VARIETIES OF CLASS-THEORETIC POTENTIALISM

NEIL BARTON AND KAMERYN J. WILLIAMS

Abstract. We explain and explore class-theoretic potentialism—the view that one can always individuate more classes over a set-theoretic universe. We examine some motivations for class-theoretic potentialism, before proving some results concerning the relevant potentialist systems (in particular exhibiting failures of the .2 and .3 axioms). We then discuss the significance of these results for the different kinds of class-theoretic potentialist.

Introduction

In this paper we examine a new kind of *potentialism* in set theory. From the off, let's state the difference between *actualism* and *potentialism*:

Set-Theoretic Actualism: There is a maximal universe of sets that is complete in the sense that we can quantify over all the sets it contains using standard first-order quantifiers and it cannot be extended.

One natural such position is universist set-theoretic actualism; the view that there is *exactly one* such universe. However, this is not necessary for actualism; one could have multiple distinct incomparable universes, each of which cannot be extended.[1] Whatever one's preference, actualism contrasts sharply with:

Set-Theoretic Potentialism: The universe of sets is not a completed totality, but rather unfolds gradually as parts either come into existence or become accessible to us.[2]

A common way of making set-theoretic potentialism mathematically precise is by viewing this 'gradual unfolding' as describing a space (or spaces) of possible worlds. Modal operators are then often introduced,

Date: 29 July 2021.

Order of authors is alphabetical. We would like to thank Carolin Antos, Julian Bacharach, John Baldwin, Salvatore Florio, Øystein Linnebo, Beau Madison Mount, Naomi Osorio-Kupferblum, Martin Pleitz, Sam Roberts... The first author is very grateful for the support of the Volkswagen Stiftung through the project *Conceptual Change in the Foundations of Mathematics*.

[1]For the arguments of this paper, nothing much hangs on the matter.

[2]This statement closely mirrors [Hamkins and Linnebo, 2018], p. 1.

and we are able to ask several kinds of questions, including (i) what holds non-modally at particular worlds, (ii) what modal principles certain worlds satisfy, and (iii) what the modal logic of different accessibility relations are. For this reason, discussion of potentialism often focuses on the nature of these set-theoretic worlds. For example, we might (inspired by [Zermelo, 1930]) view the worlds as the study of ever larger V_κ for κ inaccessible, with accessibility being coextensive with the subset relation. Another is to have the worlds be all those that can be obtained by set forcing (and moving to ground models) from some starting universe, and have one world V be accessible from another V' just in case one can force from V' to obtain V. There have been several results in this field, including isolating the modal logic of forcing ([Hamkins and Löwe, 2008]) and the study of potentialist maximality principles ([Hamkins, 2003], [Hamkins and Linnebo, 2018]). One key question (that we will deal with in detail later in this article) concerns whether the modal axioms .2 and .3 are satisfied in addition to $\mathsf{S4}$ in the relevant potentialist systems.[3] .3 indicates a kind of 'inevitability' or 'linearity' to how the worlds unfolds, and .2 indicates a form of 'convergence' present on the frame. Moreover so called 'mirroring theorems' (which allow us to move between potentialist and non-potentialist theories via a natural translation) are only known to hold on systems containing $\mathsf{S4.2}$. This has lead some authors (e.g. [Hamkins, 2018a], p. 33) to claim that the convergent forms of potentialism (i.e. those with modal logic at least $\mathsf{S4.2}$) are 'implicitly actualist'. Whatever one thinks of these specific claims, it is clear that .2 and .3 represent clear dividing lines between different potentialist systems.

This greater understanding of set-theoretic potentialism has occurred alongside an explosion in the study of *second-order set theory* (also called *class theory*).[4] These theories introduce a new kind of variable to range over classes as well as sets, which are then governed by theories such as variants of NBG and MK. Previously the key debate to be settled was whether or not the comprehension axiom for classes should be fully *impredicative* or rather whether only *predicative* comprehension was licensed by our conception of classes (see, for example, [Uzquiano, 2003]). However, recently we have discovered an

[3]It should be noted that $\mathsf{S4}$ trivially holds in any potentialist system. See §3.

[4]The nomenclature of "second-order set theory" is perhaps unfortunate, as the standard way to formulate these theories are as first-order theories with two sorts, one for sets and one for classes. Equivalently, one can formulate them in second-order logic using Henkin semantics. There is also a one-sorted version, where we only have class variables but introduce a predicate for set-hood.

entire space of possible second-order set theories with varying consequences and consistency strengths. Moreover, the dimension in which they can vary is not just along the axis of strength of comprehension. For example, class-theoretic versions of the axiom of choice are independent even of full impredicative MK (see [Gitman et al., 2019]). Moreover, *class* forcing (where the forcing partial order and generic can be proper classes) has provided us with a controlled yet flexible method for adding (sub)*classes* to a model, in a similar way to how set forcing does for (sub)*sets*.

The above observations suggest that the various kinds of set-theoretic potentialism are not the whole story. Instead, we might study the following kind of potentialism:

Class-Theoretic Potentialism: The classes of the universe do not constitute a completed totality, but rather unfold gradually as more classes either come into existence or become accessible to us.

Notice that one need not be a set-theoretic potenialist if one is a class-theoretic potentialist. One can perfectly well have the classes over a model change whilst the sets remain fixed, say if one were a set-theoretic actualist. In this paper, we explore the view that one might be a set-theoretic actualist whilst being a class-theoretic potentialist. We will argue for the following claims:

(1) Class-theoretic potentialism can be motivated on the basis of several different philosophical conceptions of classes.
(2) Whilst there are class-theoretic potentialist systems that satisfy $\mathsf{S4.3}$ and $\mathsf{S4.2}$, many exhibit failures of the .2 and .3 axioms.
(3) Depending on the desiderata that one has on class-theoretic potentialism, there are constraints placed on the *base theory* to be chosen and *constructions* allowed.

The strategy of our argument is to show how class-theoretic potentialism relates to different conceptions of classes, mathematically articulate the position and prove some results, and then discuss the relevant philosophical implications. Here's the plan in more detail:

After these introductory remarks, in §1 we'll outline some philosophical positions regarding classes that can be used to motivate class-theoretic potentialism. We'll divide these into two broad kinds: Bottom-up approaches start with some *fixed stock* of classes and then *individuate* new classes over these, whereas top-down approaches see class-theoretic potentialism as arising from *referential indeterminacy* and the ways we can *interrelate* sharpenings of the ranges of the class-theoretic variables. §§2–3 set up the key mathematical notions we shall use to examine these views, namely *potentialist systems* (structures that formalise the notion of worlds and accessibility between them) and the *modal logics* and *axioms* they satisfy. §4 proves some results about some potentialist systems, in particular showing that for weak theories below the level of $\mathsf{NBG} + \mathsf{ETR}$, given suitable assumptions we can exhibit failures of the $\mathsf{.3}$ (Theorem 18) and .2 axioms (Theorem 19), showing that some systems have non-inevitability and radical branching. We'll then discuss some implications for bottom-up approaches to classes (§5) arguing that whilst there are good motivations for handling truth predicates, global choice is problematic in this context. We then (§6) examine top-down approaches, arguing that our results are indicative of more natural cases of radical branching than is normally seen in potentialist context. §7 handles a natural objection to our approach (regarding the use of countable transitive models) one which we feel makes certain philosophical issues more perspicuous. §8 provides some concluding remarks and identifies several open questions and directions for future research.

1. Motivating Class-Theoretic Potentialism

In this section we'll outline the links between class-theoretic potentialism and a variety of conceptions of classes. Our aim is not to conclusively argue for class-theoretic potentialism, but rather to show that it is a viable position (philosophically speaking) and in fact fits nicely with a multiplicity of positions which we'll divide into *bottom-up* and *top-down* approaches. The former roughly corresponds to potentialisms linked to process of *individuation* and the latter corresponds to considerations regarding *indeterminacy of reference.*

Let's assume hereon that one is a set-theoretic actualist—one has accepted that some universe of *sets* is modally definite and cannot have sets added. For ease of expression, we'll speak as if there's just one such universe, but nothing we say will change if there are multiple such. It still behooves the set-theoretic actualist to explain what *classes* are. There are a number of options here. A mathematically popular and expressively parsimonious option is to regard talk of classes as merely shorthand for certain formulas holding within the universe. So, for instance, "$x \in \mathrm{Ord}$" can be rendered as "x is a transitive set linearly ordered by $\in$", or for a more complicated class such as a proper-class-sized embedding j, the formula "$j(x) = y$" can be rendered as some formula $\varphi(x, y)$ only referring to sets, possibly involving parameters. A disadvantage of this view is that it appears to trivialise various mathematical theorems under their natural interpretation, such as

[Kunen, 1971]'s result that there is no nontrivial $j : V \to V$ and the work by [Vickers and Welch, 2001] on embeddings from inner models to the universe.[5]

We'll now survey some of the options that have been proposed, and explain how they might motivate class-theoretic potentialism. By reviewing the literature we can extract two main strands: *bottom-up* approaches are those those view classes as given to us via some iterated process of *individuation*, and *top-down* approaches are those that view classes as existing within a potentialist framework in virtue of *indeterminateness of reference.*

1.1. **Bottom-up approaches.** The key feature of a bottom-up approach to classes is that one begins with some antecedently specified classes (e.g. the definable ones) and then builds up the classes by forming new classes from old via some process.

1.1.1. *Liberal Predicativism.* The first interpretation we'll look at is derived from the work of Parsons (e.g. [Parsons, 1974]) with subsequent development by Fujimoto (e.g. [Fujimoto, 2019]) and concerns viewing classes as *predicate extensions*:

Class Predicativism: Classes are extensions of predicates.[6]

Once this perspective has been taken, class-theoretic potentialism becomes a natural position. Simply put: Because classes are given by *language* we might think that for any language there is another that non-trivially extends it. This is borne out in the way that both Parsons and Fujimoto express their position. Take the following illustrative passage from Parsons:

> ...we do not have an independent understanding of what predicates or abstracts denote, or what class or second-order variables range over. It follows that "all extensions..." will, unless set-theoretic notions are imported, only mean "the extensions of all possible predicates". And it seems evident that the "totality" of possible predicates is irremediably potential... ([Parsons, 1974, p. 8])

[5]The point that Kunen's Theorem is trivialised when all embeddings are viewed as definable was communicated to us by Sam Roberts and is raised in [Hamkins et al., 2012], [Fujimoto, 2019] (esp. §III), and [Barton, F] (esp. §2.2).

[6]Fujimoto appears to suggest in [Fujimoto, 2019, p. 209] that classes just *are* predicates whereas Parsons characterises them as predicate extensions (see [Parsons, 1974, p. 7]]). Either way, the point remains the same: Predicates are parts of language, rather than the combinatorially characterised objects of iterative set theory.

Similar sentiments are available in Fujimoto:

> 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." ([Fujimoto, 2019], p. 211)

Both Parsons' and Fujimoto's words suggest potentialist readings, the former by referring to the introduction of classes as "irredeemably potential" and the latter by talking about how we may "admissibly *introduce*" predicates. Both, however, opt for non-modal theories, the former via a theory of classes and satisfaction and the latter adopts NBG augmented with a class-theoretic principle satisfying a version of the KF truth-theoretic axioms.[7]

Given that both Parsons and Fujimoto think that there is no definite collection of all predicates, one might think that it would be better to consider a modal class-theoretic potentialist framework in ascertaining the prospects for class predicativism. Beginning with some fixed language, we come to individuate new classes at each additional stage by adding predicates for them into our language. In this way, the classes we have are irredeemably potential, much like the stock of predicates we may admissibly introduce.

1.1.2. *Property Potentialism*[8]. The idea that classes are successively individuated by the addition of *predicates* to *language* can be tied to the thought that class membership corresponds to *property application*. The two are closely linked, since properties are often taken to be the semantic values of predicates, with an application of a property to an object corresponding to a predicate holding of that object (or a name for that object). This kind of view, when taking in unrestricted generality, leads quickly to the semantic paradoxes (cf. [Linnebo, 2006]). A restriction is therefore needed, and one suggestion (made by both [Linnebo, 2006] and [Fine, 2005a][9]) is that the application relation for properties is successively individuated. This successive individuation

[7] See [Fujimoto, 2019], p. 223 for details.

[8] We are grateful to Øystein Linnebo for suggesting this interpretation of classes as leading to class-theoretic potentialism and Sam Roberts for several helpful discussions here.

[9] [Fine, 2005a] does not expressly use property-theoretic language. However, since his view is so close to that of Linnebo's (for example allowing self-membership/application for classes) and he uses the same picture of successively specifying an application relation, we will include them together. See [Roberts, Ua] for discussion.

of the application relation can then lead to a version of class-theoretic potentialism by obtaining different domains of classes by letting each domain be those classes that are co-extensional with properties at some stage of the iteration.

In more detail, Fine conceives of us starting with only the ideological resources of set membership (and first order logic) at our disposal. We can then specify all the properties that correspond to some condition in this initial language. At the next stage, we have an enhanced understanding of what properties apply to, and so can specify which objects fall under the conditions given these expanded ideological resources. We iterate this process along the ordinals in a stage-theoretic manner taking unions at limits. Fine thinks of this process as *yielding* the ZFC sets as the extensions of properties within this framework, but we may also view ourselves as building the properties over an initial fixed stock of sets (since the class of all sets is identified by the condition of *being a set* at the very first stage).

[Linnebo, 2006]'s approach is very similar though different in motivation. Rather than being concerned with classes directly, he is concerned with providing a response to the semantic paradoxes. For this reason, he makes a distinction between sets and properties, with properties the semantic values of conditions. The exact technical details of his project needn't trouble us here (for details see [Roberts, Ua]), the important point for us is that he proposes a theory of properties on which the application relation is *successively individuated* along the ordinals. In this respect both Fine and Linnebo's theories are very similar—they have theories for non-set-like entities defined by conditions on which the application/membership relation is successively individuated. Thus the two accounts come down to essentially the same picture of (potentialist) classes: classes can be viewed as the extensions of properties at some stage of the process of individuation.

Linnebo's theory and Fine's theory are also very similar from a mathematical perspective. The strength of these theories has recently been examined by Roberts [Roberts, Ua]. These property theories are provably consistent in NBG + Π^1_1-Comprehension. More strength can be obtained by the addition of a reflection principle for properties in this context. In particular, the principle that (for φ in the language of property theory) "φ holds concerning all properties iff φ true at some stage of the iteration" yields full impredicative comprehension when we let classes be interpreted in the above manner. Moreover, these property theories are mutually interpretable with ZFC with the addition for Ord-iterated truth predicates (see [Roberts, Ub]).

Whilst both Fine and Linnebo give an *actualist* theory of properties (within which we can interpret certain fragments of class theory) their philosophical claims admit a potentialist interpretation. Talk of successively individuating the application relation (and the new classes that are available each time we do) can be naturally thought of in a potentialist manner. Indeed, one can see this potentialism concerning the classes that are the extensions of properties in the sets as being given formal codification in a stage-theoretic version of Linnebo's property theory given by [Roberts, Ua]. Philosophically speaking though, the way that the two constructions are phrased also suggest this interpretation. For example, Fine writes:

> On the usual conception of the cumulative hierarchy of Zermelo-Fraenkel set theory (ZF), we think of the membership predicate as given and of the ontology of sets or classes as something to be made out. Thus given an understanding of membership, we successively carve out the ontology of sets by using the membership predicate to specify which further sets should be added to those that are already taken to exist. Under the present approach, by contrast, we think of the ontology of classes as given and of the membership predicate as something to be made out. Thus given an understanding of the ontology of classes, we successively carve out extensions of the membership predicate by using conditions on the domain of classes to specify which further membership relationships should obtain. ([Fine, 2005a], p. 547)

Linnebo expresses a similar sentiment:

> We begin by individuating some class of set-theoretic properties. For concreteness, assume we individuate those set-theoretic properties definable...allowing for parameters referring to pure sets. Now we want to use the set-theoretic properties we have just individuated to individuate more properties. ([Linnebo, 2006], p. 173)

For both Fine and Linnebo, given this idea of successively individuating the application relation, there will be new classes appearing as the application relation progressively individuates certain properties as holding of more and more sets. For instance as we move through the first few levels of individuation, we will individuate a truth-predicate for the language of ZFC (after the first stage), and then a truth predicate for the expanded language, and so on.[10]

1.1.3. *Postulationism.* A different bottom-up approach can be obtained from Kit Fine's 'procedural postulationism' (as in [Fine, 2005b]). According to Fine, we gain knowledge of mathematical objects by *postulating* their existence. For Fine, however, postulation amounts to more than the mere postulation of a truth of a proposition, rather it concerns providing a rule for the construction of a particular entity (or entities):

> [Procedural postulationism] shares with traditional forms of postulationism, advocated by Hilbert (1930) and Poincare (1952), the belief that the existence of mathematical objects and the truth of mathematical propositions are to be seen as the product of postulation. But it takes a very different view of what postulation is. For it takes the postulates from which mathematics is derived to be imperatival, rather than indicative, in form; what are postulated are not propositions true in a given mathematical domain, but procedures for the construction of that domain. ([Fine, 2006], p. 89)

Fine links these imperatival conditions to sets of rules (what he calls *procedures* or *postulational rules*) for the construction of the domain. These procedures can be thought of as analogous to computer programs—similar to how a computer program moves a machine from one state to another, a postulational rule tells us how to go to one composition of a mathematical domain to another. Of course, the analogy cannot be completely tight, even with the relatively liberal definition of Turing computability we will not get us many of the mathematical objects we want (e.g. uncountable sets). We thus need a liberalised notion of procedure. Go too liberal though and we encounter a further problem—obviously certain imperatival rules (e.g. "Introduce the Russell set!") will be inconsistent. Fine thus attempts to sharpen his account by making explicit the kinds of rules we are allowed implement. Specifically, he introduces the following rules (letting $C(x)$ be some condition on objects):

Introduction: $!x.C(x)$

which is to be read as "introduce an object x conforming to the condition $C(x)$". This act of postulation creates an object satisfying $C(x)$ (if one does not exist already), and otherwise does nothing. The complex rules are:

[10] For details, see [Roberts, Ub].

Composition: Where β and γ are rules then so is $\beta;\gamma$, which may be read as "First execute β and then execute γ".
Conditionality: Where β is a rule and A is an indicative sentence, then $A \to \beta$ is also a rule, which may be read as "If A then do β".
Universality: Where $\beta(x)$ is a rule (that can be applied to an arbitrary object x), then so is $\forall x \beta(x)$, which may be read as "for each x (simultaneously) do β"). We also allow the rule $\forall F \beta(F)$, where F is a second-order variable ranging over any plurality of the initial domain.
Iteration: Where β is a rule, so is the operation of executing β any finite number of times (we call this operation β^*).

Fine suggests that an imperatival logic can be obtained for these conditions, but again the details are not important.[11] What is salient in this context is that his postulationism leads naturally to class-theoretic potentialism. Assume that the sets are given (which might be themselves previously constructed through postulational acts). We can then consider postulational rules for introducing classes, such as"add a truth predicate!". Such postulational acts are naturally thought of in potentialist terms. It should be noted that there are potentially many different postulational processes that might lead to different class-theoretic potentialisms. Whilst one might view postulationsism as telling us how 'the' domain of classes can be obtained via postulation, this is not essential to the position. It could be (for instance) that we instead view different postulational processes as yielding different conceptions of modal space, with no single one being privileged.

That concludes our introduction of bottom-up approaches to class-theoretic potentialism. There may be others, but this is not so important for our purposes—we just want to motivate consideration of the

[11]It should be noted that the details are not clear to us either—[Fine, 2005b] contains several promissory notes concerning technical specifics, but we do not know of anywhere that they have appeared. Nonetheless, the view (on many families of sharpenings) naturally suggests class-theoretic potentialism and that is enough for present purposes.

view, not provide a comprehensive description of its possible motivations, and there may well be others.[12] Just to review, the key facets of bottom-up approaches are:

Initial World: We obtain classes *beginning* with some *initially* specified classes and then...
Individuation: ... we *individuate* new classes over the existing classes.

1.2. **Top-down approaches.** A different route to class-theoretic potentialism is *top-down* in nature. Instead of *starting* with some antecedently given collection of classes and iterating a process of individuation, we might instead view potentialism arising out of *referential indeterminacy*. This is the core approach of *top-down* views: We state some conditions we would like domains of classes to satisfy, but it may be that there is no single domain that is thereby referred to. We can then take class-theoretic potentialism to be telling us how we may move around within these domains that satisfy our basic class-theoretic principles. More concretely, we may see the following views as motivating class-theoretic potentialism.

1.2.1. *Multiverse approaches to class-theoretic potentialism.* The first is relatively simple in nature—we may view class-theoretic potentialism as being motivated by garden-variety set-theoretic potentialism. If one thinks that any universe of set theory appears as a set in a larger universe (i.e. for universe V there is another universe V' such that $V' \in V$) *and* that any universe can be extended by set forcing,[13] then class-theoretic potentialism considers multiversally-interesting set-theoretic structures. For example given a universe V, we can always make V countable by moving to a universe V' in which V appears as a set, and then collapsing $|V|$ to ω by forcing over V' (call this universe $V'[G]$).

[12]For example, a slightly unusual way of motivating class-theoretic potentialism would be to view classes as *parts* of the universe of sets (which is trivially a part of itself). Such an approach is taken by [Welch and Horsten, 2016], but with a very Platonistic flavour. They regard the classes (i.e. parts) of V as independently existing but *non-mathematical* objects, instead opting to view them as *meta*mathematical objects. (They then use this account of classes to motivate an embedding-theoretic characterisation of a reflection principle.)

However, we might take a potentialist attitude to reference to these parts. The idea is somewhat similar to property-theoretic or predicativist account, but rather than viewing the potentialism as arising out of *individuation* or *language expansion*, we can regard it as arising out of subsequent increasingly fine-grained reference to parts. Since the view of 'parts as potential' is somewhat unusual we won't consider this possibility here, but it merits further consideration.

[13]Views of this kind include [Hamkins, 2012], [Arrigoni and Friedman, 2013], and [Scambler, 2021].

Within $V'[G]$, we can consider various class-theoretic potentialist systems (to be discussed in greater detail in §2), such as those collections $\mathcal{X}$ of subsets of V in $V'[G]$ for which $(V, \mathcal{X}) \models \mathsf{NBG}$.

For certain versions of multiversism, this kind of potentialism can be motivated on philosophical as well as mathematical grounds. If one thinks that the concept of *arbitrary set* is indeterminate (and holds a multiversism on these grounds), one is likely to hold also that our concept of *class* is also indeterminate. Thus, even if I fix some universe V as a starting position (within the multiverse), it is unlikely to be determinate exactly what classes exist over V. Similarly, for the usual set forcing potentialist, even if I fix the natural numbers, they are unlikely to hold that the reference to *all sets of natural numbers* is determinate. Thus, given a universe V in the multiverse, we can construe reference to all classes of V as referentially indeterminate and yielding a potentialist system of its own.

1.2.2. *Plurals and Potentialism.* Plural resources have been used to interpret proper-class talk (see, for example, [Uzquiano, 2003]). Often the ranges of plural variables are taken to be determinate (e.g. [Hossack, 2000], [Uzquiano, 2003]). However, this view has recently been challenged by the work of Florio and Linnebo (in [Florio and Linnebo, 2016]) who show that there are versions of Henkin semantics for plural logic, and argue that this calls into question the determinacy of plural quantification.

We won't get into the details about whether or not one should accept that plural resources are in fact indeterminate—our focus here is on considering a range of possible views that might motivate the more general idea of class-theoretic potentialism, rather than trying to settle this tricky matter. However, *if* one does accept that such resources are indeterminate, one might be able to motivate a class-theoretic potentialism. Namely, the referential indeterminacy in the plural quantification extends to referential indeterminacy about classes. Once we have this indeterminacy in the picture, it is a short step to class-theoretic potentialism, understood as the study of different *precise* interpretations of the plural variables and how we may move between these interpretations.

So, to conclude this section, in addition to bottom up approaches and their twin pillars of **Initial World** and **Individuation**, we have top-down approaches that are based on the following two ideas:

Referential Indeterminacy: Over a given universe of sets V, reference to *the classes of* V is not determinate (i.e. does not pick out a unique privileged interpretation).

Interrelation of Interpretations: Class-theoretic potentialism can be understood as interrelating these distinct possible interpretations (e.g. how one can move between them, what theories they satisfy, etc.).

2. Class-theoretic potentialist systems

With these motivations for the broad idea of class-theoretic potentialism in hand, it is time to lend some mathematical precision to their study. In this section we'll discuss some different kinds of class-theoretic potentialist systems and how some of the philosophical views relate to these potentialist systems.

2.1. **Class theoretic principles.** We will use a two-sorted approach to class theory, with *sets* and *classes* as the two types of objects. A model of class theory will be denoted $(M, \mathcal{X})$, where M is the sets and $\mathcal{X}$ is the classes. We are interested in transitive models, for whom their membership relation is the true $\in$, and will suppress the membership relation in the notation.

We call a formula in the language of class theory *elementary* if its quantifiers only occur over set variables (but class parameters are allowed).

Definition 1 (Class theories)**.** All our class theories will include ZFC for the sets. Where they differ is in their axioms for classes.[14] They also include an extensionality axiom for classes and a replacement axiom for classes—if F is a class function and a is a set then $F''a$ is a set.

- Adding the predicative comprehension schema, viz. the instances of comprehension for elementary formulas, gives *von Neumann–Gödel–Bernays class theory* NBG.
- Adding the full impredicative comprehension schema, viz. all instances of comprehension, including those with class quantifiers, gives *Morse–Kelley class theory* MK.

Beyond these two class theories certain class theoretic principles will arise in our investigation.

Definition 2. *Global choice* is the assertion that there is a global choice function for all nonempty classes. Equivalently, it may be formulated as the assertion of a bijection $\mathrm{Ord} \to V$ or the assertion of a global well-order.

[14]Though the reader should be clear that these class axioms can have consequences for the sets, e.g. implying the existence of transitive models of ZFC.

It is well-known that MK does not imply global choice, but that global choice has no consequences for sets. Given a model of class theory a generic global well-order can be added by a forcing which does not add sets.

Definition 3. *Elementary transfinite recursion* ETR is the principle asserting that transfinite recursion of elementary properties along well-founded classes have solutions.

Observe that MK proves ETR, since one can define the solution to an elementary transfinite recursion with an impredicative formula. On the other hand, ETR exceeds NBG in consistency strength, since the Tarskian truth predicate for V can be given by elementary transfinite recursion, and thereby ETR proves the consistency of ZFC.

Indeed, ETR is closely connected to truth predicates, and can equivalently be expressed as a truth-theoretic principle.

Theorem 4 (Fujimoto [Fujimoto, 2012])**.** *Over* NBG*,* ETR *is equivalent to the assertion that iterated truth predicates*[15] *of any length relative to any class parameter always exist.*

One can restrict ETR to get a hierarchy of transfinite recursion principles. If Γ is a class well-order let $\mathsf{ETR}(\Gamma)$ denote the restriction of ETR to recursions along well-founded classes of rank $\leq\Gamma$ and let $\mathsf{ETR}(<\Gamma)$ denote the restriction of ETR to rank $<\Gamma$. These principles separate from full ETR and from each other based on Γ, according to consistency strength; see [Williams, 2019] for details.

2.2. **Explaining the systems.** A *potentialist system* is a collection of structures of the same type, ordered by a reflexive and transitive relation $\subseteq$ which refines the substructure relation. They give a formalisation of a domain we can think of as dynamically growing. In this section, we will provide the basic definitions for the potentialist systems we plan on considering.

Our two philosophical approaches to class-theoretic potentialism (viz. bottom-up and top-down) correspond to two different (but related) ways of studying potentialist systems mathematically. For top-down approaches, we consider all possible collections of classes which meet some basic criteria. We begin with the following:

Definition 5. Let T be a second-order set theory, such as NBG or MK. Fix a countable $M \models \mathsf{ZFC}$. The *T-class potentialist system for M* is the collection of all countable $(M, \mathcal{X}) \models T$. Of course, these can be

[15] See the beginning of §4 for definitions and fuller discussion of truth predicates and iterated truth predicates.

identified with their second-order parts $\mathcal{X}$. The relation here is just the usual substructure relation.

For ease of writing, we will call such $\mathcal{X}$ a *T-expansion for M*.

It could be that M has no T-expansion. For example, MK has first-order consequences which go beyond ZFC and so M might have no MK-expansions due to having a bad theory, e.g. if M has no worldly cardinals. We will implicitly assume that this does not happen, and that our potentialist system is nontrivial.

Top-down approaches then can be viewed as considering different T-class potentialist systems. A top-down potentialist might, for instance, hold a view of classes that legislates they should satisfy MK. But, due to indeterminacy of reference, they cannot point to a single definite collection of classes, and instead want to consider all possible collection of classes which could be put on the sets.

Bottom-up approaches to potentialist systems, by contrast, specify rules for extension, and then consider which potentialist systems satisfy these rules. Rather than starting from the outset with a fixed potentialist system, we consider axioms governing the accessibility relation and what extensions must exist, and then ask which, if any, potentialist systems satisfy these axioms. For example, we will consider what happens when we require having extensions which arise from the addition of truth-predicates or, by taking (tame) class-forcing extensions which do not add sets. Some of the views we consider motivate further restrictions on the accessibility relation. For instance, the systems of Linnebo and Fine both have built in that the accessibility relation must be well-founded, since for them the process of property-theoretic membership individuation is well-founded.

For both approaches, an important question will be to understand the modal logic of the potentialist systems in play. We provide some definitions and background for this before we give a more fine-grained analysis.

3. Modal logics of class-theoretic potentialism

Given a potentialist system $(\mathfrak{A}, \subseteq)$ there is a natural modal interpretation: $\Box\varphi$ holds at a world $M \in \mathfrak{A}$ if φ holds in every $N \supseteq M$ and $\Diamond\varphi$ holds at a world $M \in \mathfrak{A}$ if φ holds in some $N \supseteq M$. A modal assertion is *valid* at $M \in \mathfrak{A}$ if it's true for any substitution of formulae for the propositional variables. For example, $\Box p \Rightarrow p$ is valid at every world because if φ holds in every extension of M it in particular holds at M.

A key mathematical question is then: what are the *modal validities* of $\mathfrak{A}$, the collection of modal assertions valid at every world in $\mathfrak{A}$?

An easy observation is that $\mathsf{S4}$ is valid for any potentialist system.

Definition 6. $\mathsf{S4}$ is the modal theory axiomatised by the following axioms

$$\begin{aligned} &(\mathsf{K}) && \Box(p \Rightarrow q) \Rightarrow (\Box p \Rightarrow \Box q) \\ &(\mathsf{D}) && \neg \Diamond p \Leftrightarrow \Box \neg p \\ &(\mathsf{S}) && \Box p \Rightarrow p \\ &(\mathsf{4}) && \Box p \Rightarrow \Box\Box p \end{aligned}$$

and closed under the inference rules of *modus ponens* and necessitation.

K and D hold for free for any modal logic which comes from a Kripke frame; they come from the corresponding rules for universal quantification applied to quantifying over worlds in the frame. S holds on any frame whose accessibility relation is reflexive, and $\mathsf{4}$ holds on any frame whose accessibility relation is transitive. Since potentialist systems are, by definition, reflexive and transitive they will always validate $\mathsf{S4}$.

Some other modal axioms we consider in this article are:

Definition 7.

$$\begin{aligned} &(.2) && \Diamond\Box p \Rightarrow \Box\Diamond p \\ &(.3) && (\Diamond p \wedge \Diamond q) \Rightarrow [(p \wedge \Diamond q) \vee (\Diamond p \wedge q)] \end{aligned}$$

Adding these to $\mathsf{S4}$ gives, respectively, the theories $\mathsf{S4.2}$ and $\mathsf{S4.3}$. One way to think of them is the corresponding frame conditions: .2 holds for any frame whose accessibility relation is directed and .3 holds for any frame whose accessibility relation is linear.

In the next section we will compute the exact modal validities of a few potentialist systems. Let us briefly describe the main tools, *control statements*, which are used for such arguments.

Definition 8 ([Hamkins and Löwe, 2008, Hamkins et al., 2015])**.**

- A *button* is an assertion β so that $\Diamond\Box\beta$ holds at every world. If $\Box\beta$ holds at a world M, we say β is *pushed* for M, otherwise we say β is *unpushed*. The intuition is, you can push a button, making β true forevermore, but once you push it you can never unpush it.
- A *switch* is an assertion σ so that $\Diamond\sigma$ and $\Diamond\neg\sigma$ holds at every world. The intuition is, you can toggle the truth value of σ freely back and forth.

- A *ratchet* is a finite sequence $\rho_0, \dots, \rho_n$ of buttons so that pushing ρ_i pushes ρ_j for all $j < i$. The intuition is, you can ratchet forward but never back.
- A *long ratchet* of length Γ is is a uniformly definable sequence of buttons r_ξ, indexed by $\xi < \Gamma$, so that pushing r_ξ pushes r_η for all $\eta < \xi$ and so that in no world are all buttons on the ratchet pushed. Observe that this second condition forces the ratchet to have limit length, as if there were a last button then we could push it to push all the buttons.

A collection of control statements is called *independent* if any subcollection of the control statement can be manipulated without affecting any of the other control statements.

By showing that a potentialist system admits certain control statements, we get upper bounds for their modal validities.

Theorem 9 ([Hamkins and Löwe, 2008])**.** *If a potentialist system admits arbitrarily large finite families of independent buttons and switches then its modal validities are contained within* S4.2*.*

Theorem 10 ([Hamkins et al., 2015])**.** *If a potentialist system admits arbitrarily long ratchets which are independent with arbitrarily large families of switches then its modal validities are contained within* S4.3*.*

Corollary 11. *If a potentialist system admits a long ratchet whose length Γ is closed under addition $<\omega^2$ then* S4.3 *is an upper bound for its modal validities.*

Hamkins, Leibman, and Löwe proved this result for long ratchets of length Ord, but it is simple to check that what they used is that Ord is closed under addition $<\omega^2$.

Proof sketch. By Theorem 10 it is enough to see that having a long ratchet as in the statement of the lemma can simulate arbitrarily long ratchets which are independent with arbitrarily large finite families of switches. To this end, write the unpushed indices on the long ratchet as $\xi + \omega \cdot \alpha + k$, where ξ is the supremum of the pushed indices and $k < \omega$. The α part gives a position in the ratchet (sitting at the end if α is too large) and the bits of k simulate the switches. These can be freely changed without increasing the position on the ratchet, and we can move further along the ratchet without affecting the pattern of the switches. The condition on the length Γ ensures that there is always space to do this, no matter what ξ is for the current world. □

With the set up of potentialist systems and their modal validities in hand, we can begin to examine class-theoretic potentialism mathematically and draw some philosophical conclusions on this basis. As we shall see, many potentialist systems violate the .2 and .3 axioms. These results will help us to raise some challenges for class-theoretic potentialism and help us to elucidate the position further, in particular relating these results to the philosophical motivations considered in §1 (we do so in §5 and §6).

4. Truth predicates and potentialism

In this section we present some mathematical results about class potentialist systems which involve truth predicates. Especially relevant for philosophical purposes will be failures of the .3 (Theorem 18) axiom and .2 axiom (Theorem 19) for certain systems. Most of the results are phrased in terms of certain bottom-up potentialist systems, but some of the work also applies to top-down potentialist systems; failures of modal principles for the smaller systems can sometimes be pushed up to a larger system.

Let us begin by fixing some notation and definitions. We will use capital Greek letters—e.g. Λ, Γ—to refer to class well-orders. Addition, multiplication, and exponentiation on these are defined as usual. To match the familiar notation for ordinals, we write $\xi < \Lambda$ to mean $\xi \in \mathrm{dom}(\Lambda)$. To compare elements ξ and η of Λ we write $\xi < \eta$. Given class well-orders Λ and Γ say that Λ is *closed under addition* $<\Gamma$ if whenever $\xi < \Lambda$ and $\eta < \Gamma$ we have that $\xi + \eta < \Gamma$. That is, Λ is closed under addition $<\Gamma$ if every element of Λ has an η-th successor in Λ for each $\eta < \Gamma$.[16]

Consider a fixed transitive $M \models \mathsf{ZF}$.[17] The *truth predicate* for M is the $\mathrm{Tr} \subseteq M$ which satisfies the recursive Tarskian rules for the satisfaction class for $(M, \in)$. In case we wish to emphasise for which structure

[16]Observe that if $\Lambda \cong \Lambda'$ then pushing forward along the isomorphism transforms a definition along Λ into a definition along Λ'. As such, even though in the class theoretic context we do not have access to the Mostowski collapse lemma to select representatives of well-order types, our choice of a representative is harmless.

[17]One can also consider truth predicates over even ω-nonstandard models, instead asking for a class which satisfies the Tarskian recursion for the internal notion of formula, including nonstandard formulas. The usual terminology in this context is *full satisfaction classes*, to distinguish from the externally-defined truth predicate. It is well-known that in this context full satisfaction classes need not be unique [Krajewski, 1976], though a model of NBG can have at most one full satisfaction class. Since the interest in the current article is in transitive models, we work only in that setting.

Tr is a truth predicate we will write Tr^M. Observe that Tr^M is uniquely determined by a formula which only quantifies over sets, and so any NBG-expansion of M can verify whether a class is Tr^M. Given a class $A \subseteq M$ the *truth predicate relative to* A is the unique class $\mathrm{Tr}(A) \subseteq M$ which satisfies the Tarskian recursion to be the satisfaction class for $(M, \in, A)$. Again, it is uniquely determined and can be recognised as such by any NBG-expansion of M.

Given the truth predicate $\mathrm{Tr} \subseteq M$, we can consider $\mathrm{Tr}(\mathrm{Tr})$, the truth predicate relative to Tr, and so on transfinitely. These can be unified in the single definition of an iterated truth predicate.

Working over our fixed transitive $M \models \mathsf{ZF}$, let Λ be a well-order, possibly a proper class. A Λ-*iterated truth predicate* is a class Θ of triples (ξ, φ, a) where $(\xi, \varphi, a) \in \Theta$ intuitively means that $\varphi(a)$ is true at level $\xi < \Lambda$. Here φ is a formula in the language where we added a predicate symbol $\hat{\Theta}$ for Θ. Formally, this is defined by a modified form of the Tarskian recursion, with an extra clause in the definition asserting that $(\xi, \ulcorner \hat{\Theta}(x, y, z) \urcorner, \langle \eta, \varphi, b \rangle) \in \Theta$ if and only if $\eta < \xi$ and $(\eta, \varphi, b) \in \Theta$. As with the ordinary case, we can consider iterated truth predicates relative to a class parameter A via the obvious modification.

Note that NBG suffices to prove the uniqueness of iterated truth predicates. Given two classes which satisfy the definition of being a Λ-iterated truth predicate relative to A, by predicative comprehension we can form the class of indices where they disagree. So if they disagree they must disagree at a minimal stage, from which we can derive a contradiction. We will use Tr_Λ to denote *the* Λ-iterated truth predicate and $\mathrm{Tr}_\Lambda(A)$ to denote the Λ-iterated truth predicate relative to A. If $\xi < \Lambda$ then we write Tr_ξ to mean $\mathrm{Tr}_{\Lambda \upharpoonright \xi}$.

Observe that, up to recoding, Tr_1 is the same as Tr and $\mathrm{Tr}_{\Lambda+1}$ is the same as $\mathrm{Tr}(\mathrm{Tr}_\Lambda)$, and similarly for relativised truth predicates.

The reader should also note that a class being well-founded can be expressed just by quantifying over sets. Accordingly if a class relation R is in two NBG-expansions for M then they agree on whether R is well-founded.

With these definitions in hand, let us now describe a species of class potentialist systems meant to capture the idea that we can always expand by adding truth predicates. First, a bit of notation. If $\mathcal{X}$ is a collection of classes over M and A is a class over M then let $\mathcal{X}[A] \subseteq \mathcal{P}(M)$ be the smallest NBG-expansion for M which extends

$\mathcal{X}$ and contains A.[18] Specifically, $\mathcal{X}[A]$ consists of the classes over M definable using A and finitely many classes from $\mathcal{X}$.

Definition 12. Say that a class potentialist system over $M \models \mathsf{ZF}$ is a *truth potentialist system* if it satisfies the following three properties.

(1) $(M, \mathrm{Def}(M))$ is a world, where $\mathrm{Def}(M)$ is the collection of parametrically first-order definable classes over M.
(2) If $(M, \mathcal{X})$ is a world then it satisfies NBG.
(3) If $(M, \mathcal{X})$ is a world and $A \in \mathcal{X}$ then $(M, \mathcal{X}[\mathrm{Tr}(A)])$ is a world.

We can modify the third condition to require truth predicates of a longer length, say of length $<\Lambda$. We call such a system a *length* $<\Lambda$*-length truth potentialist system.*

(3_Λ) If $(M, \mathcal{X})$ is a world, $\xi < \Lambda$, and $A \in \mathcal{X}$ then $(M, \mathcal{X}[\mathrm{Tr}_\xi(A)])$ is a world.

Observe that (ordinary) truth potentialist systems are the special case where $\Lambda = 2$. This condition can be further modified, in the obvious way, to require truth predicates along class well-orders of unbounded length. The latter situation we call an *unbounded truth potentialist system.*

Before we analyse truth potentialist systems, let us briefly remark that they place restrictions on which M we may consider.

Proposition 13 (Krajewski [Marek and Mostowski, 1975, page 475])**.** *Consider $(M, \mathcal{X})$ a transitive model of* NBG *with $\mathrm{Tr}^M \in \mathcal{X}$. Then M contains a club of ordinals α so that V^M_α is an elementary submodel of M.*

Proof. By reflection using Tr as a parameter we get club many α so that $(V^M_\alpha, \in, \mathrm{Tr} \cap V^M_\alpha)$ is a Σ_1-elementary submodel of $(M, \in, \mathrm{Tr})$. But we can express that $(V^M_\alpha, \in)$ is an elementary submodel with a Σ_1-formula referring to the truth predicate. $\square$

Thus if M admits a truth potentialist system then M must have many undefinable ordinals. This rules out, for instance, pointwise-definable models like the minimum transitive model, or models without large cardinals in inner models.[19]

[18]The reader may worry about whether there is any such NBG-expansion for M. It is right to worry, since in general there need not be. But we will confine ourselves to a context in which it is defined.

[19]For instance, the Inner Model Hypothesis of [Friedman, 2006] states that any parameter-free first-order sentence that is true in an inner model of an outer model is already true in an inner model. In the modal setting, we can think of this as the assertion (using directed modal operators) that if it is upwardly possible that it is

We next present some results about smallest truth potentialist systems over a fixed M. First, however, let us clarify in what sense a potentialist system may be smallest among a collection of systems. One way to compare potentialist systems is by containment: if $\mathfrak{A}$ and $\mathfrak{B}$ are potentialist systems then $\mathfrak{A} \subseteq \mathfrak{B}$ if every world in $\mathfrak{A}$ is a world in $\mathfrak{B}$. But this comparison is inadequate for many purposes; for instance, $\mathfrak{B}$ could have more worlds than $\mathfrak{A}$ because it breaks the worlds of $\mathfrak{A}$ into finer-grained worlds . Say that $\mathfrak{A}$ *covers* $\mathfrak{B}$ if every world in $\mathfrak{B}$ is contained in some world in $\mathfrak{A}$. If $\mathfrak{A} \subseteq \mathfrak{B}$ and $\mathfrak{A}$ covers $\mathfrak{B}$ then we say that $\mathfrak{B}$ *refines* $\mathfrak{A}$. A potentialist system is *refined* relative to a collection of systems if it has no proper refinements within the collection. Given a collection of systems, the *smallest* potentialist system in the collection, if it exists, is the refined system which is covered by every other system.

Theorem 14. *If M admits a truth potentialist system then it admits a smallest truth potentialist system. This potentialist system validates* $\mathsf{S4.3}$.

Proof. Set $\mathcal{X}_0 = \mathrm{Def}(M)$ and, for $n > 0$, set $\mathcal{X}_n = \mathcal{X}_0[\mathrm{Tr}_n]$. We claim the potentialist system consisting of these worlds, call it $\mathfrak{X}$, is as desired. By construction, each world satisfies all axioms of NBG except possibly the class replacement axiom. Because M admits some truth potentialist system, an easy induction shows that each $\mathcal{X}_n$ is contained in a world in a truth potentialist system over M, whence we conclude that $(M, \mathcal{X}_n)$ does indeed satisfy class replacement. The other properties of $\mathfrak{X}$ being a truth potentialist system are clear from the construction. In particular, to find the truth predicate relative to a class in $\mathcal{X}_n$ look in $\mathcal{X}_{n+1}$.

Now we must see that this is the smallest truth potentialist system on M. To this end, fix some other truth potentialist system $\mathfrak{A}$ on M. An easy induction shows that, for each n, some world in $\mathfrak{A}$ must contain Tr_n. So we see that $\mathfrak{A}$ covers $\mathfrak{X}$. And $\mathfrak{X}$ is refined because an easy induction shows that $\mathfrak{X}$ contains as a subset any $\mathfrak{A}$ which $\mathfrak{X}$ covers.

To see that $\mathfrak{X}$ validates $\mathsf{S4.3}$, merely observe that $\mathfrak{X}$ is linearly ordered and recall that linear orders validate $\mathsf{S4.3}$. □

This construction generalises to length $<\Lambda$ truth potentialist systems, and for lengths with the correct closure property we can exactly characterise the modal validities.

---

downwardly possible that φ, then it is already downwardly possible that φ. This principle implies that there are no worldly cardinals in M (see [Friedman, 2006], [Antos et al., 2021], and [Barton, S]).

Theorem 15. *Fix a length $\Lambda \in \mathrm{Def}(M)$ where Λ is well-founded as seen externally from V. If M admits a length $<\Lambda$ truth potentialist system then it admits a smallest one.*

Proof. The strategy is the same as before. The smallest truth potentialist system, call it $\mathfrak{X}$, will contain $\mathcal{X}_0 = \mathrm{Def}(M)$ and consist of other worlds $\mathcal{X}_\xi = \mathcal{X}_0[\mathrm{Tr}_\xi]$, where we want to take the smallest segment of ξs which makes this work. For this, we must split into cases. If Λ is closed under addition $<\Lambda$ then we take $\xi < \Lambda$. Else, we take $\xi < \Lambda \cdot \omega$. In either case, we get that if $\xi, \eta < \Lambda$ then we can find the η-iterated truth predicate for a class in $\mathcal{X}_\xi$ by looking in the world $\mathcal{X}_{\xi+\eta}$. The other properties of being a truth potentialist system are checked as in the previous result; we use that Λ really is well-founded to ensure the validity of the inductive argument that each $\mathcal{X}_\xi$ is contained in a world in some truth potentialist system over M.

Now let's see that $\mathfrak{X}$ is the smallest length $<\Lambda$ truth potentialist system on M. Fix some other such system $\mathfrak{A}$ on M. Using that Λ is externally well-founded, an inductive argument shows that each Tr_ξ must be in some world in $\mathfrak{A}$. Thus $\mathfrak{A}$ covers $\mathfrak{X}$. And $\mathfrak{X}$ is refined because an external induction, along either Λ or $\Lambda \cdot \omega$ depending on the above bifurcation, shows that $\mathfrak{X}$ contains as a subset any $\mathfrak{A}$ which $\mathfrak{X}$ covers. □

Theorem 16. *Fix a length $\Lambda \in \mathrm{Def}(M)$, where Λ is externally seen to be well-founded. Consider the smallest length $<\Lambda$ truth potentialist system $\mathfrak{X}$ on M as constructed in the previous result. Every axiom of* S4.3 *is valid in this potentialist system. Moreover if either $\Lambda \geq \omega^2$ is closed under addition $<\Lambda$ or $\Lambda \cdot \omega$ is closed under addition $<\omega^2$ then the modal validities are precisely* S4.3.

Proof. That S4.3 is a lower bound for the modal validities is again the observation that $\mathfrak{X}$ is linearly ordered. To get the upper bound, we will use Corollary 11 and demonstrate a long ratchet for this potentialist system. Let r_ξ be the assertion "Tr_ξ exists". Then $\langle r_\xi : \xi < \Lambda^* \rangle$ gives a long ratchet for this potentialist system, where Λ^* is either Λ or $\Lambda \cdot \omega$ depending on which case we are in for constructing $\mathfrak{X}$. The assumptions on Λ ensure that Λ^* is closed under addition $<\omega^2$, allowing the corollary to apply. □

Next let's see an analogous result where we do not bound the lengths of truth predicates. To carry out the argument we again need M to be correct about the well-foundedness of the would-be lengths Λ, which we encapsulate in the following definition. Say that $(M, \mathcal{X}) \models$ NBG is a *β-model* if it is correct about which of its classes are well-founded.

A short argument yields that if Ord^M has uncountable cofinality then $(M, \mathcal{X})$ is a β-model. In contrast, if Ord^M has countable cofinality then $(M, \mathcal{X})$ may fail to be a β-model.[20]

Theorem 17. *Suppose M admits an unbounded truth potentialist system whose every world is a β-model. Then M admits a smallest unbounded truth potentialist system, and this potentialist system validates exactly* $\mathsf{S4.3}$*.*

Let us remark that the condition on M is satisfied if M has an expansion to a β-model of $\mathsf{NBG} + \mathsf{ETR}$, by considering the trivial potentialist system consisting of just that one world.

Proof. We construct this smallest unbounded truth potentialist system $\mathfrak{X}$ in stages. Start with $\mathfrak{X}_0$ consisting only of $\mathcal{X}_0 = \mathrm{Def}(M)$. Given $\mathfrak{X}_n$, let $\mathfrak{X}_{n+1}$ consist of all worlds of the form $\mathcal{X}_0[\mathrm{Tr}_\Lambda]$ where Λ is a well-order in some world in $\mathfrak{X}_n$. Finally, set $\mathfrak{X} = \bigcup_{n<\omega} \mathfrak{X}_n$. By an external induction see that each $\mathfrak{X}_n$ is covered by a truth potentialist system consisting of β-models. Therefore, $\mathfrak{X}$ consists of β-models. This also establishes that each world in each $\mathfrak{X}$ satisfies class replacement, whence we know they satisfy NBG. It is clear from the construction that $\mathfrak{X}$ is closed under adding truth predicates of any length (in the current world), so we conclude $\mathfrak{X}$ is an unbounded truth potentialist system.

Showing that $\mathfrak{X}$ is smallest proceeds as before, using an external induction along the the supremum of the lengths of well-orders in worlds in $\mathfrak{X}$. Note that even if another truth potentialist system $\mathfrak{A}$ has worlds which are not β-models, an induction on ω shows it must contain worlds with the lengths in each $\mathfrak{X}_n$, and so it must be that $\mathfrak{A}$ covers $\mathfrak{X}$. The calculation of the modal validities for $\mathfrak{X}$ again uses a long ratchet of iterated truth predicates. Since the lengths of iterated truth predicates are unbounded in this potentialist system, they are closed under addition $<\omega^2$ and so we always have room from the longest length in the current world to extend for our long ratchet. $\square$

We next address the existence of global well-orders. There are a few ways one might ensure a potentialist system includes global well-orders. Observe that if M has a definable global well-order then any truth potentialist system for M satisfies the condition (1), modified to require the base world to satisfy global choice. So we get truth-potentialist systems with global well-orders which validate exactly $\mathsf{S4.3}$. However, this puts a firm restriction on the first-order theory of M,

[20] See e.g. [Williams, 2019] for constructions. The following theorem can be seen as a reworking of Theorem 5.1 of that article into the potentialist system context.

namely $M \models \exists x\ V = \mathrm{HOD}(\{x\})$. One way very well think that this extra restriction is unwarranted, and so consider the general case. Here we can get different behaviour.

Here are two ways to approach the general case. First, we could start by building up from a world of the form $(M, \mathrm{Def}(M,G))$, where G is a global well-order for M, rather than building up from the definable classes. Alternatively, we could add a new rule saying that we can expand to a larger world to add a generic global well-order. We consider both approaches.

For the first approach, the way to formulate this is to replace condition (1) in the definition of a truth potentialist system (respectively $<\Lambda$ or unbounded truth potentialist system) with the following.

(1_G) There is a world of the form $(M, \mathrm{Def}(M,G))$ where G is a global well-order for M, and all worlds extend this base world.

At first glance this may look innocuous. But, as we will now show, the choice of global well-order can affect the structure of the potentialist system.

As a warmup, let us see that some global well-orders are inter-definable with Cohen-generic classes of ordinals. For one direction of this, note that if $C \subseteq \mathrm{Ord}$ is Cohen-generic then, by density, every set is coded into the bit pattern of C. So we can define a global well-order by comparing where sets are first coded. Now note that if we take such a global well-order we can rearrange it so that it has the ordinals, in increasing order, placed precisely on the indices where C has value 1. Such a global well-order is still definable from C, and notice that if we have such a global well-order we can recover C by looking at the indices for where ordinals appear. So the two are inter-definable. Indeed, we can say more about generic global well-orders. There is a natural class forcing to add a global well-order without adding new sets. Namely, let $\mathbb{Q}$ consist of set-sized well-orders, ordered by end-extension. By density a generic H for $\mathbb{Q}$ will have all of V as its domain. And given such H we can define a Cohen-generic C by putting $i \in C$ if and only if the i-th element of H is an ordinal. So the two forcings are forcing equivalent.

Theorem 18. *There are truth potentialist systems modified to require a global well-order in the base world whose modal validities are precisely* $\mathsf{S4.2}$. *In particular,* $.3$ *is invalid for these systems.*

Proof. We construct a length $<\mathrm{Ord}$ truth potentialist system, and discuss after the proof to what extent this generalises.

Consider $(M, \mathcal{Y}) \models \mathsf{NBG} + \mathsf{ETR}(\mathrm{Ord})$ which has no global well-order. Force over this model to add a Cohen generic C. Gitman and

Hamkins showed that tame class forcing, such as Cohen forcing, preserves $\mathsf{ETR}(\mathrm{Ord})$ [Hamkins and Woodin, 2018, Theorem 16]. Consider the truth-potentialist system $\mathfrak{X}$ starting with the base world $(M, \mathcal{X}_0)$, where $\mathcal{X}_0 = \mathrm{Def}(M, C)$, and closing off under the requirement that any world can be extended by adding a ξ-iterated truth predicate relative to a parameter for any ordinal ξ. That is, $\mathfrak{X}$ consists of worlds of the form $\mathcal{X}_0[\mathrm{Tr}_\xi(A)]$ where $A \in \mathcal{X}_0$. Observe that we can meet this requirement to obtain a truth potentialist system, as all worlds reachable in this way are coded in $(M, \mathcal{Y}[C])$, and so all worlds must satisfy NBG.

We claim this truth-potentialist system has precisely $\mathsf{S4.2}$ as its modal validities. For the lower bound, it suffices to observe it is directed: if $(M, \mathcal{X}_0)$ and $(M, \mathcal{X}_1)$ are two worlds, then they both are contained within $(M, \mathrm{Def}(M, \mathrm{Tr}_\xi(C)))$ for some large enough ordinal ξ.

For the upper bound, by Theorem 9 it is enough to show that there are arbitrarily large families of independent buttons and switches. To do this, recall that $\mathrm{Add}(\mathrm{Ord}, 1)$ is equivalent to $\mathrm{Add}(\mathrm{Ord}, \omega)$. So we can split C into ω many classes C_i so that the C_i are mutually generic over $(M, \mathcal{Y})$. Further note that this splitting process is definable, so the C_i are uniformly definable from the parameter C. (Namely, you can take C_i to be formed from the bits on the coordinates equivalent to i modulo ω.) In particular, this means that given any ordinals ξ and η, if $i \neq j$ then $\mathrm{Tr}_\xi(C_i)$ is not definable from $\mathrm{Tr}_\eta(C_j)$. This is just because $\mathrm{Tr}_\eta(C_j)$ is in $(M, \mathcal{Y}[C_j])$ and C_i is generic over that model.

Fix a world $(M, \mathcal{X})$ in this potentialist system to work inside. Let λ be the supremum of the lengths of the iterated truth predicates over the C_i's which are in $\mathcal{X}$. We will use the even coordinates i for our buttons and the odd coordinates i for our switches. For the buttons, let β_i be the statement "$\mathrm{Tr}_{\lambda+1}(C_i)$ exists". For the switches, let σ_i be the statement "if ξ is the largest ordinal for which $\mathrm{Tr}_\xi(C_i)$ exists, then ξ is even". It is manifest that these are, respectively, buttons and switches. By the consequence of mutual genericity from the previous paragraph we get that they are independent; we can add a longer truth predicate relative to C_i without affecting which truth predicates exist relative to the C_j for $j \neq i$. This completes the proof. □

It may be helpful to see an explicit example of an instance of $\mathsf{.3}$ which is invalidated by this potentialist system. Suppose we are living in a world $(M, \mathcal{X})$ and define λ as in the proof. Let φ be the assertion "$\mathrm{Tr}_{\lambda+1}(C_0)$ exists but $\mathrm{Tr}_{\lambda+1}(C_1)$ does not exist" and ψ be the assertion where we swap the two coordinates, namely "$\mathrm{Tr}_{\lambda+1}(C_1)$ exists but $\mathrm{Tr}_{\lambda+1}(C_0)$ does not exist". Then φ and ψ are both possible at $(M, \mathcal{X})$.

However, if φ is true at a world then ψ is impossible at that world, and if ψ is true at a world then φ is impossible at that world, giving a failure of $.3$.

Let us remark that the definition of the independent family of infinitely many buttons and switches was uniformly definable across the worlds in the truth-potentialist system. Moreover, the only parameter needed in the definition was C itself.

Let us also comment this construction generalises to give a truth-potentialist system of any limit height. If you want height λ, then instead of allowing extensions by adding any ordinal-length iterated truth predicate, only allow extensions that keep the length below λ. Because no world in this system has the full λ-iterated truth predicate relative to C_i, there is enough space for the definition of the buttons and switches to work out.

Let's now turn to the other option for introducing global well-orders. Rather than add a condition asserting that the base world contains a global well-order, we instead allow the addition a global well-order by class forcing. That is, we want to expand our potentialist system by adding a new rule to get new worlds: if $(M, \mathcal{X})$ is a world then so is $(M, \mathcal{X}[C])$ whenever C is a Cohen subclass of Ord generic over $(M, \mathcal{X})$. We will start with the definable classes as the smallest world, and keep the old rule about being able to add truth predicates relative to extant classes.

The problem is that this is actually quite destructive. Specifically, adding a global well-order may kill off the possibility of adding a truth predicate whilst preserving the basic axioms of NBG.

Theorem 19. *Let M be a countable transitive model of ZF and let $A \subseteq M$ be a class over M so that $(M, \mathrm{Def}(M, A) \models \mathsf{NBG}$. Then there is C Cohen-generic over $(M, \mathrm{Def}(M, A))$ so that no NBG-expansion for M can contain both C and $\mathrm{Tr}(A)^M$.*

Proof. If no NBG expansion for M contains $\mathrm{Tr}(A)^M$ then we are trivially done. So suppose we are not in this case.

We claim that from $\mathrm{Tr}(A)^M$ we can define a sequence $\langle D_\alpha : \alpha \in \mathrm{Ord}\rangle$ of dense subclasses of $\mathrm{Add}(\mathrm{Ord}, 1)$ so that meeting every D_α guarantees genericity over $(M, \mathrm{Def}(M, A))$. This is because we can take D_α to be the intersection of all open dense classes definable from A with parameters from V_α. The point is, from the truth predicate we can define this sequence, because the truth predicate gives us uniform access to definability. Then D_α is open dense because the forcing is κ-distributive for every κ. And clearly meeting every D_α implies getting below every definable dense class.

Now let's use this sequence of D_αs to code a bad real into a generic. Fix a binary sequence $B : \mathrm{Ord} \to 2$ so that the set of i so that $B(i) = 1$ is cofinal in the ordinals of M and has ordertype ω. Note that no NBG-expansion for M can contain B, as B reveals that M's ordinals have countable cofinality. Define a sequence of conditions: start with $p_0 = \emptyset$. Given p_α, extend to meet D_α, where we require the extension to be of minimal length to get into D_α. Then add on the bit $B(\alpha)$ to get $p_{\alpha+1}$. And at limit stages, set $p_\lambda = \bigcup_{\alpha<\lambda} p_\alpha$. Then $C = \bigcup p_\alpha$ meets every dense class in $\mathrm{Def}(M, A)$.

Finally, note that if you have both the sequence $\langle D_\alpha \rangle$ and C, you can recover the coding points and thereby recover B. This is because, given these data, it is a definable property to see the shortest distance you need to extend to meet the next dense class. So if you had both C and the truth predicate in an NBG expansion for M, then you would also have B in the NBG expansion. This is impossible. □

Let us remark that this result immediately implies that we can kill off iterated truth predicates, since $\mathrm{Tr}_{\Lambda+1}(A)$ is inter-definable with $\mathrm{Tr}(\mathrm{Tr}_\Lambda(A))$.

5. The bottom-up approach

In this section we discuss what light the mathematical results from the previous sections shed on bottom-up approaches to class potentialism. The core point is the following: Our results indicate that various bottom-up class potentialist systems might not be convergent.

Before we discuss how this plays out in specific cases, it is helpful to think about why this is relevant for bottom-up approaches. Bottom-up approaches begin by specifying some initial starting world (i.e. **Initial World**) and then individuating new classes over this and subsequent worlds (i.e. **Individuation**). What a lack of convergence shows is that within these systems there are 'choice-points'—positions in the system where we must choose to go one way rather than another.

This has some precedent within other potentialist systems. For instance consider the following two species of *set*-theoretic potentialism: In *forcing potentialism* the worlds are composed of forcing extensions of some ground model and in *rank-extensionalism potentialism* the worlds are the countable models of ZFC, including nonstandard models, with accessibility given by rank-extension. Both exhibit branching; i.e. in each case there are worlds W and W' such that there is no common extension of both. However, the branching in rank-extensionalism potentialism is much more severe. In particular, they get different modal logics. The modal logic of forcing potentialism is $\mathsf{S4.2}$ [Hamkins and Löwe, 2008], whereas for the rank-extensionalist potentialist it is $\mathsf{S4}$ [Hamkins and Woodin, 2017]. Some of the potentialist systems we have considered exhibit explicit failures of the .2 axiom. If we imagine progressively building up the classes in such a system, we face choices of permanent consequence—statements like "There is a truth predicate for A" (where A is some class or other) can be made true (and hence also necessary), but equally can be made impossible.

The failure of .2 has two important ramifications, one philosophical and one mathematical (though perhaps they are different manifestations of the same state of affairs). On the philosophical side, non-convergence indicates a kind of further indeterminacy in our concept of class. It is not just that the classes themselves are *modally* indefinite, but there are also important choices about what is *possible* to be made within this modal space, ones that cannot be reversed.[21] The second (more technical) point to be made is that it creates obstacles for proving mirroring theorems, as in [Linnebo, 2013] or [Hamkins and Linnebo, 2018]. These results, which come in both proof-theoretic and model-theoretic versions, give a translation from a non-modal language to the modal language[22] so that proof (respectively, truth, in the case of the model-theoretic versions) in the potentialist realm corresponds to proof (respectively, truth) in an actualist realm.

The existence of a mirroring theorem for a potentialist system shows that there is a sense in which one can continue using 'actualist' theories even in the presence of a non-actualist ontology. We can move backwards and forwards seamlessly between the modal and non-modal theory, the modal theory is just able to look at the subject matter under a finer (and perhaps more ontologically honest) grain. Importantly though, current treatments of mirroring theorems require at least $\mathsf{S4.2}$ in order to go through.

There are at least two ways one might react to such non-convergence. One is to view non-convergence as a substantial *cost*—we want our notion of class, even if modal, to not contain these choice points both for philosophical cohesiveness and mathematical expedience. Another

[21] A similar point made in [Hamkins, 2018b, §7], where Hamkins argues that $\mathsf{S4}$ potentialist systems exhibit "radical branching" whereas the "directed convergence" of $\mathsf{S4.2}$ systems are closer to an actualist conception. Whilst we do not compute the exact modal logic of our systems in which .2 fails, this failure puts these systems on the radical branching side of the divide. See also the discussion in §8.2 where we conjecture that certain top-down class potentialist systems validate exactly $\mathsf{S4}$.

[22] In brief, the translation is to replace every $\exists$ in a non-modal formula with $\Diamond\exists$.

way to view them is as *interesting* but not any special cost—they indicate interesting structural properties of the relevant potentialist system (and perhaps the underlying conception of class), but this feature is unproblematic. Whether or not they are taken to be a cost or merely a point of interest may depend somewhat on one's philosophical outlook. For now, let's think about how this possible desiderata of convergence might affect the class-theoretic potentialist.

If we want convergence, smallest systems are good. We know that various kinds of classes can interfere with the addition of truth predicates, resulting in non-convergence. However, the *smallest* systems (in the sense outlined in Theorems 14 and 15) do *not* exhibit branching, and validate S4.3. (Indeed, some of these systems validate precisely S4.3.) Thus, if we want to ensure non-branching, a good way to do so is to insist that we consider smallest potentialist systems.

There is a question as to how well this response meshes with the various philosophical views essayed in §1. On the one hand, the predicativist who is only interested in adding truth predicates may have some motivation to take this position. (The case where other predicates are allowed significantly complicates things for them, and we consider this situation below when discussing Global Choice.) Whilst it is somewhat contingent upon the nature of the space of possible language expansions, it seems reasonable to assume that when we introduce individual new truth predicates into our language we do not thereby introduce further predicates beyond what is required by (i.e. definable in) the expansion. In this case, one clearly obtains the smallest such system any time one introduces a new truth predicate.

For property theories, we can simply note that the generation of properties is (by construction) limited to entities definable in a specific way. The new properties available at each additional stage are those whose application relation is definable over previous stages. One can see this as constructing a class-theoretic version of the constructible hierarchy, call it the $L(V)$ hierarchy, generating more and more classes by iterating the definability operator.[23]

For the truth-theoretic postulationist, it is part of her view that no more than is necessary be introduced to comply with a given rule.[24] It

[23]The reader should be warned, however, that there are technical hurdles in formulating this. The existence of the class $L_\Gamma(A)$ for every well-order Γ and every class A is equivalent to ETR—see [Williams, 2018, Theorem 3.14]. So if one wants to formally talk about levels of $L(V)$ in full generality one must already presuppose ETR. Nevertheless, it is a helpful intuition to keep in mind.

[24]See here [Fine, 2006], p. 93.

is thus reasonable that the potentialist system obtained for the various kinds of truth-theoretic potentialism be the smallest such.

Larger systems often admit branching. A critical point, in contradistinction to the foregoing, is that for larger systems we *do* get branching. If a certain kind of richness is needed or wanted by the bottom-up theorist, and in particular if they wish to transcend smallest systems, we often can get branching in those systems. If one thinks that branching is a cost (say because it indicates a kind of non-inevitability in how the classes unfold), then such a richness assumption seems like a dangerous desideratum.

The larger systems we considered can be seen as larger in *width*, not larger in *height*. Whilst our intention here is to bring to mind the familiar width versus height distinction for sets, the notion is different here, since all classes have the same height in the sense of ordinal rank. Here, height refers to the length of truth predicates, which one can think of as corresponding to how far one can build $L(V)$ in the classes—cf. earlier discussion in this section. The wider systems we considered were those that allowed the addition of generic global well-orders (equivalently, Cohen-generic classes of ordinals). Genericity ensures that adding these does not increase the height of a world.[25]

One might view these two observations (concerning smallest and larger systems) as a point in favour of the pictures articulated by the versions of bottom-up truth-theoretic potentialism we have considered here. If one views branching as a cost, one way to ensure branching is avoided is to consider smallest potentialist systems. As it turns out, this is precisely what the truth-theoretic versions of liberal predicativism, property theory, and postulationism motivate (the former two since they just involve adding truth predicates and closing under definability, and the latter because we add truth predicates in such a way that no more than is necessary is added). Thus for these views there is conformity between desirable properties of the potentialist system and the details of what the relevant philosophical view motivates. As we shall see, however, allowing class-forcing greatly complicates the issue.

Global Choice and class-forcing are problematic. A theme in some of our results is that Global Choice is problematic (or at least raises several questions) in the class-theoretic potentialist context. One possibility is that we could require the global well-order to be there from

[25] If you drop the requirement for genericity then global well-orders can add height. Given $(M, \mathcal{X}) \models \mathsf{NBG}$ with a global well-order, you can find a global well-order in $\mathcal{X}$ which codes any given $A \in \mathcal{X}$, say by placing A on the even indices in the order. So you could add, over the definable classes, a global well-order which codes a very long truth predicate.

the start. For many of our potentialists (e.g. the predicativist and the property theorist) the base world contains just the definable classes. To insist then that the base world contains a global well-order is just equivalent to the base world having a first-order definable well-order of the universe. This has serious first-order consequences, in particular it is equivalent to $\exists x\ V = \mathrm{HOD}(\{x\})$. We might be suspicious of our class-theoretic commitments delivering such strong set-theoretic consequences, especially given that we are thinking of building the classes *after* (in the class-theoretic potentialist's modal sense) the universe of sets has been constructed.

Another possibility is that the global well-order is generic, in the sense of class forcing. (See the discussion in §4 of how to force to add a global well-order.) Under various motivations, a potentialist might not want a global well-order which codes complicated undefinable classes, and instead want it to be "random" with respect to the definable classes.[26] This amounts to asking it to be generic; extending by a generic global well-order is adding the well-order and, through the use of forcing-names, closing off under definability from the well-order and classes in the ground model.

But in this case, as we noted in Theorem 19, there is no *prima facie* guarantee that the generic not be a bad one which kills off the addition of truth predicates. One response to this predicament is to require worlds to satisfy a theory which ensures the existence of all desired truth predicates. For example, if our first world satisfies $\mathsf{NBG} + \mathsf{ETR}$ then all the truth predicates are already there, and so a bad 'truth-killing' well-order cannot also be there. This, however, incurs the cost that the truth-theoretic potentialism is essentially trivial—all paramaterised truth predicates are there from the get-go. This contravenes the basic set up of the liberal predicativist and property theorist (though it is unclear what the situation is for the postulationist) and so would necessitate some revision of these positions.[27]

If the global well-order is to be neither definable nor generic, then what is it to be? It would be overly hasty to claim those as the only two possibilities, but the other possibilities of which we know strike us

[26]For example, a Finean postulationist might have a rule saying "introduce a global well-order". But adding a global well-order shouldn't add more than necessary; cf. Footnote 25. If other mathematical objects are desired, then they should have their own introduction rules.

[27]Moreover, there is the worry that the killing truth phenomenon can be generalised higher, to prevent expansions in worlds satisfying stronger theories. We leave fleshing out this worry to future work.

as artificial.[28] The explication of further possibilities would be a useful contribution to our understanding of classes, and we leave it to further work.

Some of the participants to the debate may have philosophical objections against the use non-definable well-orders. For example, a non-definable well-order of the sets does not *seem* to mesh particularly well with predicativist intuitions. A couple of points are relevant here.

First: Global well-orders have various applications in second-order set theory. Two examples are:

(1) In the study of determinacy of class games, it plays an essential role in moving from quasi-strategies to strategies. For example, the equivalence of ETR and clopen class determinacy [Gitman and Hamkins, 2016] requires Global Choice.
(2) The standard arguments to prove some properties of the surreal numbers, such as them forming a universal ordered field, go through Global Choice.

These examples (and others), one might think, suggest a constraint on what we want out of a potentialist class theory. The ability to have some worlds where we have the required classes to nicely interpret the reasoning of set theorists is, *ceteris paribus*, a plus. The specific case of liberalised predicativism provides an example here—part of what is at issue for them is to provide an account of classes that is predicativist in spirit, but nonetheless yields enough strength to be able to interpret parts of set theory that use non-definable classes. Fujimoto, for example, writes:

> 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. ([Fujimoto, 2019], p. 217)

Given the applications of global choice then, a class-theoretic potentialist who is more liberal in spirit should accept the possible existence of non-definable well-orders.

Second: The existence of a global well-order is an assertion about the existence of a single class and does not ascribe any complicated global structure to classes as a whole. In this respect, the existence of

[28]For example, see the sketch in Footnote 25.

a global well-order is relatively innocent as far as *liberal* predicativism is concerned. Consider the following remark:

> Liberal predicativism does well in most cases, when it comes to an axiom asserting something about a specific type of class... 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. ([Fujimoto, 2019], p. 225)

Given this, there is at least *prima facie* reason for the liberal predicativist to accept their possible existence. Similar remarks apply to the addition of any class-forcing generic (that is otherwise acceptable to the class-forcing potentialist).

The postulationist also has reason to accept the possible existence of generics, including non-definable well-orders. Since the properties of the forcing (e.g. non-atomicity), denseness, and genericity are definable in NBG we can easily formulate the postulationist condition required to introduce generics (including global well-orders) and the same is clearly true for truth predicates.

The situation is different for the property theorist. As noted earlier, the classes obtained by the property theorist are essentially those obtained by building L over V (i.e. $L(V)$). But given this, whilst some generics over our initial model are obtainable we will never get a generic that can *conflict* with a truth predicate and *arbitrary* generics are prohibited (only those that cant be obtained in some $L_\alpha(V)$ are legitimate). Thus, the property theorist rules out non-convergence by keeping a strict control on the classes that could exist. (One might, of course, view this as a cost—especially if one wants to enforce as few restrictions as possible on the classes that one is allowed to form.)

What emerges from this discussion is that there is the following tension at the heart of of bottom-up approaches. If we (i) regard non-convergence as a cost, (ii) want to allow the addition of truth predicates, and (iii) wish to allow unrestricted addition of generics, then we have a problem. The property theorist resolves this issue by rejecting (iii). This problem bites for the predicativist, and we will suggest a solution via additional modal principles (rejecting (iii)) in §8.1 (as it turns out, this strategy can also be appealed to by the postulationist). The postulationist can also 'resolve' the issue by accepting ETR as holding of the initial world, thereby trivialising (ii). We leave assessment of these options to interested readers and future work. For now, we move on to top-down approaches.

6. The top-down approach

In this section we discuss what the mathematical results from §4 say about top-down approaches to class potentialism, and in particular the interplay between what is satisfied on these pictures, **Referential Indeterminacy**, and **Interrelation of Interpretations**. First let's make clear just what assumptions are needed for the formal results to apply.

Recall that the formalisation for top-down approaches is to fix a countable model $M \models \mathsf{ZFC}$ and consider the potentialist system consisting of all expansions of M to a model of T, where T is a class theory. This potentialist system is meant to reflect the properties of the potentialist system of classes over the true universe V. (For the set multiversists: *a* universe V.) Results from §4 tell us something about this potentialist system, making some assumptions about T and M. Two main tools were used in §4: truth predicates, and class forcing. We need both to be applicable.

Let's discuss truth predicates first. Proposition 13 tells us that asking to have any world with a truth predicate puts a limitation on the choice of M. This limitation has negligible cost; it amounts to requiring that M satisfy a certain (second-order) reflection principle, and such principles are commonly taken to give basic properties of the universe of sets. Accordingly, it has negligible cost to assume V satisfies these properties.

More substantively, these tools do not apply to any choice of T. The results in § 4 were stated in terms of iterated truth predicates. There is a limit to how far this generalises. If T outright proves the existence of iterated truth predicates of any length—that is, if T proves ETR—one cannot have a nontrivial truth potentialist-like system whose worlds are models of T.

Let's now discuss forcing. This puts a limit on the worlds—we need that generics exist—which we handled by the simplifying assumption that worlds are countable. We assume this technology reflects upward to apply also to the true universe V; see §7 for further discussion of this move. It also puts restrictions on T. As discussed in §4 adding a Cohen-generic class of ordinals adds a global choice function. So these results do not speak to the class theorist who holds global choice is definitely false at every world, if any such exist.

This also rules out the inclusion of axioms that limit the classes by definability. Here's an illustrative toy example. Let T^- be NBG plus the assertion that length n iterated truth predicates exist for any finite n. In T^- we can express "every class is definable from Tr_n for some finite n". Call T the theory you get by adding this assertion to T^-. Then nontrivial forcing destroys T, and so the results in §4 do not apply to T. Indeed, T is categorical over a fixed transitive M: given a transitive model M of ZFC there is at most one T-expansion for M.

Thus, if we are considering some class theory T over the universe V such that (i) T does not require the classes to be thin (in the sense of the previous paragraph), (ii) T does not prove that arbitrary iterated truth predicates exist, and (iii) T does not have a principle implying global choice is false, then Theorems 18 and 19 both give information about the top down potentialist system for T on M. Namely, Theorem 18 gives failures of .3 in this system, and Theorem 19 gives failures of .2 in this potentialist system.

For such a potentialist, these results illuminate in a concrete way how the indeterminacy of reference underlying top-down potentialism might manifest. Exactly as in §5, weak base theories (like NBG) seem to correspond to a conception of class that is radically divergent, it is just that in this context this radical divergence is underwritten by **Referential Indeterminacy** and **Interrelation of Interpretations** rather than what is built up from **Initial World** and **Individuation**.

This observation is interesting for both our motivations for the top-down approach. Let's examine the set-theoretic multiversist first. Recall that she regards talk using some class theory T as just more set-theoretic mathematics up for reinterpretation. Fixing some appropriate M in the multiverse, each M' with $M \in M'$ will have a conception of what the T-class-potentialist system over M looks like (for some reasonable T). What our results show is that for $T =$ NBG (and appropriate M) we get failures of the .2 and .3 axioms, indicative of strong branching.

Some multiversists (e.g. [Hamkins, 2018a]) distinguish between the extreme branching of S4 and the 'inevitability' of S4.3 as well as the 'convergence' implied by S4.2, arguing that the mirroring theorem for systems containing S4.2 represents a quasi-actualist picture of the universe of sets—whilst one could use the modal theory, practically speaking nothing hangs on this since the mirroring theorem guarantees that a modal-free theory can be used. For a view to be *strongly potentialist*, one might think, non-convergent branching possibilities are required (and for this non-convergence to show up in the modal validities). There are many such varieties of multiversist-inspired set-potentialism on offer; rank-extension potentialism (discussed earlier) is one where the modal validities are S4. Critically though, known examples of reasonable kinds of potentialism with branching possibilities are limited to non-well-founded models of set theory, potentialisms with only transitive models are generally S4.2 or stronger (e.g. both forcing and countable transitive model potentialism have S4.2 as their modal validities[29]). If one then accepts (as many do) that we have an absolute understanding of well-foundedness, and that intended set-theoretic universes are all transitive/well-founded, then one seems to be (currently) linked to 'quasi-actualist' potentialist systems satisfying S4.2 and a mirroring theorem. Not so for the multiversist-inspired class-theoretic potentialist. Here, whilst we have not calculated the modal validities to be *exactly* S4, for the NBG-class-potentialist system we do have failures of .2 and .3 and hence no mirroring theorem. Moreover, these results still go through if we insist that the worlds of our potentialist systems all be transitive (indeed we can even restrict to β-models if so desired). These potentialist systems are the first to our knowledge systems of set theory to exhibit non-convergent branching even when we restrict to well-founded models.[30]

For the theorist who holds that plural quantification is indeterminate, the situation is subtle. On the one hand, plural logic (in its standard formulation) contains all impredicative instances of the plural comprehension scheme and indeed the Henkin interpretations for plural logic obtained by [Florio and Linnebo, 2016] all satisfy it (they restrict to what they call *faithful* models—those that satisfy every instance of the comprehension scheme). This can then be leveraged to provide an interpretation of MK class theory (as in [Uzquiano, 2003]). This interpretation can be carried through whether or not the range of the plural quantifiers is determinate, if it is indeterminate but nonetheless every legitimate interpretation satisfies the impredicative plural quantification scheme, this impredicativity extends immediately to obtain the impredicative class-theoretic comprehension scheme of MK within each world. As noted earlier, MK violates the presuppositions required to make our arguments go through since it trivialises truth-theoretic potentialism by implying the existence of arbitrarily iterated truth predicates. Thus, our results do not have much to tell the advocate of this kind of top-down class-potentialism. To say more, further results are needed about the MK-class-potentialist system, and we leave this as an open question; see the discussion around Question 21 for fuller details.

Though one might hold that the indeterminate plural interpretation yields MK on the basis of the 'standard' conception of the logic,

[29]See here [Hamkins and Löwe, 2008] and [Hamkins and Linnebo, 2018].

[30]Note that such occur in the context of second-order arithmetic; see §8.2.

this is far from controversial. As [Florio and Linnebo, 2016] note, often the impredicative plural comprehension scheme is motivated by the assumption that every non-full Henkin semantics for the plural quantifiers is unintended. For example Hossack writes the following regarding the determinateness of plural quantification:

> Plural set theory has no non-standard[31] models, so the indeterminacy problem does not arise for pluralism. ([Hossack, 2000], p. 440) [32]

Such an assumption leads immediately to the impredicative comprehension scheme for plurals (and hence classes) as Uzquiano notes (referring to §3.2 of [Lewis, 1991]):

> To the extent to which one accepts unrestricted plural quantification over sets as unproblematic, one will be moved by what David Lewis refers to as the evident triviality of plural comprehension, and thus one will accept all instances of plural comprehension as true. After all, one may explain, in order for an instance of comprehension to be false, there must be a formula φ such that it is neither the case that no sets satisfy it nor is it the case that some sets satisfy it. But this could never be the case. ([Uzquiano, 2003], pp. 76–77)

Of course, if we allow plural quantification to be indeterminate then we have an immediate response—an instance of a formula φ in the plural comprehension schema might be neither true nor false of some sets in virtue of there being some interpretations of the plural variables on which it is true, and other interpretations on which it is false. For example, consider the following sentence:

$$\varphi(x) = \text{“}x = x \text{ and there exists a truth predicate for } V\text{”}$$

If we do not assume that quantification is determinate and have doubts about impredicative comprehension (and so adopt NBG) then it is neither the case that no sets satisfy φ nor is it the case that some sets satisfy φ—in some worlds φ picks out V and in others it picks out $\emptyset$.[33] Throwing in impredicative plural quantification at the start simply

[31]Hossack means non-standard in the sense of non-full Henkin semantics, not in the sense of the models being ill-founded.

[32]Examples of this kind can be multiplied. For a review of the literature see [Florio and Linnebo, 2016].

[33]This example generalises beyond just NBG. E.g. if MK is our base theory then instead of asking whether truth exists we should ask whether there is a class coding an MK-extension for V, an assertion not decided by MK.

prejudices the debate in favour of MK, and the water becomes much muddier once we allow many different interpretations of the range of the plural variables.

Nonetheless, the view that Lewis' thought about pluralities is somehow *part of our conception of pluralities* is tempting. Even if we think the quantification is indeterminate, we might think that *within a world* his intuition should motivate us to accept impredicative plural quantification over that world, yielding MK locally. The thought then that this reasoning should apply schematically to *every* world—thereby trivialising truth-theoretic potentialism—merits further scrutiny. Here is not the place to adjudicate these difficult issues concerning the relationships between the philosophies of plural quantification and mathematics. However, these observations point to a substantial philosophical issue: There is a critical choice point in the selection of theory for the believer in the indeterminateness of plural quantification. Acceptance of the impredicative plural comprehension scheme despite indeterminacy has immediate mathematical ramifications, not just regarding what non-modal statements of class theory are supervaluationally valid but also the nature of the relevant potentialist systems corresponding to their plural talk regarding the sets.

7. Responding to an objection: A remark on the use of countable transitive models in studying potentialism

One tempting way of objecting to the import of our results is to point to our use of countable transitive models. For example, Theorem 19 depended on adding a generic that, once we introduce the relevant truth predicate, encodes a cofinal sequence in the ordinals. In getting this generic, however, we assumed that the ground model is countable—we view the model externally as countable and talk about the ways bits can be encoded into a particular countable sequence. One might object: For many species of class-theoretic potentialist V is uncountable, and so there is no such generic.

There are some points to be made about this objection. First, accepting this response entails you accept that apparently perfectly good parts of model theory cannot tell you about the multiverse proper (or at least their use must be justified). This in itself, is a substantial cost (without further argument) and goes against much of the practice in the field. For example the Mostowski result ([Mostowski, 1976]) that there are incompatible generics over any countable transitive model is standardly taken to show that the generic multiverse contains universes with incompatible reals, despite the dependence on countability of the models in the proof. Those who seek amalgamation within the generic multiverse generally do so via other considerations, e.g. placing a requirement on no information loss (e.g. [Steel, 2014]) or a desire for axiomatisability ([Maddy and Meadows, 2020]), rather than objecting to the use of countable models as providing the relevant the model theory.

This plays out in one's attitude to the role of the completeness theorem.[34] A key desideratum when concerned with first-order set-theoretic mathematics is being able to move seamlessly between provability and satisfaction in all set-sized first-order models. If one wishes to provide a first-order axiomatisation for your multiverse, one needs it to be the case that if φ is true in all set-sized class-theoretic multiverses, then it's provable in the multiverse theory. If there is a set model of a class-theoretic potentialism in which truth predicates are killed off, then by completeness one cannot rule out this prospect. In some sense, when studying *any* kind of potentialism that allows for the addition of classes (and hence for the range of the second-order variables to change) countable transitive models are *the* natural place in which to conduct the model theory, since this is the place where extensions of the required kind are always uncontroversially available. To deny legitimate constructions in this context is to take on a substantial explanatory burden: One must come up with principled reasons as to which extensions are legitimate and why.

Consider the following analogy with the case of the *set* forcing potentialist. She says: "The modal logic of forcing in $\mathsf{S4.2}$" whilst brandishing a copy of [Hamkins and Löwe, 2008]. Along comes a character we'll call the facetious potentialist who says: "I understand that you have a nice model for this in terms of countable transitive models, but *really* the modal logic of forcing is $\mathsf{S5}$, since there is just one non-extendible universe and all forcing is trivial." In this context, we'd rightly say that the facetious potentialist was effectively denying important potentialist principles about the plenitude of extensions. If one wants to make such a proposal, one has to come up with good *reasons* and *criteria* for saying when one extension is an acceptable extension of another and an alternative *model theory* that can be used in studying the potentialism. We do not wish to argue that such an argument is impossible, and indeed would welcome a response on behalf of the class-theoretic potentialist on this issue. It seems to us, however, that no easy solution is available.

[34]We thank Sam Roberts for this suggestion.

8. Conclusions and further directions

In this paper we've argued that there are natural interpretations of class talk over a fixed domain of sets that yield *potentialisms* of different kinds. We've also argued that it makes sense to divide these pictures into two kinds: bottom-up approaches begin with some fixed stock of classes and then individuate new ones, and top-down approaches take the modal variation of classes to arise from referential indeterminacy and the ways possible sharpenings of the ranges of the class variables relate to one another. We've proved several results about potentialist systems, in particular exhibiting failures of the .2 and .3 axioms for potentialist systems corresponding to weak theories of classes. Finally, we've discussed some philosophical payoffs of these results for the various bottom-up and top-down approaches.

This is very far from the end of the story, however, and we hope to have merely made a first-step in discussions about class-theoretic potentialism and possible responses to the challenges we have outlined. For this reason, we raise several open questions that may be of interest to others wishing to pursue this line of research. This will make the conclusion slightly longer than is usual, but we feel that identifying salient problems is important and hope that the reader will indulge us.

8.1. **Additional modal principles?**

As discussed in §6, the failure of .2 for top-down potentialism for weak class theories is particularly destructive, being witnessed by a world which cannot be further extended to add in a certain truth predicate. A top-down potentialist may very well think this catastrophic world is an artifact of the formalisation, one which does not occur in the real multiverse of classes. Her task then is to explain why this phenomenon does not occur and formulate principles prohibiting these worlds.

One approach to this latter problem is to provide additional *modal* axioms, going beyond just the resources of class theory. For instance, the following modal principle manifestly rules out the killing truth phenomenon:

$$\Box \forall X \Diamond \exists Y (\text{“}Y \text{ is a truth predicate for } X\text{”}).$$

It is easy to formulate versions of this for iterated truth predicates. And one could consider yet more modal principles to express properties of the true multiverse of classes.

Examples of this kind already exist in the case of the set-forcing potentialist. For example, *maximality principles*, assertions of the form $\Diamond\Box\varphi \Rightarrow \varphi$, have been considered in the context of set forcing potentialism; see e.g. [Hamkins, 2003] and [Hamkins and Linnebo, 2018].

An example of a different flavor, one closer in motivation to what we give here, can be found in [Steel, 2014] (with subsequent development by [Maddy and Meadows, 2020]). Steel is investigating a multiversist framework arising from set forcing. Given a countable model of set theory, it has pairs of Cohen extensions which do not *amalgamate*—there is no outer model which contains both Cohen extensions as submodels [Mostowski, 1976]. To exclude this phenomenon, Steel includes an axiom asserting that models in the generic multiverse always amalgamate. His principle is in fact higher order, referring to worlds as objects, not just to what is true of sets within each world. And one could also consider higher order principles in the context of class potentialism.

We leave the consideration of these higher order or modal principles to future work.

8.2. **An analogy to second-order arithmetic, and universal finite sequences?** A potential area for further study concerns the analogy between the use classes in the contexts of second-order arithmetic and set theory. Predicativism in mathematics often takes the totality of natural numbers as given, with the predicatively-given "classes" then being sets of natural numbers, e.g. [Feferman and Hellman, 1995, Hellman and Feferman, 2000]. There has been work addressing to what extent results about predicativism over ω generalise to predicativism over V—see e.g. [Fujimoto, 2012, Sato, 2014]. Similar to how it was formalised in the set theoretic context, one could formalise potentialism over ω by considering potentialist systems of ω-models of second-order arithmetic. To what extent does the mathematical and philosophical work about class potentialism carry over to the arithmetic context?

We also wish to mention a question arising from the analogy going in the other direction. Here, Z_2 is the theory of second-order arithmetic with full impredicative comprehension.

Theorem 20 (Hamkins–Williams). *Let T be a computably axiomatizable extension of $\mathsf{Z}_2 + \Pi^1_\infty$-Bounding. Then the modal validities of the potentialist system consisting of countable ω-models of T are precisely* $\mathsf{S4}$*, whether or not we allow parameters in formulae.*

Proof sketch. It is well-known that $\mathsf{Z}_2 + \Pi^1_\infty$-Bounding is bi-interpretable with $\mathsf{ZF}^- + V = H_{\omega_1}$, the assertion that every set is hereditarily countable. Given $(\omega, \mathcal{X})$ a model of arithmetic call the corresponding model of ZF^- its *companion model*. Such companion models must be well founded beyond ω, and if $(\omega, \mathcal{X})$ is a submodel of $(\omega, \mathcal{Y})$ then the companion model of $(\omega, \mathcal{X})$ is end-extended by the companion model of $(\omega, \mathcal{Y})$. Consider now the potentialist system consisting of these countable, ω-standard companion models, ordered by end-extension. Up to coding this is the same as considering the potentialist system of countable ω-models of T directly. An instance of [Hamkins and Williams, 2021, Theorem 6] yields that this potentialist system admits a universal finite sequence and thus its modal validities are precisely $\mathsf{S4}$.[35] □

Does this theorem generalise to the set theoretic context? More precisely:

Question 21. *Let T be* MK *plus Class Bounding and let M be a countable transitive model of* ZFC *which has a nontrivial T-class potentialist system. Does the potentialist system consisting of countable T-expansions for M have* S4 *as its modal validities?*

A positive answer to this question would imply that the fundamental branching phenomenon for top-down potentialism for weak theories also occurs for very strong theories.

References

[Antos et al., 2021] Antos, C., Barton, N., and Friedman, S.-D. (2021). Universism and extensions of V. *The Review of Symbolic Logic*, 14(1):112154.

[Arrigoni and Friedman, 2013] Arrigoni, T. and Friedman, S.-D. (2013). The Hyperuniverse Program. *Bulletin of Symbolic Logic*, 19:77–96.

[Barton, F] Barton, N. (F). Indeterminateness and 'the' universe of sets: Multiversism, potentialism, and pluralism. In Daghighi, A. S., Fitting, M., Gabbay, D., Pourmahdian, M., , and Rezus, A., editors, *Research Trends in Contemporary Logic*, Landscapes in Logic. College Publications. Forthcoming, Preprint: `https://philpapers.org/rec/BARIAT-15`.

[Barton, S] Barton, N. (S). Are large cardinal axioms restrictive? Manuscript under review. Pre-Print: `https://philpapers.org/archive/BARALC-5.pdf`.

[Ewald, 1996] Ewald, W. B., editor (1996). *From Kant to Hilbert. A Source Book in the Foundations of Mathematics*, volume I. Oxford University Press.

[Feferman and Hellman, 1995] Feferman, S. and Hellman, G. (1995). Predicative foundations of arithmetic. *Journal of Philosophical Logic*, 24(1):1–17.

[Fine, 2005a] Fine, K. (2005a). Class and membership. *Journal of Philosophy*, 102(11):547–572.

[Fine, 2005b] Fine, K. (2005b). Our knowledge of mathematical objects. In Gendler, T. Z. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, pages 89–109. Clarendon Press.

[Fine, 2006] Fine, K. (2006). Relatively unrestricted quantification. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 20–44. Oxford University Press.

[35]Theorem 6 is phrased in a general form. See the discussion at the beginning of §4 for why it applies to extensions of ZF^-.

[Florio and Linnebo, 2016] Florio, S. and Linnebo, Ø. (2016). On the innocence and determinacy of plural quantification. *Noûs*, 50(3):565–583.

[Friedman, 2006] Friedman, S.-D. (2006). Internal consistency and the inner model hypothesis. *Bulletin of Symbolic Logic*, 12(4):591–600.

[Fujimoto, 2012] Fujimoto, K. (2012). Classes and truths in set theory. *Annals of Pure & Applied Logic*, 163(11):1484–1523.

[Fujimoto, 2019] Fujimoto, K. (2019). Predicativism about classes. *Journal of Philosophy*, 116(4):206–229.

[Gitman and Hamkins, 2016] Gitman, V. and Hamkins, J. D. (2016). Open determinacy for class games. In Caicedo, A. E., Cummings, J., Koellner, P., and Larson, P., editors, *Foundations of Mathematics, Logic at Harvard, Essays in Honor of Hugh Woodin's 60th Birthday*, AMS Contemporary Mathematics. Newton Institute preprint ni15064.

[Gitman et al., 2019] Gitman, V., Hamkins, J. D., and Karagila, A. (2019). Kelley-morse set theory does not prove the class Fodor theorem. *ArXiv e-prints*. Manuscript under review.

[Hamkins et al., 2012] Hamkins, J., Kirmayer, G., and Perlmutter, N. (2012). Generalizations of the Kunen inconsistency. *Annals of Pure and Applied Logic*, 163(12):1972–1890.

[Hamkins, 2003] Hamkins, J. D. (2003). A simple maximality principle. *The Journal of Symbolic Logic*, 68(2):527–550.

[Hamkins, 2012] Hamkins, J. D. (2012). The set-theoretic multiverse. *The Review of Symbolic Logic*, 5(3):416–449.

[Hamkins, 2018a] Hamkins, J. D. (2018a). The modal logic of arithmetic potentialism and the universal algorithm.

[Hamkins, 2018b] Hamkins, J. D. (2018b). The modal logic of arithmetic potentialism and the universal algorithm. *ArXiv e-prints*, pages 1–35. Under review.

[Hamkins et al., 2015] Hamkins, J. D., Leibman, G., and Lwe, B. (2015). Structural connections between a forcing class and its modal logic. *Israel Journal of Mathematics*, 207(2):617–651.

[Hamkins and Linnebo, 2018] Hamkins, J. D. and Linnebo, O. (2018). The modal logic of set-theoretic potentialism and the potentialist maximality principles. *to appear in Review of Symbolic Logic*.

[Hamkins and Löwe, 2008] Hamkins, J. D. and Löwe, B. (2008). The modal logic of forcing. *Transactions of the American Mathematical Society*, 360(4):1793–1817.

[Hamkins and Williams, 2021] Hamkins, J. D. and Williams, K. J. (2021). The Σ_1-definable universal finite sequence. *The Journal of Symbolic Logic*, page 119.

[Hamkins and Woodin, 2017] Hamkins, J. D. and Woodin, W. H. (2017). The universal finite set. *ArXiv e-prints*, pages 1–16. manuscript under review.

[Hamkins and Woodin, 2018] Hamkins, J. D. and Woodin, W. H. (2018). Open class determinacy is preserved by forcing. *ArXiv e-prints*, pages 1–14. Under review.

[Hellman and Feferman, 2000] Hellman, G. and Feferman, S. (2000). *Challenges to Predicative Foundations of Arithmetic*, pages 317–338. Cambridge University Press.

[Hossack, 2000] Hossack, K. (2000). Plurals and complexes. *The British Journal for the Philosophy of Science*, 51(3):411–443.

[Krajewski, 1976] Krajewski, S. (1976). Non-standard satisfaction classes. In *Set theory and hierarchy theory (Proc. Second Conf., Bierutowice, 1975)*, pages 121–144. Lecture Notes in Math., Vol. 537. Springer, Berlin.

[Kunen, 1971] Kunen, K. (1971). Elementary embeddings and infinitary combinatorics. *The Journal of Symbolic Logic*, 36:407–413.

[Lewis, 1991] Lewis, D. (1991). *Parts of Classes*. Blackwell.

[Linnebo, 2006] Linnebo, Ø. (2006). Sets, properties, and unrestricted quantification. In *Absolute Generality*. Oxford University Press.

[Linnebo, 2013] Linnebo, Ø. (2013). The potential hierarchy of sets. *The Review of Symbolic Logic*, 6(2):205–228.

[Maddy and Meadows, 2020] Maddy, P. and Meadows, T. (2020). A reconstruction of Steels multiverse project. *The Bulletin of Symbolic Logic*, page 179.

[Marek and Mostowski, 1975] Marek, W. and Mostowski, A. (1975). *On extendability of models of ZF set theory to the models of Kelley-Morse theory of classes*, pages 460–542. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Mostowski, 1976] Mostowski, A. (1976). A Remark on Models of the Gödel-Bernays Axioms for Set Theory. In Müller, G. H., editor, *Sets and Classes: On the Work by Paul Bernays*, volume 84 of *Studies in Logic and the Foundations of Math.*, pages 325–340. North-Holland, Amsterdam. DOI: 10.1016/S0049-237X(09)70288-5.

[Parsons, 1974] Parsons, C. (1974). Sets and classes. *Nous*, 8(1):1–12.

[Roberts, Ua] Roberts, S. (Ua). The iterative conception of properties. unpublished manuscript.

[Roberts, Ub] Roberts, S. (Ub). Properties and truth. unpublished manuscript.

[Sato, 2014] Sato, K. (2014). Relative predicativity and dependent recursion in second-order set theory and higher-order theories. *Journal of Symbolic Logic*, 79(3):712–732.

[Scambler, 2021] Scambler, C. (2021). Can all things be counted? *Journal of Philosophical Logic*.

[Steel, 2014] Steel, J. (2014). Gödel's program. In Kennedy, J., editor, *Interpreting Gödel*. Cambridge University Press.

[Uzquiano, 2003] Uzquiano, G. (2003). Plural quantification and classes. *Philosophia Mathematica*, 11(1):67–81.

[Vickers and Welch, 2001] Vickers, J. and Welch, P. (2001). On elementary embeddings from an inner model to the universe. *The Journal of Symbolic Logic*, 66(3):1090–1116.

[Welch and Horsten, 2016] Welch, P. and Horsten, L. (2016). Reflecting on absolute infinity. *Journal of Philosophy*, 113(2):89–111.

[Williams, 2018] Williams, K. J. (2018). *The structure of models of second-order set theory*. PhD thesis, The Graduate Center, CUNY.

[Williams, 2019] Williams, K. J. (2019). Minimum models of second-order set theories. *Journal of Symbolic Logic*, 8(2):589–620.

[Zermelo, 1930] Zermelo, E. (1930). On boundary numbers and domains of sets. In [Ewald, 1996], volume 2, pages 1208–1233. Oxford University Press.

(Neil Barton) Fachbereich Philosophie, University of Konstanz, Universitätsstrasse. 10, 78464, Konstanz, Germany
E-mail address: neil.barton@uni-konstanz.de
URL: https://neilbarton.net/

(Kameryn J. Williams) Sam Houston State University, Department of Mathematics and Statistics, Box 2206, Huntsville, TX 77341-2206, USA
E-mail address: kjw059@shsu.edu
URL: http://kamerynjw.net