PREDICATIVISM ABOUT CLASSES

What are classes?[1] More precisely, what are the objects of the second sort of second-order set theory? The objects of the first sort are, of course, sets, but classes have sets as their members and behave like sets. Allegedly, the subject matter of set theory comprises all the collections there are. If classes are collections of any kind, why can't we just count them among the subject matter of set theory? A plethora of paradoxes, however, teach us that many classes are "proper" and cannot be sets. So, what are classes after all?

I. FOUR COMPETING VIEWS ABOUT CLASSES

In the present paper, we will particularly consider the following four prominent views about classes.

- *Reductionism*: Set theory need not take classes at face value, and talk of classes in set theory can and should be reduced to talk of sets.
- *Plural interpretation*: Classes are sub-pluralities of the set-theoretic universe $\mathbb{V}$.
- *Mereological interpretation*: Classes are mereological parts of $\mathbb{V}$.
- *Predicativism*: Classes are predicates of sets.[2]

The theme of this article is predicativism. Predicativism in its modern form is credited to Parsons.[3] Nowadays, however, predicativism is often

[1] I would like to thank Volker Halbach, Daniel Isaacson, Penelope Maddy, and Timothy Williamson for insightful and helpful comments. I am indebted to Ali Enayat for invaluable communications and information about the theory of satisfaction classes, and to Neil Barton and Sy-David Friedman for answering my questions about their article. I am deeply grateful to the two anonymous referees for very extensive comments and helpful suggestions, which led to a considerable improvement of the article. Finally, I would like to express my special thanks to Leon Horsten and Philip Welch for immensely fruitful discussions on the topic of this article.

[2] We use the term 'predicative' in the (literal) sense of "having the quality of predicating something" or "forming or having the function of a predicate" (*OED*) here and thus call the view at issue "predicativism." The term is, however, often used in philosophy and logic with a technical (figurative) meaning, according to which something is called predicative if it is defined, expressed, or constructed without quantification over domains that include it. If one wants to reserve the term 'predicative' for the latter technical use, one could alternatively call the view "predicationalism" or "predicational interpretation of classes."

[3] Charles Parsons, "Sets and Classes," in *Mathematics in Philosophy* (Ithaca: Cornell University Press, 1983), pp. 209–220.

0022-362X/18/0000/1

viewed negatively and sometimes even considered "obsolete" compared to other relatively new views, such as the plural interpretation. The aim of this article is to propose a form of predicativism that we call *liberal predicativism*, and argue for it against the other three views in light of recent developments in mathematical logic: a number of new developments in the use of classes in set theory, as well as in theories of classes *per se*, were recently made, and we think that it is a good time to re-evaluate predicativism in light of them. Predicativism has been considered too restrictive and unable to accommodate the use of classes in modern set theory. Most of the criticisms against predicativism so far are based on this diagnosis, but we will argue that it is true only of a specific type of predicativism and does not apply to our liberal predicativism. Our conclusion will be that predicativism is still a highly viable option, and, in particular, our liberal predicativism provides a sufficiently versatile and workable nominalist concept of classes for set theory. We will explain predicativism in detail separately in §II, and in the rest of this section we will briefly explain the other three views.

There are two different versions of reductionism. The first simply identifies classes with formulae of the first-order language $\mathcal{L}_\in$ of set theory (possibly with parameters). Accordingly, a membership statement "a set x is a member of a class X" means nothing but "Φ holds for x" for the $\mathcal{L}_\in$-formula Φ identified with the class X; quantification over classes is then shorthand for meta-theoretic quantification over $\mathcal{L}_\in$-formulae. Let us call this view *definabilism* about classes. For example, the definabilist reads "every class of ordinals is well-founded" as the following *schematic* (*meta-theoretic*) statement:

> For all $\mathcal{L}_\in$-formula Φ, the Zermelo-Fraenkel set theory ZF (or whatever theory of sets one adopts) proves that if Φ holds only for ordinals, then there is a set x such that Φ holds for x but none of the members of x.

In other words, definabilism claims that all and only classes are first-order definable in $\mathcal{L}_\in$. The second version of reductionism holds that the content of statements involving classes, particularly those involving undefinable classes, should not be taken at face value, and these statements should be understood as statements relativized to set-sized structures of the form $(V_\kappa, \in, V_{\kappa+1})$ for strongly inaccessible cardinals κ (or ordinals κ of some specific kind). Let us call this view *hermeneutic reductionism* about classes.

The plural interpretation of classes was proposed by Boolos, and it takes classes as sub-pluralities of $\mathbb{V}$.[4] Uzquiano explains that the plural

[4] George Boolos, "To Be Is to Be a Value of a Variable (or to Be Some Values of Some Variables)," *The Journal of Philosophy*, LXXXIII, 8 (August 1984): 430–49.

interpretation "construe[s] reference to classes not as singular reference to gigantic collections other than sets, but rather as plural reference to sets."[5] That is to say, the plural interpretation reads the second-order quantifier $\exists X$ of the second-order language $\mathcal{L}^2_\in$ of set theory as "there are *some sets* ..." and the membership relation $x \in X$ as "x is one of them," where "them" refers to the plurality denoted by X. Accordingly, "every class of ordinals is well-founded" is interpreted as follows:

> Whenever there are some sets such that each of them is an ordinal, there is a set such that it is one of them and every member of it is not one of them.

The plural interpretation is the most significant nominalist competitor of predicativism, and we will discuss it in more detail in §V in comparison with predicativism.

The mereological interpretation was recently proposed by Welch and Horsten.[6] They introduced it to give a philosophical foundation and content to Welch's new set-theoretic principle called the *global reflection principle* (henceforth, *GRP*):

GRP There exist a class j and an ordinal κ such that j is an elementary embedding from $(V_\kappa, \in, V_{\kappa+1})$ to $(\mathbb{V}, \in, \mathcal{C})$ with a critical point κ,

where $\mathcal{C}$ is the domain of second-order quantifiers, that is, the collection of all classes (mereological parts of $\mathbb{V}$ for them) there are.[7] The mereological interpretation is manifestly a form of realism about classes and takes classes as objects independent of sets, while the other three views try to avoid ontological commitment to any object beyond sets even if they adopt realism about sets.

There is another influential view besides the aforementioned four. *Potentialism* about sets, which has its origin in the work of Zermelo,[8] holds that the mathematical universe comprises an open-endedly and inexhaustibly growing hierarchy of "normal domains" of set theory, and

[5] Gabriel Uzquiano, "Plural Quantification and Classes," *Philosophia Mathematica*, XI, 1 (February 2003): 67–81, at p. 72.

[6] Philip Welch and Leon Horsten, "Reflecting on Absolute Infinity," *The Journal of Philosophy*, CXIII, 2 (February 2016): 89–111.

[7] The embedding asserted to exist by *GRP* is an embedding between second-order structures, and different formulations of *GRP* are possible depending on what formulae are required to be preserved thereby. In this article, we will focus on their "official" version, called $GRP_{\Sigma^0_\infty}$ by them, which asserts that the embedding preserves all *first-order* formulae (possibly with class parameters).

[8] Ernst Zermelo, "On Boundary Numbers and Domains," M. Hallett, trans., in W. Ewald, *From Kant to Hilbert: A Source Book in the Foundation of Mathematics*, vol. 2 (New York: Oxford University Press, 1996), pp. 1208–33.

quantification over sets must always be restricted to one fixed "normal domain" among many others. Potentialism thereby identifies sets and classes without falling into inconsistency by taking classes as sub*sets* of a fixed "normal domain," which are in turn members of larger "normal domains" at higher levels of the hierarchy. One problem of potentialism is that it allows for too few assertions about absolutely all sets. Another problem is that it requires an independent domain over which the "normal domains" are supposed to range, but we have had no established mathematical theorization of such a domain.[9] These problems generally apply to *multiversism*, which holds that there are more than one set-theoretic universes that set theory concerns; in contrast, *universism*, the counterview to multiversism, takes the set-theoretic universe as a unique completed totality. We will not get into the ongoing debate between universism and multiversism, which is far from settled, and we will exclude potentialism from the scope of this article. In what follows, we presuppose universism and focus on those views about classes that are compatible with universism, which include the four aforementioned views; throughout this article, the symbol $\mathbb{V}$ will denote the *single universist* universe of sets.

II. PREDICATIVISM

Predicativism takes classes as predicates of sets. Parsons explains that the basic intuition behind predicativism is that "the introduction of the notion of class answers to a general need to generalize on predicate places in the language."[10]

II.1. Strict Predicativism. The first question to ask is what predicates of sets should be counted among the domain of second-order quantifiers. Parsons is not perfectly clear on this point, but he is often taken by other philosophers to be committed to the view that class quantification is substutitional quantification over $\mathcal{L}_\in$-formulae with parameters in terms of the notion of *truth of* or *satisfaction*.[11] Let us formally explain this, essentially following Parsons's own way.[12] Either a truth predicate or a satisfaction predicate serves our purpose; we adopt a truth predicate

[9] Zermelo called a theory of the domain of "normal domains" a *meta-set theory*. For a detailed exposition of Zermelo's view on his meta-set theory, see Hans-Dieter Ebbinghaus, "Zermelo: Boundary Numbers and Domains of Sets Continued," *History and Philosophy of Logic*, XXVII, 4 (November 2006): 285–306.

[10] Parsons, "Sets and Classes," *op. cit.*, p. 211.

[11] For instance, Uzquiano, and Welch and Horsten, both construe Parsons in this way. See Uzquiano, "Plural quantification and classes," *op. cit.*, p. 79, and Welch and Horsten, "Reflecting on Absolute Infinity," *op. cit.*, p. 104.

[12] Parsons uses a theory of satisfaction equivalent to CT^- in interpreting theories of classes in the same way that the translation $\mathcal{I}$ below does. However, CT^- does not prove the $\mathcal{I}$-translations of the second-order axioms of separation and replacement. Hence, we

in this article because all the relevant technical details are found in the existing literature.[13] We first expand $\mathcal{L}_\in$ to a first-order language $\mathcal{L}_\in^+$ by adding a truth predicate T. The predicate T takes names of $\mathcal{L}_\in$-sentences with parameters as its arguments, and thus we presuppose a suitable coding scheme of the language $\mathcal{L}_\in^\infty = \mathcal{L}_\in \cup \{\mathsf{c}_a \mid a \in \mathbb{V}\}$ with constant symbols c_a for all sets $a \in \mathbb{V}$; thereby, we can express a formula $\varphi(x, b)$ with a parameter b being *true of* a by saying that T holds for the code of the $\mathcal{L}_\in^\infty$-sentence $\varphi(\mathsf{c}_a, \mathsf{c}_b)$. We define an $\mathcal{L}_\in^+$-theory CT^- as an extension of ZF with the axioms asserting the usual Tarskian inductive clauses, such as "an atomic formula $x \in y$ is true of a and b if and only if a is indeed a member of b" and "for every $\mathcal{L}_\in^\infty$-sentence σ, $\neg\sigma$ is true if and only if σ is not true." We further extend CT^- with the following two *axioms* Sep_T and $Repl_T$ (not axiom schemes!):

Sep_T For all $\mathcal{L}_\in^\infty$-formulae $\varphi(x)$ with exactly one free variable and for all sets a, there is a set b such that $b = \{d \in a \mid \varphi \text{ is true of } d\}$;

$Repl_T$ For all $\mathcal{L}_\in^\infty$-formulae $\varphi(x, y)$ with exactly two free variables and for all sets a, if $\{\langle d_0, d_1\rangle \mid \varphi \text{ is true of } d_0 \in a \text{ and } d_1\}$ defines a function, then its range is a set;

that is to say, they are truth-theoretic expressions of the second-order axioms of separation and replacement. We call the resulting theory CT_0. We can easily verify that CT_0 interprets the Von Neumann-Bernays-Gödel theory NBG of classes by translating $\forall X$ into "for all $\mathcal{L}_\in^\infty$-formulae with one free variable" and $x \in X$ into "the $\mathcal{L}_\in^\infty$-formula X is true of x"; we will denote this translation of $\mathcal{L}_\in^2$ into $\mathcal{L}_\in^+$ by $\mathcal{I}$. Furthermore, CT_0 has the same $\mathcal{L}_\in$-theorems as NBG does.[14] What we call *strict predicativism* holds that every theory S of classes is to be understood in terms of its translation $\mathcal{I}(\mathsf{S})$; for strict predicativists, NBG actually means CT_0 ($= \mathsf{CT}_0 + \mathcal{I}(\mathsf{NBG})$), and every class theory S extending NBG means $\mathsf{CT}_0 + \mathcal{I}(\mathsf{S})$. Note that strict predicativism transforms the ordinary two-sorted formalism of class theories into a one-sorted formalism by means of a truth predicate.

Both strict predicativism and definabilism intend to limit the range of classes to those definable or expressible in $\mathcal{L}_\in$, but they are different

---

change Parsons's setting and use CT_0 instead of CT^-. This change makes no essential difference to the discussion, since $\mathsf{CT}^- + \mathcal{I}(\mathsf{NBG})$ is identical with CT_0 anyway.

[13] For example, see Kentaro Fujimoto, "Classes and Truths in Set Theory," *Annals of Pure and Applied Logic*, CLXIII, 11 (November 2012): 1484–1523.

[14] This equivalence of CT_0 and NBG can be shown by the technique developed by Enayat and Visser: see Ali Enayat and Albert Visser, "New Constructions of Satisfaction Classes," in Theodora Achourioti et al. eds., *Unifying the Philosophy of Truth* (Berlin: Springer, 2015), pp. 321–35.

in two important respects. First, as opposed to definabilism, strict predicativism does not aim to eliminate reference to classes and it allows quantification over classes. Second, in model-theoretic terms, definabilism and strict predicativism coincide over ω-models of set theory, but they diverge over non-ω-models. Non-ω-models contain (codes of) nonstandard $\mathcal{L}_\in$-formulae, and the truth of such formulae can by no means be expressed in definabilism (also see fn 18).

II.2. Liberal Predicativism. Despite the widespread reading of Parsons, we doubt that Parsons indeed commits himself to strict predicativism because he also writes,

> [T]here is a sense in which the theory of classes is more general than the theory of satisfaction. ... Even if we think of classes as always the extension of predicates, the axioms of [NBG] would allow them to be extensions of predicates in any language, whether we can now specify it or not, provided they are closed under first-order logical operations. ... The usual satisfaction theory [CT_0] lacks this generality because it is based on the syntax of the language of [ZF].[15]

The form of predicativism that we propose is based on this alternative version of predicativism at which Parsons hints in the above passage. We retain the core doctrine that classes are predicates, but we do not restrict those predicates to any fixed language. As we will see in more detail in §III, modern set theory makes substantial use of classes that are not $\mathcal{L}_\in$-definable. The main problem of strict predicativism (as well as definabilism) consists in that it is formally incompatible with the use of those undefinable classes in set theory. We aim to overcome this defect of strict predicativism by "liberating" predicates from the strict predicativist constraint and allowing the introduction of a wide range of undefinable predicates as classes.

Of course, there are predicates that we should not add to our theory of sets; for example, we should not add a predicate that defines a cofinal map from ω to the class of ordinals. Let us call a predicate of sets *admissible* when it can be justifiably postulated or introduced to set theory as a definite mathematical predicate subsumed under the laws of set theory. We propose that those admissible predicates compose the domain of second-order quantifiers as the subject matter of class theory.

Our proposal is to interpret the quantifier $\exists X$ as "there exists an *admissible* predicate such that ..." or "there is a predicate we may admissibly introduce such that ..." and interpret the membership relation $x \in X$ as "the predicate X holds for x." Accordingly, "every class of ordinals is well-founded" is read as follows:

[15] Parsons, "Sets and Classes", *op. cit.*, pp. 215–16.

> No matter what predicate we admissibly introduce, if it holds only for ordinals, then there is an $\in$-minimal x such that the predicate holds for x.

We call this view *liberal predicativism* about classes. In contrast to strict predicativism, liberal predicativism preserves the ordinary two-sorted formalism of class theory. This interpretation of classes is not to be expressed as a theory of truth or satisfaction because we want to keep the domain of admissible predicates unrestricted to any formula of any particular language with any particular interpretation; otherwise, we would have to exclude all the classes from set theory that are not definable in terms of the specific predicates of a fixed language, and the resulting picture would be too restrictive for the actual practice of set theory, as we will see in §III.[16]

A general characterization of the notion of admissibility of set-theoretic predicates is expected to be difficult. What predicates are admissible depends on the view one holds about sets and the universe of sets. For example, for those who believe $\mathbb{V} = \mathbb{L}$, an (undefinable) predicate that defines a non-trivial elementary embedding from some inner model to $\mathbb{V}$ is not admissible; in contrast, if one holds the view that $\mathbb{V}$ is so rich that it has a non-trivial elementary substructure, then she may well want to introduce such a predicate for expressing her view. The predicativism about classes discussed in this article is a particular view about *what classes are.* The particular conception of classes it upholds gives partial guidance about what classes should and should not exist, but it does not solely determine exactly *what classes exist.* The latter question is also about the admissibility of predicates (for proponents of liberal predicativism) and is to be answered by a comprehensive consideration of set theory as a whole. As we will see later, a second-order axiom of set theory is often derived from a particular view about the universe of sets and/or sets themselves, and one view behind one axiom may be incompatible with another view behind another axiom. The questions of which view on the universe $\mathbb{V}$ of sets is correct, and what

[16] There is, however, an alternative formalization of the idea of liberal predicativism that renders theories of classes as theories of truth or satisfaction by interpreting a class as a formula of an expansion $\mathcal{L}$ of $\mathcal{L}_{\in}^{\infty}$ with *infinitely many completely uninterpreted* predicates; this alternative is hinted at in Parsons, "Sets and Classes," *op. cit.*, p. 216. This approach does not suffer from the ω-inconsistency phenomenon that we will discuss below, and it may appeal more to those who accept the Quinean criterion of objecthood, which we will discuss in §V, because it involves no second-order quantifiers on the face of it. However, we think that this approach only disguises second-order quantifiers because it requires a delineation of the range of possible interpretations those *uninterpreted* predicates can receive, which seems to mean nothing but the range of admissible predicates; it also results in a technically unnatural and unnecessarily cumbersome formalism. We do not get into the details here, but some details will be discussed in Kentaro Fujimoto, "On Parsons's Predicative Interpretation of Classes," in preparation.

axioms are adequate for set theory, are far from settled. We will give some sample cases of liberal predicativist justifications of second-order axioms of set theory, which suffices anyway for our current purpose of providing a predicative interpretation of classes with sufficiently rich mathematical consequences, but we do not mean that only those predicates are admissible, and we would like to leave open which predicates are admissible. In particular, we would like to emphasize that we do not intend in this article to propose any particular single formal system as *the* system of liberal predicativism.[17]

II.3. Impredicative Comprehension. We end this section by discussing one general limitation of predicativism. For a collection Γ of $\mathcal{L}^2_\in$-formulae, the Γ-comprehension axiom (henceforth, Γ-CA) is an axiom scheme asserting

$$\exists X \forall x(x \in X \leftrightarrow \Phi(x)), \text{ for all } \Phi \text{ belonging to } \Gamma.$$

The collections Π^1_n and Σ^1_n of $\mathcal{L}^2_\in$-formulae ($n \in \mathbb{N}$) are standardly defined; in particular, Π^1_0- and Σ^1_0-formulae are those with no second-order quantifier but possibly with second-order parameters, and we call them *elementary* formulae; the familiar Morse-Kelley class theory MK is defined as $\mathsf{NBG} + \{\Sigma^1_n\text{-CA} \mid n \in \mathbb{N}\}$. Following the convention, although this is slightly confusing because our theme is predicativism, we call a formula not equivalent to an elementary one *impredicative.*

It is often argued that predicativism is incompatible with impredicative comprehension, namely, Γ-CA for a collection Γ containing impredicative formulae. Indeed, strict predicativism is ω-inconsistent with impredicative comprehension, since an extension of CT_0 with the extended arithmetical induction scheme for arbitrary $\mathcal{L}^+_\in$-formulae is inconsistent with $\mathcal{I}(\Sigma^1_1\text{-CA})$. This can be shown as follows: Σ^1_1-CA produces a *truth class* X in the sense that $(\mathbb{V}, \in, X)$ is a model of CT_0; hence, if classes are $\mathcal{L}^\infty_\in$-formulae as strict predicativism holds, then we have an $\mathcal{L}^\infty_\in$-formula X that renders a truth predicate for $\mathcal{L}^\infty_\in$; next, by means of the extended arithmetical induction, we can show by induction on the complexity of $\mathcal{L}^\infty_\in$-sentences that

for all $\mathcal{L}^\infty_\in$-sentences σ, X is true of the code of σ, if and only if σ is true;

[17] In proof theory, the term "predicativism" is often used with the technical "figurative" meaning (in the sense of fn 2) and refers to a particular view about what sets of natural numbers exist; it claims that there only exist sets of natural numbers that belong to the ramified hierarchy (of sets of natural numbers) up to the Feferman-Schütte ordinal Γ_0. One could meaningfully apply this idea to the current context and claim that there only exist classes that can be constructed by a transfinite iteration of elementary comprehension. However, this would result in the exclusion of many classes in use in set theory, such as a class of ordinal indiscernibles, that transcend those "predicative" classes.

finally, by diagonalization (carried out within the truth predicate), we obtain an $\mathcal{L}_{\in}^{\infty}$-sentence λ such that

X is not true of the code of λ, if and only if λ is true,

from which we can draw a contradiction by a simple "liar" argument.[18] This ω-inconsistency results exactly from the restriction on classes to predicates of the specific language $\mathcal{L}_{\in}$.[19]

What about liberal predicativism? It is *ω-consistent* with impredicative comprehension, but we suspect that liberal predicativism is not conceptually coherent with, nor can it justify, impredicative comprehension. We can think of whatever $\mathcal{L}_{\in}^{2}$-formula as defining a predicate, but it may not be admissible. Hence, impredicative comprehension requires a justification of the admissibility of predicates whose satisfaction conditions are specified in terms of quantification over all admissible predicates. Then we face a familiar challenge: do we have a determinate totality of admissible predicates? The following passage from Parsons indicates that his answer would be negative, with which we agree:

> [I]t seems evident that the "totality" of possible predicates is irremediably potential .... To what extent can we regard such possibilities as determinate at any given point? It seems that we have no conception of a totality of all possible languages and interpretations on the basis of which we might claim such determinacy.[20]

A problem of the lack of a determinate totality of admissible predicates is that impredicative statements may then not have definite truth values. Admissible predicates ought to be justifiably introduced into set

[18] Interestingly, the existence of a truth class is ω-inconsistent but *consistent* with strict predicativism. Smith showed that a recursively saturated countable model of PA can have a truth class, say, X, and (a code of) a *nonstandard* arithmetical formula φ such that $\{n \in \mathbb{N} \mid \varphi \text{ is true of } n \text{ in terms of } X\}$ defines another truth class; it is folklore that his result can be generalized to the claim that there is a non-ω-model of CT_0 in which $\{n \in \mathbb{V} \mid \varphi \text{ is true of } a \text{ (in terms of the default truth predicate } T \text{ of } \mathsf{CT}_0)\}$ is a truth class for some nonstandard φ. See Stuart Smith, "Nonstandard Definability," *Annals of Pure and Applied Logic*, XLII, 1 (March 1989): 21–43. In contrast, the existence of a truth class is simply inconsistent with definabilism by Tarski's undefinability theorem, since definabilism renders classes as *standard* $\mathcal{L}_{\in}$-formulae.

[19] Here we present only the simplest type of derivation of ω-inconsistency to avoid unnecessary technical complication. There are different types of derivations of ω-inconsistency from different types of axioms. For instance, the proofs of ω-inconsistency mentioned in fn 23 and fn 34 are both different from the one illustrated here. However, these ω-inconsistency phenomena share the same root: they are all caused by the use of classes that are not $\mathcal{L}_{\in}$-definable. The details of the proofs of those ω-inconsistency results (of strict predicativism) asserted in this article will be included in Fujimoto, "On Parsons's Predicative Interpretation of Classes," *op. cit.*

[20] Parsons, "Sets and Classes," *op. cit.*, p. 217.

theory as a definite mathematical predicate subsumed under the laws of set theory. In particular, each of them is expected to have a definite truth value for each set as its argument; for instance, the axiom of separation asserts that $\{x \in a \mid P(x)\}$ is a set, where P demarcates the set a into two parts by prescribing which $x \in a$ is a member of this set or not, and, if P were indefinite, it would fail to give a definite demarcation, and it would be hard to justify that $\{x \in a \mid P(x)\}$ picks a unique set, which is a mathematically definite entity.[21] Hence, if a predicate P lacks a definite satisfaction condition, then P is deemed to be inadmissible.

Our lack of a conception of a determinate totality of admissible predicates might only be due to our epistemic or conceptual limitation and not necessarily imply the, say, metaphysical non-existence of such a totality or the semantic indeterminacy of impredicative assertions. However, even if there is a determinate totality of admissible predicates and if impredicative statements have definite truth values, impredicative comprehension is still highly controversial. How can we justify and on what ground can we postulate that such a totality of admissible predicates, which transcends our epistemic or conceptual capacity, has a desired structure that validates impredicative comprehension and reasoning? In particular, the notion of admissibility concerns the justifiability of the introduction of predicates and thus is a *metamathematical* notion in some sense. Then, how can we justify that an impredicative predicate partly defined in terms of such a metamathematical notion can be justifiably introduced as a definite *mathematical* predicate? It would be a highly non-trivial (and seemingly difficult) task to answer these questions in favor of impredicative comprehension.

In conclusion, it appears that impredicative comprehension is not generally congruent with the idea of liberal predicativism. Nonetheless, we want to emphasize that the argument so far does not imply that every impredicative statement is meaningless or unjustifiable with liberal predicativism. For instance, each instance of the axiom scheme Π^1_0-CA of NBG is a Π^1_2-statement, but Π^1_0-CA just expresses that admissible predicates are closed under first-order logical operations, which is perfectly acceptable for liberal predicativism.

We would welcome a convincing argument that justifies impredicative comprehension from the liberal predicativist point of view, but we are rather pessimistic about the existence of such an argument (also see §V). One of the main points of this article is that, even if we compromise

[21] For the same reason, liberal predicativism does not seem to be congruent with the idea of extending the axiom scheme of separation to arbitrary $\mathcal{L}^2_\in$-formulae.

to give up impredicative comprehension, we can still have plausible justifications of class-theoretic principles strong enough to accommodate the use of classes in modern set theory.

III. THE USE OF AND NEED FOR CLASSES IN SET THEORY

In this section, we will consider two desiderata for an appropriate interpretation of classes and argue that definabilism and strict predicativism fail to meet them and should be abandoned. The argument uses several concrete examples, with respect to which the two views fail to meet the desiderata, and we will later show in subsequent sections that, by contrast, liberal predicativism accommodates them well. We will also see at the end of this section that these examples also suggest some problems of hermeneutic reductionism. Throughout the following discussion, every theory of classes is assumed to include NBG.

The first desideratum is: (i) an appropriate interpretation of classes should not trivialize highly regarded mathematical theorems in set theory nor set-theoretic axioms under serious consideration in set theory. We will see three examples with respect to which definabilism and/or strict predicativism fail to meet this desideratum.

Example 1. Kunen proved that every elementary embedding from $\mathbb{V}$ into itself is trivial if $\mathbb{V}$ is a model of ZFC, and this theorem raised a prominent open problem, asking whether the same holds without assuming AC in $\mathbb{V}$.[22] Now, the definabilist reads Kunen's theorem as

(1) If an $\mathcal{L}_{\in}$-formula $\varphi(x, y)$ possibly with parameters defines an elementary embedding from $\mathbb{V}$ into itself in ZFC, then this elementary embedding is trivial,

and construes the open problem as asking whether (1) still holds when we replace ZFC with ZF. Suzuki gave a purely elementary proof of (1), and his proof also gives an affirmative answer to the definabilist reading of the open problem.[23] However, his proof is usually not counted as an alternative proof to Kunen's theorem, and his result is not regarded as the answer to the open problem either. This reaction to Suzuki's result from set-theorists seems to indicate that

[22] Kenneth Kunen, "Elementary Embeddings and Infinitary Combinatorics," *The Journal of Symbolic Logic*, XXXVI, 3 (September, 1971): 407–413.

[23] Akira Suzuki, "No Elementary Embedding from V into V Is Definable from Parameters," *The Journal of Symbolic Logic*, LXIV, 4 (December 1999): 1591–1594. As a matter of fact, Suzuki's proof can also be modified into an "elementary" proof of the ω-inconsistency of the strict predicativist interpretations of Kunen's theorem and the open problem.

set-theorists do not identify Kunen's theorem with (1), and definabilism is not accepted in the community of set-theorists.[24]

Example 2. *GRP* implies that the critical point κ of the postulated elementary embedding j is strongly inaccessible, and thus there is $X \subset V_\kappa$ with $(V_\kappa, \in, X) \models \mathsf{CT}_0$; therefore, *GRP* yields a truth class $j(X)$ and thus is ω-inconsistent with the strict predicative interpretation of classes. Hence, the strict predicative interpretation *trivially refutes* Welch's *GRP* as an appropriate axiom of set theory. With definabilism, in turn, we simply cannot express *GRP* because the match-up between $V_{\kappa+1}$ and $\mathcal{C}$ cannot be expressed.

Example 3. Vickers and Welch proved that if there is a proper class I of good indiscernibles for $(\mathbb{V}, \in)$, then there are an inner model $\mathbb{M}$ and an elementary embedding j from $\mathbb{M}$ into $\mathbb{V}$ with a critical point both definable in $(\mathbb{V}, \in, I)$.[25] Suzuki's aforementioned theorem can be strengthened to the $\mathcal{L}_\in$-undefinability of such a pair of $\mathbb{M}$ and j as well as to the ω-inconsistency of the existence of such a pair with strict predicativism (cf. fn 23). Hence, definabilism and strict predicativism trivialize, or significantly impair the value of, the theorem.[26]

The second desideratum is: (ii) an appropriate interpretation of classes should provide a mathematical framework in which (or, at least, should be compatible with mathematical presuppositions under which) widely accepted and/or mathematically fruitful uses of classes in set theory can be meaningfully expressed and implemented. We will see two examples below in which definabilism and/or strict predicativism fail to meet this desideratum.

Example 4. Audrito and Viale recently proposed new axioms of set theory, called *iterated resurrection axioms*, and a new notion of large cardinal, called an (α)-uplifting cardinal, for an ordinal α, which bear significant implications for generic absoluteness.[27] Both are defined in

[24] The common view on Kunen's theorem among set-theorists seems to be hermeneutic reductionism or a variant of hermeneutic reductionism that takes the real content of the theorem not as what it literally states but as various consequences of its proof collectively, such as the non-existence of an elementary embedding from $V_{\delta+2}$ into itself for any δ.

[25] John Vickers and Philip Welch, "On Elementary Embeddings from an Inner Model to the Universe," *The Journal of Symbolic Logic*, LXVI, 3 (September 2001): 1090–1116.

[26] Actually, a truth class can be easily defined in terms of a proper class I of good indiscernibles for $(\mathbb{V}, \in)$, from which the $L_\in$-undefinability and the ω-inconsistency at issue readily follow without appealing to Suzuki's theorem.

[27] Giorgio Audrito and Matteo Viale, "Absoluteness via Resurrection," *The Journal of Mathematical Logic*, XVII, 2 (December 2017): 36 pages.

terms of the existence of a winning strategy for a certain clopen *class* game. Hence, the determinacy of clopen class games is a natural presupposition for investigating them. However, Gitman and Hamkins proved that the determinacy of clopen class games is not provable in NBG and is actually equivalent modulo NBG to the axiom of *elementary transfinite recursion* (henceforth, ETR), which asserts that every elementary recursive definition along any well-founded class relation has a solution.[28] Hence, a natural background theory for the iterated resurrection axioms and (α)-uplifting cardinals is the class theory $\mathsf{NBG} + \mathrm{ETR}$, but $\mathcal{I}(\mathrm{ETR})$ is ω-inconsistent with CT_0 because ETR yields a truth class; with definabilism, ETR is simply false.[29]

Example 5. Friedman proposed a new axiom of set theory called the *Inner Model Hypothesis* (henceforth, *IMH*):

IMH If a sentence φ holds in an inner model of some outer model of $\mathbb{V}$, then it already holds in some inner model of $\mathbb{V}$.[30]

This axiom is intended to assert that the set-theoretic universe $\mathbb{V}$ is maximally rich in some sense. Antos et al. formulate a universist reading of *IMH* using class-theoretic machinery, which is meant to express that the truth in any "ideal" possible inner model can be realized in an inner model of the real actual universe.[31] A crucial technical difficulty in formulating this idea is that we have no means to directly refer to "ideal" outer models of $\mathbb{V}$. To circumvent this difficulty, they introduce the notion of $\mathbb{V}$-logic, a deductive system for languages extending $\mathcal{L}_{\in}^{\mathbb{V}} := \mathcal{L}_{\in}^{\infty} \cup \{\overline{\mathbb{V}}\}$ with a new predicate symbol $\overline{\mathbb{V}}$ for $\mathbb{V}$, which contains, besides other axioms and rules, an infinitary inference rule called the $\mathbb{V}$-*rule*:

$\mathbb{V}$-rule If a formula $\varphi(x)$ holds for c_a for all $a \in \mathbb{V}$, we can infer $(\forall x \in \overline{\mathbb{V}})\varphi(x)$.

Thereby, they reformulate *IMH* in terms of the consistency of theories in $\mathbb{V}$-logic:

[28] Victoria Gitman and Joel Hamkins, "Open Determinacy for Class Games," in Andres E. Caicedo et al. eds., *Foundations of Mathematics: Logic at Harvard: Essays in Honor of Hugh Woodin's 60th Birthday* (American Mathematical Society, 2016), pp. 121–144.

[29] As a matter of fact, regardless of the assumption of ETR, the existence of an (ω)-uplifting cardinal is contradictory in itself with the definabilist interpretation, since the asserted existence of a winning strategy implies the existence of a truth class (in NBG); similarly, the assertion is ω-inconsistent with the strict predicativist interpretation; see Kentaro Fujimoto, "An (ω)-uplifting cardinal is not first-order definable," submitted.

[30] Sy-David Friedman, "Internal Consistency and Inner Model Hypothesis," *The Bulletin of Symbolic Logic*, XII, 4 (December 2006): 591–600.

[31] Carolin Antos, Neil Barton, and Sy-David Friedman, "Universism and Extensions of V": preprint available from https://arxiv.org/abs/1708.05751.

IMH$^{\vdash_{\mathbb{V}}}$ If a $\mathbb{V}$-logic theory T satisfying some condition and a first-order sentence φ are consistent in $\mathbb{V}$-logic, then there is an inner model of $\mathbb{V}$ satisfying φ.

They justify this reformulation by appealing to a certain type of completeness theorem, which informally says that if T is consistent in $\mathbb{V}$-logic then there is an extension ("outer model") of $\mathbb{V}$ satisfying T.[32] Since derivations in $\mathbb{V}$-logic may be of proper class size, they need a theory of classes rich enough to talk about $\mathbb{V}$-logic and its consequences, and they choose $\mathsf{NBG}+\Sigma^1_1$-CA as such. However, as we have seen, Σ^1_1-CA is ω-inconsistent with strict predicativism.[33]

We have seen that definabilism and strict predicativism fail to accommodate the use of classes in modern set theory and do not meet the two desiderata, as shown by the five examples above.[34] Now, what about hermeneutic reductionism? Both Kunen's theorem and Vickers and Welch's theorem hold in $(V_\kappa, \in, V_{\kappa+1})$ for any strongly inaccessible κ; since almost all class-theoretic arguments in set theory are given within some class theory, such as MK, satisfiable in $(V_\kappa, \in, V_{\kappa+1})$, we can safely say that theorems in class theory are automatically transformed into theorems relativized to $(V_\kappa, \in, V_{\kappa+1})$. However, when it comes to postulation of axioms, hermeneutic reductionism often renders an axiom involving classes as something inequivalent and unintended. Firstly, an axiom formulated in terms of classes and its hermeneutic reductionist interpretation, such as the assertion of the existence of a models of the axiom of the form $(V_\kappa, \in, V_{\kappa+1})$ for a strongly inaccessible κ, often have different consistency strengths: the existence of a cardinal κ with $(V_\kappa, \in, V_{\kappa+1}) \models$ *GRP* has higher consistency strength than *GRP* (modulo, say, MK), for example, and the same applies to the iterated

[32] More precisely, they assume, besides the derivability relation in $\mathbb{V}$-logic, the existence of a class X that "codes" the least admissible structure $Hyp(\mathbb{V})$ containing $\mathbb{V}$ as its element, and then they construct countable "copies" M and N of $\mathbb{V}$ and $Hyp(\mathbb{V})$ by a reflection (and collapsing) argument; thereby, the consistency of T in $\mathbb{V}$-logic is transfered to the consistency in M-logic (namely, $\mathbb{V}$-logic in the sense of M) in N and then some model of T is obtained by Barwise's $\mathfrak{M}$-completeness theorem: see Barwise, *Admissible Sets and Structures* (Berlin: Springer, 1975), Chapter III.3.

[33] The notion of consistency (or derivability) in $\mathbb{V}$-logic is not $\mathcal{L}_\in$-definable in itself, since a truth predicate could otherwise be defined by defining "an $\mathcal{L}^\infty_\in$-sentence φ is true" as "$\neg\varphi^{\overline{\mathbb{V}}}$ is inconsistent (or $\varphi^{\overline{\mathbb{V}}}$ is derivable) in $\mathbb{V}$-logic from no premises," where $\varphi^{\overline{\mathbb{V}}}$ is the result of restricting all quantifiers in φ to $\overline{\mathbb{V}}$.

[34] Another interesting example is an axiom asserting that the class Ord of ordinals is "weakly compact" (in the sense that Ord has the tree property). Enayat showed that this axiom implies (and is equiconsistent with) the existence of n-Mahlo cardinals for every standard natural number n; see Ali Enayat, "Set Theory with a Class of Indiscernibles," unpublished manuscript. This axiom is also ω-inconsistent with strict predicativism.

resurrection axioms, $IMH^{\vdash\forall}$, and the definition of (α)-uplifting cardinal.[35] Secondly, and more importantly, those axioms are intended to be about the entire universe, but the hermeneutic reductionist takes them to be assertions about a particular set structure. What we learn from Examples 2, 4, and 5 is that there are some proposed axioms of set theory under serious consideration that intend to assert something about the entire universe and crucially rely upon reference to (undefinable) proper classes. Thirdly, furthermore, an axiom involving classes sometimes have different logical consequences than its relativization to $(V_\kappa, \in, V_{\kappa+1})$ has. For instance, for a strongly inaccessible κ, while it is well known that κ having the partition property, $\kappa \rightarrow (\kappa)^n_2$ for all $n < \omega$, is equivalent to κ having the tree property, both of which are second-order assertions on $(V_\kappa, \in, V_{\kappa+1})$ and equivalent to κ being weakly compact, Enayat showed that the assertion that the class Ord of ordinals has the same partition property is *not* equivalent (in NBG) to the assertion that Ord has the tree property.[36] Thus, hermeneutic reductionism does not appear to us an appropriate option, if one takes those second-order axioms of set theory seriously.

IV. HOW MUCH CAN WE DO WITH LIBERAL PREDICATIVISM?

The class-theoretic principles mentioned in Examples 1–5 essentially require undefinable classes, and the ω-inconsistency of strict predicativism with those principles is exactly caused by the restriction of predicates to definable ones. Liberal predicativism overcomes this problem by allowing a variety of undefinable predicates. With liberal predicativism, predicates (as classes) need not be definable, and we may introduce new predicates by providing them with adequate axiomatic specifications or characterizations, such as "*P* is a such and such elementary embedding" and "*Q* is a such and such truth predicate." In this way, we can enrich our vocabulary and tools for investigating the universe of sets. In this section, we will give three samples of liberal predicativist justifications of class-theoretic principles beyond NBG.

To begin with, we will argue that some weak instances of impredicative comprehension are justifiable with liberal predicativism. Consider the next principle, called the Δ^1_1-Comprehension Rule (which is not an axiom but an inference rule):

Δ^1_1-CR For all $\Phi \in \Sigma^1_1$ and $\Psi \in \Pi^1_1$, if $\forall x(\Phi(x) \leftrightarrow \Psi(x))$ is proved, then we may infer $\exists X \forall x(x \in X \leftrightarrow \Phi(x))$.

[35] This situation is parallel to the inequivalence of Vopěnka's principle and the existence of a Vopěnka cardinal, for example.
[36] Enayat, "Set Theory with a Class of Indiscernibles," *op. cit.*

Suppose the premise holds for $\Phi(x) \in \Sigma^1_1$ and $\Psi(x) \in \Pi^1_1$. We take this to mean that one has established that the equivalence of Φ and Ψ invariably holds in every possible circumstance in which all the statements about sets and classes one has accepted (and all their logical consequences) are satisfied. Each such "circumstance" can be represented by a triple $(\mathbb{V}, \in, \mathcal{C})$ for some collection $\mathcal{C}$ of admissible predicates such that $(\mathbb{V}, \in, \mathcal{C})$ is a model of the set S of the axioms that one has accepted and the theorems that one has proved. Recall that universism is presupposed throughout this article, and thus the domain $\mathbb{V}$ of sets and the membership relation on them are invariable; hence, we will denote $(\mathbb{V}, \in, \mathcal{C})$ by its only varying part $\mathcal{C}$ for simplicity. Assume that one is working with a collection $\mathcal{C}$ of the admissible predicates that one has already introduced; $\mathcal{C}$ should be a model of S, and thus Φ and Ψ are equivalent in $\mathcal{C}$. Every Σ^1_1-assertion is upward persistent with respect to second-order structures with the same first-order domain. Hence, on the one hand, if Φ is satisfied in $\mathcal{C}$, then Φ continues to be satisfied in any model $\mathcal{D}$ of S expanding $\mathcal{C}$ by whatever new admissible predicates; on the other hand, if Φ is not satisfied in $\mathcal{C}$, then the Σ^1_1-assertion $\neg\Psi$ is satisfied in $\mathcal{C}$ and thus continues to be satisfied, and so does $\neg\Phi$ by the supposition, in any model $\mathcal{D}$ of S expanding $\mathcal{C}$. Consequently, one has only to look at the admissible predicates in the given $\mathcal{C}$ and need not take into account any unknown potentially admissible predicates for determining the truth value of $\Phi(x)$, which is invariable throughout every expansion of $\mathcal{C}$ by further admissible predicates. This argument seems to give a reasonable justification of Δ^1_1-CR from the liberal predicativist point of view. Unfortunately, $\mathsf{NBG}+\Delta^1_1$-CR is conservative over NBG and does not prove, say, ETR.[37] However, if one further postulates the scheme of arithmetical induction (ω-induction) for all $\mathcal{L}^2_\in$-formulae, which seems acceptable even for liberal predicativists, we can show the existence of an admissible predicate that parametrizes any finite level or even some transfinite level (below ω^ω) of the ramified hierarchy or, equivalently, of the Tarskian hierarchy of typed truths over $\mathbb{V}$.[38]

A liberal predicative justification of any stronger instances of impredicative comprehension seems difficult. As we have explained, the idea of liberal predicativism seems to be nearly intrinsically incompatible with

[37] See Fujimoto, "Classes and Truths in Set Theory," *op. cit.*, Theorem 15.

[38] The proof is the same as the proof of the corresponding fact in second-order arithmetic. For an ordinal α, a class X parametrizing the α-th level of the Tarskian hierarchy over $\mathbb{V}$ can be formally expressed as $(\mathbb{V}, \in, X)$ being a model of the theory RT_α of ramified truth; see Fujimoto, "Classes and Truths in Set Theory," *op. cit.*, Definition 35. Δ^1_1-CR plus full ω-induction for $\mathcal{L}^2_\in$ is still short of ETR but strong enough to make the notion of an (ω)-uplifting cardinal expressible, for example.

impredicative comprehension. However, liberal predicativism is not an attempt to gain mathematical richness by means of impredicative comprehension; many undefinable classes in use in set theory cannot be defined even with the help of full impredicative comprehension anyway. One principal point of liberal predicativism is that it allows us to freely (but admissibly) introduce any predicate that can be axiomatically characterized. We have just mentioned the Tarskian hierarchy of typed truths, and a truth predicate itself is another example of a likely admissible predicate. An idea behind the admissibility of a truth predicate is the so-called "implicit commitment" in the acceptance of a formal theory: if one accepts a formal mathematical theory S, then one is implicitly committed to accepting a number of further mathematical principles, objects, and predicates that are not contained in S, and among them is a truth predicate for S in particular. If one accepts a theory S, then one is implicitly committed to the truth of S and to a truth predicate for the language of S for the sake of expressing the truth of S. This thought leads us to successively accept more and more truth predicates in the Tarskian hierarhy, and it seems fairly reasonable to accept and admissibly introduce a predicate that parametrizes the Tarskian hierarchy along any ordinal and even along any class well-ordering. The resulting axiom, asserting that there is an admissible predicate parametrizing the Tarskian hierarchy along any class well-ordering, is actually equivalent to ETR modulo NBG.[39] This seems to give a liberal predicativist justification of ETR.

By making use of a truth predicate richer than the Tarskian typed one, we can justify even stronger class-theoretic principles than ETR. Many of Feferman's works were concerned with the aforementioned problem of "implicit commitment," and he proposed his theory KF of untyped truth in an attempt to answer what we are implicitly committed to in accepting a theory S and what we can justifiably accept on the same fundamental ground as our initial acceptance of S.[40] Following Feferman's idea, we suggest that, in accepting an admissible predicate P, we can admissibly introduce an untyped truth predicate T, in the sense of KF, for the language $\mathcal{L}_T(P) := \mathcal{L}_\in \cup \{T, P\}$. Putting $\mathcal{L}_T^\infty(P) := \mathcal{L}_\in^\infty \cup \{P, T\}$, this is formally expressed by the the following axiom KF:

KF For every class X, there is a class Y such that the $\mathcal{L}_T(P)$-structure $(\mathbb{V}, \in, X, Y)$, where P and T are interpreted by X and

[39] For the precise definition of the axiom and the proof of the stated theorem, see Gitman and Hamkins, "Open Determinacy for Class Games," *op. cit.*, Theorem 9.

[40] Solomon Feferman, "Reflecting on Incompleteness," *The Journal of Symbolic Logic*, LVI, 1 (March 1991): 1–49.

Y, respectively, satisfies the following (namely, the axioms of KF with two basic predicates $\in$ and P):

(K1) Px is true (or false) of a, if and only if Pa ($\neg Pa$, resp.);

(K2) $x \in y$ is true (or false) of a and b, if and only if $a \in b$ ($a \notin b$, resp.);

(K3) An $\mathcal{L}_T^\infty(P)$-sentence σ is true, if and only if Tx is true of the code of σ;

(K4) An $\mathcal{L}_T^\infty(P)$-sentence σ is true, if and only if $\neg\neg\sigma$ is true;

(K5) An $\mathcal{L}_T^\infty(P)$-sentence $\sigma \wedge \tau$ is true (or false), if and only if both σ and τ are true (either σ or τ is false, resp.);

(K6) An $\mathcal{L}_T^\infty(P)$-sentence $\forall x \varphi(x)$ is true (or false), if and only if $\varphi(x)$ is true of all sets a ($\varphi(x)$ is false of some a, resp.);

here, by "σ is false" we mean "$\neg\sigma$ is true," and by "an $\mathcal{L}_T^\infty$-formula $\varphi(x)$ is true of a" we mean that the code of $\varphi(\mathsf{c}_a)$ is a member of Y, namely, the interpretation of the truth predicate T in the $\mathcal{L}_T(P)$-structure in question. We emphasize that the postulates (K1)–(K6) for the admissible truth predicate Y involve *no* second-order quantification.

Now, let LFP be a class-theoretic principle asserting that there is a least fixed-point of each elementary formula $\Phi(x, X)$ with all occurrences of X positive (and possibly with set and/or class parameters). The importance of LFP in the current context consists in that it enables us to express $IMH^{\vdash_{\mathbb{V}}}$ (in Example 5), since the notion of consequence in $\mathbb{V}$-logic can be inductively defined on $\mathbb{V}$.[41] Now, we have the following theorem.

Theorem 1. NBG + KF proves LFP and ETR.[42]

[41] Furthermore, NBG+LFP yields a class that "codes" $Hyp(\mathbb{V})$, namely, the least admissible structure containing $\mathbb{V}$ as a "set"; see Kentaro Fujimoto, "Truths, Inductive Definitions, and Kripke-Platek Systems over Set Theory," *The Journal of Symbolic Logic*, LXXXIII, 3 (September 2018): 868-98. As far as we can see, the relevant arguments of Antos et al. in justification of $IMH^{\vdash_{\mathbb{V}}}$, such as those mentioned in fn 32, can be thereby carried out in NBG + LFP, which derives ETR, and the full strength of Σ^1_1-CA is not necessary.

[42] We first see NBG + KF $\vdash$ FP, where FP asserts the existence of a fixed-point of every positive elementary operator. This is shown by a generalization of Corollary 3.11 of Andrea Cantini, "Notes on Formal Theories of Truth," *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, XXXV, 2 (1989): 97–130. Then, the first claim follows from Sato's theorem that NBG + FP $\vdash$ LFP: see Kentaro Sato, "Full and Hat Inductive Definitions Are Equivalent in *NBG*," *Archive for Mathematical Logic*, LIV, 1-2 (February 2015): 75–112. Finally, it is readily seen that NBG + LFP $\vdash$ ETR, and the second claim thereby follows. In second-order arithmetic, LFP is equivalent to Σ^1_1-CA, and KF is far weaker than LFP, but this equivalence breaks down in class theory, and the strength of LFP comes down to that of KF, primarily because the notion of well-foundedness is no longer Π^1_1-complete in class theory.

As a consequence, we have a liberal predicativist justification of the class-theoretic principles presupposed in the set-theoretic arguments in Examples 4 and 5.[43] Hence, the common criticism of predicativism as too restrictive only applies to strict predicativism and not to our liberal predicativism.

The type of justification of axioms exhibited so far is based on a general consideration of the mathematical act of introducing a new predicate of one's subject matter, which is not particular to set theory.[44] This "generic" type of justification is not likely to be applicable to some other axioms mentioned in Examples 1–5, such as *GRP* and $IMH^{\vdash_{\mathbb{V}}}$, which are derived from specific views about sets and $\mathbb{V}$. A justification of such a "set-theoretic" second-order axiom should involve a justification of each specific view that it upholds. The examination of such a justification for each particular case is beyond the scope of this article, but we would like to point out one crucial advantage of liberal predicativism over strict predicativism: with liberal predicativism, once one has properly justified a "set-theoretic" axiom, there is no *formal* constraint, such as the ω-inconsistency that troubles strict predicativism, for introducing relevant admissible predicates and thereby postulating the axiom. At any rate, the focus of this section is on the desiderata (i) and (ii), and the results of this section are intended to show that liberal predicativism is versatile enough to provide a sufficiently rich mathematical framework for meaningfully investigating those "set-theoretic" second-order axioms; for example, we should decide to accept or reject $IMH^{\vdash_{\mathbb{V}}}$ on the basis of a careful examination of the maximality thesis of the universe behind it, and we should *not* reject it just because the deducibility in $\mathbb{V}$-logic is undefinable.

Unfortunately, however, liberal predicativism is not a panacea. It is formally compatible with virtually all possible second-order axioms of set theory. This is simply because its formal rendering is the same as the ordinary two-sorted formalization of second-order set theory, and it requires no formal rewriting of set-theorists' usual discourse about classes, whereas definabilism, hermeneutic reductionism, and strict predicativism do. However, it upholds a particular conception of classes, and that conception might be incompatible with a view about classes that some "set-theoretic" second-order axioms of the kind we mentioned

[43] Even if one does not accept the particular conception of truth represented by the KF-axioms, there are other alternative axiomatizations of truth to achieve the same goal, such as Cantini's axioms of VF: see Andrea Cantini, "A Theory of Formal Truth Arithmetically Equivalent to ID1," *The Journal of Symbolic Logic*, LV, 1 (March, 1990): 244–259.

[44] Indeed, the ideas of Δ^1_1-CR, ETR, and the KF-truth originally arose in the context of the Feferman-Schütte predicativism in the proof theory of arithmetic, and they are given the same type of justification in that context.

uphold, even though they are formally compatible. In many cases, including Examples 1 and 3–5, there seems to be nothing wrong in interpreting classes as predicates of sets; for instance, Audrito and Viale's axiom can be interpreted with no problem as an assertion of the existence of an admissible predicate that describes a winning strategy of a certain game instructing what move to make in order to win. Liberal predicativism does well in most cases, when it comes to an axiom asserting something about a specific type of class, such as Audrito and Viale's. In contrast, if an axiom asserts something strong about the entire structure of classes or the totality of classes, liberal predicativism might be faced with a difficulty. For instance, although it does not imply impredicative comprehension, *GRP* in Example 2 requires an extraordinarily rich stock of classes and imposes a highly complicated structure on classes that goes far beyond any understanding of classes that we could gain from the notion of predicate (or even from that of mereological part) and/or through standard theories of classes such as NBG and MK.[45] What principle is conceptually compatible with liberal predicativism depends on what view one holds about the admissibility of predicates, and we do not think that liberal predicativism is absolutely conceptually incompatible with *GRP*, but it can be conceptually compatible with *GRP* only by taking a quite (and possibly excessively) strong view on the admissibility of predicates.[46]

V. NOMINALISM ABOUT CLASSES

Nominalism is a primary motivation behind predicativism. Here we mean nominalism *about classes*; we do not intend to question realism

[45] For instance, *GRP* implies that $(\mathbb{V}, \in, \mathcal{C})$ satisfies the axiom of determinacy for lightface Σ^0_α *class* games (suitably defined, say, starting from Σ^0_1-definable open classes) for any $\alpha < \omega_1$; it also implies the existence of a class $Z \in \mathcal{C}$ that codes a model of boldface $\boldsymbol{\Pi}^1_n$-determinacy for all $n \in \mathbb{N}$ in the sense that $(\mathbb{V}, \in, \{(Z)_x \mid x \in \mathbb{V}\})$ is a model of it and MK, where $(Z)_x$ denotes $\{y \mid \langle x, y\rangle \in Z\}$. Here, the collections $\boldsymbol{\Sigma}^0_\alpha$ and $\boldsymbol{\Pi}^1_n$ of class games are respectively defined in terms of the Borel hierarchy and the projective hierarchy of subclasses of the game space ${}^{<\omega}\mathbb{V}$; see Donald Martin, *Determinacy of Infinitely Long Games*, unpublished manuscript, Chapter 8. These axioms go far beyond the strength of full impredicative comprehension, and any higher-order set theory even with full impredicative comprehension at each order is still far from capturing their strength; for some details of determinacy of class games, see Sherwood Hachtman, "Determinacy Separations for Class Games," to appear in *Archive for Mathematical Logic*. We remark that $(\mathbb{V}, \in, \mathcal{C})$ satisfies those axioms, since *GRP* implies the existence of proper class many measurable Woodin cardinals above the critical point κ and thus makes all projective games in ${}^{<\omega} V_\kappa$ determined, which makes those axioms true in $(V_\kappa, \in, V_{\kappa+1})$ and thus in $(\mathbb{V}, \in, \mathcal{C})$ by the asserted elementary embedding j.

[46] If *GRP* is strengthened to the assertion of the existence of an elementary embedding preserving all *second-order* formulae, it implies full impredicative comprehension over $\mathbb{V}$ is unlikely to be acceptable from the liberal predicativist point of view.

about sets nor advocate nominalism about sets here. There are several reasons for favoring nominalism (or anti-realism, if preferred) over realism about classes, and we will discuss them in this section.

First of all, according to Ockham's razor, parsimony of ontology is in principle to be respected as far as possible. Second, if a class is another kind of object independent of sets, then the plethora of uses of classes in set theory indicate that set theory concerns two distinct kinds of objects. However, the subject matter of set theory is alleged to consist solely of sets, and the foundational significance of set theory is largely derived from the alleged fact that all of mathematics can be expressed in terms of a single kind of mathematical objects, that is, sets. Hence, the conclusion at stake would require a significant change to the shared view of set theory, and we suspect that many set-theorists would not agree that their subject matter consists of sets *and* something else, such as mereological parts of $\mathbb{V}$. Third, if classes are *mathematical* objects, it seems that nothing debars us from collecting them into sets, since the subject matter of set theory is all collections there are at least within mathematics. This results in a hierarchy of sets starting from classes as urelements stacked on top of the ordinary hierarchy of (pure) sets, that is, $\mathbb{V}$, which is a particularly problematic consequence for universism. In advocating their mereological interpretation, Welch and Horsten are well aware of this problem of realism about classes, and they cautiously allocate different ontological categories to sets and classes: they take mereological parts of $\mathbb{V}$ ("classes" for them) as objects of some *non-mathematical* kind not subsumed under the laws of mathematics, and thereby try to preclude the formation of sets of mereological parts of $\mathbb{V}$. This strategy, however, makes it completely mysterious why and how sets and classes can interact with mathematically substantial consequences. With their two-tiered view, Welch's *GRP* would be a *super-mathematical* assertion of such a mystic but mathematically productive interplay of the two kinds of entities living in two completely separate realms. The conclusion we draw from these considerations is that, for universists, nominalism is to be preferred with respect to classes as far as possible.

One possible objection to liberal predicativism as a form of nominalism about classes is that its formalism involves second-order quantifiers, and thus it ontologically commits us to extra objects beyond sets, according to the Quinean criterion of ontological commitment "to be is to be the value of a bound variable." A general discussion of the criterion of "objecthood" is beyond the scope of this article, but we do not think it necessary to accept the Quinean criterion. In regard to its formal expression of nominalism about classes, liberal predicativism takes the same strategy as the plural interpretation does, namely, that of interpreting second-order quantifiers so that they do not carry ontological commitment while keeping the two-sorted formalism.[47] In other words, liberal predicativism generalizes predicate places directly by quantifiers ranging over (admissible) predicates *as such*.[48][49]

Liberal predicativism does not license us to accept every possible second-order axiom in an ontologically noncommittal way. Particularly and repeatedly, impredicative comprehension seems unacceptable, or, at best, could only be accepted in an ontologically committal way, for liberal predicativism. In his critique of the plural interpretation, Linnebo argues that "the considerations that give content to the notion of a determinate range of arbitrary sub-pluralities belong to combinatorics and to set theory" and then concludes that, with such a combinatorial and set-theoretic understanding, pluralities are no more ontologically innocent or noncommittal than sets.[50] We agree with Linnebo's argument, and a parallel argument does apply to liberal predicativism, or predicativism in general, with impredicative comprehension; indeed, Parsons's rejection of a determinate totality of possible predicates of sets and impredicative comprehension for predicates is based on essentially the same consideration.[51]

In our view, liberal predicativism avoids ontological commitment to extra objects beyond sets, but it can only do so at the cost of impredicative comprehension. If one takes nominalism about all (or most) mathematical entities and tries to develop mathematics by means of second- or higher-order logic, then one would surely need impredicative comprehension at some point. However, we repeat that we advocate nominalism about classes and not about sets. All those uses of

[47] For more extensive discussion (and justification) of nonnominal and ontologically noncommittal interpretations of second-order quantifiers, see George Boolos, "On Second-order Logic," *The Journal of Philosophy*, LXXII, 16 (September 1975): 509–527, and Agustin Rayo and Stephen Yablo, "Nominalism Through De-nominalization," *Noûs*, XXXV, 1 (March 2001): 74–92.

[48] Parsons argues convincingly that not every predicate refers to an object; see Charles Parsons, "Objects and Logic," *The Monist*, LXV, 4 (October 1982): 491–516, at p. 502.

[49] The reader may notice a similarity between liberal predicativism and the interpretation of classes as *schematic letters*, for which any *possible* open sentence can be substituted, advocated by McGee; Vann McGee, "How We Learn Mathematical Language," *The Philosophical Review*, CVI, 1 (January 1997): 35–68. The same idea is also presented in a more mathematical form in Feferman, "Reflecting on Incompleteness," *op. cit.* With this interpretation, some class quantification can be expressible. However, this strategy seems to work only for Π^1_1 assertions and not for more complex assertions such as *GRP*. Parsons's treatment of arithmetical induction and the uniqueness of the natural number structure is based on a similar idea to theirs; see Parsons, "The Uniqueness of the Natural Numbers," *Iyuun*, XXXIX (January 1990): 13–44.

[50] Øystein Linnebo, "Plural Quantification Exposed," *Noûs*, XXXVII, 1 (March 2003): 71–92.

[51] Parsons, "Sets and Classes," *op. cit.*, pp. 216. Also see Charles Parsons, *Mathematical Thought and Its Objects* (New York: Cambridge University Press, 2008), §13 and §47.

impredicative comprehension for the development of mathematics can be incorporated into and carried out within set theory, in which impredicative comprehension is perfectly acceptable thanks to the axioms of powerset and separation; as far as the current state of mathematics is concerned, impredicative comprehension seems unnecessary anywhere beyond the realm of (the first-order part of) set theory.

To conclude this section, we briefly compare liberal predicativism and the plural interpretation. They are alleged to share the same philosophical merit of ontological innocence about classes. The proponents of the latter further contend that the plural interpretation justifies all instances of impredicative comprehension and thus MK, and thereby conclude that the plural interpretation is a superior nominalist conception of classes to the predicative interpretation.[52] However, their contention is not convincing, as Linnebo argues. Given that the two share the same difficulty in dealing with impredicative comprehension, we think that the former is rather better than the latter as a nominalist theory of classes, since there seems to be no convincing argument that justifies relatively strong class-theoretic principles such as ETR and LFP from the viewpoint of the plural interpretation with nominalism about classes.

VI. CONCLUSION

Predicativism shares the philosophical merit of ontological parsimony with reductionisms and the plural interpretation. Our liberal predicativism also shares a mathematical merit with realist interpretations of classes such as the mereological interpretation in that it accommodates various uses of classes in set theory. Predicativism is not as restrictive as has been thought, and rather provides a highly viable option to nominalists about classes. There is a limitation: predicativism does not seem to allow impredicative comprehension in general, and it might not be conceptually compatible with some types of axioms. However, we believe that the present paper at least shows that predicativism, when our liberal version is taken, is able to offer a much more versatile and workable nominalist understanding of classes than previously thought.

KENTARO FUJIMOTO

University of Bristol

[52] For example, see Uzquiano, "Plural Quantification and Classes," *op. cit.*, pp. 76–80.