Notre Dame Journal of Formal Logic
Volume 35, Number 3, Summer 1994

A Dilemma in the Philosophy of Set Theory

RALF-DIETER SCHINDLER

Abstract We show that the following conjecture about the universe V of all sets is wrong: for all set-theoretical (i.e., first order) schemata Σ true in V there is a transitive set u "reflecting" Σ in such a way that the second order statement σ corresponding to Σ is true in u. More generally, we indicate the ontological commitments of any theory that exploits reflection principles in order to yield large cardinals. The disappointing conclusion will be that our only apparently good arguments for the existence of large cardinals have bad presuppositions.

Bernays' celebrated system BL of class theory (presented in his [1]) yields all current large cardinals below measurability (see Tharp [8]). As Gloede (in his [2], Corollary 3.7) remarks, for proving the existence of all Mahlo cardinals a much weaker subsystem BL_1 of BL suffices. BL_1 comes from some standard system of class theory (e.g., the von Neumann-Bernays-Gödel system NBG including choice) by adjoining the schema of partial reflection

$$\mathrm{PR}\Pi^1_1 \quad \Phi \rightarrow \exists u(\mathrm{Trans}(u) \wedge \Phi^u), \text{ for all } \Pi^1_1\text{-formulae } \Phi,$$

where $\mathrm{Trans}(u)$ means that u is transitive and Φ^u is the relativization of Φ to u obtained in the usual manner (set quantifiers $\forall x(\exists x)$ are replaced by $\forall x \in u$ $(\exists x \in u)$, class quantifiers $\forall X(\exists X)$ by $\forall X \subset u$ $(\exists X \subset u)$).

In the following, we investigate which classes BL_1 forces to exist. Let us first sketch the philosophical motivation for this enterprise, presupposing a realistic attitude toward sets. $\mathrm{PR}\Pi^1_1$ is the weakest reflection schema giving large cardinals (see Schindler [7], Section 2, for a thorough elaboration on the "no weaker..."-part of this claim). If reflection principles were the only seemingly convincing arguments for the actual existence of large cardinals in the real world V, our belief in those cardinals would be no more justified than our belief in just those classes $\mathrm{PR}\Pi^1_1$ commits us to accept. We shall find that *predicative* classes (see below) cannot satisfy $\mathrm{PR}\Pi^1_1$ in that any model of BL_1 must have nonpredicative classes in its domain interpreting the

Received May 19, 1993; revised April 18, 1994

class variables (see our Theorems 1 and 2; however, arbitrary impredicative classes need not exist, see Theorem 3 and corollary). Thus the disillusioning result would be that anyone who does not believe in nonpredicative classes on philosophical grounds has no justification at all for believing in the existence of large cardinals.

We in fact claim that believing in the reality of large cardinals is an irrational form of behavior. Set theory without large cardinals would be a poor thing, and we do not have a bias against them, but knowledge that large cardinals really exist cannot be attributed to set theorists. By what has been said so far, to support our claim we would have to establish three subclaims:

1. Our only apparently good arguments for the actual existence of large cardinals in V come from reflection principles.
2. There are no nonpredicative classes.
3. PRΠ^1_1 presupposes the existence of nonpredicative classes.

Subclaims (1) and (2) are philosophical matters which will be dealt with here only by giving indications and references, as the present paper intends to focus on its technical part verifying (3). Nevertheless, we suppose that there are convincing arguments for (1) and (2) which we shall expose in a forthcoming paper.

Concerning Subclaim (1) There is a list along the lines of Maddy [4], pp. 501ff., categorizing the current arguments offered in favor of the existence of large cardinals. Reflection principles appear the most sophisticated realizations of the fairly vague but indisputable idea of V's being "maximal," or "inexhaustible" ("absolut grenzenlos," in Cantor's phrase): the iterative process generating V is so endless that V itself cannot be characterized by significant linguistic means. Thus reflection principles are of natural descent. And they look clearcut as well as they can be written down in formal language and so have neat applications.

On the other hand, all other kinds of arguments seem to be of a somewhat suspicious nature; viz, they all suffer from a notoriously opalescent range of applicability of the underlying "principle," whose piloting's having success seems to be conceivable only as lucky chance. It just *could* be that every level of the cumulative hierarchy has its own pecularities, it *could* be that ω is the only inaccessible number, and it *could* be that there are no weakly compacts although they can be characterized in so many different ways. This scenario only challenges us to refute its actuality, and reflection principles appear to be the only available method to do this job. Reinhardt, in his [6], p. 90, holds a similar view:

> The picture provided [i.e., the idea of the cumulative hierarchy] suffices to set up the basic axioms of set theory. It [...] does not tell us much about the transfinite sequence of ordinals [...]. Insofar as we know anything more about this, our knowledge seems to depend on so-called reflection principles.

Concerning Subclaim (2) If V denotes the universe of all sets, a predicative class is any sub-collection of V whose elements are separated by a "sound" predicate. The prototype of predicative classes are all $\{x : \Phi(x)\}$'s where Φ is set-theoretical. What distinguishes predicative classes from others (e.g., impredicative ones) is that in the defining Φ variables ranging over a totality of classes given in advance must not (and may, respectively) occur. This fact is intended to be referred to by using the epithet "sound." The underlying philosophical idea is that there are no classes "in themselves," but that classes have to be constructed step by step and at each stage by means of what has been constructed so far. This constructivistic concept of classes yields the above prototype first, and then a natural well-ordered hierarchy of predicative classes (see Schindler [7], Sections 0 and 1).

If, on the contrary, arbitrary classes existed in themselves, we could collect them together into a power hyper class $\wp$V of V. There is no reason, then, why we should not be able to repeat this process of adding a new layer above V, and thus we obtain $\wp\wp$V, $\wp\wp\wp$V, etc. But

> [...] if you are going to add a layer at the top [of V] it looks like you just forgot to finish the hierarchy. (Reinhardt [6], p. 196).

Thus, arbitrary classes existing in themselves contradict the very idea of V being the collection of all layers in the process of set formation. Therefore classes have a fundamentally different ontological status than sets, namely they exist only as extensions of sound predicates. This is the philosophical origin of our disbelief in nonpredicative classes (see also Maddy [3], p. 122).

Now let $\mathfrak{R}$ be the collection of all classes $\{x : \Phi(x, a_0, \dots, a_n)\}$ where Φ is a set-theoretical predicate and the a_i's $(i \leq n)$ are set parameters. We may well think of the prototype $\mathfrak{R}$ as the intended domain of the class variables in the class-theoretical system NBG. $\mathfrak{R}$ allows us to connect the negative result "predicative classes do not satisfy PRΠ^1_1" with the conjecture about V expressed in the abstract (and which was announced in footnote 27 of Schindler [7]). Because, if the class variables in PRΠ^1_1 range over $\mathfrak{R}$, then (fix a Π^1_1-formula Φ!) the antecedents Φ of PRΠ^1_1 states in effect a set-theoretical schema, while Φ^u states the corresponding second order assertion. Thus, as PRΠ^1_1 is false if the only classes are those in $\mathfrak{R}$, that conjecture is false, too.

Concerning Subclaim (3) We are now going to prove the following (see the proofs for concepts and notation):

Theorem 1 *For all predicative classes, $\Delta^{1,NBG}_1$-comprehension fails.*

Theorem 2 $BL_1 \vdash \Delta^{1,NBG}_1$*-comprehension.*

Theorem 3 *Let κ be weakly compact. Then $\langle L_\kappa, \Delta^1_1(\wp L_\kappa \cap L), \in\rangle \models BL_1$.*

Corollary 4 Σ^1_1*-comprehension may fail in BL_1.*

Proof: [Theorem 1] The theorem is stated somewhat vaguely. Actually, we shall show that Δ^1_1-comprehension fails if the only classes are those in $\mathfrak{R}$. To be even more precise, we shall prove that $\Delta^{1,\text{NBG}}_1$-comprehension fails, if the class variables range over the set $\Sigma_\omega\langle \text{M}, \in\rangle$ of all subsets of M set-theoretically definable over M with parameters from M, for any set model M of ZF. The method of our proof can then easily be applied to predicative classes in the sense of Schindler [7], giving a full proof of the lemma.

Without proof, we use the fact that the class of all true set-theoretical sentences is a $\Delta^{1,\text{NBG}}_1$-class. The reason for this fact is that this "truth class" can be defined as $\{x : \exists X(\Phi \wedge x \in X)\}$ where, for a given x, Φ recursively describes a unique restricted truth class. See Mostowski [5], Schindler [7], Section 1. Existence and uniqueness of the (former, unrestricted) truth class can be proved in NBG, and hence, in NBG, it equals $\{x : \forall X(\Phi \rightarrow x \in X)\}$, and so is $\Delta^{1,\text{NBG}}_1$.

Now, by Tarski's Undefinability Theorem it follows that the truth predicate for the set-theoretical language is not itself set theoretical. Therefore the unrestricted truth class is not of the form $\{x : \Phi\}$ for a set-theoretical Φ (i.e., $\Delta_1^{1,\mathrm{NBG}}$ is best possible). But this means that there is no element of $\Sigma_\omega\langle M, \in\rangle$ (i.e., no class in $\mathfrak{R}$) equal to the $\Delta_1^{1,\mathrm{NBG}}$-truth class. In other words, Δ_1^1-comprehension fails.

Proof: [Theorem 2] Assume $\mathrm{NBG} \vdash \exists X\Phi \leftrightarrow \forall X\Psi$ for set-theoretical formulae Φ, Ψ with free set variables $x, y_1 \ldots y_n$. Then, also in BL_1,

$$\exists X\Phi \leftrightarrow \forall X\Phi. \tag{1}$$

Furthermore, already in a very weak set theory, and *a fortiori* in BL_1, it is provable that $\langle u, k, \in\rangle \mathrel{\dot{\models}} \mathrm{NBG} \rightarrow \langle u, k, \in\rangle \mathrel{\dot{\models}} \ulcorner\exists X\Phi \leftrightarrow \forall X\Psi\urcorner$ for any u, k, where $\dot{\models}$ denotes the formal representation of the model relation, $\ulcorner\ldots\urcorner$ means Gödelization, and $u(k)$ is the domain of the set (class) variables. This quickly implies that, in BL_1:

$$\langle u, \wp u, \in\rangle \mathrel{\dot{\models}} \mathrm{NBG} \rightarrow \forall x y_1 \ldots y_n \in u(\exists X \subset u\Phi^u \leftrightarrow \forall X \subset u\Psi^u). \tag{2}$$

We work in BL_1 from now on. We have to show $\exists Y\, Y = \{x : \exists X\Psi\}$. By (1), this formula can be written as $\exists Y\forall x\exists X\,\Xi(Y, x, X)$, were Ξ is set-theoretical. Hence it suffices to show $\exists Y\exists Z\forall x\,\Xi(Y, x, Z_x)$, where Z codes a "choice class" $\{\langle x, y\rangle : y \in Z_x\}$ in such a way that $\exists X\,\Xi(Y, x, X) \rightarrow \Xi(Y, x, Z_x)$. Herein, as can easily be checked, $\Xi(Y, x, Z_x)$ can be written set theoretically.

Now let us consider the contraposition $\mathrm{PR}^*\Sigma_1^1$ of the self-strengthening of $\mathrm{PR}\Pi_1^1$ where $\mathrm{Trans}(u)$ is replaced by $\mathrm{Trans}(u) \wedge \langle u, \wp u, \in\rangle \mathrel{\dot{\models}} \mathrm{NBG}$. We have, as an instance of $\mathrm{PR}^*\Sigma_1^1$:

$$\begin{aligned}\forall u(\mathrm{Trans}(u) \wedge \langle u, \wp u, \in\rangle \mathrel{\dot{\models}} \mathrm{NBG} \rightarrow \exists Y \subset u\, \exists Z \subset u\, \forall x \in u(\Xi(Y, x, Z_x))^u)\\ \rightarrow \exists Y\exists Z\forall x\,\Xi(Y, x, Z_x).\end{aligned} \tag{3}$$

By our remarks, we only have to prove the antecedens of (3). $\exists Y \subset u\, \exists Z \subset u\, \forall x \in u(\Xi(Y, x, Z_x))^u$ can be transformed equivalently into $\exists Y \subset u\, \forall x \in u\, \exists X \subset u(\Xi(Y, x, X))^u$ (here we use that for any subset of u we have a class equal to this subset), and, by using (2), the latter is equivalent to $\exists Y \subset u\ Y = \{x \in u : \Phi^u\}$. But this is trivially valid, and so we have Theorem 2.

Of course, our result could be strengthened to $\mathrm{BL}_1 \vdash \Delta_1^{1,\mathrm{T}}$-comprehension, for any theory T for which there is a self-strengthening of $\mathrm{PR}\Pi_1^1$ where $\mathrm{Trans}(u)$ is replaced by $\mathrm{Trans}(u) \wedge \langle u, \wp u, \in\rangle \mathrel{\dot{\models}} \mathrm{T}$.

Proof: [Theorem 3] Let κ be weakly compact. Then $\mathcal{M} = \langle \mathrm{L}_\kappa, \wp\mathrm{L}_\kappa \cap \mathrm{L}, \in\rangle \models \mathrm{BL}_1$, and we have to show that $\langle \mathrm{L}_\kappa, \Delta_1^1(\wp\mathrm{L}_\kappa \cap \mathrm{L}), \in\rangle \models \mathrm{BL}_1$. Herein, $\Delta_1^1(\wp\mathrm{L}_\kappa \cap \mathrm{L})$ means the set of all $X \in \wp\mathrm{L}_\kappa \cap \mathrm{L}$ for which there are set-theoretical Ψ, Ξ with $\mathcal{M} \models X = \{x : \forall Y\Psi(Y, x)\} = \{x : \exists Y\Xi(Y, x)\}$. Obviously it suffices to prove that if there is $X \in \wp\mathrm{L}_\kappa \cap \mathrm{L}$ with $\mathcal{M} \models \Phi(X)$, then there is $X \in \Delta_1^1(\wp L_\kappa \cap L)$ with $\mathcal{M} \models \Phi(X)$ for any set-theoretical Φ. Our proof is an adaptation of Gödel's proof that, if V=L, then there is a Δ_2^1-well-ordering of the reals.

Let $\mathcal{M} \models \Phi(X)$ with set-theoretical Φ and $X \in \wp\mathrm{L}_\kappa\cap\mathrm{L}$. Let ZFL=ZF+V=L and let $\langle_{\mathrm{L}}$ be the Σ_1-definable canonical well-ordering for L. Our comprehension term witnessing that there is a $\Sigma_1^1(\wp\mathrm{L}_\kappa \cap \mathrm{L})$-solution of $\Phi(Z)$ is

$$\{x : \exists \mathrm{E}, Z(\langle \mathrm{V,E}\rangle \models \mathrm{ZFL} \wedge \mathrm{E} \text{ is well-founded} \wedge \forall\alpha\forall\, y(y \in \alpha \leftrightarrow y\mathrm{E}\alpha) \wedge$$
$$\text{if } z \text{ represents} Z \text{ in} \langle \mathrm{V,E}\rangle, \langle \mathrm{V,E}\rangle \models \text{“}z \text{ is } \langle_{\mathrm{L}}\text{-minimal with } \Phi(z)\text{“} \wedge x \in Z)\}. \qquad (4)$$

In the first conjunct V is a class parameter that denotes the universe of all sets (which can easily be eliminated), and E stands for a binary relation on V supposed to interpret $\in$. As a matter of fact,$\langle$ V,E$\rangle$ $\models$ZFL is $\Delta_1^{1,\mathrm{NBG}}$.

By the condensation lemma for the L-hierarchy, the second conjunct implies that $\langle \mathrm{L}_\kappa$,E$\rangle$, i.e. $\langle$ V,E$\rangle$ interpreted inside $\mathcal{M}$, can be collapsed onto $\langle \mathrm{L}_\alpha, \in\rangle$, for some α, $\kappa \leq \alpha < \kappa^+$. “E is w.-f.” can be written by quantifying over ω-sequences of elements of V, and hence is set-theoretical (compare this with “E is w.-f.” in Gödel's proof).

The third conjunct ensures that every ordinal will be collapsed onto itself.

In the fourth conjunct, “z represents Z in $\langle$V,E$\rangle$” remains to be formalized. Let $\pi : \langle \mathrm{L}_\alpha, \in\rangle \cong \langle \mathrm{L}_\kappa, \mathrm{E}\rangle$. That z represents Z in $\langle$V,E$\rangle$ means that $\pi^{-1}(z) = Z \in \mathrm{L}_\alpha$. Now, any $w \in \mathrm{L}_\kappa$ is ordinal definable in L_κ by a Δ_1^{ZFL}-formula $\Xi(\alpha_0 \ldots \alpha_n, -)$, i.e., w is unique with $\langle \mathrm{L}_\kappa,\mathrm{E}\rangle \models \Xi(\alpha_0 \ldots \alpha_n, w)$. But then w is unique with$\langle \mathrm{L}_\alpha, \in\rangle \models \Xi(\alpha_0 \ldots \alpha_n, w)$, too, by Δ_1^{ZFL}-absoluteness between $\langle \mathrm{L}_\kappa, \in\rangle$ and $\langle \mathrm{L}_\alpha, \in\rangle$. Hence $\pi^{-1}(w)$ is unique with $\langle \mathrm{L}_\kappa,\mathrm{E}\rangle \models \Xi(\pi^{-1}(\alpha_0) \ldots \pi^{-1}(\alpha_n), \pi^{-1}(w))$, i.e., with $\langle \mathrm{L}_\kappa,\mathrm{E}\rangle \models \Xi(\alpha_0 \ldots \alpha_n, \pi^{-1}(w))$, due to the third conjunct. This consideration shows that “z represents Z in $\langle$V,E$\rangle$” can be written as:

$$\forall w(w \in Z \leftrightarrow \forall \Delta_1\text{-formulae } \ulcorner\Xi\urcorner \text{ witnessing ordinal definability of } w,$$
$$\langle \mathrm{V,E}\rangle \models \text{“the unique } w \text{ with } \Xi \text{ is} \in z\text{”}). \qquad (5)$$

(5) is $\Delta_1^{1,\mathrm{NBG}}$, and so finally the comprehension formula in (4) is $\Sigma_1^1(\wp \mathrm{L}_\kappa \cap \mathrm{L})$.

We had assumed that $\mathcal{M} \models \Phi(X)$. It is now an easy exercise to show that under this circumstance there are E and Z both from $\wp \mathrm{L}_\kappa \cap \mathrm{L}$ fulfilling, in $\mathcal{M}$, the comprehension formula in (4). Moreover, such Z is unique. Hence, that formula can be rewritten in Π_1^1-form and we are done.

One could rework this last proof to show the relative consistency $\langle$M,K,$\in\rangle\models$BL$_1$ $\Rightarrow$ $\langle$M,Δ_1^1(K), $\in\rangle$ $\models$BL$_1$ under slightly weaker assumptions than M=L$_\kappa$ for κ weakly compact and K=$\wp$L$_\kappa \cap$ L. This implies that the corollary could also be proved under such weaker assumptions.

Proof: [Corollary] $\mathcal{M} = \langle \mathrm{L}_\kappa, \Delta_1^1(\wp \mathrm{L}_\kappa \cap \mathrm{L}), \in\rangle \models \mathrm{BL}_1$, if κ is weakly compact. Without proof, we use the fact that the model relation $\mathcal{M} \models \Phi$ for Σ_1^1-sentences Φ is $\Sigma_1^1(\wp \mathrm{L}_\kappa \cap \mathrm{L})$-definable. This is verified by checking that there are enough “classes” to feed the appropriate definition. We can now use that model relation to define a “universal” $\Sigma_1^1(\wp \mathrm{L}_\kappa \cap \mathrm{L})$-class U with the property that any $\Sigma_1^1(\wp \mathrm{L}_\kappa \cap \mathrm{L})$-class is a projection of U; namely, let U = $\{\langle \ulcorner\Phi\urcorner, y\rangle : \mathcal{M} \models \ulcorner\Phi(y)\urcorner\}$, where, for Φ a unary Σ_1^1-predicate, $\varphi(y)$ is the result of substituting the free variable in Φ by y. We have

$$\forall X \in \Sigma_1^1(\wp \mathrm{L}_\kappa \cap \mathrm{L})\; \exists x \in \mathrm{L}_\kappa\; X = \{y : \langle x, y\rangle \in U\}. \qquad (6)$$

If U was $\Pi_1^1(\wp \mathrm{L}_\kappa \cap \mathrm{L})$, too, i.e. U$\in$ $\Delta_1^1(\wp \mathrm{L}_\kappa \cap \mathrm{L})$, then U$'$ = $\{y : \langle y, y\rangle \notin \mathrm{U}\}$ would be $\Sigma_1^1(\wp \mathrm{L}_\kappa \cap \mathrm{L})$ and by (4) we could find $u \in \mathrm{L}_x$ with U$'$ = $\{y : \langle u, y\rangle \in$U). But then we would have u $\in$U$'$ iff $\langle u, u\rangle$ $\in$ U iff u $\notin$U$'$. Contradiction!

Our dilemma may now be summarized as follows. We cannot believe in the existence of nonpredicative classes on philosophical grounds. Our only apparently good arguments for the existence of large cardinals stem from applications of reflection principles, and in particular from PRΠ^1_1, the weakest formalization of such a principle doing the required job. Regrettably, already PRΠ^1_1 forces nonpredicative classes to exist. Hence we do not have any good reasons for believing in the existence of large cardinals, but would anyone dare to assert that there are no such cardinals at all? To plagiarize words from Hume, I cannot discover any theory which gives me satisfaction on this head.

Acknowledgment I am indebted to an anonymous referee who recommended including some hints with respect to Subclaim (1). I am indebted to Professor Peter Koepke who suggested the present way of proving Theorem 3. I gratefully acknowledge a fellowship of the Graduiertenkolleg Mathematik at the University of Bonn.

REFERENCES

[1] Bernays, P., "Zur Frage der Unendlichkeitsschemata in der Axiomatischen Mengenlehre," pp. 3–49 in *Essays on the Foundations of Mathematics*, edited by A. Fraenkel, Jerusalem, 1961.

[2] Gloede, K., "Reflection Principles and Indescribability," pp. 277–323 in *Sets and Classes*, edited by Y. Bar-Hillel, North-Holland, Amsterdam, 1976.

[3] Maddy, P., "Proper Classes," *Journal of Symbolic Logic*, vol. 48 (1983), pp. 113–139.

[4] Maddy, P., "Believing the Axioms 1," *Journal of Symbolic Logic*, vol. 53 (1988), pp. 481–511.

[5] Mostowski, A., "Some Impredicative Definitions in the Axiomatc Set-Theory," *Fundamenta Mathematicae*, vol. 37 (1950), pp. 111–124.

[6] Reinhardt, W., "Reflection Principles, Large Cardinals, and Elementary Embeddings," *Proceedings of the Symposium on Pure Mathematics*, vol. 13 Part II, (1974), pp. 189–205.

[7] Schindler, R., "Prädikative Klassen," *Erkenntnis*, vol. 39 (1993), pp. 209–241.

[8] Tharp, L., "On a Set Theory of Bernays," *Journal of Symbolic Logic*, vol. 32 (1967), pp. 319–321.

Mathematisches Institut
Universität Bonn
Beringstraße 4, D-53115 Bonn
Germany
email: UNM414@IBM.rhrz.uni-bonn.de