Varieties of class-theoretic potentialism

Neil Barton and Kameryn Williams

09 March 2022

Introduction

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.

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

In this paper, we 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.

Main Claims:

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

Contents

1 Motivating Class-Theoretic Potentialism

Bottom-up approaches. Slowly individuate more classes through some sort of 'process'.

Analogy. Linnebo/Studd/Scambler style potentialism.

Top-down approaches. Look at the modal structure we get out of indeterminacy of reference.

Analogy. Hamkins-style potentialism.

1.1 Bottom-up approaches

Liberal Predicativism

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

Class Predicativism Classes are extensions of predicates.

Then, if the predicates we have are potential, so will the classes be potential.

> ...it seems evident that the "totality" of possible predicates is irremediably potential... ([Parsons, 1974, p. 8])

Of course what counts as a legitimate predicate is going to be important (we discuss it later).

Questions?

Property Potentialism

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 (cf. [Linnebo, 2006]).

A restriction is therefore needed, and one suggestion, made by both [Linnebo, 2006] and [Fine, 2005a], is that the application relation for properties is successively individuated.

Linnebo individuates properties along the ordinals, and thereby uses a length 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.

Linnebo's theory and Fine's theory are similar from a mathematical perspective. The strength of these theories has recently been examined by Roberts (in [Roberts, Ua]).

For both Fine and Linnebo, given this idea of successively individuating the application relation, there will be new classes appearing as the application relation progressively individuates certain properties as holding of more and more sets.

For instance as we move through the first few levels of individuation, we will individuate a truth-predicate for the language of ZFC (after the first stage), and then a truth predicate for the expanded language, and so on. (For details, see [Roberts, Ub].)

Questions?

Postulationism

A different bottom-up approach can be obtained from Kit Fine's 'procedural postulationism' (as in [Fine, 2005b]).

According to Fine, we gain knowledge of mathematical objects by *postulating* their existence.

For Fine, however, postulation amounts to more than the mere postulation of a truth of a proposition, rather it concerns providing a rule for the construction of a particular entity.

Specifically, he introduces the following rules (letting $C(x)$ be some condition on objects):

Introduction $!x.C(x)$

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

The complex rules are more complicated, and won't play too much of a role in what we have here.

They contain the things you'd expect—do A then do B, do A α-many times etc.

Fine suggests that an imperatival logic can be obtained for these conditions, but again the details are not important (they are also not clear to us).

What is salient in this context is that his postulationism leads naturally to class-theoretic potentialism.

Assume that the sets are given (which might be themselves previously constructed through postulational acts).

We can then consider postulational rules for introducing classes, such as"add a truth predicate!".

Such postulational acts are naturally thought of in potentialist terms.

Obviously certain imperatival rules (e.g. "Introduce a number that is both equal and unequal to 0!") will be inconsistent.

Fine attempts to sharpen his account by making explicit the kinds of rules we are allowed implement, and we'll discuss this later.

Questions?

Others? e.g. mereological potentialism...

Just to review, the key facets of bottom-up approaches are:

Initial World We obtain classes *beginning* with some *initially* specified classes and then...

Individuation ... we *individuate* new classes over the existing classes.

1.2 Top-down approaches

A different route to class-theoretic potentialism is *top-down* in nature.

Instead of *starting* with some antecedently given collection of classes and iterating a process of individuation, we might instead view potentialism arising out of *referential indeterminacy*.

This is the core approach of *top-down* views: We state some conditions we would like domains of classes to satisfy, but it may be that there is no single domain that is thereby referred to.

We can then take class-theoretic potentialism to be telling us how we may move around within these domains that satisfy our basic class-theoretic principles.

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, then class-theoretic potentialism considers multiversally-interesting set-theoretic structures.

For example given a universe V, we can always make V countable by moving to a universe V' in which V appears as a set, and then collapsing $|V|$ to ω by forcing over V' (call this universe $V'[G]$).

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

If one thinks that the concept of *arbitrary set* is indeterminate (and holds a multiversism on these grounds), one is likely to hold also that our concept of *class* is also indeterminate.

Thus, even if I fix some universe V as a starting position (within the multiverse), it is unlikely to be determinate exactly what classes exist over V.

Questions?

Plurals and Potentialism.

Plural resources have been used to interpret proper-class talk (see, for example, [Uzquiano, 2003]).

Often the ranges of plural variables are taken to be determinate (e.g. [Hossack, 2000], [Uzquiano, 2003]).

However, this view has recently been challenged by the work of Florio and Linnebo (in [Florio and Linnebo, 2016]) who show that there are versions of Henkin semantics for plural logic, and argue that this calls into question the determinacy of plural quantification.

If one does accept that such resources are indeterminate, one might be able to motivate a class-theoretic potentialism via the plural interpretation of classes.

Simply put, 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.

Questions?

Top-down approaches are based on the following two ideas:

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

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

2 Class-theoretic potentialist systems

2.1 Class-theoretic principles

We will use a two-sorted approach to class theory, with *sets* and *classes* as the two types of objects.

A model of class theory will be denoted $(M, \mathcal{X})$, where M is the sets and $\mathcal{X}$ is the classes. We are interested in transitive models, for whom their membership relation is the true $\in$, and will suppress the membership relation in the notation.

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

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

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

Definition 2. *Global choice* is the assertion that there is a global choice function for all nonempty classes.

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

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

Theorem 4 ([Fujimoto, 2012]). *Over* NBG, ETR *is equivalent to the assertion that iterated truth predicates*[1] *of any length relative to any class parameter always exist.*

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

Note: ETR is implied by strong systems like MK.

Questions?

2.2 Explaining the systems

A *potentialist system* is a collection of structures of the same type, ordered by a reflexive and transitive relation which refines the substructure relation.

They give a formalisation of a domain we can think of as dynamically growing.

In this section, we will provide the basic definitions for the potentialist systems we plan on considering.

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

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

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

Note: We'll let context dictate when our potentialist system is nontrivial.

This idea of considering T-class potentialist systems corresponds to *top-down* approaches.

Questions?

Bottom-up approaches to potentialist systems, by contrast, specify rules for extension, and then consider which potentialist systems satisfy these rules.

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

Rather than starting from the outset with a fixed potentialist system, we consider axioms governing the accessibility relation and what extensions must exist, and then ask which, if any, potentialist systems satisfy these axioms.

For example, we will consider what happens when we have extensions which arise from the addition of truth-predicates or, by taking certain class-forcing extensions which do not add sets.

Some of the views we consider motivate further restrictions on the accessibility relation. For instance, the systems of Linnebo and Fine both have built in that the accessibility relation must be well-founded, since for them the process of property-theoretic membership individuation is well-founded.

Questions?

3 Modal logics of class-theoretic potentialism

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

An easy observation is that S4 is valid for any potentialist system.

Some other modal axioms we consider in this article are:

$(.2)\Diamond\Box p \rightarrow \Box\Diamond p$

$(.3)(\Diamond p \wedge \Diamond q) \rightarrow \Diamond[(p \wedge \Diamond q) \vee (\Diamond p \wedge q)]$

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

The failure of .2 has two important ramifications, one philosophical and one mathematical (though perhaps they are different manifestations of the same state of affairs).

On the philosophical side, non-convergence indicates a kind of further indeterminacy in our concepts.

It is not just that the sets or classes themselves are *modally* indefinite, but there are also important choices about what is *possible* to be made within this modal space, ones that cannot be reversed.

Questions?

The second (more technical) point to be made is that it creates obstacles for proving mirroring theorems, as in [Linnebo, 2013] or [Hamkins and Linnebo, 2018].

Note: Of course, there's some new technology here (that we've seen in these seminars). It's a good question as to how well these map on to the issues presently under consideration.

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, to not 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 interesting structural properties of the relevant potentialist system (and perhaps the underlying concepts), but this feature is unproblematic.

Given the information that we get from knowing a potentialist system's modal validities, we want tools which 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 us briefly describe the main tools, *control statements*, used to compute upper bounds.

Here they are for reference, but let's not go through them in detail.

Definition 6 ([Hamkins and Löwe, 2008, Hamkins et al., 2015])**.**

- A *button* is an assertion β so that $\Diamond\Box\beta$ holds at every world. If $\Box\beta$ holds at a world M, we say β is *pushed* for M, otherwise we say β is *unpushed*. The intuition is, you can push a button, making β true forevermore, but once you push it you can never unpush it.

- A *switch* is an assertion σ so that $\Diamond\sigma$ and $\Diamond\neg\sigma$ holds at every world. The intuition is, you can toggle the truth value of σ freely back and forth.

- A *ratchet* is a finite sequence $\rho_0, \ldots, \rho_n$ of buttons so that pushing ρ_i pushes ρ_j for all $j < i$. The intuition is, you can ratchet forward but never back.

- A *long ratchet* of length Γ is is a uniformly definable sequence of buttons r_ξ, indexed by $\xi < \Gamma$, so that pushing r_ξ pushes r_η for all $\eta < \xi$ and so that in no world are all buttons on the ratchet pushed. Observe that this second condition forces the ratchet to have limit length, as if there were a last button then we could push it to push all the buttons.

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

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

Theorem 7 ([Hamkins and Löwe, 2008])**.** *If a potentialist system admits arbitrarily large finite families of independent buttons and switches then its modal validities are contained within* S4.2*.*

Theorem 8 ([Hamkins et al., 2015])**.** *If a potentialist system admits arbitrarily long ratchets which are independent with arbitrarily large families of switches then its modal validities are contained within* S4.3*.*

Corollary 9. *If a potentialist system admits a long ratchet whose length* Γ *is closed under addition* $<\omega^2$ *then* S4.3 *is an upper bound for its modal validities.*

Questions?

4 Truth predicates and potentialism

OK let's see the results about *truth predicates*.

We'll see failures of the .2 and .3 axioms for certain potentialist systems.

Let us begin by fixing some notation and definitions. We will use capital Greek letters—e.g. Λ, Γ—to refer to class well-orders.

Addition, multiplication, and exponentiation on these are defined as usual.

Consider a fixed transitive $M \models \mathsf{ZF}$.

The *truth predicate* for M is the $Tr \subseteq M$ which satisfies the recursive Tarskian rules for the satisfaction class for $(M, \in)$.

In case we wish to emphasise for which structure Tr is a truth predicate we will write Tr^M.

Given a class $A \subseteq M$ the *truth predicate relative to* A is the unique class $Tr(A) \subseteq M$ which satisfies the Tarskian recursion to be the satisfaction class for $(M, \in, A)$, the expansion of $(M, \in)$ with a unary predicate symbol for A.

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

Working over our fixed transitive $M \models \mathsf{ZF}$, let Λ be a well-order, possibly a proper class. A Λ*-iterated truth predicate* is a class Θ of triples (ξ, ϕ, a) where $(\xi, \phi, a) \in \Theta$ intuitively means that $\phi(a)$ is true at level $\xi < \Lambda$.

Questions?

With these definitions in hand, let us now describe a species of class potentialist systems meant to capture the idea that we can always expand by adding truth predicates. First, a bit of notation.

If $\mathcal{X}$ is a collection of classes over M and A is a class over M then let $\mathcal{X}[A] \subseteq \mathcal{P}(M)$ be the smallest NBG-expansion for M which extends $\mathcal{X}$ and contains A.

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

Definition 10. Say that a class potentialist system over $M \models \mathsf{ZF}$ is a *truth potentialist system* if it satisfies the following three properties.

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

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

Questions?

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 $<Ord$-length truth potentialist systems correspond to the class theory one obtains from Linnebo's property theory where properties are individuated iteratively along Ord; see [Roberts, Ub].

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.

We next present some results about smallest truth potentialist systems over a fixed M.

First, however, let us clarify in what sense a potentialist system may be **smallest** among a collection of systems.

One way to compare potentialist systems is by containment: if $\mathfrak{A}$ and $\mathfrak{B}$ are potentialist systems then $\mathfrak{A} \subseteq \mathfrak{B}$ if every world in $\mathfrak{A}$ is a world in $\mathfrak{B}$.

But this comparison is inadequate for many purposes; for instance, $\mathfrak{B}$ could have more worlds than $\mathfrak{A}$ because it breaks the worlds of $\mathfrak{A}$ into finer-grained worlds .

Say that $\mathfrak{A}$ *covers* $\mathfrak{B}$ if every world in $\mathfrak{B}$ is contained in some world in $\mathfrak{A}$.

If $\mathfrak{A} \subseteq \mathfrak{B}$ and $\mathfrak{A}$ covers $\mathfrak{B}$ then we say that $\mathfrak{B}$ *refines* $\mathfrak{A}$.

A potentialist system is *refined* relative to a collection of systems if it has no proper refinements within the collection.

Given a collection of systems, the *smallest* potentialist system in the collection, if it exists, is the refined system which is covered by every other system.

Theorem 11. *If M admits a truth potentialist system then it admits a smallest truth potentialist system. This potentialist system validates* S4.3.

Proof. (Rough strategy.) Build up iterated truth predicates over $Def(M)$. □

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

Theorem 12. *Fix a length $\Lambda \in Def(M)$ where Λ is well-founded as seen externally from V. If M admits a length $<\Lambda$ truth potentialist system then it admits a smallest one.*

Proof. (Rough strategy.) Same as last time. □

Theorem 13. *Fix a length $\Lambda \in Def(M)$, where Λ is externally seen to be well-founded. Consider the smallest length $<\Lambda$ truth potentialist system $\mathfrak{X}$ on M as constructed in the previous result. Every axiom of* S4.3 *is valid in this potentialist system. Moreover if either $\Lambda \geqslant \omega^2$ is closed under addition $<\Lambda$ or $\Lambda \cdot \omega$ is closed under addition $<\omega^2$ then the modal validities are precisely* S4.3.

Proof. (Rough strategy.) That S4.3 is a lower bound for the modal validities is again the observation that $\mathfrak{X}$ is linearly ordered. Then use long ratchets to get the exact calculation. □

Theorem 14. *Suppose M admits an unbounded truth potentialist system whose every world is a β-model. Then M admits a smallest unbounded truth potentialist system, and this potentialist system validates exactly* S4.3.

Proof. (Rough strategy.) Again, construct in stages and find a long ratchet. □

We next address the existence of **global well-orders**. There are a few ways one might ensure a potentialist system includes global well-orders.

Observe that if M has a definable global well-order then any truth potentialist system, we get truth-potentialist systems with global well-orders which validate exactly S4.3.

However, this puts a firm restriction on the first-order theory of M, namely $M \models \exists x(V = HOD(\{x\}))$.

Here are two ways to approach the general case. First, we could start by building up from a world of the form $(M, Def(M, G))$, where G is a global well-order for M, rather than building up from the definable classes.

Second, we could have the global well-order added in later (so we still start with $(M, Def(M))$ but the global well-order can be added in).

Let's deal with these in turn.

Observation. Some global well-orders are interdefinable with Cohen-generic classes of well-orders.

Theorem 15. *There are truth potentialist systems modified to require a global well-order in the base world whose modal validities are precisely* S4.2*. In particular,* .3 *is invalid for these systems.*

Proof. (Rough strategy.) Chop the generic up and get a directed but non-linear system out of adding truth predicates for these pieces. □

Rather than add a condition asserting that the base world contains a global well-order, we instead allow the addition a global well-order by class forcing.

That is, we want to expand our potentialist system by adding a new rule to get new worlds: if $(M, \mathcal{X})$ is a world then so is $(M, \mathcal{X}[C])$ whenever C is a Cohen subclass of Ord generic over $(M, \mathcal{X})$.

We will start with the definable classes as the smallest world, and keep the old rule about being able to add truth predicates relative to extant classes.

The problem is that this is actually quite destructive.

Specifically, adding a global well-order may kill off the possibility of adding a truth predicate whilst preserving the basic axioms of NBG.

Theorem 16. *Let M be a countable transitive model of* ZF *and let $A \subseteq M$ be a class over M so that $(M, Def(M, A)) \models$* NBG*. Then there is C Cohen-generic over $(M, Def(M, A))$ so that no* NBG*-expansion for M can contain both C and $Tr(A)^M$.*

Proof. (Rough strategy.) Pick a 'bad' G that can be decoded into collapse from Ord to ω when we add the relevant truth predicate. □

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

Questions?

5 The bottom-up approach

Bottom-up approaches 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.

As we remarked earlier (§2) this non-convergence might be viewed as a cost or a feature, depending upon one's outlook. So, how does the 'desideratum' of convergence affect the class-theoretic potentialist?

If we want convergence, smallest systems are good. We know that various kinds of classes can interfere with the addition of truth predicates, resulting in non-convergence. However, the *smallest* systems (in the sense outlined in Theorems 11 and 12) do *not* exhibit branching, and validate S4.3.

There is a question as to how well this response meshes with the various philosophical views considered in §1. On the one hand, the predicativist who is only interested in adding truth predicates may have some motivation to take this position.

Whilst it is somewhat contingent upon the nature of the space of possible language expansions, it seems reasonable to assume that when we introduce individual new truth predicates into our language we do not thereby introduce further predicates beyond what is required by (i.e. definable in) the expansion. In this case, one clearly obtains the smallest such system any time one introduces a new truth predicate.

For property theories, we can simply note that the generation of properties is (by construction) limited to entities definable in a specific way.

The new properties available at each additional stage are those whose application relation is definable over previous stages.

One can see this as constructing a class-theoretic version of the constructible hierarchy, call it the $L(V)$ hierarchy, generating more and more classes by iterating the definability operator.

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

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

Questions?

Larger systems often admit branching.

A critical point, in contradistinction to the foregoing, is that for larger systems we *do* get branching.

If a certain kind of richness is needed or wanted by the bottom-up theorist, and in particular if they wish to transcend smallest systems, we often can get branching in those systems.

The larger systems we considered can be seen as larger in *width*, not larger in *height*.

Whilst our intention here is to bring to mind the familiar width versus height distinction for sets, the notion is different here, since all classes have the same height in the sense of ordinal rank.

Here, height refers to the length of truth predicates, which one can think of as corresponding to how far one can build $L(V)$ in the classes—cf. earlier discussion in this section.

One might view these two observations (concerning smallest and larger systems) as a point in favour of the pictures articulated by the versions of bottom-up truth-theoretic potentialism we have considered here.

If one views branching as a cost, one way to ensure branching is avoided is to consider smallest potentialist systems. As it turns out, this is precisely what the truth-theoretic versions of liberal predicativism, property theory, and postulationism motivate (the former two since they just involve adding truth predicates and closing under definability, and the latter because we add truth predicates in such a way that no more than is necessary is added).

Thus for these views there is conformity between desirable properties of the potentialist system and the details of what the relevant philosophical view motivates. As we shall see, however, allowing class-forcing greatly complicates the issue.

Questions?

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 $\exists x V = HOD(\{x\})$.

Another possibility is that the global well-order is generic, in the sense of class forcing. (See the discussion in §4 of how to force to add a global well-order.)

Under various motivations, a potentialist might not want a global well-order which codes complicated undefinable classes, and instead want it to be "random" with respect to the definable classes.

This amounts to asking it to be generic; extending by a generic global well-order is adding the well-order and, through the use of forcing-names, closing off under definability from the well-order and classes in the ground model.

But in this case, as we noted in Theorem 16, there is no *prima facie* guarantee that the generic not be a bad one which kills off the addition of truth predicates.

One response to this predicament is to require worlds to satisfy a theory which ensures the existence of all desired truth predicates. For example, if our first world satisfies $\mathsf{NBG} + \mathsf{ETR}$ then all the truth predicates are already there, and so a bad 'truth-killing' well-order cannot also be there.

This, however, incurs the cost that the truth-theoretic potentialism is essentially trivial—all paramaterised truth predicates are there from the get-go.

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.

And of course, one possibility available to the class potentialist is to simply give up on having a global well-order. However Global Choice is *mathematically useful* (cf. Fujimoto's motivations).

Moreover the introduction of a generic does not postulate 'global' considerations about the structure of all classes (e.g impredicative comprehension).

> Liberal predicativism does well in most cases, when it comes to an axiom asserting something about a specific type of class... In contrast, if an axiom asserts something strong about the entire structure of classes or the totality of classes, liberal predicativism might be faced with a difficulty. ([Fujimoto, 2019, p. 225])

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

Of course different analyses of what constitutes a legitimate postulationist conditions may arbitrate the status of generics and/or global well-orders differently. However the fact that the relevant notions are easily formalisable (and, moreover, the practice is *consistent* as long as we really do have a ground model) at least gives us *some* reason to accept their possible existence for the postulationist.

On the other hand, one might take some of our results to suggest that the postulationist requires greater clarity on what constitutes a legitimate condition.

Two rules for adding new objects might be independently consistent, but jointly incompatible. For example, the postulationist might have one rule saying "Introduce a truth predicate" and another saying "Introduce a generic global well-order".

Independently, each is fine.

But ask the genie for a global well-order and—genies being tricksters—you may get a global well-order which closes off the possibility of later adding a truth predicate, as in Theorem 16.

The situation is different for the property theorist. As noted earlier, the classes obtained by the property theorist are essentially those obtained by building L over V (i.e. $L(V)$). But given this, whilst some generics over our initial model are obtainable, we will never get a generic that can *conflict* with a truth predicate and *arbitrary* generics are prohibited (only those that cant be obtained in some $L_\alpha(V)$ are legitimate).

Thus, the property theorist rules out non-convergence by keeping a strict control on the classes that could exist. (One might, of course, view this as a cost—especially if one wants to enforce as few restrictions as possible on the classes that one is allowed to form.)

What emerges from this discussion is that there is the following tension at the heart of of bottom-up approaches. If we (i) regard non-convergence as a cost, (ii) want to allow the addition of truth predicates, and (iii) wish to allow unrestricted addition of generics, then we have a problem. The property theorist resolves this issue by rejecting (iii). This problem plausibly bites for the predicativist and the postulationist, and we will suggest a solution via additional modal principles (rejecting (iii)) in §8.1).

Questions?

6 The top-down approach

In this section we discuss what the mathematical results from §4 say about top-down approaches to class potentialism, and in particular the interplay between what is satisfied on these pictures, **Referential Indeterminacy**, and **Interrelation of Interpretations**.

Recall that the formalisation for top-down approaches is to fix a countable model $M \models \mathsf{ZFC}$ and consider the potentialist system consisting of all expansions of M to a model of T, where T is a class theory.

Let's discuss truth predicates first. We know that asking to have any world with a truth predicate puts a limitation on the choice of M. This limitation has negligible cost.

More substantively, these tools do not apply to any choice of T.

The results in §4 were stated in terms of iterated truth predicates. There is a limit to how far this generalises.

If T outright proves the existence of iterated truth predicates of any length—that is, if T proves ETR—one cannot have a nontrivial truth potentialist-like system whose worlds are models of T.

Let's now discuss forcing. This puts a limit on the worlds—we need that generics exist—which we handled by the simplifying assumption that worlds are countable.

We assume this technology reflects upward to apply also to the true universe V; see §7 for further discussion of this move.

It also puts restrictions on T. As discussed in §4 adding a Cohen-generic class of ordinals adds a global choice function.

So these results do not speak to the class theorist who holds global choice is definitely false at every world, if any such exist.

This also rules out the inclusion of axioms that limit the classes by definability.

Here's an illustrative toy example. Let T^- be NBG plus the assertion that length n iterated truth predicates exist for any finite n.

In T^- we can express "every class is definable from Tr_n for some finite n".

Call T the theory you get by adding this assertion to T^-.

Then nontrivial forcing destroys T, and so the results in §4 do not apply to T.

Indeed, T is categorical over a fixed transitive M: given a transitive model M of ZFC there is at most one T-expansion for M.

Thus, if we are considering some class theory T over the universe V such that (i) T does not require the classes to be thin (in the sense of the previous paragraph), (ii) T does not prove that arbitrary iterated truth predicates exist, and (iii) T does not have a principle implying global choice is false, then Theorems 15 and 16 both give information about the top down potentialist system for T on M.

Namely, Theorem 15 gives failures of .3 in this system, and Theorem 16 gives failures of .2 in this potentialist system.

For such a potentialist, these results illuminate in a concrete way how the indeterminacy of reference underlying top-down potentialism might manifest.

Exactly as in §5, weak base theories (like NBG) seem to correspond to a conception of class that is radically divergent, it is just that in this context this radical divergence is underwritten by **Referential Indeterminacy** and **Interrelation of Interpretations** rather than what is built up from **Initial World** and **Individuation**.

This observation is interesting for both our motivations for the top-down approach.

Let's examine the **set-theoretic multiversist** first.

Known examples of S4-style branching require non-standard models (e.g. via the universal finite sequence).

This kind of potentialism exhibits failures of .2 and .3 without any non-well-foundedness.

Questions?

For the theorist who holds that **plural quantification** is indeterminate, the situation is subtle.

On the one hand, plural logic (in its standard formulation) contains all impredicative instances of the plural comprehension scheme and indeed the Henkin interpretations for plural logic obtained by [Florio and Linnebo, 2016] all satisfy it (they restrict to what they call *faithful* models—those that satisfy every instance of the comprehension scheme).

This can then be leveraged to provide an interpretation of MK class theory (as in [Uzquiano, 2003]).

This interpretation can be carried through whether or not the range of the plural quantifiers is determinate.

As noted earlier, MK violates the presuppositions required to make our arguments go through since it trivialises truth-theoretic potentialism.

However: As [Florio and Linnebo, 2016] note, often the impredicative plural comprehension scheme is motivated by the assumption that every non-full Henkin semantics for the plural quantifiers is unintended (e.g. [Hossack, 2000]).

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

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

e.g.

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

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

Throwing in impredicative plural quantification at the start simply prejudices the debate in favour of MK, and the water becomes much muddier once we allow many different interpretations of the range of the plural variables.

Nonetheless, the view that Lewis' thought about pluralities is somehow *part of our conception of pluralities* is tempting.

Even if we think the quantification is indeterminate, we might think that *within a world* his intuition should motivate us to accept impredicative plural quantification over that world, yielding MK locally.

Questions?

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

One tempting way of objecting to the import of our results is to point to our use of countable transitive models.

One might object: For many species of class-theoretic potentialist V is uncountable, and so there is no such generic.

There are some points to be made about this objection.

Accepting this response entails you accept that apparently perfectly good parts of model theory cannot tell you about the multiverse proper (or at least their use must be justified).

This in itself, is a substantial cost (without further argument) and goes against much of the practice in the field.

Questions?

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.

8.1 Additional modal principles?

As discussed in §6, the failure of .2 for top-down potentialism for weak class theories is particularly destructive, being witnessed by a world which cannot be further extended to add in a certain truth predicate.

A top-down potentialist may very well think this catastrophic world is an artifact of the formalisation, one which does not occur in the real multiverse of classes. Her task then is to explain why this phenomenon does not occur and formulate principles prohibiting these worlds.

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

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

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

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

An example of a different flavor, one closer in motivation to what we give here, can be found in [Steel, 2014] (with subsequent development by [Maddy and Meadows, 2020]).

Steel is investigating a multiversist framework arising from set forcing. Given a countable model of set theory, it has pairs of Cohen extensions which do not *amalgamate*—there is no outer model which contains both Cohen extensions as submodels [Mostowski, 1976].

To exclude this phenomenon, Steel includes an axiom asserting that models in the generic multiverse always amalgamate.

His principle is in fact higher order, referring to worlds as objects, not just to what is true of sets within each world. And one could also consider higher order principles in the context of class potentialism.

8.2 An analogy to second-order arithmetic, and universal finite sequences?

A potential area for further study concerns the analogy between the use classes in the contexts of second-order arithmetic and set theory.

Predicativism in mathematics often takes the totality of natural numbers as given, with the predicatively-given “classes” then being sets of natural numbers, e.g. [Feferman and Hellman, 1995, Hellman and Feferman, 2000].

There has been work addressing to what extent results about predicativism over ω generalise to predicativism over V—see e.g. [Fujimoto, 2012, Sato, 2014].

Similar to how it was formalised in the set theoretic context, one could formalise potentialism over ω by considering potentialist systems of ω-models of second-order arithmetic.

To what extent does the mathematical and philosophical work about class potentialism carry over to the arithmetic context?

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

Theorem 17 (Hamkins–Williams). *Let T be a computably axiomatizable extension of $\mathsf{Z}_2 + \Pi^1_\infty$-Bounding. Then the modal validities of the potentialist system consisting of countable ω-models of T*

are precisely S4*, whether or not we allow parameters in formulae.*

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

Question 1. Let T be MK plus Class Bounding and let M be a countable transitive model of ZFC which has a nontrivial T-class potentialist system. Does the potentialist system consisting of countable T-expansions for M have S4 as its modal validities?

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[Maddy and Meadows, 2020] Maddy, P. and Meadows, T. (2020). A reconstruction of Steel's multiverse project. *The Bulletin of Symbolic Logic*, page 1–79.

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

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

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

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

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

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

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

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