \text{Acc}(a)$, and if a is of type 1 with canonical sequence (x_γ) then $\text{Prog}_\gamma \text{Acc}(x_\gamma) \rightarrow \text{Acc}(a)$. We also have, for each a , a family of deduction rules allowing inference from the premise $\text{Acc}(\bar{a})$ of the following statements:

(I) Axioms for T_a : the same as in Definition 2.2, i.e.,

$$\begin{aligned} & T_a(\text{ax}(\bar{a}, n)) \\ & T_a(\pi_1(\text{ded}(\bar{a}, n))) \wedge T_a(\pi_2(\text{ded}(\bar{a}, n))) \rightarrow T_a(\pi_3(\text{ded}(\bar{a}, n))) \\ & \text{Rd}(\bar{a}, n) \rightarrow [(\forall k)T_a(f(\bar{a}, n, i, k)) \leftrightarrow T_a(g(\bar{a}, n, i))] \\ & [b \prec \bar{a} \wedge \text{Acc}_a(b) \wedge \text{Rd}(b, n) \wedge \text{Bd}(n)] \rightarrow T_a(h(b, n)) \\ & \mathcal{A}(v_1, \dots, v_j) \leftrightarrow T_a(\ulcorner \mathcal{A}(\bar{v}_1, \dots, \bar{v}_j) \urcorner), \end{aligned}$$

with the premise $\text{Acc}_a(b)$ added to the fourth axiom. The functions appearing in these axioms are defined analogously to those in Definition 2.2, and the same condition is placed on \mathcal{A} in the final schema.

(II) Axioms for Acc_a :

$$\begin{aligned} & \text{Seq}(b, \bar{a}) \rightarrow \text{Acc}_a(b) && \text{(if } a \text{ is of type 0)} \\ & \text{Prog}_b[\text{Seq}(b, \bar{a}) \rightarrow \text{Acc}_a(b)] && \text{(if } a \text{ is of type 1)} \\ & [b \prec \bar{a} \wedge \text{Acc}_a(b) \wedge \text{Typ}_0(b)] \rightarrow T_a(\ulcorner (\forall c)[\text{Seq}(c, b) \rightarrow \text{Acc}_b(c)] \urcorner) \\ & [b \prec \bar{a} \wedge \text{Acc}_a(b) \wedge \neg \text{Typ}_0(b)] \rightarrow T_a(\ulcorner \text{Prog}_c[\text{Seq}(c, b) \rightarrow \text{Acc}_b(c)] \urcorner) \\ & [c \prec b \prec \bar{a} \wedge \text{Acc}_a(b)] \rightarrow [\text{Acc}_a(c) \leftrightarrow T_a(\ulcorner \text{Acc}_b(c) \urcorner)] \\ & b \prec \bar{a} \rightarrow [\text{Acc}(b) \leftrightarrow \text{Acc}_a(b)] \end{aligned}$$

This completes the definition of $\text{Tarski}_{\kappa^\kappa}(S)$.

As in Definition 2.3 we now inductively define $\text{Tarski}_{\kappa^\kappa}^0(Z_1^i) = Z_1^i$ and $\text{Tarski}_{\kappa^\kappa}^{n+1}(Z_1^i) = \text{Tarski}_{\kappa^\kappa}^n(\text{Tarski}_{\kappa^\kappa}^n(Z_1^i))$, and we let $\text{Tarski}_{\kappa^\kappa}^\omega(Z_1^i)$ be the union of the theories $\text{Tarski}_{\kappa^\kappa}^n(Z_1^i)$.

The system $\text{Tarski}_{\kappa^\kappa}^\omega(Z_1^i)$ is predicatively justified by taking one step up in abstraction beyond $\text{Tarski}_{\Gamma_0}(Z_1^i)$. There we had a theory involving a hierarchy of truth predicates and we formally reasoned about their acceptability. Here we have a hierarchy of acceptability predicates, each of which allows us to reason about the acceptability of truth and acceptability predicates of lower degree, and about whose acceptability we are able to formally reason. Intuitively, if a is of type 0 then the acceptability predicate at level a is supposed to affirm the acceptability of all levels belonging to the canonical sequence associated to a , and if a is of type 1 then it is supposed to affirm progressivity of the acceptability of the levels belonging to the canonical sequence associated to a .

Just as with Definition 2.2 it is possible to formulate a stricter definition which would disallow any use of T_a and Acc_a until after $\text{Acc}(\bar{a})$ has been proven. We could also set up a family of distinct Gödel numberings so that truth predicates could only refer to formulas readable by them, and similarly we could set up a family of distinct orderings of ω , one of order type a for each a , so that acceptability predicates could only refer to prior truth and acceptability predicates. As before, these changes would be merely cosmetic and would not affect the strength of the theory.

The well-ordering proof is based on two lemmas.

Lemma 2.7. *Let S be a formal theory that extends Z_1^i and satisfies all assumptions needed in Definition 2.6. Then $\text{Tarski}_{\kappa^\kappa}(S)$ proves that the statement*

$$\mathcal{A}(\alpha) \equiv (\forall a) [\text{Acc}(a) \rightarrow \text{Acc}(a + \kappa^\alpha)]$$

is progressive in α .

Proof. For any a the canonical sequence associated to $a + 1$ is $\{a\}$, so an axiom of $\text{Tarski}_{\kappa^\kappa}(S)$ asserts that $\text{Acc}(a) \rightarrow \text{Acc}(a + 1)$. This shows that $\mathcal{A}(0)$ is provable in $\text{Tarski}_{\kappa^\kappa}(S)$. At successor stages, suppose $\mathcal{A}(\alpha)$ holds and, reasoning in $\text{Tarski}_{\kappa^\kappa}(S)$, deduce that for any a satisfying $\text{Acc}(a)$ the statement $\text{Acc}(a + \kappa^\alpha \beta)$ is progressive in β . This yields $\text{Acc}(a) \rightarrow \text{Acc}(a + \kappa^{\alpha+1})$, and we infer $\mathcal{A}(\alpha + 1)$. Finally, suppose α is a limit and we have $\mathcal{A}(\gamma)$ for all $\gamma \prec \alpha$. Then for any a such that $\text{Acc}(a)$ holds we have $\text{Acc}(a + \kappa^\gamma)$ for all $\gamma \prec \alpha$, and this implies $\text{Acc}(a + \kappa^\alpha)$. So we infer $\mathcal{A}(\alpha)$. \square

Let $\delta_0 = 1$ and $\delta_{n+1} = \phi_{\Omega \cdot \delta_n}(0)$, so that $\phi_{\Omega^2}(0) = \sup_{n \in \omega} \delta_n$. Note that $\omega\kappa = \kappa$, so $(\kappa\omega)^\alpha$ equals κ^α if α is a limit and it equals $\kappa^\alpha\omega$ if α is a successor.

Lemma 2.8. *Let S be a formal theory that extends Z_1^i and satisfies all assumptions needed in Definition 2.6. Then $\text{Tarski}_{\kappa^\kappa}(S)$ plus the transfinite induction schema $\text{TI}(\mathcal{A}, \delta_n)$ for every formula \mathcal{A} in its language proves $\text{TI}(\mathcal{A}, \delta_{n+1})$ for every formula \mathcal{A} in the language of S .*

Proof. The technique is similar to that used in the proof of Lemma 2.4. Define

$$\mathcal{B}_{\alpha,b}(\mu) \equiv (\forall z) [\text{Rd}((\kappa\omega)^\alpha b, z) \wedge \text{Bd}(z) \rightarrow T_{(\kappa\omega)^\alpha b}(\ulcorner \text{Prog}[[z]] \urcorner \rightarrow \mathcal{J}(\ulcorner [z] \urcorner, \phi_{\Omega\alpha}(\mu)) \urcorner)]$$

and

$$\mathcal{C}(\alpha) \equiv (\forall b) [0 \prec b \prec \kappa^{\delta_n} \wedge \text{Typ}_0(b) \wedge \text{Acc}_{\kappa^{\delta_n}}((\kappa\omega)^\alpha b) \rightarrow T_{\kappa^{\delta_n}}(\ulcorner \text{Prog}_\mu \mathcal{B}_{\alpha,b}(\mu) \urcorner)].$$

By Lemma 2.7 and the transfinite induction hypothesis we obtain $\text{Acc}(a) \rightarrow \text{Acc}(a + \kappa^\alpha)$ for all $\alpha \prec \delta_n$. In particular, $\text{Acc}(\kappa^\alpha)$ holds for all $\alpha \prec \delta_n$, and this implies $\text{Acc}(\kappa^{\delta_n})$. We also obtain $\text{Acc}(a) \rightarrow \text{Acc}(a + \kappa^\alpha r)$ for all $\alpha \prec \delta_n$ and all $r \prec \omega$, which implies $\text{Acc}(a) \rightarrow \text{Acc}(a + \kappa^\alpha \omega)$ for all $\alpha \prec \delta_n$.

We claim that $\mathcal{C}(\alpha)$ is progressive over $\alpha \prec \delta_n$. $\mathcal{C}(0)$ is again essentially Lemma 1 on page 180 of [42]. Next, let α be a limit and suppose $\mathcal{C}(\beta)$ holds for all $\beta \prec \alpha$. Fix $0 \prec b \prec \kappa^{\delta_n}$ of type 0 and suppose $\text{Acc}_{\kappa^{\delta_n}}((\kappa\omega)^\alpha b)$. Then for every $\beta \prec \alpha$ and every x in the canonical sequence associated to b , letting $y = (\kappa\omega)^{\tilde{\alpha}}x + 1$ where $\beta + \tilde{\alpha} = \alpha$, we have $\text{Acc}_{\kappa^{\delta_n}}((\kappa\omega)^\beta y)$ and so $\mathcal{C}(\beta)$ implies progressivity in μ of the assertion that $T_{(\kappa\omega)^\beta y}(\ulcorner \text{Prog}[[z]] \urcorner \rightarrow \mathcal{J}(\ulcorner [z] \urcorner, \phi_{\Omega\beta}(\mu)) \urcorner)$ holds for all appropriate z . Since every formula readable by $T_{(\kappa\omega)^\alpha b}$ is readable by $T_{(\kappa\omega)^\beta y}$ for sufficiently large x and β and $\phi_{\Omega\alpha}$ enumerates the common values of $\phi_{\Omega\beta}$ over $\beta \prec \alpha$, this implies the desired conclusion.

Finally, suppose $\mathcal{C}(\alpha)$ holds for some $\alpha \prec \delta_n$; we must verify $\mathcal{C}(\alpha + 1)$. To do this fix $0 \prec b \prec \kappa^{\delta_n}$ of type 0 and suppose $\text{Acc}_{\kappa^{\delta_n}}((\kappa\omega)^{\alpha+1}b)$. We may assume b is a successor, $b = \tilde{b} + 1$, since progressivity of $\mathcal{B}_{\alpha+1,b}(\mu)$ for these values implies progressivity for type 0 limits by an argument similar to the one used in the previous paragraph. Now, as in the proof of Lemma 2.4, progressivity at limit values of μ follows from continuity of $\phi_{\Omega(\alpha+1)}$, and progressivity at $\mu = 0$ is proven in the same way as progressivity at successor values of μ . Therefore we consider the successor case. Fix $\mu \prec \kappa$ and, working within $T_{\kappa^{\delta_n}}$, suppose $\mathcal{B}_{\alpha+1,\tilde{b}+1}(\mu)$ holds; we must prove $\mathcal{B}_{\alpha+1,\tilde{b}+1}(\mu + 1)$.

Let $\tilde{\gamma}_0 = \phi_{\Omega(\alpha+1)}(\mu) + 1$ and inductively define $\tilde{\gamma}_{n+1} = \phi_{\Omega\alpha + \tilde{\gamma}_n}(0)$, so that $\phi_{\Omega(\alpha+1)}(\mu + 1) = \sup_{n \prec \omega} \tilde{\gamma}_n$. In order to verify $\mathcal{B}_{\alpha+1,\tilde{b}+1}(\mu + 1)$, it will therefore suffice to check $T_{(\kappa\omega)^{\alpha+1}b}(\ulcorner \text{Prog}[[z]] \urcorner \rightarrow \mathcal{J}(\ulcorner [z] \urcorner, \tilde{\gamma}_n) \urcorner)$ for arbitrary n (and appropriate z). Working within $T_{\kappa^{\delta_n}}$, we can now mimic the proof of Lemma 2.4 to show that

transfinite induction up to $\tilde{\gamma}_n$ for all formulas readable by $T_{(\kappa\omega)^{\alpha+1}\tilde{b}+(\kappa\omega)^\alpha\kappa(r+1)}$ implies transfinite induction up to $\tilde{\gamma}_{n+1}$ for all formulas readable by $T_{(\kappa\omega)^{\alpha+1}\tilde{b}+(\kappa\omega)^\alpha\kappa r}$. By the same argument used in the proof of Theorem 2.5 we then get transfinite induction up to any $\tilde{\gamma}_n$ for formulas readable by any $T_{(\kappa\omega)^{\alpha+1}\tilde{b}+(\kappa\omega)^\alpha\kappa r}$, i.e., for all formulas readable by $T_{(\kappa\omega)^{\alpha+1}b}$. In this argument $\tilde{\gamma}_n$ replaces a_n , the expression $\omega^a b$ is modified to

$$(\kappa\omega)^{\alpha+1}\tilde{b} + (\kappa\omega)^\alpha\kappa r + (\kappa\omega)^\alpha\omega^{\tilde{\alpha}}\tilde{\beta}$$

(where a and b are replaced by $\tilde{\alpha}$ and $\tilde{\beta}$), ϕ_a is modified to $\phi_{\Omega\alpha+\tilde{\alpha}}$, and the hypothesis that $\mathcal{C}(\alpha)$ holds replaces use of Lemma 1 on page 180 of [42]; otherwise the argument is identical. We conclude that $\mathcal{C}(\alpha)$ is progressive.

The remainder of the proof is similar to the proof of Lemma 2.4. \square

Theorem 2.9. Tarski $_{\kappa^\kappa}^\omega(\mathbf{Z}_1^i)$ proves transfinite induction up to any ordinal less than $\phi_{\Omega^2}(0)$ for all formulas in its language.

Proof. By reasoning identical to that used in the proof of Theorem 2.5. \square

2.10. **Tarski $_{\lambda^\omega}^\omega(\mathbf{Z}_1^i)$ and $\phi_{\Omega^\omega}(0)$.** It is possible to strengthen the construction of §2.9 to obtain a predicative well-ordering proof for notations for all ordinals less than the “small” Veblen ordinal $\phi_{\Omega^\omega}(0)$. I will only sketch the construction and I will omit all proofs.

Let $\lambda = \phi_{\Omega^\omega}(0)$ and fix a notation system for λ^{λ^ω} . Every ordinal a less than λ^{λ^ω} can be uniquely written in the form $a = \lambda^{\alpha_1}\beta_1 + \dots + \lambda^{\alpha_n}\beta_n$ with each $\alpha_i < \lambda^\omega$; each $\beta_i < \lambda$; $\alpha_1 \succeq \dots \succeq \alpha_n$; $\alpha_i = \alpha_{i+1}$ implies $\beta_i \succeq \beta_{i+1}$; and each β_i either a limit or 1. Write $\alpha_n = \lambda^{\gamma_1}\delta_1 + \dots + \lambda^{\gamma_m}\delta_m$ with similar conditions on the γ_i and δ_i (but now each $\gamma_i < \omega$). If β_n is a limit or if $\beta_n = 1$ and δ_m is a limit then we say that a is of type 0; if $\beta_n = \delta_m = 1$ then a is of type $\gamma_m + 1$. Canonical associated sequences are defined just as before in the type 0 case, for type 1 we take the canonical associated sequence to be

$$\{h(a) + \lambda^{\lambda^{\gamma_1}\delta_1 + \dots + \lambda^{\gamma_{m-1}}\delta_{m-1}} \cdot \alpha : \alpha < \lambda\},$$

and for higher types we take it to be

$$\{h(a) + \lambda^{\lambda^{\gamma_1}\delta_1 + \dots + \lambda^{\gamma_{m-1}}\delta_{m-1} + \lambda^{\gamma_m-1}\alpha} : \alpha < \lambda\},$$

where $h(a) = \lambda^{\alpha_1}\beta_1 + \dots + \lambda^{\alpha_{n-1}}\beta_{n-1}$. The Tarski $_{\lambda^\omega}(S)$ construction goes like the Tarski $_{\kappa^\kappa}(S)$ construction, with the following elaboration. Define a *bounded jumpability* predicate

$$\mathcal{J}_{a,b}^1(\mathcal{A}, c) \equiv (\forall a \preceq x < b) [\mathcal{A}(x) \rightarrow \mathcal{A}(x+c)]$$

and inductively define a *bounded k -th order jumpability* predicate $\mathcal{J}_{a,b}^k(\mathcal{A}, c)$ by

$$\mathcal{J}_{a,b}^{k+1}(\mathcal{A}, c) \equiv (\forall a \preceq x < b) [\mathcal{J}_{a,b}^k(\mathcal{A}, x) \rightarrow \mathcal{J}_{a,b}^k(\mathcal{A}, x+c)].$$

Accepting level a of the truth and acceptability hierarchy affiliated to Tarski $_{\lambda^\omega}(S)$ should then entail accepting level $h(a)$ as well as: if a is type 0, accepting all levels in its associated sequence; if a is type 1, accepting progressivity of the acceptability of the levels in its associated sequence; and if a is of type $k \geq 2$, accepting

$$\text{Prog}_{\alpha < \lambda} \mathcal{J}_{h(a), a}^{k-1}(\text{Acc}, \lambda^{\lambda^{\gamma_1}\delta_1 + \dots + \lambda^{\gamma_{m-1}}\delta_{m-1} + \lambda^{\gamma_m-1}\alpha}).$$

The axiomatization of Tarski $_{\lambda^\omega}(S)$ is similar to the axiomatization of Tarski $_{\kappa^\kappa}(S)$, with natural modifications to accomodate higher types. We define Tarski $_{\lambda^\omega}^\omega(\mathbf{Z}_1^i)$ by

induction as usual. By a more complicated but not essentially different argument from the proof of Theorem 2.9, we obtain the following result.

Theorem 2.10. $\text{Tarski}_{\lambda^\omega}^\omega(Z_1^i)$ *proves transfinite induction up to any ordinal less than $\phi_{\Omega^\omega}(0)$ for all formulas in its language.*

In this proof the analog of Lemma 2.7 asserts that for each $k \prec \omega$, $\text{Tarski}_{\lambda^\omega}(S)$ proves $\text{Acc}(\lambda^{\lambda^k})$. Then the analog of Lemma 2.8 asserts that $\text{Tarski}_{\lambda^\omega}(S)$ plus transfinite induction up to $\tilde{\delta}_n$ for all formulas in its language proves transfinite induction up to $\tilde{\delta}_{n+1}$ for all formulas in the language of S , where $\tilde{\delta}_1 = 1$ and $\tilde{\delta}_{n+1} = \phi_{\Omega^k \delta_n}(0)$. (This proof is carried out using truth and acceptability at level $\lambda^{\lambda^{k-1}}\omega$.)

As with $\text{Tarski}_{\Gamma_0}^\omega(Z_1^i)$ we can go one step further and conclude that the Veblen ordinal is itself predicatively provable. Although $\phi_{\Omega^\omega}(0)$ is not terribly large among the scale of proof-theoretically important countable ordinals, this result is still significant. Probably the most celebrated example of an allegedly impredicative mainstream theorem, Kruskal’s theorem (see, e.g., [21]), is now seen to be predicatively justified. It is equivalent over a weak base system to the well-ordering of a notation for $\phi_{\Omega^\omega}(0)$ [38].

It should not be difficult to strengthen Theorem 2.10 so as to prove the well-foundedness of a notation for the “large” Veblen ordinal $\phi_{\Omega^\omega}(0)$. But I expect that substantially larger ordinals can be accessed using predicative methods. This raises the possibility of a version of Hilbert’s program in which theories are justified via predicative, rather than finitary or intuitionistic, consistency proofs. The preceding results indicate that this program is interesting, substantial, and open to exploration. Moreover, if predicativism given the natural numbers — or, as I prefer to call it, *mathematical conceptualism* [46] — is *right*, this program is of fundamental significance for the foundations of mathematics.

1. Other concerns may include the awkwardness of predicative systems in practice and a sense that they are philosophically, as opposed to mathematically, too limiting. I address these issues in separate papers [46, 47].
2. Yet another idea is to assert that we merely *believe* that A could come to accept every statement in \mathcal{S} , but we do not *know* this. If so, it is possible that A could indeed share this belief, but without sufficient certainty to allow him to go beyond \mathcal{S} .

Whether this tactic could work depends on exactly why we have reservations about what A can accept. It is no good, for example, to say that we are not sure what an ideal predicativist can accept because there is more than one version of predicativism, since we can hardly assume that he is unable to decide which version he prefers. Nor would the argument hold up if our uncertainty were caused by not knowing whether A could accept some specific principle P , as this would be tantamount to A failing to decide between two versions of his theory.

The argument might succeed if there were an infinite sequence of principles (P_n) each one of which A might accept or reject. A case could then be made that we cannot demand that he make a simultaneous decision on the validity of every P_n . However, this now seems to be a version of the idea that A can accept each P_n individually but not the entire sequence (P_n) , which is the sort of claim I address in the main text.

3. The H system in [10] follows this description precisely. Systems of ramified analysis like Σ and R are a little more complicated in that each S_a has its own set variables X^a , and legal formulas of S_a must contain only set variables X^b with $b \preceq a$. These systems are formally more complicated than H , but they are supposed to more transparently model the intuition of a predicative universe which is only available in stages.
4. According to the proof sketched in [10] that $\Gamma_0 \leq \overline{\text{Aut}}(S)$, we can find $r \in \omega$ which is the Gödel number of a recursive function $\{r\}$ with $\{r\}(n)$ a notation for γ_n , $\{r\}(n) <_{\mathcal{O}} \{r\}(n+1)$, and

$S_0 \vdash (\forall n) \text{Prov}_{S_{\{\bar{r}\}(n)}}(\ulcorner I(\{\bar{r}\}(\overline{n+1})) \urcorner)$. Letting $\mathcal{A}(n) \equiv I(\{\bar{r}\}(n+1))$ and substituting $\{\bar{r}\}(n)$ for a in (**), a simple induction argument yields $(\forall n) I(\{\bar{r}\}(n))$ (note that S_0 supports complete induction), from which we deduce $I(\bar{a})$ with $a = 3 \cdot 5^r$.

(The expression $\mathcal{A}(\{\bar{r}\}(x))$ should be understood as an abbreviation of a formula which asserts that there exists y such that $\{r\}(x) = y$ and $\mathcal{A}(y)$. Alternatively, we can assume a language that contains symbols for all primitive recursive functions and reword the arguments — here and below — to ensure that all recursive functions in use are actually primitive recursive.)

5. In [17] Feferman refers to “the argument that the characterization of predicativity requires one to go beyond predicative notions and principles” ([17], footnote 6), which sounds like it could be a version of the general objection of §1.3. However, his response (“But the predicativist . . .”, p. 316) seems aimed merely at showing that the set of all predicatively provable ordinals is not a predicatively valid set, a view that I agree with (though not for the reason given there). This should not prevent a predicativist from understanding the assertion that every a_n is an ordinal notation, in the notation of §1.3.

One could possibly make an argument that the statement $(\forall n) I(a_n)$ cannot even be predicatively recognized as meaningful, let alone true, on the ground that the general concept of well-ordering is not available to a predicativist. Perhaps this is the point of the comment in [17]. Presumably the idea would be that each $I(a_n)$ can only be understood as a sensible assertion *once it is proven* and not before. This seems like a difficult position to defend, but in any case it would void the main argument because if one did accept that some theorems of S_{a_n} cannot be recognized as meaningful until they are actually proven, this would invalidate any use of reflection principles in the first place.

6. Note that the final \mathcal{L} in rule V, predicate substitution ([13], p. 78), should be \mathcal{L}_\exists .

7. Of course, the validity of the functional generating procedure hinges on the validity of \exists/P , so it may be significant that Feferman refers to “the correctness of P ” and not “the correctness of P in conjunction with \exists/P ”. This goes back to the question raised in §1.6 (a) about whether predicativists can trust theorems proven in \exists/P , and if not, why it makes sense for them to use this system at all.

8. For the argument to work we have to be able to imagine someone who can think, for example,

whenever it is the case that for every number n and any $\bar{\alpha}$ there exists a β satisfying $\mathcal{A}(\bar{\alpha}, n, \beta)$, for any $\bar{\alpha}$ these β 's can be unified into a single γ satisfying $\mathcal{A}(\bar{\alpha}, n, \gamma_n)$ for all n

but who cannot think

whenever for any $\bar{\alpha}$ a unique β exists satisfying $\mathcal{A}(\bar{\alpha}, \beta)$, I can introduce a functional symbol F such that $\mathcal{A}(\bar{\alpha}, F(\bar{\alpha}))$ holds for any $\bar{\alpha}$,

yet who *can* think

I can introduce a functional symbol F such that $\mathcal{A}(\bar{\alpha}, F(\bar{\alpha}))$ holds for any $\bar{\alpha}$

once he has actually proven, for any particular \mathcal{A} , the existence for any $\bar{\alpha}$ of a unique β satisfying $\mathcal{A}(\bar{\alpha}, \beta)$. This combination of abilities and deficits strikes me as incoherent.

9. Another questionable point in the “too weak” category is the restriction on allowed types in ([19], p. 81). I do not understand the justification given there, and a corresponding restriction is not made in the system sketched in [16]. In light of footnote 2 of [19], this raises the question whether $\mathcal{U}(\text{NFA})$ as described in [16] really does have proof-theoretic ordinal Γ_0 .

10. On the other hand, the minimality property of LFP (Ax 4 (ii), p. 79) is never used in [19], so this axiom could be eliminated without affecting the proof-theoretic strength of $\mathcal{U}(\text{NFA})$. The existence of not necessarily minimal fixed points might be predicatively justifiable if intuitionistic logic is used; see the discussion of inductively defined classes at the beginning of §2.6. However, this more careful analysis also reveals a fundamental impredicativity in using schematic variables, a point not discussed above; see §2.5.

11. I should point out that Feferman has in several places openly called attention to impredicative aspects of various of his systems. The impredicativity of the autonomous systems is commented on in ([13], p. 85), ([14], p. 3), and elsewhere. (“the well-ordering statement . . . on the face of it *only impredicatively justifies* the transfinite iteration of accepted principles up to a .” “. . . *prima facie* impredicative notions such as those of ordinals or well-orderings.”) The impredicativity of $P + \exists/P$ is noted in ([13], p. 92). (“In P we think of ‘ X ’ as ranging over predicates recognized to have a definite meaning; this would not seem to admit the properties expressed by formulas of \mathcal{L}_\exists .”) The impredicativity of $\text{Ref}^*(\text{PA}(P))$ is noted in ([14], pp. 41 and 42). (“one may question

substituting possibly indeterminate formulas . . . this seems to me to be the weakest point of the case for reflective closure having fundamental significance.” “this may involve some equivocation between the notions of being definite . . . and being determinate”)

12. (Perhaps also assuming that we know Y to be countable so that the subset can be extracted using principle (iii); but see below.)

13. I am deliberately avoiding the question of what sets “really are”. The argument in this paragraph suggests a quasi-physical conception according to which one could imagine actually manipulating the elements of a set. I see nothing wrong with this sort of conception, at least for subsets of ω , but it is not essential for what follows. The important point is that sets of numbers, whatever one takes them to be, should in principle always be unequivocally recognizable as such. This ought to be true on any reasonable conception (but probably would not be true, for example, if one identified “set of numbers” with “intelligible property of numbers”; see §2.5).

14. Since this statement itself refers to “all sets”, one could ask whether a predicativist could accept it without falling into contradiction, but I suppose he could have a reasonably clear idea of what is forbidden despite being unable to formally define it. In any case, the statement is made for the benefit of non-predicativists. A predicativist should not need to be explicitly forbidden from talking about “all sets” since the concept would make no sense to him and it should not even occur to him to speak this way. A similar comment could apply to the vicious-circle principle.

15. There are several predicatively equivalent versions of this condition. In intuitionistic logic with arithmetical comprehension, the numerical omniscience schema, and $\mathcal{A} \vee \neg \mathcal{A}$ for all atomic \mathcal{A} , for any $a \in \omega$ and any ordering on ω the statement (1) $(\forall X)TI(X, a)$ is equivalent to (2) the assertion that $\{b : b \prec a\}$ has no proper progressive subsets and also to (3) the assertion that for all X , if there exists $b \prec a$ in X then there is a least such b . Assuming dependent choice for arithmetical formulas, the preceding are also equivalent to (4) the assertion that every decreasing sequence in $\{b : b \prec a\}$ is eventually constant and (5) the assertion that there is no strictly decreasing sequence in $\{b : b \prec a\}$.

16. According to reference [12] it is the Δ_0 collection schema which makes the KP axioms impredicative. Footnote 7 of [12] refers to [29] for justification of this point, but the relevant comment in footnote 4 of [29] explicitly locates impredicativity in the fact that “the interpretation of the logical constants, in particular of \rightarrow , is classical”. This seems to imply that if intuitionistic logic were used then the KP axioms would be predicatively valid, so that weakening the logical axioms from classical to intuitionistic would render acceptable non-logical axioms which allow one to access ordinals beyond Γ_0 . Apparently this possibility was never pursued.

17. In an earlier version of this paper I suggested that T could alternatively be interpreted as meaning “provable in S augmented by an infinitary ω -rule”. However, this is not helpful because it is ambiguous about exactly which proof trees would be covered. If we allow all proof trees that are well-founded in the sense of admitting induction, then the predicativist should have the same difficulty accepting the Tarskian implication $T(\ulcorner \mathcal{A} \urcorner) \rightarrow \mathcal{A}$ as he has in accepting the condition (*) discussed in Section 1.4; the Tarskian hierarchies would then be impredicative in the same way as Feferman’s autonomous systems. To avoid this difficulty we would have to insist on using proof trees that are well-founded in a strong enough sense to admit an inductive proof that any theorem proven along such a tree is actually true. But this requires explicitly using the concept of truth, which is what the alternative interpretation of T was meant to avoid.

REFERENCES

- [1] A. Antonelli, Review of Solomon Feferman, *In the Light of Logic*, *Bull. Symbolic Logic* **7** (2001), 270-277.
- [2] A. Arana, Review of Solomon Feferman, *In the Light of Logic*, *Mathematical Intelligencer*, to appear.
- [3] J. Avigad, On the relationship between ATR_0 and $\widetilde{ID}_{<\omega}$, *J. Symbolic Logic* **61** (1996), 768-779.
- [4] ———, Review of Solomon Feferman, *In the Light of Logic*, *Journal of Philosophy* **96** (1999), 638-642.
- [5] ———, An ordinal analysis of admissible set theory using recursion on ordinal notations, *J. Math. Log.* **2** (2002), 91-112.

- [6] J. Avigad and R. Sommer, The model-theoretic ordinal analysis of predicative theories, *J. Symbolic Logic* **64** (1999), 327-349.
- [7] W. Buchholz, S. Feferman, W. Pohlers, and W. Sieg, *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies*, Lecture Notes in Mathematics **897**, Springer-Verlag, 1981.
- [8] T. Coquand and E. Palmgren, Intuitionistic choice and classical logic, *Arch. Math. Logic* **39** (2000), 53-74.
- [9] S. Feferman, Transfinite recursive progressions of axiomatic theories, *J. Symbolic Logic* **27** (1962), 259-316.
- [10] ———, Systems of predicative analysis, *J. Symbolic Logic* **29** (1964), 1-30.
- [11] ———, Autonomous transfinite progressions and the extent of predicative mathematics, in *Logic, Methodology and Philos. Sci. III (Proc. Third Internat. Congr., Amsterdam, 1967)*, pp. 121-135, North-Holland, 1968.
- [12] ———, Predicatively reducible systems of set theory, in *Axiomatic Set Theory (Proc. Symp. Pure Math. XIII, Vol. 2)*, pp. 11-32, 1974.
- [13] ———, A more perspicuous formal system for predicativity, in *Konstruktionen versus Positionen*, Vol. I, pp. 68-93, Walter de Gruyter, 1979.
- [14] ———, Reflecting on incompleteness, *J. Symbolic Logic* **56** (1991), 1-49.
- [15] ———, Why a little bit goes a long way: Logical foundations of scientifically applicable mathematics, in *PSA 1992*, Vol. II, pp. 442-455, 1993. Reprinted as Chapter 14 in *In the Light of Logic: Logic and Computation in Philosophy*, Oxford University Press, 1998.
- [16] ———, Gödel's program for new axioms: why, where, how and what? in *Gödel '96 (Brno, 1996)*, pp. 3-22, Lecture Notes in Logic **6**, Springer-Verlag, 1996.
- [17] ———, Comments on "Predicativity as a philosophical position" by G.R. Hellman, *Revue Internationale de Philosophie* **229** (2004), 313-323.
- [18] ———, Predicativity, in *The Oxford Handbook of the Philosophy of Mathematics and Logic*, pp. 590-624, Oxford University Press, 2005
- [19] S. Feferman and T. Strahm, The unfolding of non-finitist arithmetic, *Ann. Pure Appl. Logic* **104** (2000), 75-96.
- [20] H. M. Friedman, Internal finite tree embeddings, in *Reflections on the Foundations of Mathematics (Stanford, CA, 1998)*, pp. 60-91, Assoc. Symbol. Logic, 2002.
- [21] J. H. Gallier, What's so special about Kruskal's theorem and the ordinal Γ_0 ? A survey of some results in proof theory, *Ann. Pure Appl. Logic* **53** (1991), 199-260.
- [22] H. Gerber, An extension of Schütte's Klammersymbols, *Math. Ann.* **174** (1967), 203-216.
- [23] G. Hellman, Predicativism as a philosophical position, *Revue Internationale de Philosophie* **229** (2004), 295-312.
- [24] W. A. Howard, Some proof theory in the 1960's, in *Kreiseliana*, pp. 275-288, A. K. Peters, 1996.
- [25] G. Jäger, R. Kahle, A. Setzer, and T. Strahm, The proof-theoretic analysis of transfinitely iterated fixed point theories, *J. Symbolic Logic* **64** (1999), 53-67.
- [26] R. Kahle, Mathematical proof theory in the light of ordinal analysis, *Synthese* **133** (2002), 237-255.
- [27] G. Kreisel, Ordinal logics and the characterization of informal concepts of proof, in *Proc. Internat. Congress Math. 1958*, pp. 289-299, Cambridge Univ. Press, 1960.
- [28] ———, La Prédicativité, *Bull. Soc. Math. France* **88** (1960), 371-391.
- [29] ———, The axiom of choice and the class of hyperarithmetic functions, *Nederl. Akad. Wetensch. Proc. Ser. A* **65** [*Indag. Math.* **24**] (1962), 307-319.
- [30] ———, Foundations of intuitionistic logic, in *Logic, Methodology and Philos. Sci. (Proc. 1960 Internat. Congr.)*, pp. 198-210, Stanford University Press, 1962.
- [31] ———, Mathematical logic, in *Lectures on Modern Mathematics*, Vol. III, pp. 95-195, Wiley, 1965.
- [32] ———, Principles of proof and ordinals implicit in given concepts, in *Intuitionism and Proof Theory (Proc. Conf., Buffalo, N. Y., 1968)*, pp. 489-516, North-Holland, 1970.
- [33] S. Kripke, Outline of a theory of truth, *Journal of Philosophy* **72** (1975), 690-716.
- [34] L. W. Miller, Normal functions and constructive ordinal notations, *J. Symbolic Logic* **41** (1976), 439-459.

- [35] C. Parsons, Realism and the debate on impredicativity, 1917-1944, in *Reflections on the Foundations of Mathematics (Stanford, CA, 1998)*, pp. 372-389, Assoc. Symbol. Logic, 2002.
- [36] W. Pohlers, Pure proof theory: aims, methods and results, *Bull. Symbolic Logic* **2** (1996), 159-188.
- [37] M. Rathjen, Predicativity, circularity, and anti-foundation, in *One Hundred Years of Russell's Paradox*, pp. 191-219, de Gruyter, 2004.
- [38] M. Rathjen and A. Weiermann, Proof-theoretic investigations on Kruskal's theorem, *Ann. Pure Appl. Logic* **60** (1993), 49-88.
- [39] C. Rüede, Transfinite dependent choice and ω -model reflection, *J. Symbolic Logic* **67** (2002), 1153-1168.
- [40] K. Schütte, Predicative well-orderings, in *Formal Systems and Recursive Functions (Proc. Eighth Logic Colloq., Oxford, 1963)*, pp. 280-303, North-Holland, 1965.
- [41] ———, Eine Grenze für die Beweisbarkeit der transfiniten Induktion in der verzweigten Typenlogik, *Arch. Math. Logik Grundlagenforsch* **7** (1965), 45-60.
- [42] ———, *Proof theory*, Grundlehren der Mathematischen Wissenschaften **225**, Springer-Verlag, 1977.
- [43] S. G. Simpson, *Subsystems of Second Order Arithmetic*, Springer-Verlag, 1999.
- [44] ———, Predicativity: the outer limits, in *Reflections on the Foundations of Mathematics (Stanford, CA, 1998)*, pp. 130-136, Assoc. Symbol. Logic, 2002.
- [45] T. Strahm, Wellordering proofs for metapredicative Mahlo, *J. Symbolic Logic* **67** (2002), 260-278.
- [46] N. Weaver, Mathematical conceptualism, manuscript.
See <http://www.math.wustl.edu/~nweaver/conceptualism.html>
- [47] ———, Analysis in J_2 , manuscript.
See <http://www.math.wustl.edu/~nweaver/conceptualism.html>

DEPARTMENT OF MATHEMATICS, WASHINGTON UNIVERSITY IN SAINT LOUIS, SAINT LOUIS, MO
63130

E-mail address: nweaver@math.wustl.edu