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 develop a general method for accessing ordinals which is meant to be predicatively valid. We find that the small Veblen ordinal $\phi_{\Omega^\omega}(0)$, and probably much larger ordinals, are predicatively provable.

The precise delineation of the extent of predicative reasoning is a remarkable modern result in the foundations of mathematics. Building on ideas of Kreisel [27, 28], Feferman [11] and Schütte [41, 42] 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 [11] in particular having been cited with approval in virtually every discussion of predicativism for nearly sixty years. Γ_0 is now commonly referred to as "the ordinal of predicativity". Some publications which explicitly make this assertion are [1, 2, 3, 4, 6, 19, 20, 21, 23, 25, 26, 37, 38, 40, 45, 46].

This achievement is notable both for its technical interest 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, Skolem, and Weyl. (Historical overviews are given in [19] and [35].) Its central principle — that sets have to be "built up from below" — is, on its face, reasonable and attractive.

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 [16, 44], 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 not, predicativism could be more viable than previously thought and its current peripheral status in the philosophy of mathematics may need to be reconsidered.

Date: *February 25, 2023.*

[1]Other concerns may include questions about the ease of use of predicative systems in practice or a sense that they are philosophically, as opposed to mathematically, too limiting.

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.[2] It is my hope that this conclusion will open the way to a serious reappraisal of the significance and interest of predicativism.

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. [11], p. 13 or [43], 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 §§1.1, 1.5, and 2.3). Previous discussions of predicativism have tended to ignore these distinctions, and this will emerge as a crucial source of confusion (see §§1.1 and 1.5). To fix ideas I will use the term "well-ordered" to mean of a set X that it is equipped with a total order with respect to which, for any $Y \subseteq X$, if Y 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.[3] 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, in the same way that there is no coherent version of finitism which

[2] Similar conclusions are reached in [7].

[3] Feferman objected to this characterization when I posted the original version of this paper on the arXiv. In an email he sent me (which he gave me permission to make public) he wrote "Both Kreisel and I have raised critical questions about the proposed characterization at different stages, so it is not fair to say without qualification that we forcefully defended it." It is indeed true that he raised critical questions; see footnote 15 below. I leave it to the reader to judge whether it remains fair to say that he forcefully defended the Γ_0 thesis.

would lead one to accept every number less than 1,000 but not 1,000 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. Of course, evaluating the justice of this claim does require some understanding of predicativism.)

1.1. **The principal error.** The Γ_0 thesis is supported by a number of formal systems which have been proposed in the primary literature as models of predicative reasoning. However, I anticipate that most readers will be most familiar with the treatment of the Γ_0 thesis given in a secondary source such as the excellent graduate text [36]. So let me start with a short explanation of the problem with the analysis presented there. Then I will go on to review the formal systems which appear in the primary literature.

Just above I introduced the notion of progressivity for subsets of totally ordered sets (for every $a \in X$, if $\{b \in X : b \prec a\} \subseteq Y$ then $a \in Y$) and defined a totally ordered set to be well-ordered if progressivity of $Y \subseteq X$ entails that $Y = X$. This is easily seen to be equivalent to the usual definition in terms of every nonempty subset having a least element (and the trivial proof of this equivalence is predicatively unproblematic).

Let me also define a property P to be progressive on X if $(\forall b \prec a)P(b)$ implies $P(a)$, for all $a \in X$. This reduces to the previous setwise condition when $P(a)$ is the property "$a \in Y$" for some $Y \subseteq X$. So if X is well-ordered, any progressive property of this form will satisfy $(\forall a \in X)P(a)$. Can we draw the same conclusion for other properties? That is, if X is well-ordered and P is any progressive property, may we conclude $(\forall a \in X)P(a)$?

Yes, we can, by the simple technique of introducing the set $Y = \{a \in X : P(a)\}$. Progressivity of the property P then entails progressivity of Y in the first, setwise sense. If X is well-ordered then this yields $Y = X$, i.e., $(\forall a \in X)P(a)$.

But this is an argument that predicativists cannot generally make. Introducing the set Y requires a comprehension principle which is impredicative unless P has a special form. Predicativism famously disallows, for instance, any definition of a set of natural numbers that involves quantification over all sets of natural numbers. The comprehension principles available to predicativists are very weak.

This matters for the analysis presented in [36] and elsewhere. That analysis involves proof trees of infinite (well-ordered) heights over a straightforwardly predicative base system. The idea is that one starts with proof trees of height ε_0, say, but once one has proven that a notation for α is well-ordered — so that α is predicatively "recognized" to be an ordinal — one is allowed to use proof trees of height α. If we define $\gamma_0 = 0$ and $\gamma_{n+1} = \phi_{\gamma_n}(0)$ for $n \geq 0$, where ϕ_α are the Veblen functions, then, for all n, using a proof tree of height γ_n one can prove that a notation for γ_{n+1} is well-ordered. Thus, iterating this process, every ordinal less than $\Gamma_0 = \sup \gamma_n$ is supposed to be predicatively provable. Conversely, within the context of the infinitary systems used in [36] one cannot prove that a notation for Γ_0 is well-ordered using a proof tree of any height less than Γ_0, and this allegedly shows that Γ_0 is the limit of predicative reasoning.

An obvious question is why predicativists cannot recognize for themselves that this process yields a proof, valid for all n, that some notation for γ_n is well-ordered, and infer from this that $\Gamma_0 = \sup \gamma_n$ is well-ordered. It seems like just the sort of countable reasoning that predicativists are good at. This question was raised in [24], among other places. I will return to it in a moment. But first, let us ask how the predicativist is to infer the soundness of proof trees of height α from the fact that α is well-ordered. Why should it follow that such trees prove true theorems? It is easy to see that soundness is progressive — if every proof tree of height less than α is sound, then every proof tree of height α is sound — so one wants to make an inductive argument with the property $P(a)$ = "proof trees of height α are sound", where a is a notation for α.

But all we prove with a tree of height γ_n is that a notation for γ_{n+1} is well-ordered in the weak, setwise sense. Lacking an impredicative comprehension principle, we cannot infer the soundness of proof trees of height γ_{n+1}. The Feferman-Schütte analysis relies essentially on an impredicative inference from setwise well-ordering to induction for general properties.

I have not seen any attempt, anywhere, to actually give a predicatively valid argument that would allow one to infer the soundness of proof trees whose height is known to be (setwise) well-ordered — doubtless because the task is hopeless. However, several correspondents (but not Feferman) have suggested to me that a predicativist can simply intuit this inference without needing to prove it. This response runs into the obvious question mentioned above, about predicativists being able to recognize that for all n they can prove a notation for γ_n is well-ordered. They would then be able to get beyond Γ_0. So if one wants to stop exactly at Γ_0, one has to postulate that predicativists are able to intuit the desired inference in any particular case, but they cannot recognize that they have this general ability, as this would allow them to get past Γ_0. For every n, when it is proven that a notation for γ_n is well-ordered, the revelatory intuition that proof trees of that height are sound has to come as a surprise. This is the crevice into which defenses of the Γ_0 thesis have to squeeze.

In a nutshell: the crucial inference from "a notation for γ_n is well-ordered" to "proof trees of height γ_n are sound" requires a comprehension principle that is not available to the predicativist. But if this inference could somehow be recognized as generally valid, one would then be able to see all at once that notations of γ_n for all n are well-ordered, and infer from this that Γ_0 is well-ordered. The only way someone could get everything less than Γ_0, but not Γ_0 itself, is if they could somehow recognize the correctness of the crucial inference in each instance but not as a general principle.

A comment made in [14][4] obliquely indicates that Feferman may have, at the time of writing that paper, been aware of the difficulty. However, he did not suggest any way to resolve it, arguing instead (wrongly; see §1.7) that the new system presented in [14] could legitimately reach the desired conclusion. Over the next twenty years or so he went on to propose several distinct analyses which were supposed to justify the Γ_0 thesis. So I will have to discuss these.

[4][14], p. 85; see footnote 15 below.

1.2. **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 [11], RA^* [43], $P+\exists/P$ [14], $\mathrm{Ref}^*(\mathrm{PA}(P))$ [15], and $\mathcal{U}(\mathrm{NFA})$ [20]. 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 predicativists have proven the well-foundedness of a set of order type α they are 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 $\mathcal{A}(\overline{n})$ with $\overline{n}$ ranging over all numerals. In the first three systems all proofs are finite and the key proof principle is a "formalized ω-rule" scheme 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 $\mathcal{A}(\overline{n})$. This leads to a hierarchy of systems S_a where a is an ordinal notation and $\mathrm{S}_{a\oplus 1}$ incorporates a formalized ω-rule scheme 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 $\mathrm{Ref}^*(\mathrm{PA}(P))$ is obtained by applying a general construction Ref^* to a "schematic" form $\mathrm{PA}(P)$ of Peano arithmetic. This construction involves extending the language of $\mathrm{PA}(P)$ 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 $\mathrm{PA}(P)$. 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 $\mathrm{Ref}^*(\mathrm{S}(P))$ embodies what one "ought to accept" given that one accepts a schematic theory $\mathrm{S}(P)$, and an argument can then be made that predicativism is fundamentally based on Peano arithmetic and therefore $\mathrm{Ref}^*(\mathrm{PA}(P))$ precisely captures what a predicativist ought to accept.

Like $\mathrm{Ref}^*(\mathrm{PA}(P))$, $\mathcal{U}(\mathrm{NFA})$ 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 ([20], 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, $\mathcal{U}$(NFA) is already supposed to capture predicative reasoning.

1.3. **Outline of the critique.** All of the proposed formalizations of predicative reasoning cited in §1.2 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.[5]

Nonetheless, each of the formal systems in §1.2 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 §1.4; 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 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. $\mathrm{Ref}^*(\mathrm{PA}(P))$ makes truth claims about schematic predicates which do not make sense unless one assumes an impredicative comprehension axiom. $\mathcal{U}$(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 $\mathcal{U}$(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 §1.10.

There are also more subtle instances of impredicativity which occur in the use of self-applicative schematic predicates in $\mathrm{Ref}^*(\mathrm{PA}(P))$ and $\mathcal{U}$(NFA); see §2.4.

[5] Unless, of course, they were deliberately engineered to get that answer. One may suspect this of all of them aside from the autonomous systems; it is most obvious in the case of $P + \exists/P$.

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

1.4. **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, perhaps, 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 maybe the disciples of some foundational philosophy can indeed see, as we do, that there exists a proof that they would accept for each statement in $\mathcal{S}$, but they 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 them to take this step). Or it might be possible to identify some special limitation in their belief system which prevents them from grasping the validity of all of $\mathcal{S}$ at once despite their ability to accept each statement in $\mathcal{S}$ individually.[6]

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 their perspective. But it is more difficult to reconcile this with the claim that we *do* know that A 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 scheme 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 predicativists would accept each proof in the scheme but they cannot see this.

In fact this is highly implausible, for the following reason. 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 these proofs are all essentially the same. It is therefore hard to understand why someone who is presumed to grasp induction on ω (and even, allegedly, in "schematic" form [15, 17, 20]) would not be able to infer the single assertion that a_n is an ordinal notation for all n.

[6]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 legitimize going 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; in that case, there would simply be more than one kind of ideal predicativist, and the objection would apply to them all separately. In any case, the arguments which have been brought forward to justify the Γ_0 thesis all very explicitly purport to show that for all n the predicativist can *prove* that a notation for γ_n is well-ordered.

It is reasonable to expect that if predicativists understand 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 they *can* infer the statement that every a_n is an ordinal notation. As this would enable them 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 they 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 in the next section, totally unpersuasive.

I now turn to the systems introduced in §1.2.

1.5. **The finitary autonomous systems.** The initial idea behind the finitary autonomous systems Σ, H, and R is that if predicativists trust some formal system for second order arithmetic, such as ACA, say (see [44]), then they should accept not only the theorems of the system itself, but also additional statements such as Con(ACA) which reflect the fact that the axioms are true. Feferman [10] analyzed several such "reflection principles" and found the strongest of them to be the formalized ω-rule scheme

$$(\forall n)\,[\mathrm{Prov}(\ulcorner \mathcal{A}(\overline{n})\urcorner) \to \mathcal{A}(n)]\,,$$

where $\ulcorner \mathcal{A}(\overline{n})\urcorner$ is the Gödel number of $\mathcal{A}(\overline{n})$ and Prov formalizes "is the Gödel number of a provable formula" (here, provable in ACA).

Having accepted this scheme, the argument runs, predicativists are then committed to a stronger system consisting of ACA plus the ω-rule scheme, and they should therefore now accept a version of the formalized ω-rule scheme 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 predicativists should accept the system indexed by a when and only when they have a prior proof that a is an ordinal notation.[7]

Feferman [11] 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 $a_1, \ldots, a_n$ 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 ($1 \le i < n$). 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. I already mentioned that in Section 1.1. 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

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

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

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

[7] The H system in [11] 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.

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 $\mathrm{Prov}_{\mathrm{S}_b}(\ulcorner\mathcal{A}\urcorner)$ 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 predicativists could somehow infer the soundness of S_a then they 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, predicativists are (somehow) always able to make the deduction

$$\text{from} \quad I(\overline{a}) \quad \text{and} \quad \mathrm{Prov}_{\mathrm{S}_a}(\ulcorner\mathcal{A}(\overline{n})\urcorner), \qquad \text{infer} \quad \mathcal{A}(\overline{n}), \qquad (*)$$

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

$$(\forall a)(\forall n)\,[I(a) \wedge \mathrm{Prov}_{\mathrm{S}_a}(\ulcorner\mathcal{A}(\overline{n})\urcorner) \to \mathcal{A}(n)] \qquad (**)$$

for any formula $\mathcal{A}$?

As a straightforward consequence of [11], one can use $(**)$ to prove $I(\overline{a})$ with a some standard notation for Γ_0.[8] The claim must therefore be that predicativists can recognize each instance of $(*)$ to be valid but cannot recognize the validity of the general assertion $(**)$. In other words, whenever they have proven that a is an ordinal notation they can infer the statement that all theorems of S_a hold, but they do *not* accept the general 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 to be able to make the inference $(*)$ in any instance.

---

[8]According to the proof sketched in [11] that $\Gamma_0 \leq \overline{\mathrm{Aut}}(\mathrm{S})$, 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 $\mathrm{S}_0 \vdash (\forall n)\,\mathrm{Prov}_{\mathrm{S}_{\{\overline{r}\}(n)}}(\ulcorner I(\{\overline{r}\}(\overline{n+1}))\urcorner)$. Letting $\mathcal{A}(n) \equiv I(\{\overline{r}\}(n+1))$ and substituting $\{\overline{r}\}(n)$ for a in $(**)$, a simple induction argument yields $(\forall n)\,I(\{\overline{r}\}(n))$ (note that S_0 supports complete induction), from which we deduce $I(\overline{a})$ with $a = 3 \cdot 5^r$.

(The expression $\mathcal{A}(\{\overline{r}\}(x))$ 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 use 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.)

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 §1.5 (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 predicativists, "truth" and "predicative truth" are the same thing, so it would not make sense to suggest that they 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 ([12], 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 §1.4 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_{γ_n+2}, let alone all the way to $N_{\gamma_{n+1}}$, because the construction of N_{γ_n+2} uses sets in N_{γ_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 predicativists could establish the existence of N_ω. Starting with $N_0 = \emptyset$, they can use the basic step to directly pass to N_1, then to N_2, and so on, so that for each $n \in \omega$ they can give a finite proof of the existence of N_n. But in order to accept the existence of N_ω they have to somehow globally grasp that N_n exists for all n *without* sequentially proving their existence one at a time. Presumably they 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 they *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 is surely not predicatively entailed by $I(a)$ (cf. the end of §2.3). 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 completely hypothetical because the alleged conversion process has not been identified.

(d) *Feferman's position.* In [14] 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 [$+\exists/P$]" ([14], p. 85). Similar comments appear in ([15], p. 3) and ([19], p. 24). It is certainly true that the systems $P + \exists/P$, $\mathrm{Ref}^*(\mathrm{PA}(P))$, and $\mathcal{U}$(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 [15] 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.[9]

1.6. **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, predicativists do not actually reason within any of these infinitary systems, but they believe that in principle these systems 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 predicativists to accept the well-foundedness of (some notation for) α *only if they knew this fact*. But the only way they could know what is provable in an infinitary system is via some kind of meta-argument about what is provable in that system. This immediately suggests that they should be able to get beyond Γ_0 by performing a single act of reflection on the finitary system in which they actually reason.

We can now see that just as in the case of the finitary autonomous systems one is faced with a dilemma: (1) why should predicativists 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 they can draw this inference for any particular α, why do they 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.

[9]In [18] Feferman refers to "the argument that the characterization of predicativity requires one to go beyond predicative notions and principles" ([18], footnote 6), which sounds like it could be a version of the general objection of §1.4. 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 §1.4.

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 grounds that the general concept of well-ordering is not available to a predicativist. Perhaps this is the point of the comment in [18]. 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.

Working only in Peano arithmetic, predicativists 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, they need some new principle going beyond Peano arithmetic. Augmenting PA with an axiom scheme 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 and larger ordinals. Expressing the principle as a deduction rule scheme rather than as an axiom scheme 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 additional act of reflection. One is virtually forced to assert that whenever predicativists prove that a set is well-founded they are then able to infer the validity of proof trees of that height via an unformalizable leap of intuition. I am not aware of any rational basis for such a claim.

1.7. **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 [14] 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 $\exists/P$ ([14], p. 76) and in the rules IV and V ([14], p. 78) that at least part of the premise must 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 (the allowed formulas of $\exists/P$ and rule (V)), 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 $\exists/P$. This is problematic because agreeing that P is indeed dispensable would obviate the need for the functional defining axioms altogether and thereby void Feferman's justification for not allowing a predicativist to get beyond Γ_0 by reflecting on the validity of $P + \exists/P$ (see §1.7 (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. ([14], 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 ([14], 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.[10] 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. ([14], 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 Σ^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 §1.4. 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 ([14], 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 $\mathcal{A}(\vec{\alpha}, \beta)$

[10] Note that the final $\mathcal{L}$ in rule V, predicate substitution ([14], p. 78), should be $\mathcal{L}_\exists$.

(with $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n)$, 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.[11]

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 [14]. The problem is that it tends to conflict with rule VII (relative explicit definition) and axiom VIII (unification) ([14], 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.[12]

1.8. **The system** $\mathrm{Ref}^*(\mathrm{PA}(P))$**.** The Ref^* construction applies to any schematic formal theory, but the case of interest for us is schematic Peano arithmetic $\mathrm{PA}(P)$. 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 scheme replaced by the single axiom

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

and there is an additional deduction rule scheme allowing substitution of arbitrary formulas for P. Now if $\mathrm{S}(P)$ is any schematic theory then $\mathrm{Ref}^*(\mathrm{S}(P))$ is a theory in the language of $\mathrm{S}(P)$ augmented by two predicate variables T and F whose axioms are the axioms of $\mathrm{S}(P)$ 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.

[11]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 §1.7 (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.

[12]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 $\vec{\alpha}$ there exists a unique β satisfying $\mathcal{A}(\vec{\alpha}, n, \beta)$, for any $\vec{\alpha}$ these β's can be unified into a single $\vec{\gamma}$ satisfying $\mathcal{A}(\vec{\alpha}, n, \gamma_n)$ for all n

but who cannot think

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

yet who *can* think

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

once they have actually proven, for any particular $\mathcal{A}$, the existence for any $\vec{\alpha}$ of a unique β satisfying $\mathcal{A}(\vec{\alpha}, \beta)$. This combination of abilities and deficits is patently incoherent.

(a) *Too strong.* I will discuss the clear impredicativity of schematic predicate variables in §2.4. Leaving that issue aside for now, the first point to make here is that the key axiom which distinguishes $\mathrm{Ref}^*(\mathrm{PA}(P))$ from the much weaker system $\mathrm{Ref}(\mathrm{PA}(P))$, axiom 3.2.1 $(\mathrm{i})^{(P)}$ ([15], 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 ([15], 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

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

a special case of 3.2.1 $(\mathrm{i})^{(P)}$, cannot be given the latter interpretation since the number $\ulcorner P(\overline{n})\urcorner$ 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 $(\mathrm{i})^{(P)}$, i.e., as indicating membership in a fixed set, then the substitution rule $P - \mathit{Subst} : L(P)/L(P,T,F)$ ([15], Definition 3.3.2 (iii)) can only be justified by an impredicative comprehension principle (cf. [15], p. 8).

Feferman characterizes axiom 3.2.1 $(\mathrm{i})^{(P)}$ as "relativizing T and F to P" ([15], 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 $(\Pi^0_1 - \mathrm{CA})_{<\Gamma_0} \leq \mathrm{Ref}^*(\mathrm{PA}(P))$). In fairness, I should point out that this problem is noted in ([15], §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 $\mathrm{Ref}^*(\mathrm{PA}(P))$ 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 $(\mathrm{i})^{(P)}$ and the substitution rule $P - \mathit{Subst} : L(P)/L(P,T,F)$.

(b) *Too weak.* The Ref^* construction is described in [15] 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 [15] might leave the impression that the statements $\mathcal{A}$ such that $\mathrm{Ref}^*(\mathrm{PA}(P))$ 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 $(\forall n)\,T(\{\overline{r}\}(n))$ 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. Why shouldn't we add this formula as an axiom?

The more careful formulation that the self-truth axioms "correspond directly to the informal notion of grounded truth and falsity" ([15], p. 42) is well-taken, but we must not confuse this with the claim that the self-truth axioms *capture* the informal notion.

Now consider the claim that in general $\mathrm{Ref}^*(\mathrm{S}(P))$ encapsulates what one "ought to accept" given that one has accepted $\mathrm{S}(P)$ ([15], 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 $\mathrm{Ref}^*(\mathrm{PA}(P))$ encapsulates what one ought to accept given that one has accepted $\mathrm{PA}(P)$. We may suppose that predicativists do not realize (and indeed, *ought not accept*) that their commitment to Peano arithmetic obliges them to accept every theorem of $\mathrm{Ref}^*(\mathrm{PA}(P))$, although this is in fact the case. This leads us back to the question posed in §1.4. Evidently we are dealing with a claim that predicativists can affirm each theorem of $\mathrm{Ref}^*(\mathrm{PA}(P))$ individually but cannot accept this system globally.

This point is not explicitly addressed in [15], 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 [15] that the passage from $\mathrm{S}(P)$ to $\mathrm{Ref}(\mathrm{S}(P))$ should not be iterated because this would "vitiate 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 $\mathrm{Ref}^*(\mathrm{PA}(P))$ by the single statement $(\forall n)\, T(\{\overline{r}\}(n))$ described above.

Surely predicativists *can* justify adding this statement if they are able to generally recognize that every statement proven true by $\mathrm{Ref}^*(\mathrm{PA}(P))$ is indeed true. Given that $\mathrm{Ref}^*(\mathrm{PA}(P))$ is finitely axiomatized and that the predicativist is presumed to accept each theorem of $\mathrm{Ref}^*(\mathrm{PA}(P))$ 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 [15] 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 $\mathrm{Ref}^*(\mathrm{PA}(P))$ 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 $(\forall n)\, T(\{\overline{r}\}(n))$. 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.

1.9. **The system $\mathcal{U}(\mathrm{NFA})$.** Distinct, not obviously equivalent, versions of $\mathcal{U}(\mathrm{NFA})$ are presented in [17] and [20]. I give priority to the later version in [20].

(a) *Too weak.* Like Ref^*, $\mathcal{U}$ is presented in [20] as a general construction ("unfolding") which can be applied to any schematic formal system $\mathrm{S}(P)$. As usual, granting that acceptance of $\mathrm{S}(P)$ justifies acceptance of every theorem of $\mathcal{U}(\mathrm{S}(P))$, we can ask why it fails to justify accepting a formalized ω-rule scheme referring to theorems of $\mathcal{U}(\mathrm{S}(P))$. This question is not addressed in either [17] or [20]; the closest I can find to an answer is the following passage in [17]:

> [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. ([17], p. 10)

Here reference is made to the fact that $\mathcal{U}$ uses partial operations, which is apparently seen as having fundamental significance. It is true that the question of which partial operations of $\mathcal{U}$(NFA) are total is (unsurprisingly) not decidable, though this in itself seems a questionable basis for forbidding us from proceeding beyond $\mathcal{U}$(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 scheme applied to $\mathcal{U}$(NFA), which would seem to require only that one accept $\mathcal{U}$(NFA) is sound.[13]

(b) *Way too strong.* $\mathcal{U}$(NFA) is actually impredicative in three 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 ([20], 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 Ref*(PA(P)), use of the substitution rule (Subst) ([20], p. 82) would have to presume impredicative comprehension.

The really striking impredicativity of $\mathcal{U}$(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 [8]. This not only vitiates any claim of $\mathcal{U}$(NFA) to model predicative reasoning, it more broadly undermines the idea that $\mathcal{U}$(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 [8], which is far stronger than $\mathcal{U}$(NFA).[14]

The third impredicativity arises from the use of a schematic predicate symbol P with the assumption that predicativists can understand and reason about formulas containing P, so that the allowed substitutions for P would include formulas containing P. See §2.4.

1.10. **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?

---

[13]Another questionable point in the "too weak" category is the restriction on allowed types in ([20], p. 81). I do not understand the justification given there, and a corresponding restriction is not made in the system sketched in [17]. In light of footnote 2 of [20], this raises the question of whether $\mathcal{U}$(NFA) as described in [17] really does have proof-theoretic ordinal Γ_0.

[14]On the other hand, the minimality property of LFP (Ax 4 (ii), p. 79) is never used in [20], so this axiom could be eliminated without affecting the proof-theoretic strength of $\mathcal{U}$(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 §3.

First, I stated that each of the formal systems of §1.2 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 §1.5). 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. $\mathrm{Ref}^*(\mathrm{PA}(P))$ 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 $\mathrm{Ref}^*(\mathrm{PA}(P))$. $\mathcal{U}(\mathrm{NFA})$ 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 §1.4 are brief, vague, and unpersuasive. The only such responses of which I am aware are Kreisel's answers in [27], [31], and [32] (see §1.5 (a), (b), (c)) and Feferman's answer about $P + \exists/P$ in [14] (see §1.7 (c)). In [27] and [14] 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 $\mathrm{Ref}^*(\mathrm{PA}(P))$ and $\mathcal{U}(\mathrm{NFA})$, as far as I am aware the objection has not even been discussed in the literature, except tangentially by an argument in [15] that the Ref^* construction should not be iterated.

Finally, I said that each of the formal systems of §1.2 is manifestly impredicative in some way. The most blatant example of this is the least fixed point operator in $\mathcal{U}(\mathrm{NFA})$, but in all three of $P + \exists/P$, $\mathrm{Ref}^*(\mathrm{PA}(P))$, and $\mathcal{U}(\mathrm{NFA})$ there is a basic impredicativity involving the ability to substitute second order formulas for free set variables. In $\mathrm{Ref}^*(\mathrm{PA}(P))$ and $\mathcal{U}(\mathrm{NFA})$ 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.[15]

As I mentioned in §1.3, 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 §2.3).

[15]I should point out that Feferman has in several places called attention to impredicative aspects of various of his systems. The impredicativity of the autonomous systems is commented on in ([14], p. 85), ([15], 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 ([14], 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 $\mathrm{Ref}^*(\mathrm{PA}(P))$ is noted in ([15], 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.")

There is still a question as to why so many different attempted fomalizations 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 most obvious in the case of $P + \exists/P$, with its completely unclear and unmotivated distinction between the validity of theorems proven in P versus those proven in $\exists/P$. 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 were already thought to have attained the correct answer. This could have made it difficult to free oneself from a conclusion that had already been established. However, a properly functioning scientific community should be expected to debate and criticize major ideas, not to passively accept them. This does not seem to have been done in a serious way in the present case.

In hindsight, it is perfectly clear what happened. When Feferman and Schütte first published their result about Γ_0 and autonomous systems, in the early 1960s, neither they nor Kreisel were aware of the induction versus recursion issue. Kreisel recognized the problem sometime before 1968 and Feferman sometime before 1979, but there was no public acknowledgement that the analysis was fatally flawed. Instead, Feferman spent the next 25 years trying to develop alternative routes to the same conclusion, but never really succeeding. The foundations community accepted the conclusion without challenge because it was simple and convenient.

2. The principles of predicativism

In the first part of this paper I criticized several formal systems which have been put forward as models of predicative reasoning. Part of the problem there is a lack of clarity as to what kinds of reasoning are legitimately available to the predicativist. This is especially seen in the $P + \exists/P$ systems which bizarrely disallow almost, but not quite all, use of second order quantifiers. So I would now like to make some general comments on what kinds of reasoning predicativism, as I understand it, endorses.

According to my understanding of predicativism, the key informal motivating principles are (i) the universe of sets is only available in stages, with each stage building on previous ones, and (ii) constructions, computations, definitions, etc., of length ω are legitimate.

2.1. **Predicatively valid logic.** Classical logic is not well suited to reasoning about a variable universe which is only available in stages. Since the general extension process by which new sets are recognized cannot be completely formalized, we do not expect every assertion about sets to necessarily have a well-defined truth value. Rather, we should regard the family of true statements as another 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. Predicativists accept derivations of length ω and intuitionists do not. 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. Predicatively the law of excluded middle is initially suspect for any statement that quantifies over all subsets of ω.

On the other hand, we do regard statements relativized to any well-defined partial universe as having definite truth values. For example, we can be sure 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 [9], leading to the suggestion that predicativists can adopt the *numerical omniscience scheme*

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

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

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 a set of numbers whose definition quantifies over all sets of numbers, but if we accepted $(\forall n)(\mathcal{A}_n \vee \neg\mathcal{A}_n)$ then we ought to be able to form S, for the same reason that we accept the numerical omniscience scheme. This shows that $\mathcal{A} \vee \neg\mathcal{A}$ must not be assumed to hold in every case.

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 definite truth values. To keep this distinction clear I will say an assertion is *definite* if it has a definite truth value and *meaningful* if its sense is understood. Thus, every definite assertion must be meaningful and every meaningful assertion is potentially definite.

2.2. **Second order quantification.** First order (numerical) quantification is predicatively unproblematic. The legitimacy of second order quantification is perhaps less clear since we do not regard the power set of ω as a fixed, 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 [14].

To what extent, if any, are second order quantifiers acceptable? First, because the concept "set of numbers" is predicatively cogent, 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 [14].

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 of numbers for which the assertion fails because its truth is inherent in the concept "set of numbers". As another example, given any $X \subseteq \omega$, we can surely affirm the 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 [19], 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 [14], 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 and then verifying the claimed relation between X, Y, and Z. 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.

A more natural suggestion 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.*

2.3. **Predicative well-ordering.** In the previous section I explained why a second order induction statement is predicatively legitimate. 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. However, once we have proven $\mathcal{A}(0)$ we must at least agree that this instance is definitely true. If, moreover, we have also 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 §1.7 (b) that it is generally not valid to substitute arbitrary, possibly indefinite, 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 indefinite formulas can never appear in true statements. Predicativists should accept complete induction (provided they are using intuitionistic logic) since they 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 definite.

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 affirm relatively strong versions of the statement that ω is well-ordered. If we agree with the conclusions of §2.2 then statements of transfinite induction of the form $(\forall X)\,\mathrm{TI}(X,a)$ (transfinite induction up to a on a subset of ω equipped with a total order $\preceq$) are predicatively meaningful.[16] Here I use the abbreviations

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

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 $(\forall X)\,\mathrm{TI}(X,a)$, we cannot in general infer $\mathrm{TI}(\mathcal{A},a)$ ($\equiv \mathrm{Prog}(\mathcal{A}) \to (\forall b \prec a)\mathcal{A}(b)$ where $\mathrm{Prog}(\mathcal{A}) \equiv (\forall b)[(\forall c \prec b)\mathcal{A}(c) \to \mathcal{A}(b)]$) for arbitrary formulas $\mathcal{A}$. The latter scheme is genuinely stronger because $(\forall X)\mathrm{TI}(X,a)$ only asserts induction for *sets* that are by assumption well-defined, whereas $\mathrm{TI}(\mathcal{A},a)$ can hold if $\mathcal{A}$ is not definite, and it can even be used to prove that $\mathcal{A}(b)$ *is* definite for all $b \prec a$. It may in fact be the case that whenever there is a predicatively valid proof of $(\forall X)\,\mathrm{TI}(X,a)$ there is also a proof of $\mathrm{TI}(\mathcal{A},a)$ for any meaningful formula $\mathcal{A}$. However, there is no clear reason why this should be the case, and simply passing from the weaker statement to the stronger one seems to me clearly predicatively illegitimate.

2.4. **Schematic assertions.** Even the complete induction scheme 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 meaningful way, then a predicativist should accept the induction scheme for all formulas of the new language too.

This issue is addressed in [15] and [17] 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 meaningful way, this does seem to fully capture a predicative understanding of induction on ω —

[16]There are several predicatively equivalent versions of this condition. In intuitionistic logic with arithmetical comprehension, the numerical omniscience scheme, 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) $(\forall X)\mathrm{TI}(X,a)$ 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 $\{b : b \prec a\}$.

but only from an external perspective, not to the predicativist. The problem is this. If we can accept as meaningful a formula $\mathcal{A}(P)$ which contains a schematic predicate symbol P that ranges over all formulas that we can accept as meaningful, then in particular it should apply to the case when P is replaced by $\mathcal{A}(P)$ itself. Even the meaningfulness of the original statement $\mathcal{A}(P)$ is predicatively doubtful if it could itself appear as one of its own substitution instances, just the kind of "vicious circle" that predicativism forbids. This shows yet another way in which the systems $\mathrm{Ref}^*(\mathrm{PA}(P))$ and $\mathcal{U}(\mathrm{NFA})$ are 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 "meaningful predicate" is not itself predicatively meaningful and that it is therefore not possible for a predicativist to legitimately make assertions about all meaningful predicates (cf. §2.2). 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 a 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 $(\forall n)(I_X(n) \to \mathcal{A}(n))$. 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 $\mathrm{TI}(P, a)$ 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 with intuitionistic logic, KP^i, also has a superficial predicative plausibility, but it too fails to be predicative. Here the culprit is the KP^i foundation scheme. For a statement of the KP^i 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 scheme 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^i and would implicitly make reference to the universe being defined. Thus, in order to verify the foundation scheme we would need to build up the KP^i 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.[17]

3. Truth theories

As I discussed in §§1.3 and 1.10, all of the formal systems of §1.2 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 techniques employed by those systems. My goal in this section is to develop new methods of producing predicative well-ordering proofs.

Without using some kind of reflection principle I doubt that predicativists can get very far beyond ε_0. 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 [15]. 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 §2.4 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. To the contrary, the concept of belonging to an inductively defined class does not seem predicatively objectionable on its own; for example, according to §2.3 the assertion "n is a Church-Kleene ordinal notation" is predicatively meaningful (though presumably not definite). 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 [15] 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 [15] that could

[17] According to reference [13] it is the Δ_0 collection scheme which makes the KP^i axioms impredicative. Footnote 7 of [13] 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 $\to$, is classical". This seems to imply that if intuitionistic logic were used then the KP^i 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.

be accepted as predicatively legitimate. However, such a theory would have proof-theoretic ordinal only in the neighborhood of ε_0. 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.[18]

The idea will be to work our way up the ordinals, proving not just well-ordering of their notations in the usual setwise sense, but well-ordering in the predicatively much stronger sense of supporting induction for all properties expressible in the language. In order to avoid circularities, which in general are predicatively highly problematic, any such proof will take place in a Tarskian "metasystem" which contains a truth predicate for the target system. But in order to turn this into an iterative process it is necessary to implement a hierarchy of truth predicates.

I will be working with formal systems of second order arithemtic, and the easiest way to implement a truth predicate for such a system is to introduce a predicate symbol T which takes a natural number as its argument, with the intended meaning of $T(n)$ being "the formula with Gödel number n is true". (In everything that follows, "predicate" will mean a formula with exactly one free number variable and no free set variables.) But I will pass over technicalities such as Gödel numbers in this paper. If the argument were made in greater detail we would need to be able, within the system, to perform basic syntactic constructions on numerically encoded strings. E.g., if $\ulcorner\mathcal{A}\urcorner$ and $\ulcorner\mathcal{B}\urcorner$ are the codes for two formulas $\mathcal{A}$ and $\mathcal{B}$, and $\ulcorner\mathcal{A}\wedge\mathcal{B}\urcorner$ is the code for their conjunction, then we require the ability to calculate $\ulcorner\mathcal{A}\wedge\mathcal{B}\urcorner$ from $\ulcorner\mathcal{A}\urcorner$ and $\ulcorner\mathcal{B}\urcorner$. More precisely, we must be able to express and reason about a numerical function of two arguments which yields $\ulcorner\mathcal{A}\wedge\mathcal{B}\urcorner$ when given the inputs $\ulcorner\mathcal{A}\urcorner$ and $\ulcorner\mathcal{B}\urcorner$. Probably primitive recursive arithmetic would suffice for everything we need to do along these lines. I will omit these kinds of details from the present account.

The truth predicates I will be discussing are never self-applicative; they only apply to the sentences of some target language to which they do not belong. But they are substantial; for instance, within the truth theory of the target system S we will be able to prove (a formalization of) the statement that every theorem provable in S is true. Since we can also prove in this truth theory that, say, $0 = 1$ is not true, we will then be able to derive that $0 = 1$ is not a theorem of S. In other words, the truth theory of S will always prove that S is consistent, which we know that S, if it really is consistent, cannot do. This shows that truth theories increase deductive strength.

This raises the question of what more can be achieved by iterating the construction. We would expect to produce a hierarchy of truth theories of increasing deductive strength, a quality that can be measured by asking what the provable ordinals of the system are. Here we say that α is a *provable ordinal* of a system S if (1) there is a Turing machine that outputs (in suitably encoded form) a well-ordering of $\mathbb{N}$ which is isomorphic to α, and (2) we can prove in S that the output of this Turing machine is a well-ordering of $\mathbb{N}$. (Again, in what follows we will actually prove well-ordering in the strong sense of supporting transfinite induction up to α

[18]The argument given below fixes an error, pointed out to me by Asger Ipsen, in a proof that appeared in an earlier version of the manuscript.

for all predicates in the language.) The *ordinal strength* of S is the supremum of its provable ordinals.[19]

The subtlety here is that, given an arbitrary hierarchy of truth theories, there may be a limit as to how much of the hierarchy we trust. Intuitively, we convince ourselves that a theory in the hierarchy is sound by a transfinite induction argument, meaning that before accepting its theorems we would need its index to be, not just an ordinal, but a provable ordinal, in a strong enough sense to justify taking this step. As we work our way up the hierarchy, we expect to accumulate ever-larger provable ordinals, leading us to accept ever-higher theories in the hierarchy. This should give some (oversimplified) understanding of how our well-ordering proofs are going to work.

3.1. **The system Tar**(S)**.** Throughout this discussion S will be a theory in a language of second order arithmetic which extends ACA^i, the system ACA equipped with intuitionistic logic. We use intuitionistic logic for the reasons articulated in Section 2. (The analysis would still go through if classical logic were used, though, or even if we worked with first, not second, order systems of arithmetic.)

The simplest way to augment S with a truth predicate would be to add a predicate symbol T, together with, for each formula $\mathcal{A}$ of the language of S, the axiom $\mathcal{A} \leftrightarrow T(\ulcorner\mathcal{A}\urcorner)$. This would implement Tarski's "T-scheme", so that it would qualify as a truth predicate in Tarski's sense, but it would accomplish almost nothing. We would have produced a conservative extension of the original theory, yielding no new theorems expressible in the original, unaugmented language. Even simple assertions like "for all sentences $\mathcal{A}$ and $\mathcal{B}$, if $\mathcal{A}$ and $\mathcal{B}$ are both true then so is $\mathcal{A} \wedge \mathcal{B}$" could not be proven, as only finitely many instances of the T-scheme could be used in any proof, which clearly is not sufficient to draw this conclusion.

We need to be able to reason about truth globally. So I define Tar(S) to be the formal system system whose language L_T is the language L of S augmented by a single predicate symbol T, and whose nonlogical axioms are the nonlogical axioms of S together with

- a single axiom which affirms[20] $T(\ulcorner\mathcal{A}\urcorner)$ for all the axioms $\mathcal{A}$ of S
- a single axiom which affirms

$$T(\ulcorner\mathcal{A}\urcorner) \wedge T(\ulcorner\mathcal{B}\urcorner) \rightarrow T(\ulcorner\mathcal{C}\urcorner)$$

  for any $\mathcal{A}, \mathcal{B}, \mathcal{C} \in L$ for which there is a deduction rule of S that infers $\mathcal{C}$ from $\mathcal{A}$ and $\mathcal{B}$
- for every formula $\mathcal{A} = \mathcal{A}(x)$ in L_T with free number variable x (and possibly other free variables not shown), the induction axiom

$$(\mathcal{A}(0) \wedge (\forall n)(\mathcal{A}(n) \rightarrow \mathcal{A}(n+1))) \rightarrow (\forall n)\mathcal{A}(n)$$

[19]The reason for working in second order arithmetic is that the standard definition of well-ordering cannot be expressed in first order terms. However, "supports transfinite induction for all predicates in the language" is perfectly good in the first order setting. So actually our analysis could be carried out using first order arithmetic, with Peano arithmetic replacing ACA^i.

[20]Literally, the statement "$(\forall n)T(f(n))$" for some primitive recursive function f which enumerates the Gödel numbers of the axioms of S. But I will not make any more comments of this sort in what follows.

- the *ω-rule*, a single axiom which states that for every predicate $\mathcal{A}$ in L (i.e., every formula in L with exactly one free number variable) we have

$$(\forall n)T(\ulcorner\mathcal{A}(\hat{n})\urcorner) \leftrightarrow T(\ulcorner(\forall n)\mathcal{A}(n)\urcorner),$$

  where $\hat{n}$ represents a canonical constant term that evaluates to n
- the T-scheme, which consists of one axiom of the form

$$\mathcal{A} \leftrightarrow T(\ulcorner\mathcal{A}\urcorner)$$

  for each sentence $\mathcal{A}$ in L.

The intention is to interpret $T(\ulcorner\mathcal{A}\urcorner)$ as affirming the truth of $\mathcal{A}$, or of the universal closure of $\mathcal{A}$ if it has any free variables.

3.2. **Justifying Tar**(S)**.** How do we justify accepting Tar(S), given that we accept S?

By "accept" I mean something like: we understand that we have the right, indeed that we are rationally compelled, to affirm every theorem of S. Why should we feel the same way about Tar(S)?

The question seems a little murky because Tar(S) invokes the notion of "truth", which is surprisingly subtle, and still the subject of philosophical debate. I have my own explanation of the nature of truth, but I will not call on it here. Instead, I will argue that a basic grasp of infinitary reasoning would allow us to accept Tar(S) without requiring prior familiarity with any notion of truth.

Indeed, predicativism (given the natural numbers) allows reasoning which admits countably infinite sets and constructions — and sentences and derivations. This alone is enough to justify passing from S to Tar(S), provided the language of S has only countably many sentences. Because then we could enumerate these formulas as $\mathcal{A}_1, \mathcal{A}_2, \ldots$ and informally define the desired predicate T by the infinite conjunction

$$(T(\ulcorner\mathcal{A}_1\urcorner) \leftrightarrow \overline{\mathcal{A}}_1) \wedge (T(\ulcorner\mathcal{A}_2\urcorner) \leftrightarrow \overline{\mathcal{A}}_2) \wedge \cdots$$

(where $\overline{\mathcal{A}}$ denotes the universal closure of $\mathcal{A}$). The axioms of Tar(S) could then be straightforwardly justified. For instance, consider "$T(\ulcorner\mathcal{A}\urcorner)$, for every axiom $\mathcal{A}$ of S". This, of course, demands that we have already accepted all the axioms of S, but given this, and given the definition of T, we can infer $T(\ulcorner\mathcal{B}_1\urcorner) \wedge T(\ulcorner\mathcal{B}_2\urcorner) \wedge \cdots$ where the $\mathcal{B}_n$'s enumerate the axioms of S, and from there the single statement "$T(\ulcorner\mathcal{A}\urcorner)$, for every axiom $\mathcal{A}$ of S".

A similar justification can be given for the second axiom, which states that deduction under T is valid. For the induction scheme, which is applied to the language of Tar(S), not just of S, we do not need to assume anything about T. All we have to do is imagine being given the premises $\mathcal{A}(0)$ and $(\forall n)(\mathcal{A}(n) \to \mathcal{A}(n+1))$, for some formula $\mathcal{A}$ with exactly one free variable, in the language of Tar(S). We can easily see how to construct a proof from this of $\mathcal{A}(1)$, then a proof of $\mathcal{A}(2)$, etc., yielding finally (with one final transfinite inference) $(\forall n)\mathcal{A}(n)$. (Cf. §2.4.)

Given our definition of T, the ω-rule is straightforwardly justifiable along similar lines. This completes my sketch of a justification of Tar(S), given S. No prior ideas about "truth" were needed to do this. I suppose all the material that follows could be reworked in terms of countably infinite logic, without mentioning truth at all, but I have not tried to do this.

I mentioned above that it will be a theorem of Tar(S) that every theorem of S is true. This is proven by induction on the number of steps in the derivation of a theorem of S. The theorems provable in a single step are precisely the axioms of S, which we know are all true, and the truth of theorems provable at any subsequent step then follows inductively from the fact that T respects the deduction rules of S. The assertion that $0 = 1$ is not true is an immediate consequence of the forward implication in '$T(\ulcorner 0 = 1 \urcorner) \leftrightarrow 0 = 1$".

3.3. **The system Tarski**(S)**.** Now let us iterate the Tar(S) construction. Start with a computable total ordering $\preceq$ of $\mathbb{N}$. The intention is to iterate along a well-ordering of $\mathbb{N}$, but we do not assume in advance that $\preceq$ is well-ordered. We do need to know that it is a total ordering, that it has a least element, that every element has an immediate successor, and so on. We must be able to express operations of addition, multiplication, and exponentiation which correspond to the usual ordinal operations. One can prove in Peano arithmetic that any of the usual notation systems for the standard "large" countable ordinals has all the properties we need.

In the sequel, any notation for sum, product, or exponent will always refer to these (presumptively) ordinal operations, not to the usual arithmetical operations on $\mathbb{N}$. Also, I will use the notation $\widetilde{\mathbb{N}}$ for "$\mathbb{N}$ equipped with the ordering $\preceq$", and (hopefully without too great a risk of confusion) use the symbols "0", "1", ..., "ω", ..., "Γ_0", ... to refer both to the abstract ordinals and to their representatives in $\widetilde{\mathbb{N}}$.

The language of Tarski(S), which I denote L, will be the language of S augmented by a family of predicate symbols T_a, for $a \in \widetilde{\mathbb{N}}$, and one additional predicate symbol Acc. The idea is that the T_a's should be the truth predicates in the hierarchy and that Acc(a) represents our acceptance of the hierarchy up to a (with respect to the ordering $\preceq$).

For each $a \in \widetilde{\mathbb{N}}$ we let L_a be the language of S augmented by only the predicate symbols T_b with $b \prec a$. It is the formulas of L_a to which we want to apply T_a.

We would like a hierarchy of formal systems S_a with $\mathrm{S}_{a+1} = \mathrm{Tarski}(\mathrm{S}_a)$ and $\mathrm{S}_a = \mathrm{Tarski}(\bigcup_{b \prec a} \mathrm{S}_b)$ at limits. This could be achieved by a recursive construction, but it is simpler (and also a slightly different, better result) to define them all at once, in the following way.

For each $a \in \widetilde{\mathbb{N}}$ let S_a be the formal system whose language is L_{a+1} (so it includes the predicate T_a) and whose nonlogical axioms are the nonlogical axioms of S together with axioms about T_a:

- the statement "$T_a(\ulcorner \mathcal{A} \urcorner)$ for all the axioms $\mathcal{A}$ of S"
- the statement "$T_a(\ulcorner \mathcal{A} \urcorner) \wedge T_a(\ulcorner \mathcal{B} \urcorner) \to T_a(\ulcorner \mathcal{C} \urcorner)$ for all $\mathcal{A}, \mathcal{B}, \mathcal{C} \in L_a$ for which there is a deduction rule that infers $\mathcal{C}$ from $\mathcal{A}$ and $\mathcal{B}$"
- a single axiom which says that T_a holds of every instance of the induction scheme for formulas in L_a
- the statement "$(\forall n) T_a(\ulcorner \mathcal{A}(\hat{n}) \urcorner) \leftrightarrow T_a(\ulcorner (\forall n) \mathcal{A}(n) \urcorner)$ for every predicate $\mathcal{A}$ in L_a"
- for every sentence $\mathcal{A}$ in L_a, the truth definition

$$\mathcal{A} \leftrightarrow T_a(\ulcorner \mathcal{A} \urcorner)$$

and axioms about T_b for $b \prec a$:

- the single statement "for all $b \prec a$ and every sentence $\mathcal{A} \in L_b$ we have $T_a(\ulcorner \mathcal{A} \leftrightarrow T_b(\ulcorner \mathcal{A} \urcorner) \urcorner)$".
- for each $b \prec a$, the statement "$T_b(\ulcorner \mathcal{A} \urcorner) \leftrightarrow T_a(\ulcorner \mathcal{A} \urcorner)$ for all $\mathcal{A} \in L_b$"

S_a is the theory whose acceptance is expressed in Tarski(S) by $\mathrm{Acc}(a)$. The appropriate axiom for $\mathrm{Acc}(\cdot)$ is

$$(\forall b \prec a)\mathrm{Acc}(b) \leftrightarrow \mathrm{Acc}(a).$$

I will call this condition "progressivity*" and write "$\mathrm{Prog}^*(\mathrm{Acc})$". (Note that it is stronger than the usual notion of progressivity, which affirms only the forward implication. This choice simplifies the analysis given below a little.) Indeed, S_0 is precisely Tar(S), and for any $a \succ 0$, although S_a is not exactly the same as $\mathrm{Tar}(\bigcup_{b\prec a}\mathrm{S}_b)$ (because of the final "$T_b(\ulcorner\mathcal{A}\urcorner) \leftrightarrow T_a(\ulcorner\mathcal{A}\urcorner)$ for all $\mathcal{A} \in L_b$" scheme), essentially the same argument can be used to justify accepting it given that we have accepted $\bigcup_{b\prec a}\mathrm{S}_b$. So there is a straightforward justification of progressivity* of Acc.

It is easy to see that for every $b \prec a$ it is a theorem of S_a that all the axioms of S_b are true. Together with the truth of all deduction rules for formulas in L_a, this yields the following fact.

Proposition 3.1. *For every $b \prec a$, it is a theorem of S_a that every theorem of S_b is true.*

Now I am ready to describe the axioms of Tarski(S). Its language was defined earlier; its nonlogical axioms will consist of

- the nonlogical axioms of S
- the axiom $\mathrm{Prog}^*(\mathrm{Acc})$
- the induction scheme for all formulas in the language of S augmented by Acc
- for each a and each axiom $\mathcal{A}$ of S_a the statement "$\mathrm{Acc}(\hat{a}) \to \mathcal{A}$".

Since the meanings of the T_a predicates are not initially specified, I exclude them from the induction scheme. For the record, I think this precaution is philosophically unnecessary, but for my purposes there is no harm in being conservative, as none of the results proven below is affected by this restriction. We could be even more conservative and forbid any reasoning about formulas containing T_a until $\mathrm{Acc}(\hat{a})$ has been proven, by, for each a and each axiom $\mathcal{A}$ of S_a, replacing the "$\mathrm{Acc}(\hat{a}) \to \mathcal{A}$" axiom with a deduction rule that infers $\mathcal{A}$ from $\mathrm{Acc}(\hat{a})$. This also would not affect the validity of anything that follows in any essential way.

3.4. **The Veblen hierarchy.** We can now start to think about the provable ordinals of Tarski(S). Let us do this in terms of the Veblen hierarchy (ϕ_α).

We have fixed a computable total ordering $\preceq$ of $\mathbb{N}$. Now fix an element $c_0 \in \widetilde{\mathbb{N}}$ which is greater than ω and is stable under ordinal exponentiation, i.e., if $a, b \prec c_0$ then $a^b \prec c_0$. Again, we don't have to know that $\{c \in \widetilde{\mathbb{N}} : c \prec c_0\}$ is well-ordered. Recall that a predicate $\mathcal{A}$ is progressive* if for all $a \in \widetilde{\mathbb{N}}$

$$(\forall b \prec a)\mathcal{A}(b) \leftrightarrow \mathcal{A}(a);$$

I will also say it is *progressive* up to* c_0 if $(\forall b \prec a)\mathcal{A}(b) \leftrightarrow \mathcal{A}(a)$ for all $a \prec c_0$.

For each $a, c \prec c_0$ define a predicate $\mathcal{B}_a^c$ in the language $L_{\omega^a\cdot c+1}$ by

$$\mathcal{B}_a^c(b) := (\forall \mathcal{A} \in L^p_{\omega^a\cdot c})T_{\omega^a\cdot c}(\ulcorner \mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(\phi_a(b))\urcorner)$$

where L_a^p denotes the set of numerical predicates in L_a, i.e., formulas with exactly one free number variable and no free set variables. For fixed a, the predicates $\mathcal{B}_a^c$ say essentially the same thing as c varies: that any progressive* predicate holds of $\phi_a(b)$. The only difference is the languages to which they are applied. This is necessary because we will want to apply some $\mathcal{B}_a^c$'s to others.

Let $\mathcal{C}$ be the predicate in the language L_{c_0+1}

$$\mathcal{C}(a) := (\forall c \prec c_0)T_{c_0}(\ulcorner \mathrm{Prog}^*(\mathcal{B}_a^c)\urcorner).$$

Recalling that $\mathcal{B}_a^c$ and $\mathcal{B}_a^{c'}$ differ only in the languages to which they are applied, the crude intuition for $\mathcal{C}(a)$ is that it says the assertion that anything progressive* holds of $\phi_a(b)$ is itself progressive* in b. The key theorem we have to prove is that $\mathcal{C}$ is itself progressive*!

(The reason the strong "progressivity*" version of progressivity defined above helps is that, under this definition, if $\mathcal{A}$ is progressive* and $\mathcal{A}$ holds of a, then the "shifted" predicate $\mathcal{A}^{\to a}(\cdot) = \mathcal{A}(a + \cdot)$ is also progressive*. Our version of progressivity ensures that if $\mathcal{A}(a)$ fails then $\mathcal{A}(a')$ also fails for all $a' \succ a$, so this shifting can only occur within the initial, truly progressive* part.)

The proof of Theorem 3.5 below is a little long because each a and each b can be zero, a successor, or a limit, leading to nine separate cases which are nearly all genuinely different. (For limit values of b, the argument is essentially the same regardless of what a is.) For the sake of readability I have split up the proof into three lemmas. The crucial case appears in Lemma 3.4, when a and b are both successors. (It is only here that the "$\omega^a \cdot c$" expression in $\mathcal{B}_a^c$ becomes important. The point is that the index $\omega^{a+1} \cdot c$ will always be a limit of indices of the form $\omega^a \cdot c'$, even when c is a successor.)

In the following proofs we will assume $\mathrm{Acc}(\hat{c}_0)$, which means that all the axioms for T_{c_0} are available. In particular, we have the T-scheme for T_{c_0}. So we can pass freely between $\mathcal{A}$ and $T_{c_0}(\ulcorner\mathcal{A}\urcorner)$ for any sentence $\mathcal{A}$ of L_{c_0}. What we cannot do is to reason under T_{c_0} about predicates such as $\mathcal{C}$ which do not belong to L_{c_0}. Note that in the definition of $\mathcal{C}(a)$ given above we cannot use the ω-rule to import the quantifier $\forall c \prec c_0$ into T_{c_0}, because $\mathcal{B}_a^c$ is a distinct predicate for every a and c, so an expression like "$(\forall c \prec c_0)\mathrm{Prog}^*(\mathcal{B}_a^c)$" would not be a well-formed formula. But if we are working under T_{c_0}, we can simultaneously reason about $\mathcal{B}_a^c$ for arbitrary values of c. Indeed, this is just the sort of thing truth predicates are good for — allowing us to reason schematically. I have included some comments about which truth predicate we are reasoning under at which point, but the essence of the proofs can be understood by ignoring all references to truth and just accepting that we can reason schematically.

The next three lemmas are proven in Tarski(S).

Lemma 3.2. $\mathrm{Acc}(\hat{c}_0)$ *implies* $\mathcal{C}(0)$.

Proof. Working in Tarski(S), assume $\mathrm{Acc}(\hat{c}_0)$. Fix $c \prec c_0$; reasoning under T_{c_0}, we must prove $\mathrm{Prog}^*(\mathcal{B}_0^c)$, i.e., we must prove progressivity* in b of the statement

(1) $$(\forall \mathcal{A} \in L_c^p)T_c(\ulcorner \mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(\omega^b)\urcorner).$$

For $b = 0$ this is trivial; it just says that for any $\mathcal{A} \in L_c^p$ it is true (according to T_c) that progressivity* of $\mathcal{A}$ implies that $\mathcal{A}$ holds of $\omega^0 = 1$. The way we prove this in Tarski(S) is by first proving in Tarski(S) — what can actually be done in ACA^i — a formalization of the statement that for any predicate $\mathcal{A}$ in L_c there is a proof in the system S_c of the sentence "$\mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(1)$". We can do this by writing out a proof template and then verifying its validity for any $\mathcal{A}$. Combining this with Proposition 3.1 (and using $\mathrm{Acc}(\hat{c}_0)$), we get $(\forall \mathcal{A} \in L_c^p)T_{c_0}(\ulcorner \mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(1)\urcorner)$, and then $(\forall \mathcal{A} \in L_c^p)T_c(\ulcorner \mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(1)\urcorner)$. I wanted to spell all this out once, but I will omit similar arguments in the sequel.

For limit values of b, assume we are given $(\forall b' \prec b)T_c(\mathcal{B}_0^c(b'))$; that is, assume

$$T_c(\ulcorner \mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(\omega^{b'})\urcorner)$$

for all $\mathcal{A} \in L_c^p$ and all $b' \prec b$. Using the ω-rule to bring the quantifier "$\forall b' \prec b$" under T_c, we get that for every $\mathcal{A} \in L_c^p$ it is true (according to T_c) that progressivity* of $\mathcal{A}$ implies $\mathcal{A}(\omega^{b'})$ for all $b' \prec b$. Since $\omega^b = \sup_{b' \prec b} \omega^{b'}$, we infer from this (for arbitrary $\mathcal{A}$, under T_c) that progressivity* of $\mathcal{A}$ implies $\mathcal{A}(\omega^b)$.

Finally, if $b = b' + 1$ is a successor, assume that for every predicate $\mathcal{A}$ in L_c it is true (according to T_c) that progressivity* of $\mathcal{A}$ implies that $\mathcal{A}$ holds of $\omega^{b'}$. Since $\mathrm{Prog}^*(\mathcal{A}) \wedge \mathcal{A}(\omega^{b'})$ implies $\mathrm{Prog}^*(\mathcal{A}^{\to \omega^{b'}})$ (as noted a few paragraphs before this lemma), it is then true that $\mathcal{A}^{\to \omega^{b'}}$ holds of $\omega^{b'}$, i.e., $\mathcal{A}$ holds of $\omega^{b'} + \omega^{b'}$. Continuing inductively, we get the truth for all $n \prec \omega$ of $\mathcal{A}(\omega^{b'} \cdot n)$, and progressivity* of $\mathcal{A}$ then implies that $\mathcal{A}$ holds of $\omega^{b'} \cdot \omega = \omega^{b'+1} = \omega^b$. We have shown that $\mathcal{B}_0^c(b')$ implies $\mathcal{B}_0^c(b' + 1)$. This completes the proof that (1) is progressive* in b and establishes $\mathcal{C}(0)$. □

Lemma 3.3. $\mathrm{Acc}(\hat{c}_0)$ *implies that for any limit* $a \prec c_0$ *we have* $(\forall a' \prec a)\mathcal{C}(a') \to \mathcal{C}(a)$.

Proof. Working in Tarski(S), assume $\mathrm{Acc}(\hat{c}_0)$. Fix a limit $a \prec c_0$ and assume $(\forall a' \prec a)\mathcal{C}(a')$. This means that we are given the truth, according to T_{c_0}, of $\mathrm{Prog}^*(\mathcal{B}_{a'}^{c'})$ for every $a' \prec a$ and $c' \prec c_0$, and we must show that for every $c \prec c_0$ it is true according to T_{c_0} that $\mathcal{B}_a^c$ is progressive*.

Fix $c \prec c_0$. First we check the case $b = 0$, where we have to verify $\mathcal{B}_a^c(0)$. Since we are assuming, for every $c' \prec c_0$ and $a' \prec a$, that it is true (according to T_{c_0}) that $\mathcal{B}_{a'}^{c'}$ is progressive*, in particular we have for all such c' and a' the truth of $\mathcal{B}_{a'}^{c'}(0)$. Thus, for every predicate $\mathcal{A}$ in some $L_{\omega^{a'} \cdot c'}$ with $a' \prec a$ and $c' \prec c_0$ — which is to say, every predicate $\mathcal{A}$ in L_{c_0} — we have that progressivity* of $\mathcal{A}$ implies the truth of $\mathcal{A}(\phi_{a'}(0))$ for all $a' \prec a$. In particular, for any predicate $\mathcal{A} \in L_{\omega^a \cdot c}$ it is true according to T_{c_0}, and hence according to $T_{\omega^a \cdot c}$, that progressivity* of $\mathcal{A}$ implies that A holds of $\sup_{a' \prec a} \phi_{a'}(0) = \phi_a(0)$. This verifies $\mathcal{B}_a^c(0)$. That was the $b = 0$ case.

Next we will show, reasoning under T_{c_0}, that if b is a limit and we have $\mathcal{B}_a^c(b')$ for all $b' \prec b$, then we also have $\mathcal{B}_a^c(b)$. At this point we have two induction hypotheses going, one for $a' \prec a$ and one for $b' \prec b$. To verify $\mathcal{B}_a^c(b)$, let $\mathcal{A}$ be a predicate in $L_{\omega^a \cdot c}$. Then $\mathcal{B}_a^c(b')$ for all $b' \prec b$ tells us that according to $T_{\omega^a \cdot c}$, progressivity* of $\mathcal{A}$ implies that $\mathcal{A}$ holds of $\phi_a(b')$ for all $b' \prec b$. So according to $T_{\omega^a \cdot c}$, progressivity* of $\mathcal{A}$ implies that $\mathcal{A}$ holds of $\sup_{b' \prec b} \phi_a(b') = \phi_a(b)$, as desired. We did not even need the induction hypothesis on a in this part.

The final case is for successor values of b. Say $b = b' + 1$. Reasoning under T_{c_0}, we have to verify that $\mathcal{B}_a^c(b')$ implies $\mathcal{B}_a^c(b' + 1)$. For this we must use both (i) the induction hypothesis on a (for every $c' \prec c_0$ and $a' \prec a$ the predicate $\mathcal{B}_{a'}^{c'}$ is progressive*) and (ii) the induction hypothesis on b (we have $\mathcal{B}_a^c(b')$). Now $\mathcal{B}_a^c(b')$ says that every progressive* $\mathcal{A} \in L_{\omega^a \cdot c}$ holds of $\phi_a(b')$, and $\phi_a(b')$ is a fixed point of every $\phi_{a'}$ with $a' \prec a$, so we can also say that every progressive* $\mathcal{A} \in L_{\omega^a \cdot c}$ holds of $\phi_{a'}(\phi_a(b'))$ for all $a' \prec a$. This means that $\mathcal{B}_{a'}^{c'}(\phi_a(b'))$ holds for all $a' \prec a$ and $c' \prec c_0$ such that $L_{\omega^{a'} \cdot c'} \subseteq L_{\omega^a \cdot c}$.

But for any $a' \prec a$, we can find a'' such that $a' + a'' = a$, and then by putting $c' = \omega^{a''}\cdot c$, we get $\omega^{a'}\cdot c' = \omega^a\cdot c$. So by the conclusion reached in the last paragraph, for any $a' \prec a$ there exists $c' \prec c_0$ such that $L_{\omega^{a'}\cdot c'} = L_{\omega^a\cdot c}$ and $\mathcal{B}^{c'}_{a'}(\phi_a(b'))$ holds. And since every such $\mathcal{B}^{c'}_{a'}$ is progressive*, by (i), it follows that $\mathcal{B}^{c'}_{a'}(\phi_a(b')+1)$ also holds.

We are now in a position to affirm that for every $a' \prec a$, every progressive* $\mathcal{A} \in L_{\omega^a\cdot c}$ lies within the scope of some $\mathcal{B}^{c'}_{a'}$ which satisfies $\mathcal{B}^{c'}_{a'}(\phi_a(b')+1)$, so that $\mathcal{A}$ holds of $\phi_{a'}(\phi_a(b')+1)$. If a' is a successor, $a' = a''+1$, then this means that $\mathcal{A}$ holds of the next fixed point of $\phi_{a''}$ after $\phi_a(b')$. Thus, by progressivity*, $\mathcal{A}$ holds of the supremum of these fixed points over all $a'' \prec a$, which equals $\phi_a(b'+1) = \phi_a(b)$. That verifies $\mathcal{B}^c_a(b)$ and completes the proof of $\mathcal{C}(a)$. □

Lemma 3.4. $\mathrm{Acc}(\hat{c}_0)$ *implies that for any* $a \prec c_0$, $\mathcal{C}(a)$ *implies* $\mathcal{C}(a+1)$.

Proof. Working in Tarski(S), assume $\mathrm{Acc}(\hat{c}_0)$. Fix $a \prec c_0$ and assume $\mathcal{C}(a)$, and on the way to proving $\mathcal{C}(a+1)$, fix $c \prec c_0$. Reasoning under T_{c_0}, we must prove that $\mathcal{B}^c_{a+1}$ is progressive*, i.e., that for every b, if $\mathcal{B}^c_{a+1}(b')$ holds for all $b' \prec b$ then $\mathcal{B}^c_{a+1}(b)$ also holds.

If $b = 0$ then we simply have to show that $\mathcal{B}^c_{a+1}(0)$ holds, i.e., that any progressive* $\mathcal{A} \in L_{\omega^{a+1}\cdot c}$ holds of $\phi_{a+1}(0)$. Since we have assumed $\mathcal{C}(a)$, we know that for any $c' \prec c_0$, according to T_{c_0} the predicate $\mathcal{B}^{c'}_a$ is progressive*. In particular, for each $n \prec \omega$ the predicate $\mathcal{B}^{c+n}_a$ is progressive*. Thus $\mathcal{B}^{c+n}_a$ holds of 0, so every progressive* predicate in $L_{\omega^a(c+n)}$ holds of $\phi_a(0)$, and hence of $\phi_a(0)+1$. In particular, $\mathcal{B}^{c+n-1}_a$ holds of $\phi_a(0)+1$. So every progressive* predicate in $L_{\omega^a(c+n-1)}$ holds of $\phi_a(\phi_a(0)+1)$. Inductively, every progressive* predicate in $L_{\omega^a(c+n-k)}$ holds of $\phi^k_a(\phi_a(0)+1)$, which means (taking $k = n$) that every progressive* predicate in $L_{\omega^a\cdot c}$ holds of $\phi^n_a(\phi_a(0)+1)$. Since n was arbitrary, it follows that every progressive* predicate in $L_{\omega^a\cdot c}$ holds of $\sup_{n\prec\omega}\phi^n_a(\phi_a(0)+1) = \phi_{a+1}(0)$. This finishes the $b = 0$ case.

Next, assuming that for some limit b we have $\mathcal{B}^c_{a+1}(b')$ for all $b' \prec b$, we must verify $\mathcal{B}^c_{a+1}(b)$. This is easy, because we are given that any progressive* predicate in $L_{\omega^{a+1}\cdot c}$ holds of $\phi_{a+1}(b')$ for all $b' \prec b$, and so by progressivity* it holds of $\sup_{b'\prec b}\phi_{a+1}(b') = \phi_{a+1}(b)$. The corresponding cases in Lemmas 3.2 and 3.3 were handled in the same way.

Finally, letting b be arbitrary, we must make the inference from $\mathcal{B}^c_{a+1}(b)$ to $\mathcal{B}^c_{a+1}(b+1)$. So assume $\mathcal{B}^c_{a+1}(b)$ and fix $\mathcal{A} \in L^p_{\omega^{a+1}\cdot c}$. We can write $\omega^{a+1}\cdot c = \omega^a\cdot\omega c$, and since ωc is a limit, this shows that $\mathcal{A}$ must belong to $L_{\omega^a\cdot c'}$ for some $c' \prec \omega c$. Moreover, since ωc is a limit, we have $c'+n \prec \omega c$ for all $n \prec \omega$.

Now for each $n \prec \omega$ the predicate $\mathcal{B}^{c'+n}_a$ belongs to $L_{\omega^{a+1}\cdot c}$, and is progressive* since we are assuming $\mathcal{C}(a)$, so the hypothesis of $\mathcal{B}^c_{a+1}(b)$ yields $\mathcal{B}^{c'+n}_a(\phi_{a+1}(b))$, and then by progressivity* we get $\mathcal{B}^{c'+n}_a(\phi_{a+1}(b)+1)$. Applying this to $\mathcal{B}^{c'+n-1}_a$, which lies in $L_{\omega^a(c'+n-1)+1} \subseteq L_{\omega^a(c+n)}$, then yields $\mathcal{B}^{c'+n-1}_a(\phi_a(\phi_{a+1}(b)+1))$, and inductively we finally get $\mathcal{B}^{c'}_a(\phi^n_a(\phi_{a+1}(b)+1))$. Since $\mathcal{A} \in L_{\omega^a\cdot c'}$, if $\mathcal{A}$ is progressive* this entails $\mathcal{A}(\phi^{n+1}_a(\phi_{a+1}(b)+1))$, and as n was arbitrary $\mathcal{A}(\phi_{a+1}(b+1))$ follows. We have established $\mathcal{B}^c_{a+1}(b)$, as desired. □

Putting these lemmas together yields the following theorem.

Theorem 3.5. Tarski(S) *proves that* $\mathrm{Acc}(\hat{c}_0)$ *implies* $\mathcal{C}$ *is progressive* up to* c_0.

Recall that we define $\gamma_0 = 0$ and $\gamma_{n+1} = \phi_{\gamma_n}(0)$ for $n \geq 0$, so that $\Gamma_0 = \sup_n \gamma_n$.

In what follows the term "transfinite induction up to a for S" will refer to the scheme consisting of all sentences of the form "$\mathrm{Prog}^*_{\hat{a}}(\mathcal{A}) \to (\forall b \prec \hat{a})\mathcal{A}(b)$" as $\mathcal{A}$ ranges over the predicates in the language of S, where "$\mathrm{Prog}^*_{\hat{a}}(\mathcal{A})$" means "$\mathcal{A}$ is progressive* up to $\hat{a}$".

Corollary 3.6. *Let $n \geq 2$. Then* Tarski(S) *plus transfinite induction up to γ_n for* Tarski(S) *proves transfinite induction up to γ_{n+1} for* S*.*

Proof. Fix n. Since Acc is progressive*, transfinite induction up to γ_n for Tarski(S) yields $(\forall a \prec \gamma_n)\mathrm{Acc}(a)$, and then, with one further application of progressivity*, $\mathrm{Acc}(\gamma_n)$. Theorem 3.5 then yields, in Tarski(S) plus $\mathrm{Acc}(\gamma_n)$, that $\mathcal{C}$ is progressive* up to γ_n, and now the hypothesis on transfinite induction up to γ_n yields $(\forall a \prec \gamma_n)\mathcal{C}(a)$. In particular, we have $(\forall a \prec \gamma_n)T_{c_0}(\mathcal{B}^0_a(0))$.

Thus for any predicate $\mathcal{A}$ in L, we have

$$T_{c_0}(\mathrm{Prog}^*(\mathcal{A}) \to (\forall a \prec \gamma_n)\mathcal{A}(\phi_a(0))),$$

hence

$$\mathrm{Prog}^*(\mathcal{A}) \to (\forall a \prec \gamma_n)\mathcal{A}(\phi_a(0))$$

hence

$$\mathrm{Prog}^*(\mathcal{A}) \to (\forall b \prec \gamma_{n+1})\mathcal{A}(b)$$

since $\sup_{a \prec \gamma_n} \phi_a(0) = \gamma_{n+1}$. We have proven transfinite induction up to γ_{n+1} for S. □

3.5. **Iterating the Tarski**(S) **construction.** Theorem 3.5 doesn't get us very far by itself, since it requires a strong assumption about Acc. What we must do now is to "reflect" on the legitimacy of the Tarski construction. We have agreed that, for arbitrary S, if we accept S then we should accept Tarski(S), and so we should also accept $\mathrm{Tarski}^2(\mathrm{S}) = \mathrm{Tarski}(\mathrm{Tarski}(\mathrm{S}))$, then $\mathrm{Tarski}^3(\mathrm{S}) = \mathrm{Tarski}(\mathrm{Tarski}^2(\mathrm{S}))$, and so on. We could now formulate a "higher order" construction that iterates the $\mathrm{S} \mapsto \mathrm{Tarski}(\mathrm{S})$ construction along $\preceq$, which would be justified in the same way we justified Tarski(S); I will return to this idea in the next section. At this point we only need to do this for all finite n.

Theorem 3.7. $\mathrm{Tarski}^\omega(\mathrm{ACA}^i) = \bigcup_{n \prec \omega} \mathrm{Tarski}^n(\mathrm{ACA}^i)$ *proves* $\mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(\hat{a})$ *for every $a \prec \Gamma_0$ and every predicate $\mathcal{A}$ in its language.*

Proof. We make the argument for any second order theory S that extends ACA^i. Fix $n \prec \omega$. The main idea is to prove, in $\mathrm{Tarski}^n(\mathrm{S})$, transfinite induction up to $\gamma_2 = \varepsilon_0$ for $\mathrm{Tarski}^{n-1}(\mathrm{S})$, and then inductively apply Corollary 3.6 to get transfinite induction up to γ_{k+1} in $\mathrm{Tarski}^{n-k}(\mathrm{S})$, for $1 \leq k \leq n$, yielding finally transinite induction up to γ_{n+1} for S. At that point the full statement of the theorem will follow easily.

The base case of transfinite induction up to γ_2 is similar to Lemma 3.2, but different enough to merit being written out. Taking $\mathrm{S}' = \mathrm{Tarski}^{n-1}(\mathrm{S})$, we want to prove, in $\mathrm{Tarski}^n(\mathrm{S}) = \mathrm{Tarski}(\mathrm{S}')$, transfinite induction up to γ_2 for S'. To do this we introduce a *jump predicate* $J_\mathcal{A}$, for every predicate $\mathcal{A}$ in the language of S', defined by

$$J_\mathcal{A}(b) := (\forall a)(\mathcal{A}(a) \to \mathcal{A}(a + \omega^b)).$$

This is a different predicate for each $\mathcal{A}$. Now, for any $\mathcal{A}$, we start by proving in S' that progressivity* of $\mathcal{A}$ implies progressivity* of $J_\mathcal{A}$. Fix $\mathcal{A}$ and assume it is progressive*. Then $J_{\mathcal{A}}(0)$ says that $(\forall a)(\mathcal{A}(a) \to \mathcal{A}(a+1))$, and this is an immediate consequence of progressivity* of $\mathcal{A}$. If b is a limit, then $J_{\mathcal{A}}(b')$ for all $b' \prec b$ says that for any a we have $\mathcal{A}(a) \to \mathcal{A}(a+\omega^{b'})$ for all $b' \prec b$, and this yields $\mathcal{A}(a) \to \mathcal{A}(a+\omega^b)$ for all a, by progressivity* of $\mathcal{A}$ again. Finally, assuming $J_{\mathcal{A}}(b)$, we must prove $J_{\mathcal{A}}(b+1)$. Here we use the reasoning that $J_{\mathcal{A}}(b)$ tells us that $\mathcal{A}(a)$ implies $\mathcal{A}(a+\omega^b)$, and also that $\mathcal{A}(a+\omega^b)$ implies $\mathcal{A}(a+\omega^b+\omega^b)$, and so on; that is, an induction argument shows that $J_{\mathcal{A}}(b)$ implies $(\forall a)(\mathcal{A}(a) \to \mathcal{A}(a+\omega^b \cdot n))$ for all $n \prec \omega$. One final appeal to progressivity* of $\mathcal{A}$ then yields $J_{\mathcal{A}}(b) \to J_{\mathcal{A}}(b+1)$.

For a given predicate $\mathcal{A}$ in the language of S′, we have shown, in S′, that $\mathrm{Prog}^*(\mathcal{A}) \to \mathrm{Prog}^*(J_{\mathcal{A}})$. Now, reasoning in the one-step truth theory Tarski(S′), we can see that $T(\mathrm{Prog}^*(\mathcal{A}))$ implies $T(\mathrm{Prog}^*(J_{\mathcal{A}}))$ implies $T(\mathrm{Prog}^*(J_{J_{\mathcal{A}}}))$ implies $\cdots$, for any $\mathcal{A}$. Thus, if $\mathcal{A}$ is progressive* then $\mathcal{A}(1)$ is true, $J_{\mathcal{A}}(1)$ is true (which implies that $\mathcal{A}(\omega)$ is true), $J_{J_{\mathcal{A}}}(1)$ is true (which implies that $J_{\mathcal{A}}(\omega)$ is true, which implies that $\mathcal{A}(\omega^\omega)$ is true), and so on. Inductively we get $(\forall n \prec \omega)T(\mathcal{A}(\omega^{(n)}))$, where $\omega^{(1)} = \omega$ and $\omega^{(n+1)} = \omega^{\omega^{(n)}}$. So $T(\mathrm{Prog}^*(\mathcal{A}))$ implies $T(\mathcal{A}(\varepsilon_0))$.

That was in Tar(S′). But in Tarski(S′) we have Prog*(Acc), so in particular we have Acc(0), which means that we can reason in the one-step truth theory Tar(S′). Thus Tarski^n(S) proves transfinite induction up to $\gamma_2 = \varepsilon_0$ for Tarski^{n-1}(S).

From this point, we can apply Corollary 3.6 $n-1$ times to obtain transfinite induction up to γ_{n+1} for S, where n was arbitrary. But Tarski^ω(S) literally equals $\mathrm{Tarski}^\omega(\mathrm{Tarski}^k(\mathrm{S}))$ for any k, so this argument actually yields transfinite induction up to γ_{n+1} for Tarski^k(S), for any k and any n. We conclude that Tarski^ω(S) proves $\mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(\gamma_n)$ for all n and all predicates $\mathcal{A}$ in its language. □

3.6. **Going even further.** The system $\mathrm{Tarski}^\omega(\mathrm{ACA}^i)$ is somewhat unnatural because once we understand the passage from S to Tarski(S) we can imagine iterating it along any ordinal, not just ω. Thus we get $\mathrm{Tarski}(\mathrm{ACA}^i)$, $\mathrm{Tarski}^2(\mathrm{ACA}^i) = \mathrm{Tarski}(\mathrm{Tarski}(\mathrm{ACA}^i))$, …, $\mathrm{Tarski}^\omega(\mathrm{ACA}^i) = \bigcup_{n<\omega} \mathrm{Tarski}^n(\mathrm{ACA}^i)$, …, $\mathrm{Tarski}^{2\omega}(\mathrm{ACA}^i)$, etc.

Let us consider the system $\mathrm{Tarski}^{\omega^2}(\mathrm{ACA}^i)$. I must now introduce some notation. Following [34], let an *r-normal* function be a strictly increasing continuous function from some ordinal r to itself, and given such a function ψ let ψ_1 be the function that enumerates its fixed points, ψ_2 the function that enumerates the fixed points of ψ_1, and so on. This yields a hierarchy of r-normal functions ψ_α. When $\psi(\alpha) = \omega^\alpha$, this is just the usual Veblen hierarchy (ϕ_α).

But we can now move one step up and let ψ^1 be the r-normal function which enumerates the values $\psi(0)$, $\psi_1(0)$, …, $\psi_\omega(0)$, …. Let us call this new function the *Veblenization* of the original function ψ. This induces a new hierarchy (ψ^α) where $\psi^{\alpha+1}$ is the Veblenization of ψ^α, and when α is a limit, ψ^α enumerates the intersection of the ranges of ψ^β for $\beta \prec \alpha$.

If we take $\psi(a) = \omega^a$, then ψ enumerates the sequence $(\phi_a(0))$ and ψ^1_1 enumerates the sequence (Γ_a). Setting $\gamma^0 = 0$ and $\gamma^{n+1} = \psi^{\gamma_n}(0)$, we get the Ackermann ordinal $\psi_{\Omega^2}(0)$ as the limit of the sequence (γ^n).

We have already done the technical work needed to analyze $\mathrm{Tarski}^{\omega^2}(\mathrm{ACA}^i)$. The key point is that Corollary 3.6 actually shows something stronger: not only can we go from γ_n to γ_{n+1}, we can actually prove, in Tarski(S) plus transfinite induction up to $\psi^1(b)$ for Tarski(S) and progressivity* in a of transfinite induction up to $\psi(a)$ for S, that we have transfinite induction up to $\psi^1(b+1)$ for S, for any r-normal function ψ and any $b \in \widetilde{\mathbb{N}}$.

Set $\mathrm{S}' = \mathrm{Tarski}^\omega(\mathrm{S})$. The preceding yields, in S′ plus transfinite induction up to $\psi^1(b)$ for S′ and progressivity* in a of transfinite induction up to $\psi(a)$ for S, that we have transfinite induction up to $\psi^1(b+1)$ for S′. So we can prove in Tar(S′) the single statement that transfinite induction up to $\psi^1(b)$ for S′ is progressive* in b, for any r-normal function ψ for which we have progressivity* in a of transfinite induction up to $\psi(a)$ for S.

This straightforwardly leads to an analog of Corollary 3.6 which states that Tarski(S′) plus transfinite induction up to γ^n for Tarski(S′) and progressivity* in a of transfinite induction up to $\psi(a)$ for S yields transfinite induction up to γ^{n+1} for S′. Then the proof of Theorem 3.7 can be adopted to show the following.

Theorem 3.8. $\mathrm{Tarksi}^\omega((\mathrm{ACA}^i)') = \mathrm{Tarski}^{\omega^2}(\mathrm{ACA}^i)$ *proves* $\mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(\hat{a})$ *for every* $a \prec \phi_{\Omega^2}(0)$ *and every predicate in its language.*

The Ackermann ordinal $\phi_{\Omega^2}(0)$ is predicatively provable. This should not be surprising to anyone with a basic familiarity with this ordinal. It clearly is "built up from below" and therefore ought to be predicative.

Reasoning in a similar way, we can push this further:

Theorem 3.9. $\mathrm{Tarski}^{\omega^\omega}(\mathrm{ACA}^i)$ *proves* $\mathrm{Prog}^*(\mathcal{A}) \to \mathcal{A}(\hat{a})$ *for every* $a \prec \phi_{\Omega^\omega}(0)$ *and every predicate in its language.*

So the small Veblen ordinal $\phi_{\Omega^\omega}(0)$ is also predicatively provable. Again, this should not be a surprise.

As was mentioned in Section 3.3, we can consider the sequence $\mathrm{Tarski}^\alpha(\mathrm{ACA}^i)$ along a totally ordered set which we may suspect, but do not initially know, to be well-ordered. Acceptability along this new iteration will be measured by a new predicate Acc_1, such that $\mathrm{Acc}_1(a)$ is supposed to indicate the correctness of $\mathrm{Tarski}^\alpha(\mathrm{ACA}^i)$ where a is a notation for α. Each step of this iterated construction passes from some system S to Tarski(S), so each of them has its own separate Acc predicate. Incorporating the new, higher order, acceptability predicate Acc_1, we get a new, higher order, iterated Tarskian truth theory $\mathrm{Tarski}_1(\mathrm{ACA}^i)$. I conjecture that $\mathrm{Tarski}_1^\omega(\mathrm{ACA}^i)$ proves transfinite induction up to the large Veblen ordinal $\phi_{\Omega^\Omega}(0)$, but I have not tried to prove this.

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.

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] A. Avron, Poincaré-Weyl's predicativity: going beyond Γ_0, manuscript.
[8] 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.
[9] T. Coquand and E. Palmgren, Intuitionistic choice and classical logic, *Arch. Math. Logic* **39** (2000), 53-74.
[10] S. Feferman, Transfinite recursive progressions of axiomatic theories, *J. Symbolic Logic* **27** (1962), 259-316.
[11] ———, Systems of predicative analysis, *J. Symbolic Logic* **29** (1964), 1-30.
[12] ———, 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.
[13] ———, Predicatively reducible systems of set theory, in *Axiomatic Set Theory (Proc. Symp. Pure Math. XIII, Vol. 2)*, pp. 11-32, 1974.
[14] ———, A more perspicuous formal system for predicativity, in *Konstruktionen versus Positionen*, Vol. I, pp. 68-93, Walter de Gruyter, 1979.
[15] ———, Reflecting on incompleteness, *J. Symbolic Logic* **56** (1991), 1-49.
[16] ———, 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.
[17] ———, 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.
[18] ———, Comments on "Predicativity as a philosophical position" by G.R. Hellman, *Revue Internationale de Philosophie* **229** (2004), 313-323.
[19] ———, Predicativity, in *The Oxford Handbook of the Philosophy of Mathematics and Logic*, pp. 590-624, Oxford University Press, 2005
[20] S. Feferman and T. Strahm, The unfolding of non-finitist arithmetic, *Ann. Pure Appl. Logic* **104** (2000), 75-96.
[21] H. M. Friedman, Internal finite tree embeddings, in *Reflections on the Foundations of Mathematics (Stanford, CA, 1998)*, pp. 60-91, Assoc. Symbol. Logic, 2002.
[22] 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.
[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. Jager, 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, *Proof Theory: An Introduction*, Springer LNM 1407, 1989.

[37] ———, Pure proof theory: aims, methods and results, *Bull. Symbolic Logic* **2** (1996), 159-188.

[38] M. Rathjen, Predicativity, circularity, and anti-foundation, in *One Hundred Years of Russell's Paradox*, pp. 191-219, de Gruyter, 2004.

[39] M. Rathjen and A. Weiermann, Proof-theoretic investigations on Kruskal's theorem, *Ann. Pure Appl. Logic* **60** (1993), 49-88.

[40] C. Rüede, Transfinite dependent choice and ω-model reflection, *J. Symbolic Logic* **67** (2002), 1153-1168.

[41] K. Schütte, Predicative well-orderings, in *Formal Systems and Recursive Functions (Proc. Eighth Logic Colloq., Oxford, 1963)*, pp. 280-303, North-Holland, 1965.

[42] ———, Eine Grenze für die Beweisbarkeit der transfiniten Induktion in der verzweigten Typenlogik, *Arch. Math. Logik Grundlagenforsch* **7** (1965), 45-60.

[43] ———, *Proof theory*, Grundlehren der Mathematischen Wissenschaften **225**, Springer-Verlag, 1977.

[44] S. G. Simpson, *Subsystems of Second Order Arithmetic*, Springer-Verlag, 1999.

[45] ———, Predicativity: the outer limits, in *Reflections on the Foundations of Mathematics (Stanford, CA, 1998)*, pp. 130-136, Assoc. Symbol. Logic, 2002.

[46] T. Strahm, Wellordering proofs for metapredicative Mahlo, *J. Symbolic Logic* **67** (2002), 260-278.

Department of Mathematics, Washington University in Saint Louis, Saint Louis, MO 63130

Email address: nweaver@math.wustl.edu