

# VARIETIES OF CLASS-THEORETIC POTENTIALISM

NEIL BARTON 

IFIKK, UNIVERSITETET I OSLO  
and

KAMERYN J. WILLIAMS  
SAM HOUSTON STATE UNIVERSITY

**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 potentialists.

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

**Set-Theoretic Actualism:** There is a maximal completed universe of sets.

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.<sup>1</sup> 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.<sup>2</sup>

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, 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 [53]) view the worlds as the study of ever larger  $V_\kappa$  for  $\kappa$  inaccessible, with accessibility being coextensive with the subset relation. Another species of set-theoretic potentialism 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

---

Received: November 10, 2021.

2020 *Mathematics Subject Classification*: Primary 00A30, 03A05, 03E70.

*Keywords*: class theory, potentialism, truth predicates.

<sup>1</sup> For the arguments of this paper, nothing much hangs on the matter.

<sup>2</sup> This statement closely mirrors [25, p. 1]



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 [26] and the study of potentialist maximality principles [19, 25]. 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 S4 in the relevant potentialist systems.<sup>3</sup> The modal axiom .3 indicates a kind of ‘inevitability’ or ‘linearity’ to how the worlds unfold, and .2 indicates a form of ‘convergence’ present on the frame. Moreover mirroring theorems (which allow us to move between potentialist and non-potentialist theories via a natural translation) are known to hold on systems containing S4.2. This has led some authors (e.g., [22, p. 33]) to claim that the convergent forms of potentialism (i.e., those with modal logic at least 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 *class theory* (also called *second-order set theory*). These theories introduce a new kind of variable to range over classes as well as sets, which are then governed by theories such as GB, MK, and their variants. 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, [49]). However, recent literature has investigated an entire space of possible class theories with varying consequences and consistency strengths, analogous to the study of subsystems of second-order arithmetic. 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 [16]).

These developments suggest that the various kinds of set-theoretic potentialisms 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.

One need not be a set-theoretic potentialist 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’ll look at varieties of class-theoretic potentialisms under both set-theoretic actualism and set-theoretic potentialism. 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) Class potentialist systems can exhibit a variety of modal structures, with some validating S4.3 and S4.2, while others will fail to validate .2 or .3.
- (3) The various motivations for class-theoretic potentialism suggest different ways of modelling the position and distinct pictures of the modal behaviour of classes.

Here’s the plan: In Section 2 we’ll outline some philosophical positions regarding classes that can be used to motivate class-theoretic potentialism. We’ll divide these into

<sup>3</sup> It should be noted that S4 trivially holds in any potentialist system. See Section 4, which also has definitions of all these modal axioms.

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. Sections 3 and 4 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. Section 5 proves some results about some potentialist systems, in particular showing that for weak theories below the level of  $\text{GB} + \text{ETR}$ , given suitable assumptions we can exhibit failures of the .3 (Theorem 17) and .2 axioms (Theorem 21), showing that some systems have non-inevitability and radical branching. We'll then discuss some implications for bottom-up approaches to classes (Section 6), arguing that whilst there are good motivations for handling truth predicates, global choice is problematic in this context, and care is required in setting up the 'right' model theory. We then (Section 7) examine top-down approaches, arguing that our results are indicative of more natural cases of radical branching than is normally seen in many potentialist contexts. Section 8 provides some concluding remarks and identifies several open questions and directions for future research.

**§2. Motivating class-theoretic potentialism.** A mathematically popular and expressively parsimonious conception of classes is to regard talk of classes as merely shorthand for certain formulas holding within the universe. So, for instance, ' $x \in \text{Ord}$ ' 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's [33] result that there is no nontrivial  $j : V \rightarrow V$  and the work by Vickers and Welch [50] on embeddings from inner models to the universe.<sup>4</sup> Another problem point is that some recent work in class theory, such as work on the strength of the class forcing theorem,<sup>5</sup> makes necessary use of undefinable classes such as truth predicates. For this reason, we'll set this interpretation to one side.

Instead, 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 or exhaustively survey the views that might support it. Rather we aim to show that it is a viable position (philosophically speaking) and fits

<sup>4</sup> Briefly: if a nontrivial  $j : V \rightarrow V$  is definable without parameters, then so is its critical point  $\text{crit } j$ , the least ordinal moved by  $j$ . But elementary embeddings must fix all sets definable without parameters, so we would get that  $j(\text{crit } j) = \text{crit } j$ , contrary to the definition of critical point. A small extra argument gets that we further cannot have  $j$  definable with (set) parameters, by looking at such  $j$  with minimal critical point (for details see [48]). A similar argument works for consideration of  $j : M \rightarrow V$  (as in [50]).

The point that Kunen's theorem appears to be trivialised when all embeddings are viewed as definable was communicated to us by Sam Roberts and is raised in [18], [12] (esp. Section III), and [2] (esp. Section 2.2). Of course there is much to say here, it is not as though the definabilist about classes *cannot* give a nontrivial interpretation of these theorems (for example, Kunen's theorem can be interpreted as showing that there is no  $j : V_{\lambda+2} \rightarrow V_{\lambda+2}$ ), but the point remains that a non-definabilist has a nice interpretation of the theorem that the definabilist lacks.

<sup>5</sup> See [15, 30, 31].

nically with a several different positions, and indeed is even tenable if one is a set-theoretic actualist. We'll divide these into *bottom-up* and *top-down* approaches. The former roughly correspond to processes of *individuation* and the latter to considerations regarding *indeterminacy of reference*. At this stage, we'll keep things relatively informal, before discussing the details of each proposal once the mathematical machinery is in place.

**2.1. 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.

*2.1.1. Liberal predicativism.* The first interpretation we'll look at is derived from the work of Parsons (e.g., [41]) with subsequent development by Fujimoto (e.g., [12]) and views classes as *predicate extensions*:

**Class Predicativism:** Classes are extensions of predicates.<sup>6</sup>

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 nontrivially extends it. In particular, if we think that we can always extend by a truth predicate for any given language, we will obtain an infinite hierarchy of languages. 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...[41, p. 8]

Similar sentiments are available in Fujimoto, who argues that predicates can be 'admissibly introduced' and agrees with Parsons that there is no determinate totality of admissible predicates [12, pp. 212, 214–215].

Both Parsons' and Fujimoto's words suggest potentialist readings. However, they each opt for non-modal theories. Parsons uses a theory of classes and satisfaction and Fujimoto adopts GB augmented with a class-theoretic principle satisfying a version of the KF truth-theoretic axioms.<sup>7</sup>

Given that both Parsons and Fujimoto think there is no definite collection of all predicates, it is natural 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 by adding predicates for them into our language. The classes for them are thus 'irremediably potential', much like the stock of predicates we may admissibly introduce.

<sup>6</sup> We should note that this liberalised sense of predicativism contrasts somewhat with the usual sense of the term as developed historically by Poincaré, Russell, and Weyl. Thanks to Laura Crosilla for pointing out this issue.

<sup>7</sup> See [12, p. 223] for details.

2.1.2. *Property Potentialism*.<sup>8</sup> A somewhat similar approach is to view classes as arising from *property application* relation. This kind of view, when taken in unrestricted generality, leads quickly to semantic paradoxes via the consideration of a property  $P$  that applies to all and only the non-self-applicable things ( $P$  is then self-applicable iff it is not self-applicable, exactly as in Russell's Paradox or the paradox of heterologicality). A restriction is therefore needed, and one suggestion, made by both Linnebo [35] and Fine [9], is that the application relation for properties is successively individuated. Linnebo individuates properties along the ordinals, and thereby uses a length  $\text{Ord}$  iteration, while Fine considers longer iterations. This successive individuation 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.<sup>9</sup>

Whilst both Fine and Linnebo use a non-modal framework for 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 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 [42]. Philosophically speaking though, the way that the two constructions are phrased also suggest this interpretation. For example, Linnebo writes:

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. [35, p. 173]

Fine expresses a similar sentiment, in the context of contrasting his approach with the traditional cumulative hierarchy picture:

...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. [9, p. 547]

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

<sup>8</sup> 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.

<sup>9</sup> Whilst Fine [9] does not expressly use property-theoretic language he uses the same picture of successively specifying an application relation, and so we include them together. Linnebo's theory and Fine's theory are also similar from a mathematical perspective. The strength of these theories has recently been examined by Roberts (in [42]). These property theories are provably consistent in  $\text{GB} + \Pi_1^1\text{-Comprehension}$ . More strength can be obtained by the addition of a reflection principle for properties in this context. Moreover, some of these property theories are mutually interpretable with ZFC with the addition of  $\text{Ord}$ -iterated truth predicates (see [43]).

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.

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 view, not provide a comprehensive description of its possible motivations.

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.

We will save some further discussion of *Initial World* and *Individuation* for when we have some more mathematical tools in play. For now, let's move on to *top-down* approaches.

**2.2. Top-down approaches.** Instead of *starting* with some given collection of classes and iterating a process of individuation, we might instead view potentialism as 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.

**2.2.1. 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,<sup>10</sup> 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 forcing over  $V'$  to add a bijection  $V \rightarrow \omega$  (call this universe  $V'[G]$ ), via the usual collapse forcing to make a set countable. Within  $V'[G]$ , we can consider class-theoretic potentialist systems consisting of  $V$  equipped with different possible collections of classes (to be discussed in greater detail in Section 3).

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 the set-multiversist fixes some universe  $V$  as a starting position (within the multiverse), it is unlikely to be determinate exactly what classes exist over  $V$ . We can then construe reference to all classes of  $V$  as referentially indeterminate and yielding a potentialist system of classes.

**2.2.2. Plurals and potentialism.** Plural resources have been used to interpret proper class talk (see, for example, [49]). Often the ranges of plural variables are taken to be determinate (e.g., [32, 49]). However, this view has recently been challenged by the

<sup>10</sup> Views of these kinds include [1, 20, 46].

work of Florio and Linnebo (in [10]) who show that there are versions of Henkin semantics for plural logic, and argue that this calls into question the determinacy of plural quantification.

It is out of the scope of this paper to settle whether or not one should accept that plural resources are in fact indeterminate. Our focus here is on examining 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 motivate a class-theoretic potentialism via the plural interpretation of classes. If we hold that there is referential indeterminacy in the ranges of the plural variables, and we're interpreting classes as pluralities over our fixed universe, then the referential indeterminacy concerning plurals transfers immediately 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. Much as a *set-theoretic* multiversist can extract a *set-theoretic* potentialism from their multiversism,<sup>11</sup> so a theorist who is effectively a multiversist about the *pluralities* that exist (and hence a multiversist about *classes* too) can extract a *class-theoretic* potentialism from their kind of multiversism.

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 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 and what theories they satisfy).

**§3. 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 class-theoretic potentialist systems and their relation to bottom-up and top-down approaches.

**3.1. Class-theoretic principles.** We'll use a two-sorted approach to class theory, with *sets* and *classes* as the two types of objects.<sup>12</sup> A model of class theory will be denoted  $(M, \mathcal{X})$ , where  $M$  is the sets and  $\mathcal{X}$  is the classes. Our focus will be transitive models, for whom their membership relation is the true  $\in$ , and we will suppress the membership relation in the notation.<sup>13</sup>

<sup>11</sup> See the introduction of [25] for a brief discussion of the move from set-theoretic multiversism to set-theoretic potentialism.

<sup>12</sup> For the reader who wants a more thorough exposition of class theories than we give, we recommend [52, sect. 2] for a modern treatment.

<sup>13</sup> The reason for this restriction is most researchers in this area view nonstandard models as unintended, ersatz models. For the reader who does not share this view, we remark that some of our technical results about countable transitive models can be reworked to apply to nonstandard models. This requires more care and complicates the exposition, so we leave it to such a reader to check what generalises to that context.

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.<sup>14</sup> 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 Gödel–Bernays class theory GB.
- 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 sets. Equivalently, it may be formulated as the assertion of a global well-order. Let GBC denote GB + global choice.

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. See [8] for a classic treatment of this construction.

**DEFINITION 3.** Elementary transfinite recursion ETR is the principle asserting that transfinite recursions of elementary properties along well-founded classes have solutions. (See [14, sect. 4] for a more precise formulation.)

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 GB 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 [11]). *Over GB, ETR is equivalent to the assertion that for any class  $A$  and any class well-order  $\Gamma$  the length  $\Gamma$  iterated truth predicate relative to the parameter  $A$  exists.<sup>15</sup>*

We caution the reader that for this theorem to be true it must include class well-orders far longer than Ord. It's straightforward to define a well-order of ordertype Ord + 1, or Ord + Ord, or Ord<sup>2</sup>, following similar constructions on  $\omega$ . Stronger class theoretic principles allow one to define longer and longer well-orders, akin to how stronger subsystems of second-order arithmetic allow the definition of longer and longer countable well-orders.

One can restrict ETR to get a hierarchy of transfinite recursion principles. If  $\Gamma$  is a class well-order let ETR( $\Gamma$ ) denote the restriction of ETR to recursions along well-founded classes of rank  $\leq \Gamma$  and let ETR( $< \Gamma$ ) denote the restriction of ETR to rank

<sup>14</sup> Though these class axioms can have consequences for the sets, e.g., Con(ZFC).

<sup>15</sup> See the beginning of Section 5 for definitions and fuller discussion of truth predicates and iterated truth predicates.



$<\Gamma$ . These principles separate from full ETR and from each other based on  $\Gamma$ , according to consistency strength; see [52] for details.

The existence of models of any of these class theories is beyond ZF in consistency strength. But the large cardinal strength is very mild. All constructions we study are below the level of one inaccessible cardinal.

**3.2. Potentialist 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. Potentialist systems 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) will correspond to two different ways of studying potentialist systems mathematically. In both cases, we will consider the sets as a fixed totality, with the potential nature entirely in the classes.

For bottom-up approaches, the potentialist systems will be specified by rules which specify the base world and how worlds may be extended. We will look at extension rules based on adding truth predicates and certain class forcing extensions; we leave other extension rules to future work. We then ask which if any potentialist systems satisfy these rules. For example, we will show that the rules corresponding to liberal predicativism and property potentialism admit smallest systems satisfying those rules.

For top-down approaches, we consider all possible collections of classes which meet some basic criteria. For  $M$  a model of ZFC, say that a *top-down class potentialist system for  $M$*  is the collection of all worlds  $(M, \mathcal{X}) \models T$  satisfying a fixed property, where  $T$  is a class theory. For ease of writing we call such  $\mathcal{X}$  a  *$T$ -expansion for  $M$* . We will consider two choices for  $M$  and this fixed property, corresponding to different philosophical commitments. Corresponding to set multiversism we will look at countable transitive  $M$  and the top-down class potentialist system for  $M$  consisting of all countable  $(M, \mathcal{X}) \models T$ . Additionally, we will make extra assumptions on  $M$  to ensure there actually are  $T$ -expansions for  $M$  (cf. Proposition 12). Corresponding to set universalism we will look at  $M = V_\kappa$  for some inaccessible  $\kappa$  and the top-down class potentialist system for  $M$  consisting of all  $(M, \mathcal{X}) \models T$  with  $|\mathcal{X}| = \kappa$ . This restriction avoids the trivial case of allowing the world  $(M, \mathcal{P}(M))$ , which would be morally nothing more than class theoretic actualism.

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

**§4. 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  $\varphi$  holds in every  $N \supseteq M$  with  $N \in \mathfrak{A}$  and  $\Diamond\varphi$  holds at a world  $M \in \mathfrak{A}$  if  $\varphi$  holds in some  $N \supseteq M$  with  $N \in \mathfrak{A}$ . We will look only at modal formulae which are obtained from propositional modal formulae by substituting class theoretic formulae for the propositional variables. Quantifiers are allowed in the class theoretic formulae we substitute, but we do not work with quantified modal logic per se. A propositional modal formula is *valid* at  $M \in \mathfrak{A}$  if it's true for any substitution of class theoretic formulae for the propositional variables. For example,  $\Box p \Rightarrow p$  is valid at every world because if  $\varphi$  holds in every extension of  $M$  it in particular holds at  $M$ .<sup>16</sup>

<sup>16</sup> For the reader unfamiliar with modal logic looking for a general resource we recommend [3, 5]. For the application of modal logic to set theoretic potentialism we recommend [25, 26].

In the particular case of class-theoretic potentialist systems, the extension relation  $(M, \mathcal{X}) \subseteq (M, \mathcal{Y})$  is purely in the classes—the sets are fixed. So  $(M, \mathcal{X}) \models \Diamond\varphi$  if there is a larger system of classes which satisfies  $\varphi$  and  $(M, \mathcal{X}) \models \Box\varphi$  if every larger system of classes satisfies  $\varphi$ . For these systems, it will not be possible to express  $\Box\varphi$  nor  $\Diamond\varphi$  in the language of class theory (contrast with the modal logic of forcing [26] where the modal operations are expressible in the language of set theory).

A key mathematical question is then: what are the *modal validities* of  $\mathfrak{A}$ , the collection of modal assertions valid at every world in  $\mathfrak{A}$ ? And one can refine this question, by asking whether this is allowing or disallowing parameters in formulas changes the validities, or by asking whether it varies by world.

**DEFINITION 5.** *S4 is the propositional modal theory obtained by closing under the inference rules of modus ponens and necessitation (viz. if  $\varphi$  is a theorem then so is  $\Box\varphi$ ), starting with the following axioms, for any propositional modal formulae  $\varphi$  and  $\psi$ :*

- (K)  $\Box(\varphi \Rightarrow \psi) \Rightarrow (\Box\varphi \Rightarrow \Box\psi),$
- (Dual)  $\neg\Diamond\varphi \Leftrightarrow \Box\neg\psi,$
- (T)  $\Box\varphi \Rightarrow \varphi,$
- (4)  $\Box\varphi \Rightarrow \Box\Box\varphi.$

K and Dual hold 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. T holds on any frame whose accessibility relation is reflexive, and 4 holds on any frame whose accessibility relation is transitive. Potentialist systems are by definition reflexive and transitive, so they always validate S4.

Some other modal axioms we consider in this article are:

**DEFINITION 6.**

- (.2)  $\Diamond\Box\varphi \Rightarrow \Box\Diamond\varphi,$
- (.3)  $(\Diamond\varphi \wedge \Diamond\psi) \Rightarrow \Diamond[(\varphi \wedge \Diamond\psi) \vee (\Diamond\varphi \wedge \psi)].$

*Adding these to S4 gives, respectively, the theories S4.2 and S4.3.*

It is straightforward to check that .2 is valid for any frame whose accessibility relation is directed and .3 holds for any frame whose accessibility relation is linear. The converses, however, do not hold. Counterexamples to both can be found in the modal logic of forcing. The modal logic of forcing allowing any poset is S4.2 [26], but by an old result of Mostowski [40] there are forcing extensions with no common width-extension to a model of ZFC. And the modal logic of collapse forcings is S4.3 [23], but this potentialist system is not linearly ordered since you could pick different generics for the same collapse forcing. Nonetheless, these frame conditions are a fruitful heuristic for thinking about these modal axioms.

We wish to briefly discuss this Mostowski result, which was originally formulated in the context of class forcing with models of GB. This result will imply that many of class potentialist systems we consider are not directed, and its proof foreshadows the argument for Theorem 21. We state a special case of his more general result; see the original paper [40] or the recent [17] for further extensions.

**THEOREM 7 (Mostowski).** *Let  $(M, \mathcal{X}) \models \text{GB}$  be countable and transitive. Then there are  $C, B \subseteq \text{Add}(\text{Ord}, 1)^M$  generic over  $(M, \mathcal{X})$  so that from  $C$  and  $B$  can be defined over  $M$*

a real which codes an order of ordertype  $\text{Ord}^M$ . Consequently, no GB-expansion for  $M$  can contain both  $C$  and  $B$ .

*Proof Sketch.* Let  $r \subseteq \omega$  be such a real. The strategy is to place the bits of  $r$  in the Cohen generic  $C$  so that given both generics we can recover the coding points. The strategy for this is to alternate between building up  $C$  and  $B$  to meet the next generic in the list. At stage  $n$ , extend the partial construction of  $C$  to meet the next dense set, then place the  $n$ -th bit of  $r$ . Meanwhile, extend the partial construction of  $B$  by adding 0s to have the same length as what we've built up for  $C$ . Then alternate: further extend the partial construction of  $B$  to meet the next dense set, while putting 0s on the end of the approximation for  $C$  to have the same length. Given the generics  $C$  and  $B$  built up in this way, you can recover the coding points by following the alternating blocks of 0s. Thereby, any GB-expansion for  $M$  which contained both  $B$  and  $C$  would think that  $M$  has countable cofinality, but that would contradict the class replacement axiom  $\square$

The top-down class potentialist system consisting of all countable GB-expansions for countable transitive  $M$  will fail to be directed, by this theorem. More generally, any class potentialist system over a countable transitive universe of sets which allows extensions by Cohen generic classes ordinals will also fail to be directed. Nevertheless, it is possible for such a potentialist system to still validate .2 and exhibit directedness in its modal truths (see Theorem 17). On the other hand, we will see that the top-down class potentialist systems of all countable GB-expansions fails to validate .2 (see Theorem 21). Contrast this to the modal logic of (set) forcing, where the Mostowski theorem applies but the modal validities are nonetheless S4.2.

As we discuss elsewhere the validity of .2 provides a dividing line for potentialist systems. To use the language of [22, sect. 7], systems for which .2 is invalid exhibit a 'radical branching', whereas systems for which .2 is valid exhibit 'directed convergence'. Additionally, the validity of .2 enables mirroring theorems, as in [36] or [25].

Given the information that we get from knowing a potentialist system's modal validities, we want tools that allow us to calibrate them. Lower bounds are relatively easy to determine—one simply shows that one can cook up worlds of the required kind—but showing that *no more* is satisfied (i.e., upper bounds) is substantially harder. Let's briefly describe the main tools, *control statements*, used to compute upper bounds.

DEFINITION 8 [23, 26].

- A *button* is an assertion  $\beta$  so that  $\diamond\Box\beta$  holds at every world. If  $\Box\beta$  holds at a world  $M$ , we say  $\beta$  is *pushed* for  $M$ , otherwise we say  $\beta$  is *unpushed*. The intuition is, you can push a button, making  $\beta$  true forevermore, but once you push it you can never unpush it.
- A *switch* is an assertion  $\sigma$  so that  $\diamond\sigma$  and  $\diamond\neg\sigma$  hold at every world. The intuition is, you can toggle the truth value of  $\sigma$  freely back and forth.
- A *ratchet* is a finite sequence  $\rho_0, \dots, \rho_n$  of buttons so that pushing  $\rho_i$  pushes  $\rho_j$  for all  $j < i$ . The intuition is, you can ratchet forward but never back.
- A *long ratchet* of length a limit well-order  $\Gamma$  is a uniformly definable sequence of buttons  $r_\xi$ , indexed by  $\xi < \Gamma$ , so that pushing  $r_\xi$  pushes  $r_\eta$  for all  $\eta < \xi$  and so that in no world are all buttons on the ratchet pushed.

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** [26]. *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** [23]. *Suppose a potentialist system admits a long ratchet whose length  $\Gamma$  is closed under addition of ordinals  $<\omega^2$ . Then, the modal validities of any world in which the ratchet is not fully pushed are contained within S4.3.*<sup>17</sup>

With the setup 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 Section 2 (we do so in Sections 6 and 7).

**§5. Truth predicates and potentialism.** In this section we present some mathematical results about class potentialist systems which involve truth predicates. 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. The more philosophically inclined reader can skim this section; in the subsequent sections we discuss the philosophical significance of the results in this section.

Let us begin by fixing some notation and definitions. We will use capital Greek letters 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  $\xi \in \text{dom}(\Lambda)$ . To compare elements  $\xi$  and  $\eta$  of  $\Lambda$  we write  $\xi < \eta$ . Given class well-orders  $\Lambda$  and  $\Gamma$  say that  $\Lambda$  is *closed under addition*  $<\Gamma$  if whenever  $\xi < \Lambda$  and  $\eta < \Gamma$  we have that  $\xi + \eta < \Gamma$ . That is,  $\Lambda$  is closed under addition  $<\Gamma$  if every element of  $\Lambda$  has an  $\eta$ -th successor in  $\Lambda$  for each  $\eta < \Gamma$ . Note that whether a class is well-founded can be expressed just by quantifying over sets. Accordingly any two GB-expansions for  $M$  agree on whether a common relation is well-founded.

On the other hand, a model of GB may think that a class is well-founded while externally it is seen to actually be ill-founded. Say that  $(M, \mathcal{X}) \models \text{GB}$  is a  $\beta$ -model if it is correct about which of its classes are well-founded. A short argument yields that if  $\text{Ord}^M$  has uncountable cofinality then  $(M, \mathcal{X})$  is a  $\beta$ -model. In contrast, if  $\text{Ord}^M$  has countable cofinality then  $(M, \mathcal{X})$  may fail to be a  $\beta$ -model. Indeed, being a  $\beta$ -model is a strong property; for example, Marek and Mostowski showed [38, theorem 3.2] that if  $\beta$  is the least ordinal height of a  $\beta$ -model of MK, then there are transitive models of MK of height  $< \beta$ .

Consider a fixed transitive  $M \models \text{ZF}$ . The *truth predicate* for  $M$  is the class  $\text{Tr} \subseteq M$  which satisfies the recursive Tarskian truth conditionals for  $(M, \in)$ . In case we wish to emphasise for which structure  $\text{Tr}$  is a truth predicate we will write  $\text{Tr}^M$ . Observe that the property of being  $\text{Tr}^M$  is uniquely determined by a formula which only quantifies

<sup>17</sup> Hamkins, Leibman, and Löwe formulated this result for long ratchets of length  $\text{Ord}$ , but it is simple to check that what they used is that  $\text{Ord}$  is closed under addition of ordinals  $<\omega^2$ .

over sets, and so any GB-expansion of  $M$  can verify whether a class is  $\text{Tr}^M$ . Given a class  $A \subseteq M$  the *truth predicate relative to  $A$*  is the unique class  $\text{Tr}(A) \subseteq M$  which satisfies the recursive Tarski truth conditionals for  $(M, \in, A)$ , the expansion of  $(M, \in)$  with a unary predicate symbol for  $A$ . Again, it is uniquely determined and can be recognised as such by any GB-expansion of  $M$ .

Given the truth predicate  $\text{Tr} \subseteq M$ , we can consider  $\text{Tr}(\text{Tr})$ , the truth predicate relative to  $\text{Tr}$ . This gives expanded resources from which more can be defined. For example, the theory  $\text{GB}+$  ‘ $\text{Tr}$  exists’ proves the consistency of ZFC—restricting  $\text{Tr}$  to sentences gives a consistent completion of ZFC. Since ZFC and GB are equiconsistent this theory proves the consistency of GB. But this theory cannot prove its own consistency. However,  $\text{GB}+$  ‘ $\text{Tr}(\text{Tr})$  exists’ does prove the consistency of  $\text{GB}+$  ‘ $\text{Tr}$  exists’ by a similar argument. It won’t prove its own consistency, but for that we could use  $\text{Tr}(\text{Tr}(\text{Tr}))$ . We can continue to build this hierarchy of truth about truth about truth (and so on) higher, even into the transfinite. This hierarchy can be unified in the single definition of an iterated truth predicate.

Working over our fixed transitive  $M \models \text{ZF}$ , let  $\Lambda$  be a well-order. A  $\Lambda$ -iterated truth predicate is a class  $\Theta$  of triples  $(\xi, \varphi, a)$  where  $(\xi, \varphi, a) \in \Theta$  intuitively means that  $\varphi(a)$  is true at level  $\xi < \Lambda$ . Here  $\varphi$  is a formula in the language where we added a predicate symbol  $\Theta$  which is interpreted by  $\Theta$ . Formally, this is defined by a modified form of the Tarskian recursion, with an extra clause in the definition asserting that  $(\xi, \ulcorner \Theta(x, y, z) \urcorner, \langle \eta, \varphi, b \rangle) \in \Theta$  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. See [15, sect. 9] for precise definitions.

GB suffices to prove the uniqueness of iterated truth predicates. Given two classes which satisfy the definition of being a  $\Lambda$ -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  $\text{Tr}_\Lambda$  to denote the  $\Lambda$ -iterated truth predicate and  $\text{Tr}_\Lambda(A)$  to denote the  $\Lambda$ -iterated truth predicate relative to  $A$ . If  $\xi < \Lambda$  then we write  $\text{Tr}_\xi$  to mean  $\text{Tr}_{\Lambda \upharpoonright \xi}$ . Observe that, up to recoding,  $\text{Tr}_1$  is the same as  $\text{Tr}$  and  $\text{Tr}_{\Lambda+1}$  is the same as  $\text{Tr}(\text{Tr}_\Lambda)$ , and similarly for relativised truth predicates. In general,  $\text{Tr}_\Gamma(\text{Tr}_\Lambda)$  is inter-definable with  $\text{Tr}_{\Lambda+\Gamma}$ .

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 GB-expansion for  $M$  which extends  $\mathcal{X}$  and contains  $A$ . Specifically,  $\mathcal{X}[A]$  consists of the classes over  $M$  definable using  $A$  and finitely many classes from  $\mathcal{X}$ .

**DEFINITION 11.** *Say that a class potentialist system over  $M \models \text{ZF}$  is a truth potentialist system if it satisfies the following three properties.*

- (1)  $(M, \text{Def}(M))$  is a world, where  $\text{Def}(M)$  is the collection of parametrically first-order definable classes over  $M$ .
- (2) If  $(M, \mathcal{X})$  is a world then it satisfies GB.
- (3) If  $(M, \mathcal{X})$  is a world and  $A \in \mathcal{X}$  then  $(M, \mathcal{X}[\text{Tr}(A)])$  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  $< \Lambda$ -length truth potentialist system.

(3<sub>Λ</sub>) If  $(M, \mathcal{X})$  is a world,  $\xi < \Lambda$ , and  $A \in \mathcal{X}$  then  $(M, \mathcal{X}[\text{Tr}_\xi(A)])$  is a world.

Observe that truth potentialist systems simpliciter, as in Definition 11, are the special case where we require iterated truth predicates of length  $< 2$ . 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'll call an unbounded truth potentialist system.

The intent of this definition is to give a modal interpretation of views such as Fujimoto's or Linnebo's. Unbounded truth potentialist systems correspond to Fujimoto's proposal, allowing arbitrary length iterated truth predicates. And  $< \text{Ord}$ -length truth potentialist systems correspond to the class theory one obtains from Linnebo's property theory where properties are individuated iteratively along  $\text{Ord}$  (see [43]). But we have formulated things in a more general context, allowing different lengths. The point is, we see the same phenomena in the general setting, so small changes in their views won't produce a different outcome.

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

**PROPOSITION 12** (Krajewski [38, p. 475]). *Consider  $(M, \mathcal{X})$  a transitive model of GB with  $\text{Tr}^M \in \mathcal{X}$ . Then  $M$  contains a club of ordinals  $\alpha$  so that  $V_\alpha^M$  is an elementary submodel of  $M$ .*

*Proof.* By reflection we get club many  $\alpha$  so that  $(V_\alpha^M, \in, \text{Tr} \cap V_\alpha^M)$  is a  $\Sigma_1$ -elementary submodel of  $(M, \in, \text{Tr})$ . But we can express that  $(V_\alpha^M, \in)$  is an elementary submodel with a  $\Sigma_1$ -formula referring to the truth predicate. □

This result rules out, for instance, pointwise-definable models like the minimum transitive model (see [24]) or models without worldly cardinals from being the sets for a truth potentialist system. Indeed, such models have no  $T$ -expansion for any  $T$  strong enough to prove the existence of the truth predicate as a class.

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 doesn't get at the desired phenomenon. For instance,  $\mathfrak{A}$  could consist of a single model of GB with an  $n$ -iterated truth predicate for each  $n$ , and  $\mathfrak{B}$  could be obtained from  $\mathfrak{A}$  by refining its one world into many by adding  $\omega$  many new worlds where the  $n$ -th world only has up to the  $n$ -iterated truth predicate. Then  $\mathfrak{A} \subseteq \mathfrak{B}$ , but we wouldn't say that  $\mathfrak{A}$  is smaller than  $\mathfrak{B}$ . To avoid this sort of issue: 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.

That clarification out of the way, let's see that there are smallest truth potentialist systems. We state this as three results for the three different variants of the definition.

**THEOREM 13.** *If  $M$  admits a truth potentialist system then it admits a smallest truth potentialist system. This potentialist system validates S4.3.*

**THEOREM 14.** *Fix a length  $\Lambda \in \text{Def}(M)$  where  $\Lambda$  is well-founded. If  $M$  admits a length  $< \Lambda$  truth potentialist system then it admits a smallest one. Every axiom of S4.3 is valid*

in this smallest 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.

**THEOREM 15.** *Suppose  $M$  admits an unbounded truth potentialist system whose every world is a  $\beta$ -model. Then  $M$  admits a smallest unbounded truth potentialist system, and this potentialist system validates exactly S4.3.*

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

These three theorems come from the same construction, with slight tweaks for each case, so we will only give it once, for Theorem 14. The reader can easily adjust the proof to give the other two theorems.

*Proof.* We will build the smallest  $<\Lambda$  truth potentialist system, call it  $\mathfrak{T}$ . The key point is that  $\text{Tr}_\xi(\text{Tr}_\eta)$  is inter-definable with  $\text{Tr}_{\eta+\xi}$ , so if we build up from the base world  $\mathcal{X}_0 = \text{Def}(M)$  we only have to add  $\text{Tr}_\xi$  for enough  $\xi$ . Define  $\mathfrak{T}$  to consist of worlds  $\mathcal{X}_\xi$ , where  $\mathcal{X}_0 = \text{Def}(M)$  and, more generally,  $\mathcal{X}_\xi = \mathcal{X}_0[\text{Tr}_\xi]$ . We want to take the shortest segment of  $\xi$ 's which makes this work. For this, we must split into cases. If  $\Lambda$  is closed under addition  $<\Lambda$  then we take all  $\xi < \Lambda$  for  $\mathcal{X}_\xi$  to be in  $\mathfrak{T}$ . Else, we take all  $\xi < \Lambda \cdot \omega$ . We claim that  $\mathfrak{T}$  is the smallest  $<\Lambda$  truth potentialist system.

It is immediate by construction that each world satisfies all axioms of GB except possibly the class replacement axiom. Consider now any  $<\Lambda$ -truth potentialist system  $\mathfrak{A}$  over  $M$ . An easy induction on  $\xi$  shows that each  $\mathcal{X}_\xi$  is contained in a world in  $\mathfrak{A}$ . This establishes that  $\mathfrak{T}$  is covered by  $\mathfrak{A}$  and thus each  $(M, \mathcal{X}_\xi)$  does indeed satisfy class replacement; after all, a failure of class replacement in a smaller collection of classes for  $M$  would also be a failure in a larger collection of classes. A similar induction shows that if  $\mathfrak{T}$  covers  $\mathfrak{A}$  then each world in  $\mathfrak{A}$  is one of the  $\mathcal{X}_\xi$ . Thus,  $\mathfrak{T}$  is refined. Finally, that we can always find iterated truth predicates in a larger world is the key observation above. To find  $\text{Tr}_\xi(A)$  for a class  $A \in \mathcal{X}_\eta$ , look in the world  $\mathcal{X}_{\eta+\xi}$ ; this world has  $\text{Tr}_{\eta+\xi}$  and, using that  $A$  is definable from  $\text{Tr}_\eta$ , we can use this longer truth predicate to define  $\text{Tr}_\xi(A)$ .

To see that  $\mathfrak{T}$  validates S4.3, merely observe that  $\mathfrak{T}$  is linearly ordered and recall that linear orders validate S4.3. To get the upper bound, we will use Theorem 10 and demonstrate a long ratchet for this potentialist system. Let  $r_\xi$  be the assertion ‘ $\text{Tr}_\xi$  exists’. Then  $\langle r_\xi : \xi < \Lambda^* \rangle$  gives a long ratchet for this potentialist system, where  $\Lambda^*$  is either  $\Lambda$  or  $\Lambda \cdot \omega$  depending on which case we are in for constructing  $\mathfrak{T}$ . The assumptions on  $\Lambda$  ensure that  $\Lambda^*$  is closed under addition  $<\omega^2$ , allowing the theorem to apply. □

**REMARK 16.** *For Theorems 14 and 15 the purpose of the assumptions about  $\Lambda$  being well-founded and about  $\beta$ -models is to allow the inductions on  $\xi$  to go through. These inductions take place in the meta-theory, so we need the  $\xi$  to be well-founded as seen from the meta-theory.*

We next address potentialist systems with worlds that satisfy GBC—that is, worlds that additionally have a global well-order. We emphasise that the previous results hold for any transitive  $M$ , whereas for results about global well-orders we will use more specific assumptions. The two cases we will focus on is when  $M$  is countable and when  $M = V_\kappa$  for some inaccessible cardinal  $\kappa$ . These two cases correspond to two possible approaches to class potentialism (see the discussion in Sections 6 and 7).

There are a few ways one might ensure a potentialist system gives worlds which satisfy global choice. If  $M$  has a definable global well-order then in any truth potentialist system for  $M$  every world satisfies GBC. So we get truth-potentialist systems which validate exactly S4.3 and whose every world satisfies GBC. However, this puts a firm restriction on the first-order theory of  $M$ , namely  $M \models \exists x V = \text{HOD}(\{x\})$ .<sup>18</sup> One may very well think that this extra restriction is unwarranted, and so consider the general case. Here we can get different behaviour.

Here are two approaches. First, we could start with a base world that satisfies GBC. If this base world is to be as small as possible, it must be of the form  $(M, \text{Def}(M, G))$ , where  $G$  is a global well-order for  $M$ . With this approach every world in the potentialist system will satisfy GBC. Alternatively, we could add a new rule saying that we can expand to a larger world to add a global well-order. Some worlds, such as the base world of definable classes, may not satisfy global choice, but you can always extend to a world satisfying GBC.

For the first approach, the way to formulate this is to replace condition (1) in the definition of a truth potentialist system with the following.

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

To start let's see that some (though not all) global well-orders are inter-definable with Cohen-generic classes of ordinals. For one direction, if  $C \subseteq \text{Ord}$  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. Given 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. This rearrangement 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 it doesn't add sets because  $\mathbb{Q}$  is  $<\kappa$ -closed for every  $\kappa$ . 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. Thus, the two forcings are forcing equivalent.

Starting with a global well-order in the base world can affect the modal validities. This phenomenon happens in both the countable case and the  $V_\kappa$  case. For the sake of readability we state it as two theorems.

**THEOREM 17.** *Let  $M$  be a countable transitive model of ZFC and let  $\Lambda$  be a limit-length well-order over  $M$ . Suppose that  $M$  is the first-order part of a model of  $\text{GB} + \text{ETR}(\Lambda)$ . There are  $<\Lambda$  truth potentialist systems, but modified to require a global well-order in the base world, over  $M$  whose modal validities are precisely S4.2, where we allow parameters. In particular,  $\exists$  is invalid for these systems.*

*Proof.* Consider  $(M, \mathcal{Y}) \models \text{GB} + \text{ETR}(\Lambda)$  where  $\mathcal{Y}$  is countable. Force over this model to add a Cohen generic  $C$ . If  $M$  is countable such exists because there are only

<sup>18</sup> This is like the well-known fact [39] that  $V = \text{HOD}$  is equivalent to having a parameter-free definable global well-order; the only difference with  $V = \text{HOD}(\{x\})$  is that  $x$  gets used as a parameter in the definition of the global well-order.



countably many dense sets to meet. If  $M$  is a  $V_\kappa$ , note that  $\text{Add}(\text{Ord}, 1)^M = \text{Add}(\kappa, 1)$  is  $<\kappa$ -closed and since there are only  $\kappa$  many dense sets in  $\mathcal{Y}$  we can meet them to find a generic. Gitman and Hamkins showed that tame class forcing, such as Cohen forcing, preserves  $\text{ETR}(\Lambda)$  [28, theorem 16].

Consider the truth-potentialist system  $\mathfrak{T}$  starting with the base world  $(M, \mathcal{X}_0)$ , where  $\mathcal{X}_0 = \text{Def}(M, C)$ , and closing off under the requirement that any world can be extended by adding a  $\xi$ -iterated truth predicate relative to a parameter for any  $\xi < \Lambda$ . That is,  $\mathfrak{T}$  consists of worlds of the form  $\mathcal{X}_0[\text{Tr}_\xi(A)]$  where  $A \in \mathcal{X}_0$ . 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 GB.

We claim this truth-potentialist system has precisely 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  $(M, \text{Def}(M, \text{Tr}_\xi(C)))$  for some large enough  $\xi < \Lambda$ .

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  $\text{Add}(\text{Ord}, 1)$  is equivalent to  $\text{Add}(\text{Ord}, \omega)$ . So we can split  $C$  into  $\omega$  many classes  $C_i$  so that the  $C_i$  are mutually generic over  $(M, \mathcal{Y})$ . Further, this splitting process is definable, so the  $C_i$  are uniformly definable from the parameter  $C$ —you can take  $C_i$  to be formed from the bits on the coordinates equivalent to  $i$  modulo  $\omega$ . In particular, this means that given any  $\xi, \eta < \Lambda$ , if  $i \neq j$  then  $\text{Tr}_\xi(C_i)$  is not definable from  $\text{Tr}_\eta(C_j)$ . This is just because  $\text{Tr}_\eta(C_j)$  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  $\mu$  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  $\beta_i$  be the statement ‘ $\text{Tr}_{\mu+1}(C_i)$  exists’. For the switches, let  $\sigma_i$  be the statement ‘if  $\xi$  is the largest ordinal for which  $\text{Tr}_\xi(C_i)$  exists, then  $\xi$  is even’. As a 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.  $\square$

**THEOREM 18.** *Let  $M = V_\kappa$  for inaccessible  $\kappa$  and let  $\Lambda$  be a limit-length well-order over  $M$ . There are  $<\Lambda$  truth potentialist systems, but modified to require a global well-order in the base world, over  $M$  whose modal validities are precisely S4.2, where we allow parameters. In particular, .3 is invalid for these systems.*

*Proof Sketch.* The same construction goes through. The only thing to check is that we can find the desired generic  $C$ . First, by taking an elementary submodel of  $(M, \mathcal{P}(M))$  we get  $(M, \mathcal{Y}) \models \text{GB} + \text{ETR}(\Lambda)$  where  $|\mathcal{Y}| = |M| = \kappa$ . We can force over this model to add a Cohen generic  $C$ . This is because  $\text{Add}(\text{Ord}, 1)^M = \text{Add}(\kappa, 1)$  is  $<\kappa$ -closed and since there are only  $\kappa$  many dense sets in  $\mathcal{Y}$  we can meet them to find a generic  $C$ .  $\square$

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

true at a world then  $\psi$  is impossible at that world, and if  $\psi$  is true at a world then  $\varphi$  is impossible at that world, giving a failure of .3.

**REMARK 19.** *This construction can also be used to give unbounded truth potentialist systems, i.e., where the iterated truth predicates can be of any length, modified to have a global well-order in the base world whose modal validities (allowing parameters) are precisely S4.2. In the  $V_\kappa$  case this requires no extra assumption, while in the countable transitive model case this requires that  $M$  be the first-order part of a model of  $\text{GB} + \text{ETR}$ .*

This construction also has implications for top-down potentialist systems. We state this only for the countable case, as we give a more precise calculation of the modal validities in the  $V_\kappa$  case in Corollary 28.

**COROLLARY 20.** *Fix countable transitive  $M \models \text{ZFC}$  and suppose  $M$  is the sets of some model of  $\text{ETR}(\Lambda)$  for some limit length well-order  $\Lambda$ . Consider the top-down potentialist system consisting of all countable  $\text{GB}$ -expansions of  $M$ . Then this potentialist system does not validate .3, allowing parameters in formulae.*

*Proof.* The same assertion  $\varphi$  witnesses the failure of .3. Indeed, the same construction of independent buttons and switches to get S4.2 as an upper bound applies. □

In Corollary 25 we will see that .2 may also fail.

We now turn to the other option for getting worlds which satisfy global choice. Rather than adding a condition asserting that the base world contains a global well-order, we instead add a rule allowing the addition of a global well-order. Let's first focus on the stronger assumption that the global well-order is added by forcing. That is, we add 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  $\text{Ord}$  generic over  $(M, \mathcal{X})$ . Of course, which Cohen generics exist will depend on the choice of model. If  $(M, \mathcal{X})$  is countable, this is just the well-known Rasiowa–Sikorski lemma that you can always meet countably many dense sets. In the  $M = V_\kappa$  case, if  $\mathcal{X}$  has cardinality  $\kappa$  then we can build a generic in  $\kappa$  many sets using the  $<\kappa$ -closure of  $\text{Add}(\text{Ord}, 1)^M = \text{Add}(\kappa, 1)$ . 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.

If  $M$  is countable this is quite destructive. Specifically, adding a global well-order may kill off the possibility of adding a truth predicate whilst preserving the basic axioms of  $\text{GB}$ .<sup>19</sup>

**THEOREM 21.** *Let  $M$  be a countable transitive model of  $\text{ZFC}$  and let  $A \subseteq M$  be a class over  $M$  so that  $(M, \text{Def}(M, A)) \models \text{GB}$ . Then there is  $C$  Cohen-generic over  $(M, \text{Def}(M, A))$  so that no  $\text{GB}$ -expansion for  $M$  can contain both  $C$  and  $\text{Tr}(A)^M$ .*

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

We claim that from  $\text{Tr}(A)^M$  we can define a sequence  $\langle D_\alpha : \alpha \in \text{Ord} \rangle$  of dense subclasses of  $\text{Add}(\text{Ord}, 1)$  so that meeting every  $D_\alpha$  guarantees genericity over  $(M, \text{Def}(M, A))$ . This is because we can take  $D_\alpha$  to be the intersection of all open dense classes definable from  $A$  with parameters from  $V_\alpha$ . The point is, from the truth

---

<sup>19</sup> The non-relativised version of this theorem can be found as [51, theorem 1.57].

predicate we can define this sequence, because the truth predicate gives us uniform access to definability. Then  $D_\alpha$  is open dense because the forcing is  $\kappa$ -distributive for every  $\kappa$ . And clearly meeting every  $D_\alpha$  implies getting below every definable dense class.

Now let's use this sequence of  $D_\alpha$ s to code a bad real into a generic. Fix a binary sequence  $B : \text{Ord} \rightarrow 2$  so that the set of  $i$  so that  $B(i) = 1$  is cofinal in the ordinals of  $M$  and has ordertype  $\omega$ . Note that no GB-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  $p_\alpha$ , extend to meet  $D_\alpha$ , where we require the extension to be of minimal length to get into  $D_\alpha$ . Then add on the bit  $B(\alpha)$  to get  $p_{\alpha+1}$ . And at limit stages, set  $p_\lambda = \bigcup_{\alpha < \lambda} p_\alpha$ . Note that the sequence  $\langle p_\alpha : \alpha < \lambda \rangle$  is an element of  $M$ —and thus  $p_\lambda \in M$ —by application of the class replacement axiom in a GB-expansion for  $M$  containing  $\text{Tr}(A)^M$ , using that any initial segment of  $B$  is in  $M$ . Then  $C = \bigcup p_\alpha$  meets every dense class in  $\text{Def}(M, A)$ .

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 a GB-expansion for  $M$ , then you would also have  $B$  in the GB-expansion. This is impossible.  $\square$

This result immediately implies we can kill off iterated truth predicates.

**COROLLARY 22.** *Let  $M$  be a countable transitive model of ZFC, let  $A \subseteq M$  be a class over  $M$  so that  $(M, \text{Def}(M, A)) \models \text{GB}$ , and let  $\Lambda \in \text{Def}(M, A)$  be a class well-order. Then there is  $C$  Cohen-generic over  $(M, \text{Def}(M, \text{Tr}_\Lambda(A)))$  so that no GB-expansion for  $M$  can contain both  $C$  and  $\text{Tr}_{\Lambda+1}(A)^M$ .*

*Proof.* Assume we are in the nontrivial case where some GB-expansion of  $M$  contains  $\text{Tr}_{\Lambda+1}(A)^M$ . Now apply the theorem to  $\text{Tr}(\text{Tr}_\Lambda(A))$  to produce a generic  $C$ . Because  $\text{Tr}(\text{Tr}_\Lambda(A))$  is inter-definable with  $\text{Tr}_{\Lambda+1}(A)$ , the generic  $C$  cannot be in a GB-expansion for  $M$  which contains this iterated truth predicate.  $\square$

This killing truth phenomenon rules out the possibility of getting a truth potentialist system with allowing extensions to add a generic global well-orders, in the case of working over a countable transitive model.

**COROLLARY 23.** *Let  $M$  be a countable transitive model of ZFC and let  $\Lambda$  be a limit length well-order over  $M$  so that  $M$  has a nontrivial  $<\Lambda$  truth potentialist system. Consider the smallest  $<\Lambda$  truth potentialist system over  $M$ , call it  $\mathfrak{A}$ . Consider the potentialist system, call it  $\mathfrak{B}$ , you get by expanding  $\mathfrak{A}$  by adding in all extensions of worlds in  $\mathfrak{A}$  by adding a generic global well-order. Then, for any world in  $\mathfrak{B}$  which came from  $\mathfrak{A}$ , (1) the .2 axiom is invalid in  $\mathfrak{B}$  at every world which came from  $\mathfrak{A}$ , where we allow parameters; and (2) there are worlds in  $\mathfrak{B}$  with classes for which it is impossible to find a truth predicate in a larger model of GB over  $M$ . That is, this larger potentialist system cannot be expanded to become a truth potentialist system.*

*Proof.* Because  $\mathfrak{A}$  is the smallest  $<\Lambda$  truth potentialist system over  $M$ , each world in  $\mathfrak{A}$  has classes of the form  $\mathcal{X}_\xi = \text{Def}(M, \text{Tr}_\xi)$  for some  $\xi < \Lambda$ . Fix  $\xi$  and consider the corresponding world in  $\mathfrak{A}$ . Then there is  $C \subseteq \text{Ord}^M$  Cohen-generic over  $(M, \mathcal{X}_\xi)$  so that no GB-expansion for  $M$  can contain both  $C$  and  $\text{Tr}_{\xi+1}$ . Let  $\varphi$  be the formula asserting that  $\text{Tr}_{\xi+1}$  exists. Then  $\varphi$  is possibly necessary at  $\mathcal{X}_\xi$ , since it is necessary at  $\mathcal{X}_{\xi+1}$  but

is not necessarily possible, since it is impossible in  $\mathcal{X}_\xi[C]$ . This establishes (1) of the corollary. For (2), the theorem tells us that  $\text{Tr}(C)$  cannot exist in any GB-expansion of  $M$ .  $\square$

**REMARK 24.** *If, instead of only allowing extensions adding generic global well-orders, we include extensions adding any global well-order, we still get the same phenomenon. This is because such a potentialist system would include extensions by generic global well-orders, so the truth-killing theorem would still apply.*

For a similar reason, this result affects the modal validities for top-down potentialist systems over a countable transitive model.

**COROLLARY 25.** *Fix countable transitive  $M \models \text{ZFC}$  and consider the top-down class potentialist system  $\mathfrak{A}$  for  $M$  consisting of countable  $(M, \mathcal{X}) \models \text{GB}$ . If  $(M, \mathcal{X})$  is a world in  $\mathfrak{A}$  which contains a class  $A$ , does not contain its truth predicate  $\text{Tr}(A)$ , but does have an expansion to a world which contains  $\text{Tr}(A)$ , then the .2 axiom is invalid at  $(M, \mathcal{X})$ , allowing  $A$  as a parameter. In particular, if  $A$  is parameter-free definable, then .2 is invalid at  $(M, \mathcal{X})$  without allowing parameters.*

For example, if  $M$  has a GB-expansion which contains  $\text{Tr}$ , then  $(M, \text{Def}(M))$  is such a world.

*Proof.* The statement ‘ $\text{Tr}(A)$  exists’ is possibly necessary but by the theorem is not necessarily possible.  $\square$

On the other hand, if  $M = V_\kappa$  for  $\kappa$  inaccessible then this sort of destructive behaviour is impossible.

**OBSERVATION 26.** *Suppose  $M = V_\kappa$  where  $\kappa$  is inaccessible and consider  $(M, \mathcal{X}) \models \text{GB}$ . If  $C$  is any Cohen generic class of ordinals over  $(M, \mathcal{X})$  then there is an extension  $(M, \mathcal{Y}) \models \text{GB}$  of  $(M, \mathcal{X})$  which contains  $C$  and a truth predicate relative to any class in  $\mathcal{X}$ . Indeed, we may pick  $\mathcal{Y}$  to contain  $\text{Tr}_\xi(A)$  for any  $\xi, A \in \mathcal{X}$ .*

*Proof.* One can construct  $\mathcal{Y}$  in a minimal manner, but let’s use a blunt tool. Let  $(M, \mathcal{Y})$  be an elementary submodel of  $(M, \mathcal{P}(M))$  which contains  $\mathcal{X}$  and  $A$ . Done.  $\square$

**PROPOSITION 27.** *Suppose  $M = V_\kappa$  where  $\kappa$  is inaccessible and let  $\mathfrak{A}$  be the smallest  $<\Lambda$  truth potentialist system over  $M$  for some limit length  $\Lambda$ . Consider a potentialist system  $\mathfrak{B}$  created by taking  $\mathfrak{A}$  by adding in extensions of worlds in  $\mathfrak{A}$  by adding a global well-order, then closing off by adding  $<\Lambda$  iterated truth predicates. Then  $\mathfrak{B}$  validates exactly S4.2, allowing parameters.*

*Proof Sketch.* The observation implies that  $\mathfrak{B}$  is directed as a partial order—given  $\mathcal{X}$  and  $\mathcal{Y}$  from  $\mathfrak{B}$  then the closure of  $\mathcal{X} \cup \mathcal{Y}$  under definability will be a world in  $\mathfrak{B}$ —and so it validates S4.2. To show that S4.2 is also an upper bound, we use the same strategy as in Theorem 18. If  $(M, \mathcal{X})$  is a world in  $\mathfrak{B}$  containing a class  $C$  which is Cohen generic over a smaller world in  $\mathfrak{B}$ , then there is a bound to the length of the iterated truth predicates relative to  $C$  which are in  $\mathcal{X}$ . This gives the necessary space to define independent buttons and switches by looking at iterated truth predicates relative to fragments  $C_i$  of  $C$ .  $\square$

Similar considerations apply to top-down potentialist systems.

**COROLLARY 28.** *Suppose  $M = V_\kappa$  where  $\kappa$  is inaccessible and let  $\mathfrak{A}$  be the top-down class potentialist system for  $M$  consisting of all  $(M, \mathcal{X}) \models \text{GB}$  with  $|\mathcal{X}| = \kappa$ . Then the modal validities for  $\mathfrak{A}$  are exactly S4.2, allowing parameters. Indeed, every world in  $\mathfrak{A}$  extends to a world whose validities are exactly S4.2, again allowing parameters.*

*Proof Sketch.* That S4.2 is a lower bound is because  $\mathfrak{A}$  is directed as a partial order. Consider any world  $(M, \mathcal{X})$  in  $\mathfrak{A}$ . Pick  $X \subseteq M^2$  so that  $\mathcal{X}$  is coded by  $X$ , meaning  $\mathcal{X} = \{(X)_i : i \in M\}$ . Then  $\mathcal{Y} = \text{Def}(M, X)$  gives a world in  $\mathfrak{A}$  which extends  $\mathcal{X}$ . Extend  $(M, \mathcal{Y})$  to  $(M, \mathcal{Y}[C])$ , where  $C$  is Cohen-generic over  $\mathcal{Y}$ . The same strategy as before yields that the modal validities at  $(M, \mathcal{Y}[C])$  are exactly S4.2.  $\square$

**§6. 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. Our results indicate that various bottom-up class potentialist systems might not be convergent. Why this is relevant for bottom-up approaches? Well, these begin by specifying some initial starting world (i.e., *Initial World*) and then individuate 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.

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.

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 mathematical concepts of set and class, even if modal, not to contain these choice points both for philosophical cohesiveness and mathematical expedience. Another way to view them is as *interesting* but not any special cost—they indicate structural properties of the relevant potentialist system (and perhaps the underlying concepts), 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. This highlights some important choice points for the bottom-up theorist.

**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 13 and 14) 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. This insistence is natural with a generative view of classes, as from [12] or derived from [35]. Having a smallest potentialist system amounts to being able to say: ‘This is the minimal amount needed to realise these generative conditions’. So, for example, Theorem 15 gives that there is a smallest potentialist system which corresponds to a modal reworking of Fujimoto’s liberal predicativism. Similarly, Theorem 14, with the length  $\Lambda = \text{Ord}$ , says that there is a smallest potentialist system corresponding to Linnebo’s property theory.<sup>20</sup>

<sup>20</sup> We thank Øystein Linnebo for emphasising this point.

How well does the insistence on smallest systems mesh with the two views? 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. Roberts [42, 43] has analysed these systems as corresponding to what classes are definable from longer and longer iterated truth predicates. Formulated in the potentialist perspective, this is looking at the smallest truth potentialist systems constructed in Section 5, where the lengths of iterated truth predicates one allows may depend on other philosophical commitments.

**Wider systems can admit branching.** A critical point, in contradistinction to the foregoing, is that for larger systems we *do* get branching over a countable transitive model. 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 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 lengths of truth predicates. 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. If you drop the requirement for genericity then global well-orders can add height. Given  $(M, \mathcal{X}) \models \text{GBC}$ , 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, or any other arbitrarily complicated class.

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 and property theory motivate (since they just involve adding truth predicates and closing under definability). 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 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  $\exists x V = \text{HOD}(\{x\})$ . (See the discussion in Section 5.) 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 Section 5 of how to force to add a global well-order.) 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. 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. This has the interesting feature that such a well-order cannot (unlike truth predicates) be viewed as the introduction of a *unique* class with a certain property—the introduction of one Cohen-generic necessitates the addition of many. However, it can still be viewed a individuation of a certain kind—we ask for a generic object and then close under definability. We leave it open whether such 'indeterminate individuation' should be acceptable to the kinds of bottom-up theorists we've considered.<sup>21</sup>

If we are allowed to add such a generic well-order then, as we noted in Theorem 21, 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  $\text{GB} + \text{ETR}$  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 parameterised truth predicates are there from the get-go. This contravenes the basic set-up of the liberal predicativist and property theorist and so would necessitate some revision of these positions.

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 as artificial.<sup>22</sup> The explication of further possibilities would be a useful contribution to our understanding of classes, and we leave it to further work.

And of course, one possibility available to the class potentialist is to simply give up on having a global well-order. One might think this is natural for those bottom-up approaches based on liberalised forms of definability. Choice principles for sets draw motivation from a combinatorial view of sets, so if one has a non-combinatorial view of classes one might expect to not have choice principles for classes.<sup>23</sup> Interestingly, for some of the bottom-up approaches we consider, this is not so immediate.

<sup>21</sup> This is not unique to the class forcing case—it applies to set-forcing too (e.g., in the addition of a single Cohen real over a model of ZFC).

<sup>22</sup> See the earlier point that any class can be coded into a global well-order.

<sup>23</sup> We thank Øystein Linnebo for pressing us on this point.

Let's start by noting that a class potentialist who wishes to prohibit global well-orders pays a mathematical cost. Namely, their class theory is unable to directly support the mathematical work which requires global choice. For example:

- (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 [13] 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.<sup>24</sup>

On the one hand, these examples are somewhat niche when compared to the large number of equivalences with standard (set) AC and many fundamental statements of mathematics. But on the other hand, the equivalences with AC also rely on creatures of the higher infinite. For example, AC is equivalent over ZF to the assertion that every vector space has a basis [4]. But the vector spaces used in this equivalence are not the familiar ones from undergraduate linear algebra, and are necessarily quite large in cardinality. It is consistent that AC fails and yet every vector space in, say,  $V_{\omega+\omega}$  has a basis. It strikes us as artificial to be sceptical of uses of choice at the level of classes but not be sceptical of uses at the level of sets of high von Neumann rank, though we acknowledge that there is more debate to be had.

One might think that 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, wants to ‘provide a mathematical framework in which...widely accepted and/or mathematically fruitful uses of classes can be meaningfully expressed and implemented’ [12, p. 217]. Given that there are mathematical uses for global choice, it thus seems that there's some pressure to accept the possible existence of global well-orders for Fujimoto. Insofar as other approaches have similar naturalistic assumptions, they're also subject to similar pressures.<sup>25</sup>

Given this naturalistic pressure, it's sensible to consider how the possible existence of global well-orders meshes with the philosophical positions we've considered. For the liberal predicativist, we should note that 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. Fujimoto writes:

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. [12, p. 225]

Fujimoto is claiming here that the ‘bad’ principles (from the liberal predicativist perspective) are those that apply global structure to the classes. Such bad examples

<sup>24</sup> See [6, sect. 9]. Note, however, that it remains an open question whether global choice is necessary for this result; see the discussion on this MathOverflow question [21].

<sup>25</sup> Of course one could always reject the use of such classes as ‘legitimate’ parts of mathematics, our only contention is that it is a cost to do so.



include impredicative comprehension, where the use of unbounded quantifiers (and their alternations) can result in complex global structure. By contrast, the introduction of a truth predicate (a canonical example of a ‘good’ class for the liberal predicativist) simply involves introducing a single class and then closing under definability. Given this characterisation, there is at least *prima facie* reason for the liberal predicativist to accept the possible existence of a global well-order—it is an assertion about a single class that also meshes well with the liberal predicativist’s desire for mathematical freedom. If one wishes to reject the possibility of such classes for the liberal predicativist, reasons should be given as to why. We do not rule out that there may be such principles, but we do not see any obvious candidates. Similar remarks apply to the addition of other class-forcing generics—these are assertions about the introduction of a single class (and then closure under definability), and so seem acceptable from the liberal predicativist perspective.

The situation is different for the property theorist. As noted earlier, their classes are essentially those obtained by iterating the definability predicate—equivalently, looking at what is definable from longer and longer iterated truth predicates. The property theorist thus 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.

**The choice of model theory affects convergence.** Many of our results showing non-convergence required the use of countable transitive models. For example, Theorem 21 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. For example, the class-theoretic potentialists we’ve presented who are also set-theoretic actualists accept this. In fact, out of the views we’ve considered it’s only some brands of set-theoretic multiversist/potentialist who accept that  $V$  can be made countable.

This contrasts sharply with the version of the model theory where we consider potentialist systems over  $V_\kappa$  for  $\kappa$  inaccessible. For, Observation 26 says that the killing truth phenomenon does not arise for such  $V_\kappa$ —we can always add iterated truth predicates after adding a global well-order. And Proposition 27 says that there is a natural system obtained by adding a global well-order and then closing off under the addition of iterated truth predicates that validates S4.2. If one thinks that such a  $V_\kappa$  is a better model of the universe than a countable transitive model—say because one is a set-theoretic actualist but class-theoretic potentialist—then one can avoid the killing truth phenomenon (and indeed justify S4.2) by adopting the appropriate model theory. But this shows that real care must be taken in providing a model-theoretic set-up; proceeding with a countable transitive model and no restriction on the generics allowed (an oft-made move) will result in (possibly unpleasant) unintended artefacts of the models employed appearing in one’s conception of class-theoretic potentialism.

What emerges from this discussion is that there is the following tension at the heart 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 (as with countable transitive models), then we have a problem. The property theorist resolves this issue by rejecting (iii), as does the advocate of the model theory using an

inaccessible rank  $V_\kappa$  instead of a countable transitive model. This problem is thus a pressing issue (only) for the predicativist who accepts the use of countable transitive models, and we will suggest a different solution via additional modal principles (also rejecting (iii)) in Section 8.1.

**§7. The top-down approach.** In this section we discuss what the mathematical results from Section 5 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.

Our formalisation for top-down approaches is to fix a model  $M \models \text{ZFC}$  and consider a potentialist system consisting of expansions of  $M$  to a model of  $T$ , where  $T$  is a class theory. Corresponding to the set multiversist view we consider countable transitive  $M$  and all countable  $T$ -expansions for  $M$ . Corresponding to set universalism we consider  $M = V_\kappa$  for  $\kappa$  inaccessible and  $T$ -expansions for  $M$  of size  $\kappa$ . Results from Section 5 tell us something about these potentialist systems, making some assumptions about  $T$  and  $M$ . Two main tools were used in Section 5: truth predicates, and class forcing. We need both to be applicable.

Let's discuss truth predicates first. Proposition 12 tells us that asking to have any world with a truth predicate puts a limitation on the choice of  $M$ . For the  $V_\kappa$  case it is already implied by  $\kappa$  being inaccessible and for the countable transitive model case 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 Section 5 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. As discussed in Section 5 adding a Cohen-generic class of ordinals adds a global choice function. This puts a restriction on the worlds allowed—if one thinks that such generics can be added, one cannot hold that Global Choice fails at every world. The assumption that generics can be added also rules out the inclusion of axioms that limit the classes by definability.<sup>26</sup>

In sum, what we need from  $T$  for the results of Section 5 to apply is (i)  $T$  does not limit the classes by definability, (ii)  $T$  does not prove that arbitrary iterated truth predicates exist, and (iii)  $T$  does not have a principle implying global choice is false. These results illuminate in a concrete way how the indeterminacy of reference underlying top-down potentialism might manifest.

**Radical branching in the set-theoretic multiverse.** 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  $M'$  which has  $M$  as an element and thinks it's countable will have

<sup>26</sup> Here's an illustrative toy example. Let  $T$  be GB plus the assertions that length  $n$  iterated truth predicates exist for any finite  $n$  and that every class is definable from some  $\text{Tr}_n$ . Any nontrivial forcing adds a generic not in the ground model and thus not definable from any  $\text{Tr}_n$ , thus destroying  $T$ . So the results in Section 5 do not apply to  $T$ .

a conception of what the  $T$ -class-potentialist system over  $M$  looks like (for some reasonable  $T$ ). For  $T$  near the level of GB (and appropriate  $M$ ) we get failures of the .2 and .3 axioms, indicative of strong branching (see Corollaries 20 and 25).

Some multiversists (e.g., Hamkins [22]) distinguish between the extreme branching of S4 and the ‘inevitability’ of S4.3 as well as the ‘convergence’ implied by S4.2. The argument proceeds from the mirroring theorems for systems containing S4.2—whilst one could use the modal theory, the mirroring theorem guarantees that a modal-free theory can be used. Hamkins argues that theories with S4.2 are thus ‘quasi-actualist’—practically speaking nothing of substance hangs on whether we think of the universe modally or non-modally. 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 potentialisms 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<sup>27</sup>). 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) bound to ‘quasi-actualist’ potentialist systems satisfying S4.2 and a mirroring theorem. Not so for the multiversist-inspired class-theoretic potentialist. Here, the GB-class-potentialist systems have failures of .2 and .3 and hence no mirroring theorem (using standard technology). These potentialist systems are the first to the best of our knowledge systems of set theory to exhibit non-convergent branching even when we restrict to well-founded models.<sup>28</sup>

**Plural indeterminacy and impredicativity.** 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 [10] 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 [49]). This interpretation works whether or not the range of the plural quantifiers is determinate; if the ranges are 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 30 for fuller details.

Though one might hold that the indeterminate plural interpretation yields MK on the basis of the ‘standard’ conception of the logic, this is controversial. As Florio

<sup>27</sup> See [25, 26].

<sup>28</sup> Note that such occur in the context of second-order arithmetic (see Section 8.2).

and Linnebo [10] note, often the impredicative plural comprehension scheme is motivated by the assumption that every non-full Henkin semantics for the plural quantifiers is unintended.<sup>29</sup> Such an assumption leads immediately to the impredicative comprehension scheme for plurals and, by extension, classes (see [49, pp. 76–77] and [34, sect. 3.2]).

Of course, if we allow plural quantification to be indeterminate then we have an immediate response—an instance of a formula  $\varphi$  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:

$$\varphi(x) = 'x = x \text{ and the Class Fodor Principle}'.$$

Whether the Class Fodor Principle holds is independent of MK [16]. If we do not assume that quantification is determinate, and there are worlds in which Class Fodor holds and others in which Class Fodor fails, then it is neither the case that no sets satisfy  $\varphi$  nor is it the case that some sets satisfy  $\varphi$ —in some worlds  $\varphi$  picks out  $V$  and in others it picks out  $\emptyset$ .

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 who is happy using countable transitive models. Acceptance of the impredicative plural comprehension scheme despite indeterminacy has immediate mathematical ramifications, not just regarding what non-modal statements of class theory are true at every world but also the nature and validities of the relevant potentialist systems.

**The underlying ontology affects branching.** Let's assume that one does accept the plural interpretation, but allows that weak theories (like GB) can be thereby motivated. Let's also assume, however, that one is a set-theoretic *actualist*. For such a theorist, we contend, the model theory that uses some inaccessible  $V_\kappa$  is much better for representing the class-theoretic potentialism on offer, certainly in comparison to a countable transitive model. As we noted in Corollary 28, the corresponding top-down potentialist system validates S4.2 even if we consider an interpretation with a global well-order added by a Cohen-generic class, we can always consider interpretations that have a truth predicate for this class. So the divergence does not occur as it does for the set-theoretic multiversist who accepts that *any* universe can be made countable. The different model theories suggested by the different underlying motivations thus have a profound influence upon the nature of the relevant class-theoretic potentialism.

<sup>29</sup> See [10] for further references.

**§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.

**8.1. Additional modal principles?** As discussed in Section 7, 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 way to do so is in the model theory—adopting a model theory that uses inaccessible  $V_\kappa$  instead of countable transitive models.

A different (but related) approach 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:

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

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 these kinds 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., [19, 26]). An example of a different flavor, one closer in motivation to what we give here, can be found in [47] (with subsequent development by Maddy and Meadows [37]). 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 [40]. 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 of 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., [7, 29]). There has been work addressing to what extent results about predicativism over  $\omega$  generalise to predicativism over  $V$  (see, e.g., [11, 44]). Similar to how it was formalised in the set theoretic context, one could formalise potentialism over  $\omega$  by considering potentialist systems of  $\omega$ -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 and  $\Sigma^1_\infty$ -AC<sub>0</sub> is the choice schema for definable collections of sets; consult [46, chap. VII.6] for a definition. If  $(\omega, \mathcal{X})$  and  $(\omega, \mathcal{Y})$  are models of second-order arithmetic then  $\mathcal{X}$  is a  $\beta$ -submodel of  $\mathcal{Y}$ , written  $(\omega, \mathcal{X}) \subseteq_\beta (\omega, \mathcal{Y})$ , if  $\mathcal{X}$  is a subset of  $\mathcal{Y}$  and the two models agree about which of  $\mathcal{X}$ 's relations are well-founded.

**THEOREM 29** (Hamkins–Williams). *Let  $T$  be a computably axiomatizable extension of  $Z_2 + \Sigma^1_\infty$ -AC<sub>0</sub>. Then the modal validities of the potentialist system consisting of countable  $\omega$ -models of  $T$  ordered by  $\subseteq_\beta$  are precisely S4, whether or not we allow parameters in formulas.*

*Proof sketch.* It is well-known that  $Z_2 + \Sigma^1_\infty$ -AC<sub>0</sub> is bi-interpretable with  $ZF^- + V = H_{\omega_1}$ , the assertion that every set is 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  $\omega$ , and if  $(\omega, \mathcal{X})$  is a  $\beta$ -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,  $\omega$ -standard companion models, ordered by end-extension. Up to coding this is the same as considering the potentialist system of countable  $\omega$ -models of  $T$  directly. An instance of [27, theorem 6] yields that this potentialist system admits a universal finite sequence and thus its modal validities are precisely S4.<sup>30</sup> □

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

**QUESTION 30.** *Let  $T$  be MK plus Class Bounding and let  $M$  be a countable transitive model of ZFC which has a nontrivial top-down class potentialist system consisting of  $T$ -expansions for  $M$ . Does the potentialist system consisting of countable  $T$ -expansions for  $M$  ordered by the substructure relation<sup>31</sup> 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.

**Acknowledgment.** We would like to thank Carolin Antos, Julian Bacharach, Laura Crosilla, John Baldwin, Salvatore Florio, Øystein Linnebo, Beau Madison Mount, Naomi Osorio-Kupferblum, Martin Pleitz, Sam Roberts, and audiences in Konstanz and New York for comments and discussion.

**Funding.** The first author is very grateful for the support of the VolkswagenStiftung via the project *Forcing: Conceptual Change in the Foundations of Mathematics* and

<sup>30</sup> Theorem 6 is phrased in a general form. See the discussion at the beginning of Section 4 for why it applies to extensions of  $ZF^-$ .

<sup>31</sup> For models of class theory with the same sets every submodel is a  $\beta$ -submodel for free.

Norges forskningsrådet (the Research Council of Norway) for their support via the project *Infinity and Intensionality: Towards A New Synthesis* (Grant No. 314435).

#### BIBLIOGRAPHY

- [1] Arrigoni, T., & Friedman, S.-D. (2013). The hyperuniverse program. *Bulletin of Symbolic Logic*, **19**, 77–96.
- [2] Barton, N. (2021). Indeterminateness and ‘the’ universe of sets: Multiversism, potentialism, and pluralism. In Daghighi, A. S. and Rezus, A., editors. *Selected Topics from Contemporary Logics, Landscapes in Logic*. Georgia: College Publications, pp. 105–183.
- [3] Blackburn, P., de Rijke, M., and Venema, Y. (2002). *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge: Cambridge University Press.
- [4] Blass, A. (1984). Existence of bases implies the axiom of choice. In J. E. Baumgartner, D. A. Martin, and S. Shelah, editors. *Axiomatic Set Theory (Boulder, CO, 1983)*, Contemporary Mathematics, vol. 31. Providence: American Mathematical Society, pp. 31–33.
- [5] Cocchiarella, N. B., & Freund, M. A. (2008). *Modal Logic: An Introduction to its Syntax and Semantics*. Oxford: Oxford University Press.
- [6] Ehrlich, P. (2012). The absolute arithmetic continuum and the unification of all numbers great and small. *Bulletin of Symbolic Logic*, **18**(1), 1–45
- [7] Feferman, S., & Hellman, G. (1995). Predicative foundations of arithmetic. *Journal of Philosophical Logic*, **24**(1), 1–17.
- [8] Felgner, U. (1971). Comparison of the axioms of local and universal choice. *Fundamenta Mathematicae*, **71**(1), 43–62.
- [9] Fine, K. (2005). Class and membership. *Journal of Philosophy*, **102**(11), 547–572.
- [10] Florio, S., & Linnebo, Ø. (2016). On the innocence and determinacy of plural quantification. *Noûs*, **50**(3), 565–583.
- [11] Fujimoto, K. (2012). Classes and truths in set theory. *Annals of Pure & Applied Logic*, **163**(11), 1484–1523.
- [12] ———. (2019). Predicativism about classes. *Journal of Philosophy*, **116**(4), 206–229.
- [13] 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*, Contemporary Mathematics, vol. 690. Providence: American Mathematical Society.
- [14] ———. (2017). Open determinacy for class games. In Caicedo, A. E., Cummings, J., Koellner, P., and Larson, P. B., editors. *Foundations of Mathematics: Logic at Harvard Essays in Honor of W. Hugh Woodin’s 60th Birthday*, Contemporary Mathematics, vol. 690. Providence: American Mathematical Society, pp. 121–144.
- [15] Gitman, V., Hamkins, J. D., Holy, P., Schlicht, P., & Williams, K. J. (2020). The exact strength of the class forcing theorem. *Journal of Symbolic Logic*, **85**(3), 869–905.
- [16] Gitman, V., Hamkins, J. D., & Karagila, A. (2019). Kelley–Morse set theory does not prove the class Fodor principle. *Fundamenta Mathematicae*, **254**(2021), 133–154.
- [17] Habič, M. E., Hamkins, J. D., Klausner, L. D., Verner, J., & Williams, K. J. (2019). Set-theoretic blockchains. *Archive for Mathematical Logic*, **58**(7–8), 965–997

- [18] Hamkins, J., Kirmayer, G., & Perlmutter, N. (2012). Generalizations of the Kunen inconsistency. *Annals of Pure and Applied Logic*, **163**(12), 1972–1890
- [19] Hamkins, J. D. (2003). A simple maximality principle. *Journal of Symbolic Logic*, **68**(2), 527–550
- [20] ———. (2012). The set-theoretic multiverse. *Review of Symbolic Logic*, **5**(3), 416–449
- [21] ———. (2016). Is the universality of the surreal number line a weak global choice principle? MathOverflow. Available from: <https://mathoverflow.net/q/227849> (version: 2017-04-13).
- [22] ———. (2018). The modal logic of arithmetic potentialism and the universal algorithm. Preprint, [arXiv:1801.04599](https://arxiv.org/abs/1801.04599), pp. 1–35. Under review.
- [23] Hamkins, J. D., Leibman, G., & Löwe, B. (2015). Structural connections between a forcing class and its modal logic. *Israel Journal of Mathematics*, **207**(2), 617–651
- [24] Hamkins, J. D., Linetsky, D., & Reitz, J. (2013). Pointwise definable models of set theory. *Journal of Symbolic Logic*, **78**(1), 139–156.
- [25] Hamkins, J. D., & Linnebo, O. (2018). The modal logic of set-theoretic potentialism and the potentialist maximality principles. *Review of Symbolic Logic*, **15**, 1–35.
- [26] Hamkins, J. D., & Löwe, B. (2008). The modal logic of forcing. *Transactions of the American Mathematical Society*, **360**(4), 1793–1817
- [27] Hamkins, J. D. and Williams, K. J. (2021). The  $\Sigma_1$ -definable universal finite sequence. *Journal of Symbolic Logic*, **87**, 783–801.
- [28] Hamkins, J. D. and Woodin, W. H. (2018). Open class determinacy is preserved by forcing. Preprint, [arXiv:1806.11180](https://arxiv.org/abs/1806.11180), pp. 1–14. Under review.
- [29] Hellman, G., & Feferman, S. (2000). *Challenges to Predicative Foundations of Arithmetic*. Cambridge: Cambridge University Press, pp. 317–338.
- [30] Holy, P., Krapf, R., Lücke, P., Njegomir, A., & Schlicht, P. (2016). Class forcing, the forcing theorem, and Boolean completions. *Journal of Symbolic Logic*, **81**(4), 1500–1530
- [31] Holy, P., Krapf, R., & Schlicht, P. (2018). Characterizations of pretameness and the Ord-cc. *Annals of Pure and Applied Logic*, **169**(8), 775–802
- [32] Hossack, K. (2000). Plurals and complexes. *British Journal for the Philosophy of Science*, **51**(3), 411–443
- [33] Kunen, K. (1971). Elementary embeddings and infinitary combinatorics. *Journal of Symbolic Logic*, **36**, 407–413.
- [34] Lewis, D. (1991). *Parts of Classes*. Oxford: Blackwell.
- [35] Linnebo, Ø. (2006). Sets, properties, and unrestricted quantification. In Uzquiano, G. and Rayo, A., editors. *Absolute Generality*. Oxford: Oxford University Press, pp. 149–178.
- [36] ———. (2013). The potential hierarchy of sets. *Review of Symbolic Logic*, **6**(2), 205–228.
- [37] Maddy, P., & Meadows, T. (2020). A reconstruction of Steel’s multiverse project. *Bulletin of Symbolic Logic*, **26**, 118–169.
- [38] Marek, W., & Mostowski, A. (1975). *On Extendability of Models of ZF Set Theory to the Models of Kelley–Morse Theory of Classes*. Berlin–Heidelberg: Springer, pp. 460–542.



- [39] McAloon, K. (1971). Consistency results about ordinal definability. *Annals of Mathematical Logic*, **2**(4), 449–467
- [40] 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*, Studies in Logic and the Foundations of Mathematics, vol. 84. Amsterdam: North-Holland, pp. 325–340. [https://doi.org/10.1016/S0049-237X\(09\)70288-5](https://doi.org/10.1016/S0049-237X(09)70288-5)
- [41] Parsons, C. (1974). Sets and classes. *Nous*, **8**(1), 1–12
- [42] Roberts, S. (MSa). The iterative conception of properties. Unpublished manuscript.
- [43] ———. (MSb). Properties and truth. Unpublished manuscript.
- [44] 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.
- [45] Scambler, C. (2021). Can all things be counted? *Journal of Philosophical Logic*, **50**, 1079–1106.
- [46] Simpson, S. G. (2009). *Subsystems of Second Order Arithmetic*, Perspectives in Logic. New York: Association for Symbolic Logic.
- [47] Steel, J. (2014). Gödel’s program. In Kennedy, J., editor. *Interpreting Gödel*. Cambridge: Cambridge University Press.
- [48] Suzuki, A. (1999). No elementary embedding from  $V$  into  $V$  is definable from parameters. *Journal of Symbolic Logic*, **64**(4), 1591–1594.
- [49] Uzquiano, G. (2003). Plural quantification and classes. *Philosophia Mathematica*, **11**(1), 67–81.
- [50] Vickers, J., & Welch, P. (2001). On elementary embeddings from an inner model to the universe. *Journal of Symbolic Logic*, **66**(3), 1090–1116.
- [51] Williams, K. J. (2018). The structure of models of second-order set theory. PhD Thesis, The Graduate Center, CUNY.
- [52] ———. (2019). Minimum models of second-order set theories. *Journal of Symbolic Logic*, **8**(2), 589–620
- [53] Zermelo, E. (1930). On boundary numbers and domains of sets. In Ewald, W. B., editor. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Volume II*. Oxford: Oxford University Press, pp. 1208–1233.

IFIKK, UNIVERSITETET I OSLO

POSTBOKS 1020, BLINDERN

0315 OSLO, NORWAY

E-mail: [n.a.barton@ifikk.uio.no](mailto:n.a.barton@ifikk.uio.no)

URL: <https://neilbarton.net/>

BARD COLLEGE AT SIMON’S ROCK

DIVISION OF SCIENCE, MATHEMATICS, AND COMPUTING

84 ALFORD ROAD

GREAT BARRINGTON, MA 01230, USA

E-mail: [kwilliams@simons-rock.edu](mailto:kwilliams@simons-rock.edu)

URL: <https://kamerynjw.net>