Journal for the Philosophy of Mathematics

OPEN ACCESS

Citation: INCURVATI, Luca (2025). On Class Hierarchies. *Journal for the Philosophy of Mathematics*. 2: 45-74. doi: 10.36253/jpm-3459

Received: April 15, 2025

Accepted: June 22, 2025

Published: December 30, 2025

ORCiD
LI: 0000-0001-7381-7378

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Firenze University Press
www.fupress.com

On Class Hierarchies

LUCA INCURVATI

Department of Philosophy and Institute for Logic, Language and Computation, University of Amsterdam.
Email: l.incurvati@uva.nl

Abstract: In her seminal article 'Proper Classes', Penelope Maddy introduced a novel theory of classes validating the naïve comprehension rules. The theory is based on a step-by-step construction of the extension and anti-extension of the membership predicate, which mirrors Kripke's construction of the extension and anti-extension of the truth predicate. Maddy's theory has been criticized by Øystein Linnebo for its 'rampant indeterminacy' and for making identity among classes too fine-grained. In this paper, I present a theory of classes which, while building on Maddy's theory, avoids its rampant indeterminacy and allows for identity among classes to be suitably coarse-grained. I begin by presenting a bilateral natural deduction system for Maddy's theory, which improves on her axiomatization in several respects. I then go on to show how to avoid the rampant indeterminacy by using supervaluational schemes in the construction of the extension and anti-extension of the membership predicate and how to augment the proof theory with corresponding, motivated rules. It turns out that whilst a van Fraassen-style supervaluational scheme suffices to avoid the basic problem of rampant indeterminacy, a supervaluational scheme based on maximally consistent extensions is needed for a proper treatment of identity.

Keywords: Classes, hierarchies, supervaluation, Maddy, identity, maximal consistency, bilateral logic.

1. Introduction

Despite early doubts about their legitimacy—doubts that, in the philosophical literature, continue to this day—sets have become the bread and butter of mathematics. And despite doubts about its cogency and limitations, the iterative conception of set—according to which sets are the objects obtained by iterated applications of the *set of* operation—has become the received view of sets in mathematics and philosophy.[1] Even if one accepts this received view, one might well wonder: are there collections other than sets?

[1] For an extended defence of the iterative conception against rival conceptions of set, see Incurvati, 2020.

George Boolos forcefully advocated a negative answer—a 'settist' position, as he liked to call it. In response to Charles Parsons's (1974) proposal for a theory of sets *and* classes, he famously complained:

> Wait a minute! I thought that set theory was supposed to be a theory about all, 'absolutely' all, the collections that there were and that 'set' was synonymous with 'collection'. (Boolos, 1974, 35)

Those that defend a positive answer—those who argue that we should countenance collections other than sets—typically do so on the grounds that they are needed to perform tasks that sets are unable or less suited to accomplish. Indeed, it is in part because of his settist convictions that Boolos embarked upon his programme of providing a plural interpretation of second-order logic: our apparent talk about what appear to be collections that cannot form a set, such as the collection of all ordinals and the collection of all sets, should be seen as plural talk in disguise.

One of the philosophers to have most clearly presented some of the reasons for admitting collections other than sets is Penelope Maddy. In her seminal article 'Proper classes' (Maddy, 1983), she marshalled a series of mathematical and philosophical considerations in support of the existence of classes and indeed proper classes—collections that are too big to form a set. On the basis of these considerations, Maddy went on to introduce a novel theory of classes, whose mathematical properties she investigated further in later work (Maddy, 2000).[2]

Maddy's theory is based on a step-by-step construction of the extension and anti-extension of the membership predicate, which mirrors Kripke's (1975) construction of the extension and anti-extension of the truth predicate. The theory has been criticized by Øystein Linnebo for its 'rampant indeterminacy' and for making identity among classes too fine-grained. In this paper, I present a theory of classes that builds on Maddy's theory but avoids its rampant indeterminacy and allows for identity among classes to be suitably coarse-grained.

I begin by providing an overview of the roles that classes have been invoked to play and isolating some key desiderata on a theory of classes to which these roles give rise. I then present Maddy's theory of classes and provide a bilateral natural deduction system for the theory, which improves on Maddy's own axiomatization in several respects. I go on to describe the problem of rampant indeterminacy for Maddy's theory, and I show how it can be avoided by using supervaluational schemes in the construction of the extension and anti-extension of the membership predicate. It turns out, however, that whilst a van Fraassen-style supervaluational scheme suffices to avoid the basic problem of rampant indeterminacy, a supervaluational scheme based on maximally consistent extensions is needed for a proper treatment of identity. For both supevaluational theories, I also provide bilateral natural deduction systems. Indeed, the proof theory will play a central role in motivating the treatment of identity and in answering a challenge of Greg Restall (2010) to any coarse-grained account of identity in naïve class theory.

2. Desiderata on classes

Classes have been invoked to play a variety of roles. A detailed examination and assessment of these roles is beyond the scope of this paper. My goal is to provide an overview of the most

[2] As Maddy (2024) points out, the publication date of Maddy, 2000 is misleading: it was written in the mid-80s. It should also be noted that Maddy has since then come closer to endorsing a settist position herself, on the grounds that classes do not appear to provide any real mathematical contribution to the theory of sets. Nonetheless, she is open to the possibility of classes being needed for different endeavours, such as formal semantics. See Maddy, 2024, 82–83.

prominent roles, with the aim of lending support to the desiderata on a theory of classes we will rely on.[3]

In general, classes have been called upon to play roles that sets have been deemed unsuited to serve. First, it has been argued that classes are *genuinely used* in set-theoretic practice (Barton and Williams, 2024; Linnebo, 2006; Parsons, 1974). Upon opening a set theory textbook, one immediately comes across terms such as V, L and Ω, which intuitively stand for collections (namely, the collection of all sets, the collection of all constructible sets and the collection of all ordinals) that are too big to form a set and should therefore be considered classes. It is however common place to regard this apparent talk of classes as a a way of stating facts that only involve sets. For instance, when a set theorist writes '$x \in \Omega$', this ought to be taken as a shorthand for 'x is a transitive set linearly ordered by $\in$'. To use Quinean terminology (Quine, 1948), classes are convenient myths, to be paraphrased away. However, it is not clear how far the paraphrase strategy can be taken. A case that is especially difficult to handle concerns the formulation and treatment of certain large cardinal axioms and reflection principles. For instance, the paraphrase strategy applied to non-trivial elementary embeddings would seem to trivialize Kunen's celebrate theorem that there is no non-trivial elementary embedding of the universe onto itself (Fujimoto, 2019; Hamkins et al., 2012).[4]

Second, it has been argued that classes are needed to *make sense of* set-theoretic practice. A case is point is the debate on unrestricted quantification (Florio, 2014). In standard model theory, an interpretation is an ordered pair consisting of the domain and an interpretation function, where both of these are sets. But according to the iterative conception, there is no set of all sets. Hence, there can be no interpretation whose domain contains all sets. But this is problematic. For one thing, if the intended interpretation of set theory is the one in which the domain contains all sets, and if truth is truth on the intended interpretation, then it would seem impossible to account for the apparent truth of 'Every set has a power set' or 'No set contains all sets'. For another, if we define logical truth as truth on all interpretations, then a sentence whose quantifiers range over all sets could be logically true and yet false in the universe of sets (Kreisel, 1967).

The conclusion that some (e.g., Lear, 1977; Parsons, 1974) have been willing to draw is that the set-theoretic quantifiers never range over absolutely all sets, and so there is not a single, intended interpretation of set theory. Others have tried to resist this conclusion by jettisoning the assumption that the domain of quantification is a set. But if not a set, what is it? One option is that the domain is not a single entity at all: it is a plurality of things—*the sets* (Boolos, 1985; Rayo and Uzquiano, 1999). Another option is that it is not an object but a higher-order entity of some kind—perhaps a Fregean concept (Rayo and Williamson, 2003). Yet another option is that the domain is a collection-like object other than a set—a class. Classes might therefore allow us to preserve both the idea that the set-theoretic quantifiers range over all sets and that the domain of quantification is an object, as standard model-theoretic semantics would have it.

Third, it has been argued that classes are needed for a treatment of cardinality in line with our basic intuitions about numbers (Maddy, 1983). Cardinal numbers are numbers that represent the size of collections. According to Cantor, a theory of size should be based on the concept of one-to-one correspondence: the cardinality of two sets should be the same just in case they can be put into one-to-one correspondence—a version of Hume's Principle (Frege, 1884). It is

[3] See also Schindler, 2019 for discussion of the roles that classes have been invoked to play.
[4] Another difficult case concerns work on class forcing. See Gitman et al., 2020; Holy et al., 2016.

then natural to take the cardinality of a set a to *just be* the collection of sets that can be put into one-to-one correspondence with a. This definition immediately delivers Hume's Principle, and it is known as the Frege–Russell definition of cardinal number (Frege, 1884; Russell, 1903; see Incurvati, 2020, 33 for discussion). However, the collection of sets that can be put into one-to-one correspondence with a given set is not, in general, a set. It could be a class, however. Classes, therefore, provide a way of rehabilitating a Frege–Russell treatment of cardinality.[5] Similar considerations apply, *mutatis mutandis*, to the development of ordinal numbers by replacing the notion of one-to-one correspondence with the notion of an isomorphism.

Fourth, it has been argued that classes are needed for the development of natural language semantics. In standard model-theoretic semantics, the predicate 'is wise' in 'Andrea is wise' is standardly taken to pick out a function representing a property. However, in English we also have nominalizations of predicates (such as 'wisdom' in 'Wisdom is a virtue'), which would seem to purport to refer to properties in nominal position. It is widely acknowledged that nominalizations are very useful, and there are strong reasons for thinking they are not always eliminable (see Button and Trueman, 2024). One may be tempted to take this to show that properties are objects, and theories that treat properties as objects have been devised (Aczel, 1980; Bealer, 1982). However, there is a longstanding philosophical tradition, going back to Frege (1892) and recently revived by Jones (2016) and Trueman (2021), for regarding the idea of treating properties as objects as incoherent. But even if properties are not objects, there seems still to be the need for objects which serve as first-order proxies for properties and be the referents of nominalizations (Chierchia and Turner, 1988, Incurvati, 2020: Ch. 7). These objects cannot be sets since, among the predicates of our language, there are predicates such as 'is a set' or 'is an ordinal', whose nominalizations cannot refer to sets. Thus, the argument concludes, the referents of nominalizations must be classes.

Having described some key roles that classes have been invoked to play, we can now list a numbers of basic desiderata sanctioned by these roles. First, our discussion should make it clear that there should be 'big' classes such as the class of all ordinals or the class of all sets. Ideally, we would want to validate the naïve comprehension principle that to every condition there corresponds a class of all and only the things satisfying that condition.

Second, classes should be *logical* collections: they are characterized by reference to some predicate, concept or property, which determines its members. This is in contrast with sets, which are *combinatorial* collections: they are characterized not by reference to some predicate, concept, or property, but by reference to their members in a more direct fashion.[6] For one thing, as our discussion has made clear (especially the discussion of the fourth role that classes have been invoked to play), there seems to be a tight connection between classes and predicates or properties, which ought to be vindicated by what we take classes to be. For another, as Maddy (1983, 122) emphasizes, it should be clear why classes are not just 'another stage of sets we forgot to include'. Providing a theory of classes as logical collections allows us to clearly distinguish them from iterative sets, which are combinatorial collections.

Third, as Maddy (1983, 120–123) also demands, classes should be real, well-defined entities. Our discussion of the first role of classes makes it clear why, metaphysical considerations aside,

[5] This is not to say that other strategies are not available. In particular, it is still possible to approximate the Frege–Russell definition within iterative set theory by considering the collection of sets that can be put into one-to-one correspondence with a that occur as low as possible in the cumulative hierarchy. This is known as the Scott-Tarski definition of a cardinal. See Incurvati, 2020, 80.

[6] On the distinction between logical and combinatorial collections, see Maddy, 1990, 121 and Incurvati, 2020, 31.

this is a sensible desideratum on a theory of classes. For if classes are not real entities to be eliminated via paraphrase, it becomes hard to make sense of some aspects of set-theoretic practice such as Kunen's theorem about elementary embeddings.

Finally, if classes are real objects, it should be possible to provide a satisfactory criterion of identity for classes. No entity without identity, after all. Now, *qua* logical collections, it may be unreasonable or even wrong to demand that classes ought to be extensional: identity among classes ought not to be too coarse-grained. But identity among classes ought not be too fine-grained either. I will not belabour this point further, because we will return to it in greater detail below.

3. Maddy's theory of classes

We can now describe Maddy's (1983; 2000) theory of classes. Her key idea is to address the logical paradoxes by allowing certain membership claims involving classes to be indeterminate. Kripke's (1975) theory of truth famously allows certain truth claims to be indeterminate. He outlines his theory by presenting a construction in which the extension and the anti-extension of the truth predicate (that is, the collection of things to which the predicate definitely applies and the collection of things to which the predicate definitely does not apply) are built in stages. Maddy presents her theory of classes by specifying a construction in which the extension and the anti-extension of the membership relation are similarly built in stages. To emphasize the analogies with Kripke's construction, which will be relevant below, I will follow the presentation of Maddy's theory recently provided by Linnebo (2024).

We begin by specifying the object language of the theory. Ultimately, we are going to want to use membership relations involving only sets, which we take as given, to determine membership relations involving classes as well as sets. One option would be to think of these as two different relations; another option is to treat them as the same relation holding between different kinds of entities. Maddy opts for the second option.[7] Thus, we start with the language of first-order logic with identity extended with a membership symbol η (reserving the symbol $\in$ for membership in the meta-language).

Next, we want to have terms for the objects of our theory. We have a hat operator which produces a class term when applied to an open formula. So we have class terms such as $\hat{x}(x = x)$ and $\hat{x}(x \not\eta x)$. To increase the expressive power of the language, Maddy adds a class constant $\bar{V}$, standing for the universe of sets V. On Maddy's account, V is simply the real universe of sets, so it will be a class in the meta-theory. For reasons that will become clear below, it is in fact preferable to let $\bar{V}$ denote a set from the point of view of the meta-theory, which serves as the universe of sets and hence a class from the point of view of the object theory. A natural choice, and the one which we shall stick to, is to let $\bar{V}$ denote V_κ, where κ is the first inaccessible cardinal. Accordingly, our meta-theory will consist of ZFC plus the assertion that there exists an inaccessible cardinal. Finally, Maddy assumes that for each $a \in V$ there is a constant $\bar{a}$.

Formally, the terms and formulae of the language $\mathcal{L}_\eta$ are defined as follows:

(i) All constants and variables are terms.
(ii) If t and t' are terms, then $t = t'$ and $t \eta t'$ are formulae.
(iii) If F and G are formulae and x is a variable, then $(F \wedge G)$, $\neg F$ and $\forall x\, F$ are formulae.

[7] As Linnebo (2024, 72) notes, not much hinges on the choice from a technical point of view, and it would be possible to reformulate (with some adjustments) the discussion to follow using the first option.

(iv) If F is a formula and x is among the free variables of F, then $\hat{x}F$ is a term.

Sentences are closed formulae. In general, F and G stand for (possibly open) formulae, whereas A and B always stand for sentences. We let T be the set of all terms of $\mathcal{L}_\eta$, T^* be the set of all closed terms (a subset of T), C be the set of all class terms and C^* be the set of all closed class terms (a subset of C). Disjunction, the conditional, the biconditional, and the existential quantifier are defined as usual. That is, $F \vee G$ is defined as $\neg(\neg F \wedge \neg G)$, $F \to G$ is defined as $\neg(F \wedge \neg G)$, $F \leftrightarrow G$ is defined as $(F \to G) \wedge (G \to F)$, and $\exists x F$ is defined as $\neg \forall x \neg F$.

Our next step is to provide a model theory for the language. We represent an extension σ^+ and anti-extension σ^- of η as sets of ordered pairs of closed terms, that is subsets of $T^* \times T^*$. Any pair σ of such an extension and anti-extension can be regarded as a model of the Strong Kleene Logic K_3 provided that $\sigma^+ \cap \sigma^- = \emptyset$. In particular, we can recursively define a satisfaction relation $\models$ indicating what the model thinks is true of sets and classes, and an anti-satisfaction relation $=\!\mid$ indicating what the model thinks is false of them:

- $\sigma \models t = t'$ iff $t = t'$ for all $t, t' \in T^*$;
- $\sigma =\!\mid t = t'$ iff $t \neq t'$ for all $t, t' \in T^*$;
- $\sigma \models t \eta t'$ iff $\langle t, t' \rangle \in \sigma^+$ for all $t, t' \in T^*$;
- $\sigma =\!\mid t \eta t'$ iff $\langle t, t' \rangle \in \sigma^-$ for all $t, t' \in T^*$;
- $\sigma \models \neg A$ iff $\sigma =\!\mid A$;
- $\sigma =\!\mid \neg A$ iff $\sigma \models A$;
- $\sigma \models A \wedge B$ iff $\sigma \models A$ and $\sigma \models B$;
- $\sigma =\!\mid A \wedge B$ iff $\sigma =\!\mid A$ or $\sigma =\!\mid B$;
- $\sigma \models \forall x\, F$ iff $\sigma \models F[t/x]$ for all $t \in T^*$.
- $\sigma =\!\mid \forall x\, F$ iff $\sigma =\!\mid F[t/x]$ for some $t \in T^*$.

It is straightforward to show that, similarly to the case of Kripke's theory of truth, satisfaction and anti-satisfaction are monotonic in the following sense: if σ_2 is an *expansion* of σ_1 (in symbols: $\sigma_1 \sqsubseteq \sigma_2$, where $\sigma_1 \sqsubseteq \sigma_2$ iff $\sigma_1^+ \subseteq \sigma_2^+$ and $\sigma_1^- \subseteq \sigma_2^-$), then, for any A, $\sigma_1 \models A$ only if $\sigma_2 \models A$, and $\sigma_1 =\!\mid A$ only if $\sigma_2 =\!\mid A$.

The last step is that of building a sufficiently encompassing extension/anti-extension pair (EA-pair for short). At stage zero, we have all positive and negative membership facts given from set theory and the intuitive meaning of $\bar{V}$. At limit stages, we take unions as customary. At successor stages, just as in the case of Kripke's construction, we apply a jump operation J taking us from an EA-pair σ to another in a monotonic fashion, i.e. so that $J(\sigma)$ is an expansion of σ. Maddy opts for a jump operation which is the exact analogue of Kripke's jump operation for truth: if a term t satisfies a formula F, we add $\langle t, \hat{x}F \rangle$ to the extension of η; if t anti-satisfies F, we add it to the anti-extension.

Definition 3.1 (Maddy jump)**.** Given an EA-pair $\sigma = \langle \sigma^+, \sigma^- \rangle$, the *Maddy jump* J_{M} is the operation such that, for all $t \in T^*$, $J_{\mathrm{M}}(\sigma^+) = \{\langle t, \hat{x}F \rangle \mid \sigma \models Ft\}$ and $J_{\mathrm{M}}(\sigma^-) = \{\langle t, \hat{x}F \rangle \mid \sigma =\!\mid Ft\}$.

We are now ready to recursively define the *Maddy hierarchy*.

Definition 3.2 (The Maddy hierarchy)**.**

$$\sigma_0^{+,\mathrm{M}} = \{\langle \bar{a}, \bar{b}\rangle \mid V_\kappa \models a \in b\} \cup \{\langle \bar{a}, \bar{V}\rangle \mid a \in V_\kappa\};$$
$$\sigma_0^{-,\mathrm{M}} = \{\langle \bar{a}, \bar{b}\rangle \mid V_\kappa \models a \notin b\} \cup \{\langle t, \bar{a}\rangle \mid a \in V_\kappa\} \cup \{\langle t, \bar{V}\rangle\} \text{ for all } t \in C^*;$$
$$\sigma_{\alpha+1}^{+,\mathrm{M}} = J^{+}{}_{\mathrm{M}}(\sigma_\alpha);$$
$$\sigma_{\alpha+1}^{-,\mathrm{M}} = J^{-}{}_{\mathrm{M}}(\sigma_\alpha);$$
$$\sigma_\lambda^{+,\mathrm{M}} = \bigcup_{\alpha<\lambda} \sigma_\alpha^{+,\mathrm{M}} \text{ if } \lambda \text{ is a limit ordinal;}$$
$$\sigma_\lambda^{-,\mathrm{M}} = \bigcup_{\alpha<\lambda} \sigma_\alpha^{-,\mathrm{M}} \text{ if } \lambda \text{ is a limit ordinal.}$$

By a theorem of Flagg (who established a conjecture of Tait's, see Maddy, 2000, 315, fn. 22), the construction of the Maddy hierarchy where V is a set in the meta-theory reaches a fixed point σ^{M} at the first admissible ordinal greater than all the ordinals in V.[8] We can therefore consider σ^{M} as giving the extension and anti-extension of η according to Maddy's theory of classes.

4. A bilateral system for Maddy's theory

At the end of her 1983 paper, Maddy posed the following question:

Question 4.1 (Maddy)**.** How and to what extent can the theory of V^* [the universe of sets and classes in Maddy's theory] be axiomatized?

I am going to provide a sound and complete axiomatization of the fixed point σ^{M} of the Maddy hierarchy. The axiomatization builds on the axiomatization provided by Maddy (2000) (with suggestions from Myhill), but also differs from it in important respects. First, its background logic is bilateral, in that it uses signs for the speech acts of assertion and rejection. This will allow me to formulate the logic in a natural deduction system with rules that are harmonious and separable. Second, the axiomatization allows one to *reason* using Maddy's theory of classes, whereas Maddy's axiomatization only allows one to compute what is in σ^{M}. Concretely, this will put us in a position to provide a general model theory for which the natural deduction system is sound and complete. Finally, the natural deduction system will pave the way towards the theories of classes that I will develop below to address the problems with Maddy's theory.

We begin by characterizing the language of the system. As announced, we are working towards a bilateral system. So we extend the language $\mathcal{L}_\eta$ to the signed language $\mathcal{L}_\eta^S$ by including the signs $+$ and $-$, standing, respectively, for the speech acts of assertion and rejection. The notions of term, formula and sentence are defined as in $\mathcal{L}_\eta$, but, in addition, we also have the notion of a signed sentence, which will be anything obtained by prefixing a sentence with a $+$ or a $-$.

We can now lay down the rules of the natural deduction system. The model-theoretic clauses for conjunction tell us that a conjunction is satisfied just in case both of its conjuncts are, and anti-satisfied just in case at least one of its conjunct is. The proof-theoretic side of this coin,

[8] By contrast, as Maddy (2000, 308) notices and was proved by Tait, the construction has no fixed point if V is the universe of sets from the point of view of the meta-theory.

in terms of assertion and rejection, is captured by the following rules (where φ is a signed sentence):

$$(+\wedge\text{I.})\ \frac{+A \qquad +B}{+A \wedge B} \qquad (+\wedge\text{E.}_1)\ \frac{+A \wedge B}{+A} \qquad (+\wedge\text{E.}_2)\ \frac{+A \wedge B}{+B}$$

$$(-\wedge\text{I.}_1)\ \frac{-A}{-A \wedge B} \qquad (-\wedge\text{I.}_2)\ \frac{-B}{-A \wedge B} \qquad (-\wedge\text{E.})\ \frac{-A \wedge B \qquad \begin{array}{c}[-A]\\ \vdots \\ \varphi\end{array} \qquad \begin{array}{c}[-B]\\ \vdots \\ \varphi\end{array}}{\varphi}$$

The model-theoretic clauses for satisfaction and anti-satisfaction of negated and universally quantified sentences can be similarly mirrored by the following natural deduction rules for the assertion and rejection of those sentences:

$$(+\neg\text{I.})\ \frac{-A}{+\neg A} \qquad (+\neg\text{E.})\ \frac{+\neg A}{-A} \qquad (-\neg\text{I.})\ \frac{+A}{-\neg A} \qquad (-\neg\text{E.})\ \frac{-\neg A}{+A}$$

$$(+\forall\text{I.})\ \frac{+F[t/x]}{+\forall x\, F}\ \text{for all } t \in T^* \qquad (+\forall\text{E.})\ \frac{+\forall x\, F}{+F[t/x]}\ \text{for any } t \in T^*$$

$$(-\forall\text{I.})\ \frac{-F[t/x]}{-\forall x\, F}\ \text{for some } t \in T^* \qquad (-\forall\text{E.})\ \frac{-\forall x\, F \qquad \begin{array}{c}[-F[t/x]]\\ \vdots \\ \varphi\end{array}}{\varphi}\ \text{for any } t \in T^*$$

Note that the quantifiers rules are simply infinitary counterparts of the conjunction rules. The infinitary character of these rules should not be entirely surprising, given that we are attempting to characterize the fixed point of a hierarchy whose first stage includes all positive and negative membership facts settled by V_κ. Indeed, the situation is analogous to the situation for truth. In that case, the first stage of the relevant hierarchies (such as the Kripke hierarchy and supervaluational versions thereof) includes all facts settled by arithmetic. For this reason, axiomatizations of the fixed point of this hierarchies typically include some version of the ω-rule (see Incurvati and Schlöder, 2023; Meadows, 2015).

We now turn to the treatment of identity. Identity is a thorny issue in the context of naïve theories of classes, and we shall return to it below. For now, recall that Maddy's model-theoretic clauses for identity tell us that $t = t$ is always satisfied and $t = t'$ is always anti-satisfied. These clauses can be mirrored in the proof theory by the following introduction rules.

$$(+{=}\text{I.})\ \frac{}{+t = t} \qquad (-{=}\text{I.})\ \frac{}{-t = t'}\ \text{for any distinct } t, t' \in T^*$$

How could we formulate suitable elimination rules for identity? The introduction rules tell us that nothing is required to infer $+t = t$ and $-t = t'$. Accordingly, nothing can be inferred from those statements either. Nonetheless, given that these are *all* the rules for identity we also know that we can never infer $-t = t$ and $+t = t'$. So assuming that the logic of the meta-theory encompasses minimal logic, it follows that anything can be inferred from them.

$$(-{=}\text{E.})\ \frac{-t=t}{\varphi} \qquad (+{=}\text{E.})\ \frac{+t=t'}{\varphi}\ \text{for any distinct } t,t'\in T^*$$

Kripke's construction was designed to validate the naïve truth rules allowing us to move between the assertion of A and that of 'A is true'. Similarly, Maddy's construction was designed to validate the naïve class rules allowing us to move between the assertion of Ft and the assertion of 't is in $\hat{x}F$'. Similar considerations apply for the case in which Ft and 't is in $\hat{x}F$' are rejected. We therefore add the following rules:

$$(+\eta\text{I.})\ \frac{+Ft}{+t\eta\hat{x}F} \qquad (+\eta\text{E.})\ \frac{+t\eta\hat{x}F}{+Ft} \qquad (-\eta\text{I.})\ \frac{-Ft}{-t\eta\hat{x}F} \qquad (-\eta\text{E.})\ \frac{-t\eta\hat{x}F}{-Ft}$$

The rules laid down so far (indeed, the introduction rules alone) would suffice to characterize the fixed point of the Maddy hierarchy. Recall, however, that we want to provide a deductive system that allows us to reason using Maddy's theory. Recall, moreover, that any EA-pair can be considered a K_3 model. But K_3 is an explosive logic, in which anything follows from a contradiction. In a bilateral context, this suggests adding the following rule of bilateral explosion, as Ryan Simonelli (*forthcoming*) has suggested:

$$(\text{Bilateral Explosion})\ \frac{+A \qquad -A}{\varphi}$$

Note that the identity elimination rules are an immediate consequence of the identity introduction rules and Bilateral Explosion. Bilateral Explosion does not serve to specify the meaning of a particular logical constant, but rather to characterize the interaction between the speech acts of assertion and rejection in the proof theory. Principles governing the interaction between speech acts are known as coordination principles in the bilateral and multilateral literature (Incurvati and Schlöder, 2023; Rumfitt, 2000; Smiley, 1996).

Let Maddian Bilateral Logic (MBL for short) be the system consisting of the above introduction and elimination rules for negation, conjunction, the universal quantifier, identity and membership together with the Bilateral Explosion rule. To be able to characterize the fixed point of the Maddy hierarchy, one last ingredient is needed: we need to give a theory providing all the facts about membership among sets and about membership in $\bar{V}$ from which the construction of the Maddy hierarchy begins. So let $V_\kappa \models A$ be defined as usual for As in the language $\mathcal{L}_\in$ of set theory. True Set Theory (TST for short) is then the union of the following sets of signed sentences:

(i) $\{+\bar{a}\eta\bar{b} \mid V_\kappa \models a \in b\}$
(ii) $\{-\bar{a}\eta\bar{b} \mid V_\kappa \models a \notin b\}$
(iii) $\{+\bar{a}\eta\bar{V} \mid a \in V_\kappa\}$
(iv) $\{-t\eta\bar{a} \mid a \in V_\kappa\}$ for all $t \in C^*$
(v) $\{-t\eta\bar{V}\}$ for all $t \in C^*$

We are now in a position to prove that, over TST, MBL axiomatizes the fixed point of the Maddy hierarchy. We begin by showing the soundness of MBL with respect to the model-theoretic consequence relation induced by σ^{M}.

Theorem 4.2. *Let Γ be a set of signed sentences and φ a signed sentence. Suppose that $\Gamma \vdash_{\mathsf{MBL}} \varphi$ and that for all ψ in Γ, $\sigma^{\mathrm{M}} \models A$ if $\psi = +A$, and $\sigma^{\mathrm{M}} \mathrel{=\!\!|} A$ if $\psi = -A$. Then, $\sigma^{\mathrm{M}} \models B$ if $\varphi = +B$ and $\sigma^{\mathrm{M}} \mathrel{=\!\!|} B$ if $\varphi = -B$.*

Proof. The identity introduction rule are sound because for any distinct $t, t' \in T^*$, $\sigma^{\mathrm{M}} \models t = t$ and $\sigma^{\mathrm{M}} \mathrel{=\!\!|} t \neq t'$. Moreover, it is easy to verify that if the premises of the rules for the connectives, the universal quantifier and η are (anti-)satisfied by σ^{M}, then so is the conclusion. (For the η-introduction rules, we use the definition of the Maddy jump and the fact that σ^{M} is a fixed point.) Finally, the Rejection rule is sound because by construction of the Maddy hierarchy, for no A it is ever the case that $\sigma^{\mathrm{M}} \models A$ and $\sigma^{\mathrm{M}} \mathrel{=\!\!|} A$. □

Since by construction $\sigma^{\mathrm{M}} \models A$ for all $+A \in \mathsf{TST}$ and $\sigma^{\mathrm{M}} \mathrel{=\!\!|} A$ for all $-A \in \mathsf{TST}$, Theorem 4.2 immediately yields that, over TST, MBL is sound with respect to σ^{M}.

Theorem 4.3. *For every sentence A, if $\mathsf{TST} \vdash_{\mathsf{MBL}} +A$, then $\sigma^{\mathrm{M}} \models A$, and if $\mathsf{TST} \vdash_{\mathsf{MBL}} -A$, then $\sigma^{\mathrm{M}} \mathrel{=\!\!|} A$.*

I now prove that, over TST, MBL is also complete with respect to σ^{M}.

Theorem 4.4. *For every sentence A, if $\sigma^{\mathrm{M}} \models A$, then $\mathsf{TST} \vdash_{\mathsf{MBL}} +A$, and if $\sigma^{\mathrm{M}} \mathrel{=\!\!|} A$, then $\mathsf{TST} \vdash_{\mathsf{MBL}} -A$.*

Proof. I prove that if $\sigma^{\mathrm{M}}_{\alpha} \models A$, then $\mathsf{TST} \vdash_{\mathsf{MBL}} +A$, and if $\sigma^{\mathrm{M}}_{\alpha} \mathrel{=\!\!|} A$, then $\mathsf{TST} \vdash_{\mathsf{MBL}} -A$. The theorem will then follow by ordinal induction.
Base case. Immediate from the definition of TST.
Inductive step. Our induction hypothesis is that, for every A, if $\sigma^{\mathrm{M}}_{\alpha} \models A$, then $\mathsf{TST} \vdash_{\mathsf{MBL}} +A$, and if $\sigma^{\mathrm{M}}_{\alpha} \mathrel{=\!\!|} A$, then $\mathsf{TST} \vdash_{\mathsf{MBL}} -A$. We need to prove that, for every A, if $\sigma^{\mathrm{M}}_{\alpha+1} \models A$, then $\mathsf{TST} \vdash_{\mathsf{MBL}} +A$, and if $\sigma^{\mathrm{M}}_{\alpha+1} \mathrel{=\!\!|} A$, then $\mathsf{TST} \vdash_{\mathsf{MBL}} -A$. We proceed by induction on the complexity of A. We only cover the positive cases, since the negative cases proceed similarly, using the negative introduction rule for the relevant logical constant, rather than the positive one.

A is of the form $t\eta t'$. Suppose $\sigma^{\mathrm{M}}_{\alpha+1} \models t\eta t'$. If t' is a set term or $\bar{V}$, we have that $\mathsf{TST} \vdash_{\mathsf{MBL}} +t\eta t'$ by the fact that $\sigma_{\alpha+1} \sqsupseteq \sigma_0$. So suppose that t' is of the form $\hat{x}F$ for some F. Then, by construction of the Maddy hierarchy, $\sigma^{\mathrm{M}}_{\alpha} \models Ft$. By the induction hypothesis of the ordinal induction, it follows that $\mathsf{TST} \vdash_{\mathsf{MBL}} +Ft$. But then, $\mathsf{TST} \vdash_{\mathsf{MBL}} +t\eta\hat{x}F$ since MBL includes the ($+\eta$I.) rule.

A is of the form $t = t'$. If $\sigma^{\mathrm{M}}_{\alpha+1} \models t = t'$, then, by the model-theoretic clauses for identity, t is the same term as t'. But then, $\mathsf{TST} \vdash_{\mathsf{MBL}} +t = t'$ since MBL includes the ($+=$I.) rule.

A is of the form $B \wedge C$. If $\sigma^{\mathrm{M}}_{\alpha+1} \models B \wedge C$, then, by the model-theoretic clauses for conjunction, $\sigma^{\mathrm{M}}_{\alpha+1} \models B$ and $\sigma^{\mathrm{M}}_{\alpha+1} \models C$. But then, by the induction hypothesis of the induction on complexity, $\mathsf{TST} \vdash_{\mathsf{MBL}} +B$ and $\mathsf{TST} \vdash_{\mathsf{MBL}} +C$. Since MBL includes the ($+\wedge$I.) rule, it follows that $\mathsf{TST} \vdash_{\mathsf{MBL}} +B \wedge C$.

A is of the form $\neg B$. If $\sigma^{\mathrm{M}}_{\alpha+1} \models \neg B$, then, by the model-theoretic clauses for negation, $\sigma^{\mathrm{M}}_{\alpha+1} \mathrel{=\!\!|} B$. But then, by the induction hypothesis of the induction on complexity, $\mathsf{TST} \vdash_{\mathsf{MBL}} -B$. Since MBL includes the ($+\neg$I.) rule, it follows that $\mathsf{TST} \vdash_{\mathsf{MBL}} +\neg B$.

A is of the form $\forall x\, F$. If $\sigma^{\mathrm{M}}_{\alpha+1} \models \forall x\, F$, then, by the model-theoretic clauses for the universal quantifier, for all t, $\sigma^{\mathrm{M}} \models F[t/x]$. By the induction hypothesis of the induction on complexity, $\mathsf{TST} \vdash_{\mathsf{MBL}} +F[t/x]$. By the ($+\forall$I.) rule, it follows that $\mathsf{TST} \vdash_{\mathsf{MBL}} +\forall x\, F$.

Limit case. Suppose that $\sigma^{\mathrm{M}}_{\lambda} \models A$ with λ a limit ordinal. Since $\sigma^{\mathrm{M}}_{\lambda} = \bigcup_{\alpha<\lambda} \sigma^{\mathrm{M}}_{\alpha}$, there is some $\beta < \lambda$ such that $\sigma^{\mathrm{M}}_{\lambda} \models A$. By the induction hypothesis of the ordinal induction, $\mathsf{TST} \vdash_{\mathsf{MBL}} +A$. The negative case is analogous. □

Having established that we can axiomatize the fixed point of the Maddy hierarchy, we now provide a general model theory for MBL. An interpretation $\mathcal{I}$ consists of a set of objects $\mathcal{D}$ and an interpretation function $\mathcal{F}$. The function $\mathcal{F}$ assigns elements of $\mathcal{D}$ to closed terms and an EA-pair σ to η where σ^+ and σ^- consist of ordered pairs of elements of $\mathcal{D}$. The satisfaction and anti-satisfaction clauses for identity, the connectives and the universal quantifier are as above. An interpretation $\mathcal{I}$ assigning an EA-pair σ is then said to be η-admissible if, for all A, $\mathcal{I} \models Ft$ just in case $\langle \mathcal{F}(t), \mathcal{F}(\hat{x}F)\rangle \in \sigma^+$, and $\mathcal{I} \mathrel{=\!\!|} Ft$ just in case $\langle \mathcal{F}(t), \mathcal{F}(\hat{x}F)\rangle \in \sigma^-$.

We can now show that MBL is sound and complete with respect to the class of η-admissible models. For soundness, the argument used in Theorem 4.2 above suffices, except that we now use the restriction to η-admissible models to establish the soundness of the η-rules. For completeness, we first need a few additional definitions. We say that a set of formulae Γ is *bilaterally prime* (*b-prime* for short) just in case if $-A \wedge B \in \Gamma$ then either $-A \in \Gamma$ or $-B \in \Gamma$. Moreover, for a given deductive system S, we say that a set Γ is *bilaterally consistent$_S$* (*b-consistent$_S$* for short) if for no A is it the case that both $\Gamma \vdash_{\mathsf{S}} +A$ and $\Gamma \vdash_{\mathsf{S}} -A$. We can then prove the following model existence result:

Lemma 4.5. *Let Γ be a b-consistent$_{\mathsf{MBL}}$ set of signed sentences and let Γ^* be the extension of its closure under derivability in* MBL *to a b-prime set. Then there is an interpretation $\mathcal{I}$ such that $+A \in \Gamma^*$ iff $\mathcal{I} \models A$, and $-A \in \Gamma^*$ iff $\mathcal{I} \mathrel{=\!\!|} A$.*

Proof. Let $\mathcal{I} = \langle \mathcal{D}, \mathcal{F}\rangle$ be the canonical term model for Γ^* and set $\mathcal{F}(\eta) = \langle \tau^+, \tau^-\rangle$ where τ^+ is $\{\langle \bar{a}, \bar{b}\rangle \mid +a\eta b \in \Gamma^*\}$ and τ^- is $\{\langle \bar{a}, \bar{b}\rangle \mid -a\eta b \in \Gamma^*\}$. Then, for all A, we have that $+A \in \Gamma^*$ iff $\mathcal{I} \models A$, and $-A \in \Gamma^*$ iff $\mathcal{I} \mathrel{=\!\!|} A$. The proof is by induction on the complexity of A.

A is of the form $a\eta b$. $+a\eta b \in \Gamma^*$ iff $\langle \bar{a}, \bar{b}\rangle \in \tau^+$ iff $\mathcal{I} \models a\eta b$. The case of $-a\eta b$ is analogous.

A is of the form $a = b$. $+a = b \in \Gamma^*$ iff $\bar{a} = \bar{b}$ iff $\mathcal{I} \models a = b$. The case of $-a = b$ is analogous.

A is of the form $B \wedge C$. $+B \wedge C \in \Gamma^*$ iff $+B \in \Gamma^*$ and $+C \in \Gamma^*$ (since Γ^* is MBL-closed) iff $\mathcal{I} \models B$ and $\mathcal{I} \models C$ (by the induction hypothesis) iff $\mathcal{I} \models B \wedge C$. Similarly, $-B \wedge C \in \Gamma^*$ iff $-B \in \Gamma^*$ or $-C \in \Gamma^*$ (since Γ^* is b-prime) iff $\mathcal{I} \mathrel{=\!\!|} B$ or $\mathcal{I} \mathrel{=\!\!|} C$ (by the induction hypothesis) iff $\mathcal{I} \mathrel{=\!\!|} B \wedge C$.

A is of the form $\neg B$. $+\neg B \in \Gamma^*$ iff $-B \in \Gamma^*$ (since Γ^* is MBL-closed) iff $\mathcal{I} \mathrel{=\!\!|} B$ (by the induction hypothesis) iff $\mathcal{I} \models \neg B$ (by the model-theoretic clauses for negation). Similarly, $-\neg B \in \Gamma^*$ iff $+B \in \Gamma^*$ iff $\mathcal{I} \models B$ iff $\mathcal{I} \mathrel{=\!\!|} \neg B$.

A is of the form $\forall x\, F$. $+\forall x\, F \in \Gamma^*$ iff for every $t \in T^*$, $F[t/x] \in \Gamma^*$ (since Γ^* is MBL-closed) iff $\mathcal{I} \models Ft$ (by the induction hypothesis) iff $\mathcal{I} \models \forall x\, F$ (by the fact that there is a term for every set and class). Similarly, $-\forall x\, F \in \Gamma$ iff $F[t/x] \in \mathsf{MBL}$ for every $t \in T^*$ (since Γ^* is MBL-closed and MBL includes the ($-\forall$E.) rule) iff $\mathcal{I} \mathrel{=\!\!|} Ft$ for every $t \in T^*$ iff $\mathcal{I} \mathrel{=\!\!|} \forall x\, F$. □

We are now in a position to prove completeness. To make the statement of the theorem simpler, it will be useful to extend our model-theoretic notion to cover the signed language. If A is a sentence, we write $\mathcal{I} \models_B +A$ wherever $\mathcal{I} \models A$, and $\mathcal{I} \models_B -A$ wherever $\mathcal{I} \mathrel{=\!\!|} A$. If φ is a signed sentence, we then write $\Gamma \models_B \varphi$ just in case if, for all η-admissible $\mathcal{I}$, if $\mathcal{I} \models_B \psi$ for all $\psi \in \Gamma$, then $\mathcal{I} \models_B \varphi$.

Theorem 4.6. *Let Γ be a set of signed sentences and φ a signed sentence. If $\Gamma \models_B \varphi$, then $\Gamma \vdash_{\mathsf{MBL}} \varphi$.*

Proof. Suppose that $\Gamma \nvdash_{\mathsf{MBL}} \varphi$. It follows that Γ is b-consistent$_{\mathsf{MBL}}$ and so, by Lemma 4.5, there is an interpretation $\mathcal{I}$ and a set of signed sentences Γ^* such that $\mathcal{I} \models A$ iff $+A \in \Gamma^*$, and such that $\mathcal{I} \dashv A$ iff $-A \in \Gamma^*$. Since $\Gamma \subseteq \Gamma^*$, clearly, $\mathcal{I} \models_B \psi$ for all $\psi \in \Gamma$. Moreover, $\mathcal{I} \not\models_B \varphi$ since $\varphi \notin \Gamma^*$. So $\Gamma \not\models_B \varphi$. □

In her original paper, Maddy introduces a notation intended to allow us to talk about the situation in which an EA-pair σ does not decide a certain membership fact. Using our notation, we can similarly define:

- $\sigma \perp\!\!\!\perp A$ iff neither $\sigma \models A$ nor $\sigma \dashv A$.

With this notation, we can now write that $\sigma^M \perp\!\!\!\perp \hat{x}(x \not\eta x)\eta\hat{x}(x \not\eta x)$, which expresses model-theoretically the fact that Maddy theory's of sets and classes is agnostic over whether the Russell class belongs to itself. We cannot yet express this fact proof-theoretically. We can do so by moving from a bilateral to a trilateral setting and augmenting the expressive power of MBL with an additional force marker $?$, denoting the speech act expressing agnosticism about a certain matter. This speech act is governed by the following coordination principles:

$$\text{(Trilateral Explosion}_1\text{)}\ \frac{+A \qquad ?A}{\varphi} \qquad \text{(Trilateral Explosion}_2\text{)}\ \frac{-A \qquad ?A}{\varphi}$$

$$\text{(Trilateral Excluded Fourth)}\ \frac{\begin{matrix}[+A] \\ \vdots \\ \varphi\end{matrix} \qquad \begin{matrix}[?A] \\ \vdots \\ \varphi\end{matrix} \qquad \begin{matrix}[-A] \\ \vdots \\ \varphi\end{matrix}}{\varphi}$$

The Trilateral Explosion$_i$ principles state that it is incoherent to express agnosticism towards a subject matter while at the same time expressing belief or disbelief towards that very same subject matter. Trilateral Excluded Fourth captures the idea that there are three attitudes that can be taken towards a certain content: belief, disbelief or agnosticism. An alternative, and formally equivalent route, would be to take the agnostic attitude to be expressible by weakly asserting (Incurvati and Schlöder, 2019) and weakly rejecting (Incurvati and Schlöder, 2017) the same content (Ferrari and Incurvati, 2022) and lay down corresponding coordination principles.

Now consider the system obtained by extending MBL with the Trilateral Explosion$_i$ principles and Trilateral Excluded Fourth. It is straightforward to prove in this system that $?\hat{x}(x \not\eta x)\eta\hat{x}(x \not\eta x)$.[9] Indeed, one can show that the system is sound and complete with respect to the general model theory from the previous section augmented with $\perp\!\!\!\perp$ defined as above.

5. Rampant indeterminacy

Maddy's theory of classes satisfies several of the desiderata we discussed above. It validates the unrestricted comprehension rules and hence implies the existence of big classes such as the universal class (i.e. the class containing everything there is) and the Russell class (i.e. the class of all collections that do not belong to themselves). It takes classes to be real, well-defined entities. And it takes classes to be logical collections, since they are characterized in terms of their defining condition, thereby distinguishing them from sets.

[9] In the alternative route using weak assertion and weak rejection, it would be straightforward to prove that $\hat{x}(x \not\eta x)\eta\hat{x}(x \not\eta x)$ is both weakly asserted and weakly rejected.

However, there are some serious problems with Maddy's theory, which have been stressed in recent work by Linnebo (2024). As Linnebo puts it, Maddy's 'account is very "gappy": there is a *lot* of indeterminacy' (p. 74, emphasis in the original). But in K_3, a universal generalization is indeterminate just in case at least one of its instances is. As a result, it very difficult for a universal generalization to be determinately true in Maddy's theory of classes. This leads to two specific problems, which Linnebo focuses on.

The first concerns Maddy's attempt to develop a theory of Frege–Russell numbers within her theory of classes. Within such a theory, two collections (sets or classes) are equinumerous if and only if there exists a one-to-one correspondence between their members. It follows that two collections are *not* equinumerous if and only every relation fails to be such a correspondence. This is a generalization, which is then subject to the problem of indeterminacy. As a result, it is very hard to prove that two collections are not equinumerous in Maddy's theory. Indeed, as Maddy (2000, 312) herself proves, if the statement that two non-empty collections are equinumerous is not satisfied in her original theory of classes, then it is neither satisfied nor anti-satisfied. It follows that we cannot even prove that $\{\emptyset\}$ is not equinumerous with $\{\emptyset, \{\emptyset\}\}$. Linnebo takes this 'rampant indeterminacy' to be difficult to accept, and Maddy herself takes the result to be troublesome.

The second problem concerns identity in Maddy's theory. Recall that, for any two terms, the model-theoretic clauses for identity declare their referents to be determinately identical just in case the terms themselves are identical, and determinately different just in case the terms themselves are different. This means, in particular, that if F and G are different formulae, then the classes $\hat{x}F$ and $\hat{x}G$ will also be different: Maddy's theory identifies classes in far too fine-grained a manner and hence fails to satisfy our fourth desideratum on a theory of classes.

To repair the situation, one might try to single out a coarser equivalence relation among collections, in terms of which identity among classes could then be defined. Maddy (2000, 305) considers the option of defining identity among collections in terms of the relation $\simeq$, defined as follows.

Definition 5.1. For $t, t' \in T$, $t \simeq t' \equiv_{\text{def}} \forall z(z\eta t \leftrightarrow z\eta t')$

As Maddy observes, this proposal might appear at first sight to give us what we want, since it identifies coextensive sets and classes, such as $\bar{a}$ and $\hat{x}(x\eta\bar{a})$, and certain classes which would have been declared to be different by the original syntactic definition of identity, such as $\hat{x}(x\eta\bar{a})$ and $\hat{x}(x\eta\bar{a} \wedge x\eta\bar{a})$.

However, says Linnebo, the proposal fails because of the problem of rampant indeterminacy to which universal generalizations are subject. As Maddy (2000, 305) points out, σ^{M} can satisfy $u\eta t \leftrightarrow u\eta t'$ only if it is decided about $u\eta t$ or $u\eta t'$. It follows that the definition of identity as $\simeq$ entails that two classes can only be identical if every membership claim concerning them must be either satisfied or anti-satisfied. Indeed, $t \simeq t$ becomes a way of expressing that t is total: $\sigma^M \models t \simeq t$ just in case $\sigma^M \models \forall z(z\eta t \vee z \not\eta t)$. It follows that Maddy's theory is undecided about whether $\hat{x}(x \not\eta x)$ is identical to $\hat{x}(x \not\eta x \wedge x \not\eta x)$. Indeed, the theory is undecided about whether $\hat{x}(x \not\eta x)$ is identical to itself: under the proposed definition, identity is not even reflexive.

Linnebo considers two possible ways of improving on Maddy's account along broadly Maddian lines. The first, which he has developed in joint work with Leon Horsten (Horsten and Linnebo, 2016), consists in building up not an EA-pair for the membership relation, but rather for the identity predicate, so that we have that $\hat{x}F = \hat{x}G$ if and only if $\forall x(Fx \leftrightarrow Gx)$. The construction proceeds by adding $\langle \hat{x}F, \hat{x}G \rangle$ to the extension of the identity predicate whenever $\forall x(Fx \leftrightarrow Gx)$ is satisfied, and by adding $\langle \hat{x}F, \hat{x}G \rangle$ to the anti-extension of the identity predicate whenever $\forall x(Fx \leftrightarrow Gx)$ is anti-satisfied. However, as Linnebo points out, the construction is severely limited, in that it only works for classes whose defining formula does not contain occurrences of the membership predicate.

The second attempt Linnebo considers is due to Jönne Kriener (2014). Kriener combines ideas from Maddy and from Horsten and Linnebo, in that he builds up EA-pairs for both the membership relation and the identity predicate. The resulting theory, however, has important limitations with regards to the classes whose existence it admits. For instance, the natural way of defining the singleton class of a given class (that is, as $\hat{x}x = \hat{y}F$) fails. Kriener himself takes his results to cast doubt on the viability of a theory of classes based on a hierarchical approach along Kripkean lines.

Linnebo concludes that the prospects for a Maddian account of classes are dim:

> The picture that emerges is thus one of a failed research program. At the outset, Maddy's idea of building on Kripke's highly successful account of truth seemed very promising. And as we have seen, there are natural ways to transpose the account to the case of classes. Unfortunately, the resulting theories are not very attractive; and the situation does not improve materially on any of the more natural ways to modify Maddy's account. (Linnebo, 2024, 75)

In the remainder of this paper, I will argue that, contrary to appearances, the prospects for a hierarchical approach to classes along broadly Kripkean-Maddian lines are in fact bright. There is a very natural way of modifying Maddy's account so as to solve the problem highlighted by Linnebo.

6. Supervaluations

The rampant indeterminacy observed by Linnebo is indeed a problematic feature of Maddy's theory. The problem is a familiar one: it already affects Kripke's theory of truth and is due to the fact that the underlying logic of Maddy's theory, just like Kripke's, is K3.

A natural strategy to deal with the problem is to adopt a different scheme for handling membership gaps. In his original paper, Kripke (1975, 711–712) already suggested the possibility of handling truth-value gaps using a scheme based on Bas van Fraassen's (1966) notion of supervaluation. van Fraassen originally introduced the idea of supervaluation to handle truth-value gaps generated by the use of empty singular terms. In the subsequent literature, much attention has been devoted to the application of the supervaluational strategy to the case of vagueness (Fine, 1975; Incurvati and Schlöder, 2022; Keefe, 2000).

In the case of truth, the idea is to handle truth-value gaps by assigning a sentence A to the extension (anti-extension) of the truth predicate at the next stage of the construction just in case all admissible expansions of the EA-pair of the truth predicate at the current stage classically satisfy (anti-satisfy) A. Transposing this to the class-theoretic setting, the idea is to handle membership gaps by assigning an ordered pair $\langle t, \hat{x}F \rangle$ consisting of an object and a class to the extension (anti-extension) of the membership predicate at the next stage of the construction just in case all admissible expansions the EA-pair of the membership predicate at the current stage classically satisfy (anti-satisfy) Ft. This gives rise to the following template for class supervaluational schemes, where the satisfaction relation $\models$ and the antisatisfaction relation $\Dashv$ are recursively defined as before, except that we now require that, for every EA-pair σ and for every sentence A, either $\sigma \models A$ or $\sigma \Dashv A$:

Definition 6.1 (Supervaluational-jump template)**.** Given an EA-pair $\sigma = \langle \sigma^+, \sigma^- \rangle$, the *supervaluational jump* J_s is the operation such that, for all $t \in T^*$,

$$J_s(\sigma^+) = \{\langle t, \hat{x}F \rangle \mid \text{ for all } \tau \text{ that are s-admissible expansions of } \sigma, \tau \models Ft\}, \text{ and}$$

$$J_s(\sigma^-) = \{\langle t, \hat{x}F \rangle \mid \text{ for all } \tau \text{ that are s-admissible expansions of } \sigma, \tau \Dashv Ft\}.$$

Different supervaluational schemes are then obtained by specifying when an EA-pair is an admissible expansion of another EA-pair. The first supervaluational scheme considered by Kripke is a straightforward adaptation of van Fraassen's original notion of supervaluation to the case of truth. In particular, the idea is to take an admissible expansion of an EA-pair σ to be one that does not satisfy (anti-satisfy) any sentence that is anti-satisfied (satisfied) by σ. We can proceed in an analogous manner in the case of classes: an admissible expansion is one that respects the already established membership facts.

Definition 6.2 (MvF-admissible expansion)**.** An EA-pair $\tau = \langle \tau^+, \tau^- \rangle$ is a *MvF-admissible expansion* of an EA-pair $\sigma = \langle \sigma^+, \sigma^- \rangle$ iff (i) $\tau \sqsupseteq \sigma$, (ii) $\tau^+ \cap \sigma^- = \emptyset$, and (iii) $\tau^- \cap \sigma^+ = \emptyset$.

By replacing the Maddy jump J_M with the Maddy–van Fraassen jump J_{MvF} in the definition of the Maddy hierarchy, we obtain the *basic supervaluational class hierarchy*. It is easy to check that (still taking $\bar{V}$ to denote V_κ where k is the first inaccessible) the construction of the basic supervaluational class hierarchy reaches a fixed point σ^{MvF}. We can therefore consider σ^{MvF} as giving the extension of η according to the Maddy–van Fraassen theory of classes—the theory of classes obtained by adopting the van Fraassen supervaluational scheme within the context of a theory of classes along broadly Maddian lines.

I now want to present a natural deduction system for the Maddy–van Fraassen theory of classes. It is well known that the addition of the following rule of Bilateral Excluded Middle to the asserted and rejected rules for negation and conjunction of MBL and Bilateral Explosion delivers a bilateral version of classical propositional logic:

$$\text{(Bilateral Excluded Third)}\ \frac{\begin{matrix}[+A] \\ \vdots \\ \varphi\end{matrix} \qquad \begin{matrix}[-A] \\ \vdots \\ \varphi\end{matrix}}{\varphi}$$

For the principles of Bilateral Explosion and Bilateral Excluded Third are interderivable with the following coordination principles (see, e.g., del Valle-Inclan, 2023, 382), which are sufficient to provide a sound and complete axiomatization of the classical propositional calculus in the presence of the rules for negation and conjunction of MBL (Incurvati and Schlöder, 2023; Rumfitt, 2000; Smiley, 1996), where '(SR$_i$)' abbreviates 'Smileian reductio$_i$'.

$$\text{(Rejection)}\ \frac{+A \qquad -A}{\bot} \qquad \text{(SR}_1\text{)}\ \frac{\begin{matrix}[+A] \\ \vdots \\ \bot\end{matrix}}{-A} \qquad \text{(SR}_2\text{)}\ \frac{\begin{matrix}[-A] \\ \vdots \\ \bot\end{matrix}}{+A}$$

Now, a central supervaluational intuition is that classical reasoning is in order when no rules are used that can gender indeterminacy. And in our current context, this indeterminacy can only arise because of the application of the η-rules. This suggests restricting the application of these rules within the coordination principle of Bilateral Excluded Third, in order to obtain a supervaluationistically acceptable version thereof. We thereby obtain the following:

$$\text{(BET*)}\ \frac{\begin{matrix}[+A] \\ \vdots \\ \varphi\end{matrix} \qquad \begin{matrix}[-A] \\ \vdots \\ \varphi\end{matrix}}{\varphi}\ \text{if the subderivations of } \varphi \text{ use no } \eta\text{-rules}$$

Let MvFBL be the system obtained by adding (BET*) to MBL and restricting the subderivations in the ($-\wedge$E.) and ($-\forall$E.) rules in the same manner (that is, to subderivations that make no use η-rules). I am now going to prove that analogous results hold between MvFBL and the Maddy–van Fraassen theory of classes as those that obtained between MBL and Maddy's theory of classes.

I begin by proving that, over TST, MvFBL axiomatizes the fixed point of the basic supervaluational hierarchy. I first prove that MvFBL is sound with respect to the consequence relation induced by σ^{MvF}.

Theorem 6.3. *Let Γ be a set of signed sentences and φ a signed sentence. Suppose that for all ψ in Γ, $\sigma^{\text{MvF}} \models A$ if $\psi = +A$, and $\sigma^{\text{MvF}} \dashv A$ if $\psi = -A$. Then, if $\Gamma \vdash_{\mathsf{MvFBL}} \varphi$, $\sigma^{\text{MvF}} \models B$ if $\varphi = +B$ and $\sigma^{\text{MvF}} \dashv B$ if $\varphi = -B$.*

Proof. The arguments in the proof of Theorem 4.2 carry over to the present case except for the ($-\wedge$E.) and ($-\forall$E.) rules, the (BET*) principle and the η-rules. For the ($-\wedge$E.) and ($-\forall$E.) rules and the (BET*) principle, it suffices to note that all of their instances are instances of classically valid arguments (see Incurvati and Schlöder, 2017). It remains to check the η-rules. We only cover the positive cases because the negative ones proceed in a completely analogous fashion by replacing the satisfaction relation with the anti-satisfaction one.

For the ($+\eta$I.) rule, suppose that $\sigma^{\text{MvF}} \models Ft$. By monotonicity, this means that for every $\tau \sqsupseteq \sigma^{\text{MvF}}$, $\tau \models Ft$. So, in particular, for every τ which is an MvF-admissible expansion of σ^{MvF}, $\tau \models Ft$. By the definition of the Maddy–van Fraassen supervaluational jump, it follows that $\langle t, \hat{x}F\rangle \in J_{\text{MvF}}(\sigma^{\text{MvF}})$. Since σ^{MvF} is a fixed point, it follows that $\sigma^{\text{MvF}} \models t\eta\hat{x}F$.

For the ($+\eta$E.) rule, suppose that $\sigma^{\text{MvF}} \models t\eta\hat{x}F$. By the definition of J_{MvF}, this means that for every τ which is an MvF-admissible expansion of σ^{MvF}, $\tau \models Ft$. Since σ^{MvF} is an admissible expansion of itself, we have, in particular, that $\sigma^{\text{MvF}} \models Ft$. □

As in the case of Maddy's theory of classes, by construction we have that $\sigma^{\text{MvF}} \models A$ for all $+A \in \mathsf{TST}$ and $\sigma^{\text{MvF}} \dashv A$ for all $-A \in \mathsf{TST}$. Hence, Theorem 6.3 immediately yields that, over TST, MvFBL is sound with respect to σ^{MvF}.

Theorem 6.4. *For every sentence A, if $\mathsf{TST} \vdash_{\mathsf{MvFBL}} +A$, then $\sigma^{\text{MvF}} \models A$, and if $\mathsf{TST} \vdash_{\mathsf{MvFBL}} -A$, then $\sigma^{\text{MvF}} \dashv A$.*

I now turn to the proof that, over TST, MvFBL is complete with respect to σ^{MvF}. For a given deductive system S, we say that Γ is maximally b-consistent$_{\mathsf{S}}$ if it is b-consistent$_{\mathsf{S}}$ and either

$+A \in \Gamma$ or $-A \in \Gamma$. Moreover, for every system S, we let S^* be S without the η-rules. Using Zorn's Lemma, we can prove the following:[10]

Lemma 6.5. *Every b-consistent*$_{\mathsf{MvFBL}^*}$ *set of signed sentences can be extended to a maximally b-consistent*$_{\mathsf{MvFBL}^*}$ *set.*

We can now prove that every maximally b-consistent$_{\mathsf{MvFBL}^*}$ extension of TST closed under derivability in MvFBL has a model:

Theorem 6.6. *Suppose that $\Gamma \supseteq \mathsf{TST}$ is a maximally b-consistent*$_{\mathsf{MvFBL}^*}$ *set of signed sentences closed under derivability in* MvFBL, *and let τ be the EA-pair $\langle \tau^+, \tau^- \rangle$, where τ^+ is $\{\langle \bar{a}, \bar{b} \rangle \mid +a\eta b \in \Gamma\}$ and τ^- is $\{\langle \bar{a}, \bar{b} \rangle \mid -a\eta b \in \Gamma\}$. Then, for all sentences A, we have that $+A \in \Gamma$ iff $\tau \models A$, and $-A \in \Gamma$ iff $\tau \mathrel{=\!\!\mid} A$.*

Proof. By induction on the complexity of A.

A is of the form $t\eta t'$. By definition of τ.

A is of the form $t = t'$. Suppose $+t = t' \in \Gamma$ but $\tau \not\models t = t'$. Then t is not the same term as t' and hence, since Γ is closed under MvFBL-derivability and MvFBL includes the $(-{=}\text{I.})$ rule, $-t \neq t' \in \Gamma$. But this contradicts the assumption that Γ is b-consistent$_{\mathsf{MvFBL}^*}$. For the other direction, if $\tau \models t = t'$, then, by the model-theoretic clauses for identity, t is the same term as t'. But then, $+t = t' \in \Gamma$, since MvFBL includes the $(+{=}\text{I.})$ rule. The negative case is analogous.

A is of the form $B \wedge C$. Suppose $+B \wedge C \in \Gamma$. Then $+B \in \Gamma$ and $+C \in \Gamma$, since Γ is closed under MvFBL-derivability and MvFBL includes the $(+\wedge\text{I.})$ rule. By the induction hypothesis, $\tau \models B$ and $\tau \models C$. By the model-theoretic clauses for conjunction, $\tau \models B \wedge C$. The reverse direction is analogous. For the negative case, suppose $-B \wedge C \in \Gamma$. Then $-B \in \Gamma$ or $-C \in \Gamma$, since Γ is maximally b-consistent$_{\mathsf{MvFBL}^*}$. By the induction hypothesis, $\tau \mathrel{=\!\!\mid} B$ or $\tau \mathrel{=\!\!\mid} C$. By the model-theoretic clauses for conjunction, $\tau \mathrel{=\!\!\mid} B \wedge C$. For the other direction, suppose $\tau \mathrel{=\!\!\mid} B \wedge C$. Then $\tau \mathrel{=\!\!\mid} B$ or $\tau \mathrel{=\!\!\mid} C$. By the induction hypothesis, $-B \in \Gamma$ or $-C \in \Gamma$. Since Γ is closed under MvFBL-derivability and MvFBL includes the $(-\wedge\text{I.})$ rule, it follows that $-B \wedge C \in \Gamma$.

A is of the form $\neg B$. If $+\neg B \in \Gamma$, then $-B \in \Gamma$, since MvFBL includes the $(+\neg\text{I.})$ rule. By the induction hypothesis, it follows that $\tau \mathrel{=\!\!\mid} B$ and hence that $\tau \models \neg B$. The reverse direction is analogous, and so are the negative cases.

A is of the form $\forall x\, F$. If $+\forall x\, F \in \Gamma$, then, for all $t \in T^*$, $+F[t/x] \in \Gamma$, since MvFBL includes the $(+\forall\text{I.})$ rule. By the induction hypothesis, it follows that, for all $t \in T^*$, $\tau \models F[t/x]$, and, by the model-theoretic clauses for the universal quantifier, that $\tau \models \forall x\, F$. The reverse direction is analogous, and so are the negative cases. □

We are now in a position to prove completeness.

Theorem 6.7. *For every sentence A, if $\sigma^{\mathrm{MvF}} \models A$, then $\mathsf{TST} \vdash_{\mathsf{MvFBL}} +A$, and if $\sigma^{\mathrm{MvF}} \mathrel{=\!\!\mid} A$, then $\mathsf{TST} \vdash_{\mathsf{MvFBL}} -A$.*

Proof. I prove that if $\sigma_\alpha^{\mathrm{MvF}} \models A$, then $\mathsf{TST} \vdash_{\mathsf{MvFBL}} +A$, and if $\sigma_\alpha^{\mathrm{MvF}} \mathrel{=\!\!\mid} A$, then $\mathsf{TST} \vdash_{\mathsf{MvFBL}} -A$. The proof proceeds exactly like the proof of Theorem 4.4, except for the case in which A is of the form $t\eta t'$ in the induction on the complexity of A in the inductive step of the ordinal induction.

[10] In the context of a countable language, it would be possible to construct the maximally b-consistent set using a step-by-step procedure. See Incurvati and Schlöder, 2023, 568–569 for an application of the method in the theory of truth.

So suppose that $\sigma^{\mathrm{MvF}}_{\alpha+1} \models t\eta t'$. If t' is a set term or $\bar{V}$, we have that $\mathsf{TST} \vdash_{\mathsf{MvFBL}} +t\eta t'$ by the fact that $\sigma^{\mathrm{MvF}}_{\alpha+1} \sqsupseteq \sigma^{\mathrm{MvF}}_{0}$. So let t' be of the form $\hat{x}F$ for some F. Now suppose that $\mathsf{TST} \nvdash_{\mathsf{MvFBL}} +t\eta\hat{x}F$. This means that the deductive closure of $\mathsf{TST} \cup -t\eta\hat{x}F$ under MvFBL-derivability is b-consistent$_{\mathsf{MvFBL}^*}$ and so it has a maximally b-consistent$_{\mathsf{MvFBL}^*}$ extension Γ. Let $\tau = \langle \tau^+, \tau^- \rangle$, where τ^+ is $\{\langle \bar{a}, \bar{b}\rangle \mid +a\eta b \in \Gamma\}$ and τ^- is $\{\langle \bar{a}, \bar{b}\rangle \mid -a\eta b \in \Gamma\}$. I now show that τ is an MvF-admissible expansion of $\sigma^{\mathrm{MvF}}_{\alpha}$.

First, suppose that there was a formula $t\eta\hat{x}G$ such that $\sigma^{\mathrm{MvF}}_{\alpha} \models t\eta\hat{x}G$ but $\tau \not\models t\eta\hat{x}G$. By the induction hypothesis, it follows that $\mathsf{TST} \vdash_{\mathsf{MvFBL}} +t\eta\hat{x}G$. But then, by construction of τ, $\tau \models t\eta\hat{x}G$, which contradicts the original supposition. Hence, (i) $\tau \sqsupseteq \sigma^{\mathrm{MvF}}_{\alpha}$. Next, suppose that there is a $\langle t, \hat{x}G\rangle$ such that $\langle t, \hat{x}G\rangle \in \tau^+$ and $\langle t, \hat{x}G\rangle \in \sigma^{\mathrm{MvF},-}_{\alpha}$. By definition of τ, it follows that $+t\eta\hat{x}G \in \Gamma$. And by the induction hypothesis, it follows that $\mathsf{TST} \vdash_{\mathsf{MvFBL}} -t\eta\hat{x}G$ and so that $-t\eta\hat{x}G \in \Gamma$. But since Γ is b-consistent$_{\mathsf{MvFBL}^*}$, we have that $+t\eta\hat{x}G \notin \Gamma$, which by construction of τ means that $\langle t, \hat{x}G\rangle \notin \tau^+$. This contradicts our original supposition. Hence, (ii) $\tau^+ \cap \sigma^{\mathrm{MvF},-}_{\alpha} = \emptyset$. A similar reasoning shows that (iii) $\tau^- \cap \sigma^{\mathrm{MvF},+}_{\alpha} = \emptyset$.

By supposition, we have that $-t\eta\hat{x}F \in \Gamma$. Since Γ is closed under MvFBL-derivability, this means that $-Ft \in \Gamma$ too. By Theorem 6.6 it then follows that $\tau \dashv Ft$. So there is an MvF-admissible expansion of $\sigma^{\mathrm{MvF}}_{\alpha}$ which does not satisfy Ft. This contradicts the supposition that $\sigma^{\mathrm{MvF}}_{\alpha+1} \models t\eta\hat{x}F$. So if $\sigma^{\mathrm{MvF}}_{\alpha+1} \models t\eta\hat{x}F$, then $\mathsf{TST} \vdash_{\mathsf{MvFBL}} +t\eta\hat{x}F$. □

I now provide a general model theory for MvFBL. An interpretation $\mathcal{I}$ consists of a set of objects $\mathcal{D}$ and an interpretation function $\mathcal{F}$. The function $\mathcal{F}$ assigns elements of $\mathcal{D}$ to closed terms and an EA-pair σ to η where σ^+ and σ^- consists of ordered pairs of elements of $\mathcal{D}$. The satisfaction and anti-satisfaction clauses for identity, the connectives and the universal quantifier are as above. We then need to extend the definition of model-theoretic consequence to cover the signed language. If A is a sentence, we write $\mathcal{I}_\sigma \models_S +A$ wherever, for all τ that are MvF-admissible expansions of σ, $\mathcal{I}_\tau \models A$, and $\mathcal{I}_\sigma \models_S -A$ wherever, for all for all τ that are MvF-admissible expansions of σ, $\mathcal{I}_\tau \dashv A$. If Γ is a set of signed sentences and φ a signed sentence, we then write $\Gamma \models_S \varphi$ just in case if, for all $\mathcal{I}$, if $\mathcal{I} \models_S \psi$ for all $\psi \in \Gamma$, then $\mathcal{I} \models_S \varphi$.

We can now show that MvFBL is sound and complete with respect to the class of η-admissible models. For soundness, the argument used in Theorem 6.3 above suffices, except that we now use the restriction to η-admissible models to establish the soundness of the η-rules. For completeness, we prove the following model existence result:

Lemma 6.8. *Let Γ be a b-consistent$_{\mathsf{MvFBL}^*}$ set of signed sentences and let Γ^* be a maximally b-consistent$_{\mathsf{MvFBL}^*}$ extension of its closure. Then there is an interpretation $\mathcal{I}$ such that $+A \in \Gamma^*$ iff $\mathcal{I} \models A$, and $-A \in \Gamma^*$ iff $\mathcal{I} \dashv A$.*

Proof. Let $\mathcal{I} = \langle \mathcal{D}, \mathcal{F}\rangle$ be the canonical term model for Γ^* and set $\mathcal{F}(\eta) = \langle \tau^+, \tau^-\rangle$ where τ^+ is $\{\langle \bar{a}, \bar{b}\rangle \mid +a\eta b \in \Gamma^*\}$ and τ^- is $\{\langle \bar{a}, \bar{b}\rangle \mid -a\eta b \in \Gamma^*\}$. Then, for all A, we have that $+A \in \Gamma^*$ iff $\mathcal{I} \models A$, and $-A \in \Gamma^*$ iff $\mathcal{I} \dashv A$. The proof is by induction on the complexity of A.

A is of the form $a\eta b$. $+a\eta b \in \Gamma^*$ iff $\langle \bar{a}, \bar{b}\rangle \in \tau^+$ iff $\mathcal{I} \models a\eta b$. The case of $-a\eta b$ is analogous.

A is of the form $a = b$. $+a = b \in \Gamma^*$ iff $\bar{a} = \bar{b}$ iff $\mathcal{I} \models a = b$. The case of $-a = b$ is analogous.

A is of the form $B \wedge C$. $+B \wedge C \in \Gamma^*$ iff $+B \in \Gamma^*$ and $+C \in \Gamma^*$ (since Γ^* is MBL-closed) iff $\mathcal{I} \models B$ and $\mathcal{I} \models C$ (by the induction hypothesis) iff $\mathcal{I} \models B \wedge C$. Similarly, $-B \wedge C \in \Gamma^*$ iff $-B \in \Gamma^*$ or $-C \in \Gamma^*$ (since Γ^* is maximally b-consistent$_{\mathsf{MvFBL}^*}$) iff $\mathcal{I} \dashv B$ or $\mathcal{I} \dashv C$ (by the induction hypothesis) iff $\mathcal{I} \dashv B \wedge C$.

A is of the form $\neg B$. $+\neg B \in \Gamma^*$ iff $-B \in \Gamma^*$ (since Γ^* is MvFBL-closed) iff $\mathcal{I} \dashv B$ (by the induction hypothesis) iff $\mathcal{I} \models \neg B$ (by the model-theoretic clauses for negation). Similarly, $-\neg B \in \Gamma^*$ iff $+B \in \Gamma^*$ iff $\mathcal{I} \models B$ iff $\mathcal{I} \dashv \neg B$.

A is of the form $\forall x\, F$. $+\forall x\, F \in \Gamma^*$ iff for every $t \in T^*$, $F[t/x] \in \Gamma^*$ (since Γ^* is MvFBL-closed) iff $\mathcal{I} \models Ft$ (by the induction hypothesis) iff $\mathcal{I} \models \forall x\, F$ (by the fact that there is a term for every set and class). Similarly, $-\forall x\, F \in \Gamma$ iff $F[t/x] \in$ MvFBL for every $t \in T^*$ (since Γ^* is MvFBL-closed and MvFBL includes the ($-\forall$E.) rule) iff $\mathcal{I} \dashv Ft$ for every $t \in T^*$ iff $\mathcal{I} \dashv \forall x\, F$. □

We can now establish completeness.

Theorem 6.9. *Let* Γ *be a set of signed sentences and* φ *a signed sentence. If* $\Gamma \models_S \varphi$*, then* $\Gamma \vdash_{\mathsf{MvFBL}} \varphi$*.*

Proof. Suppose that $\Gamma \nvdash_{\mathsf{MvFBL}} \varphi$. It follows that Γ is b-consistent$_{\mathsf{MvFBL}^*}$ and so, by Lemma 6.8, there is an interpretation $\mathcal{I}_\sigma$ and a set of signed sentences Γ^* such that $\mathcal{I}_\sigma \models A$ iff $+A \in \Gamma^*$, and such that $\mathcal{I}_\sigma \dashv A$ iff $-A \in \Gamma^*$. Let $\mathcal{I}'_\tau$ be the interpretation obtained by taking the set of all these $\mathcal{I}_\sigma$ and defining τ as in Theorem (just above). It is easy to check that τ is maximal superset of the set of all admissible expansions of σ. Hence, using Lemma 6.8, $\mathcal{I}_\tau \models_S \psi$ for all $\psi \in \Gamma^*$ and hence for all $\psi \in \Gamma$ (since $\Gamma \subseteq \Gamma'$). Moreover, $\mathcal{I}_\tau \not\models_S \varphi$ since $\varphi \notin \Gamma^*$. So $\Gamma \vdash_{\mathsf{MvFBL}} \varphi$. □

7. Identity

The Maddy–van Fraassen approach goes a long way towards addressing the problem of rampant indeterminacy. For the supervaluational character of the theory means that more generalizations will be declared true than in Maddy's original theory, due to the fact that instances of those generalizations are satisfied or antisatisfied that were not in MBL.

For example, it is indeterminate whether the Russell class belongs to itself: MvFBL proves neither $+\hat{x}(x \not\eta x)\eta\hat{x}(x \not\eta x)$ nor $+\hat{x}(x \not\eta x) \not\eta \hat{x}(x \not\eta x)$. Indeed, when augmented with a force marker for the speech act expressing agnosticism and corresponding coordination principles (as we did in the case of MBL), MvFBL proves $?\hat{x}(x \not\eta x)\eta\hat{x}(x \not\eta x)$. Nonetheless, MvFBL proves that every thing either belongs or does not belong to the Russell class, that is $+\forall y(y\eta\hat{x}(x \not\eta x) \vee y \not\eta \hat{x}(x \not\eta x))$.

The supervaluational character of the Maddy–van Fraassen approach also opens up the way to a Frege–Russell treatment of cardinality. As we saw above, Maddy had proved that if the statement that two non-empty collections are equinumerous is not satisfied in her original theory of classes, then it is neither satisfied nor anti-satisfied. To understand the proof, it will be helpful to state in full the Frege–Russell definition of equinumerosity adopted by Maddy.

Definition 7.1. For $t, t' \in T$, $t \approx t'$ abbreviates

$$\exists z(\forall u \forall v \forall w((\langle u, v\rangle \eta z \wedge \langle u, w\rangle \eta z \supset v = w) \wedge (\langle u, v\rangle \eta z \wedge \langle w, v\rangle \eta z \supset u = w)) \wedge$$
$$\forall u((u\eta t \supset \exists v(v\eta t' \wedge \langle u, v\rangle \eta z)) \wedge (u\eta t' \supset \exists v(v\eta t \wedge \langle v, u\rangle \eta z))))$$

Maddy's proof then goes as follows. To establish that $t \not\approx t'$, we need to show that every member of T^* falsifies one of the conjuncts of the definition of equinumerosity. Maddy then shows that if t and t' are non-empty, this cannot be. A crucial step in the proof is that the two conjuncts of the first clause, namely

$$((\langle u,v\rangle\eta z \wedge \langle u,w\rangle\eta z \supset v = w) \wedge (\langle u,v\rangle\eta z \wedge \langle w,v\rangle\eta z \supset u = w))$$

can never be falsified because there are classes such that, for any collection, the theory does not decide whether the collection belongs to them. (For instance, for every t, $\sigma^M \perp\!\!\!\perp t\eta\hat{x}(x \not\eta y)$.) However, this step fails in the Maddy–van Fraassen theory of classes, since a material conditional $A \to B$ can be false even though both A and B are indeterminate. Similar considerations apply to the reasoning Maddy employs to establish that the two conjuncts of the second clause can never be falsified.

The Maddy–van Fraassen approach does not go all the way towards addressing the problem of rampant indeterminacy, however. Recall that Linnebo's second problem concerned identity and in particular the fact that identity among classes is too fine-grained on Maddy's theory, so that whenever F and G are different formulae, then $\hat{x}F$ and $\hat{x}G$ are different classes. As we saw, a natural option to deal with the problem, one already considered by Maddy, is to define identity in terms of the equivalence relation $\simeq$, where $t \simeq t' \equiv_{\text{def}} \forall z(z\eta t \leftrightarrow z\eta t')$. While promising, this definition failed to even be reflexive because of the problem of rampant indeterminacy. In the Maddy–van Fraassen theory, by contrast, the relation behaves structurally as we want. In particular, $\sigma^{\text{MvF}} \models \forall z(z\eta t \leftrightarrow z\eta t)$ and so $\sigma^{\text{MvF}} \models t \simeq t$. Indeed, it is easy to check that $\simeq$ is an equivalence relation in the Maddy–van Fraassen theory of classes, as desired.

Nonetheless, the problem of identity is far from being solved. For we still have that $\sigma^{\text{MvF}} \perp\!\!\!\perp \forall z(z\eta\hat{x}(x \not\eta x) \leftrightarrow z\eta\hat{x}(x \not\eta x \wedge x \not\eta x))$. It follows that, like Maddy's theory, the Maddy–van Fraassen theory is undecided about whether $\hat{x}(x \not\eta x)$ is identical to $\hat{x}(x \not\eta x \wedge x \not\eta x)$. The reason, essentially, is that, in the definition of the Maddy–van Fraassen jump, we are supervaluating over all expansions consistent with what has already been established about the extension and anti-extension of the membership relation. These expansions include ones in which $\hat{x}(x \not\eta x)$ and $\hat{x}(x \not\eta x \wedge x \not\eta x)$ have different membership constituencies. Should the identity problem persist, Linnebo's negative assessment of the prospects for a hierarchical approach to classes based on a Kripke-style construction would after all be justified.

From the way I have introduced it, one might form the impression that the identity problem only besets Maddy's original theory and its supervaluational cousin. This impression would be mistaken. For the problem is really a problem about how to treat identity within the context of a naïve theory of classes—that is, a theory of classes that validates the unrestricted η-rules.

The issue is brought into sharp relief by Restall (2010), who presents a result that spells trouble for Hartry Field's (2008) theory of properties and indeed any naïve theory of properties or classes. Restall's result, which generalizes a theorem of Roland Hinnion (Hinnion and Libert, 2003), shows that there are limits to how finely classes can be identified. In particular, suppose that we take identity among classes to be governed by the following introduction and elimination rules:

$$(+{=}\,\text{I.})\ \frac{\begin{matrix}[+t\eta\hat{x}F] \\ \vdots \\ +t\eta\hat{x}G\end{matrix} \qquad \begin{matrix}[+t\eta\hat{x}G] \\ \vdots \\ +t\eta\hat{x}F\end{matrix}}{+\hat{x}F = \hat{x}G}\ \text{if the subderivations of } +t\eta\hat{x}G \text{ and } +t\eta\hat{x}F \text{ use no side premisses}$$

$$(+{=}\,\text{E.})\ \frac{+t\eta\hat{x}F \qquad +\hat{x}F = \hat{x}G}{+t\eta\hat{x}G}$$

Then, we can derive a contradiction using only the asserted η-rules and assuming the existence of a sentence from whose assertion anything follows (which we do have in MBL and hence MvFBL, since both system includes the principle of Bilateral Explosion).[11] Essentially, what is happening is that the resources of class theory allow us to derive a Curry-style result without using a conditional, whose logical complexity is simulated by using the class-term operator $\hat{}$.

The difficulties surrounding extensionality in the context of a naïve theory of classes have long been known (Gilmore, 1974; Grišin, 1982; White, 1979). One might however have been tempted to set those difficulties aside on the grounds that full-blown extensionality for classes, understood as logical collections, might not be desirable anyway: while the set of chordates and the set of renates, to use Quine's (1951) famous example, are the same, the corresponding classes or properties are different. Restall's result makes it clear that this reaction would be hasty. For the identity rules used in his derivation do not require classes to be extensional, since the subderivation in the introduction rule is restricted so as to rule out side premisses. More than mere extensional equivalence is needed for two classes to be declared identical. Indeed, Restall takes ($+=$ I.) to encode the idea that (rephrasing things in our framework) if asserting $t\eta\hat{x}F$ commits one to asserting $t\eta\hat{x}G$ on purely logical grounds, and if asserting $t\eta\hat{x}G$ commits one to asserting $t\eta\hat{x}F$ on purely logical grounds, then one is already committed to asserting that $\hat{x}F$ and $\hat{x}G$ are identical. Restall goes on to argue that this intensional criterion of identity of classes is hard to reject:

> To reject [($+=$ I.)] is to reject a coarse account of properties. It is to not only accept a finer individuation of properties where logically equivalent statements pick out distinct properties, but to hold that any attempt at defining coarse properties in terms of fine ones must fail. But at what point does that construction break down? It seems like a straightforward construction of equivalence classes or their representatives. To reject [($+=$ I.)] is also to leave open the vexed question of what the identity conditions for properties can be, and should be. Just what is a property that requires that they be more finely individuated than logic requires? Field's model construction doles out properties in exactly the measure of our language. But it would be bizarre to think that this is an adequate picture of properties. How convenient it would be if properties fit our language like hand to glove? But which language? My language now? Or yours? How are we to get to an adequate understanding of the properties picked out by this naïve theory of properties? (Restall, 2010, pp. 442–443)

Restall is certainly correct that an identity criterion for classes which closely tracks their syntax is unsatisfactory. This holds for the criterion of identity which can be extracted from the model-theoretic construction of Field (2008), which is the focus of Restall's paper, as much as it does for Maddy's original definition of identity.

It is less clear, however, that '[t]o reject [($+=$ I.)] is to reject a coarse account of properties' (or classes). In his response (Field, 2010), Field interprets Restall's ($+=$ I.) as proof-theoretically encoding the model-theoretic principle that if $t\eta\hat{x}F \models_w t\eta\hat{x}G$ and $t\eta\hat{x}G \models_w t\eta\hat{x}F$, then $\models_w \hat{x}F = \hat{x}G$ (where $A \models_w B$ is *weak entailment*: in every model in which A is satisfied, so is B). But, Field

[11] Restall presents his derivation in a sequent calculus setting, which makes explicit the structural assumptions. In that context, the derivation uses the rules of Reflexivity and Cut.

says, this is implausible. If we let L be a Liar sentence, and F be '$L \wedge x$is a kangaroo' and G be '$L \wedge x$is a cockroach', then $t\eta\hat{x}F \models_w t\eta\hat{x}G$ and $t\eta\hat{x}G \models_w t\eta\hat{x}F$ in Field's theory of truth, but we would not want to conclude that $\hat{x}F = \hat{x}G$.

In more recent work, Restall has returned to the issue in collaboration with Shawn Standefer and Rohan French (Standefer et al., 2020). They point out that Restall's result is, first and foremost, a proof-theoretic result, based on the proof-theoretic principle that if I can derive $t\eta\hat{x}G$ from $t\eta\hat{x}F$ (appealing to other assumptions) and *vice versa*, then I can derive $\hat{x}F = \hat{x}G$. Field assumes that if I can derive B from A appealing to no other assumptions, all I learn is that A weakly entails B. But perhaps I learn more. Perhaps I learn that A strongly entails B, where A strongly entails B just in case $\models_w A \to B$. Given the failure of the Deduction Theorem for the conditional Field uses and as suggested by the terminology, weak entailment does not imply strong entailment.

Now, as Standefer et al. (2020) note, it is hard to tell what is needed proof-theoretically in order to establish a strong entailment, since Field has not provided a proof theory for his theory of truth, but only a model theory. We can do so, however, for the Maddy–van Fraassen theory of classes. To this end, it will be helpful to return to the definition of identity as $\simeq$. It is easy to see that, in the context of the supervaluationist setting of the Maddy–van Fraassen theory, this definition sanctions the following rules:

$$(+\simeq \text{I.})\ \frac{\begin{matrix}[+t\eta\hat{x}F] \\ \vdots \\ +t\eta\hat{x}G\end{matrix} \qquad \begin{matrix}[+t\eta\hat{x}G] \\ \vdots \\ +t\eta\hat{x}F\end{matrix}}{+\hat{x}F \simeq \hat{x}G}\ \text{if the subderivations of } +t\eta\hat{x}G \text{ and } +t\eta\hat{x}F \text{ use no } \eta\text{-rules}$$

$$(+\simeq \text{E.})\ \frac{+t\eta\hat{x}F \qquad +\hat{x}F \simeq \hat{x}G}{+t\eta\hat{x}G}$$

Thus, there is at least a sense, brought out by the proof theory, in which defining identity as $\simeq$ in the context of the Maddy–van Fraassen theory is both more permissive and more restrictive than the definition of identity for classes/properties proposed by Restall. It is more permissive in that it does allow side premisses in the subderivations of the introduction rule for asserted identity. It is more restrictive in that it does not allow the use of the η-rules in the subderivations. In general, to derive $+B$ from $+A$ without using any other assumptions does not suffice in MvFBL to establish that A strongly entails B where strong entailment is defined using the material conditional (that is, $\sigma^{\text{MvF}} \models A \to B$). What is rather needed is that the derivation from $+A$ to $+B$ does not use the η-rules.[12]

Using our proof-theoretic resources, we can then clearly explain what more is needed in our setting to establish a strong entailment. However, we still have not solved the problem that the proposed criterion of identity among classes appears far too restrictive. For it is still the case that we cannot conclude in MvFBL that $\hat{x}(x \not\eta x)$ and $\hat{x}(x \not\eta x \wedge x \not\eta x)$ are identical. So one might suspect that Restall was after all right that to reject $(+= \text{I.})$ is to reject a coarse account of classes. And Linnebo would after all be vindicated in thinking that the problem of identity for Maddy's

[12] Weak entailment and strong entailment in Field's system correspond to global and local consequence in a supervaluationist setting. For more on the relationship between global and local consequence, where the latter is characterized using the material conditional, see Incurvati and Schlöder, 2022.

theory should lead us to conclude that a hierarchical approach to classes based on a Kripke-style construction is a failed research programme.

8. Maximal consistency

Fortunately, we can give a coarse account of classes while remaining within the remit of a hierarchical approach to classes based on a Kripke-style construction. To develop such an account, a proof-theoretic point of view will again prove useful. The fact that $(+\simeq \text{I.})$ disallows the use of the η-rules in its subderivations sheds light on why we cannot conclude that $\hat{x}(x \not\eta x)$ and $\hat{x}(x \not\eta x \wedge x \not\eta x)$ are identical. For it is rather natural to try to establish that $\hat{x}(x \not\eta x)$ and $\hat{x}(x \not\eta x \wedge x \not\eta x)$ via the following derivation:

$$\dfrac{\dfrac{\dfrac{\dfrac{[+t\eta\hat{x}(x \not\eta x)]}{+t \not\eta t}\,(+\eta\text{ E.}) \quad \dfrac{[+t\eta\hat{x}(x \not\eta x)]}{+t \not\eta t}\,(+\eta\text{ E.})}{+t \not\eta t \wedge t \not\eta t}\,(+\wedge\text{ I.})}{+t\eta\hat{x}(x \not\eta x \wedge x \not\eta x)}\,(+\eta\text{ I.}) \quad \dfrac{\dfrac{\dfrac{[+t\eta\hat{x}(x \not\eta x \wedge x \not\eta x)]}{+t \not\eta t \wedge t \not\eta t}\,(+\eta\text{ E.})}{+t \not\eta t}\,(+\wedge\text{ E.})}{+t\eta\hat{x}(x \not\eta x)}\,(+\eta\text{ I.})}{+t\eta\hat{x}(x \not\eta x) \simeq t\eta\hat{x}(x \not\eta x \wedge x \not\eta x)}\,(+\simeq\text{ I.})$$

However, this derivation is not valid in MvFBL because it makes use of the η-rules in the subderivations of the $\simeq$ introduction rule. Now, the ban on applications of η-rules in subderivations was motivated on the grounds that, from a supervaluationist standpoint, classical reasoning is in order when no rules are used that can engender indeterminacy. But from a supervaluationist standpoint, classical reasoning would also seem to be in order when, even though the rules are used, they are used only to access and therefore exploit the logical structure of the defining formula F of some class $\hat{x}F$. For this structure and what follows from it are unaffected by the way in which the indeterminacy might be resolved. We can formally capture this idea by weakening the restriction on the $\simeq$ introduction rule so that the use of the η-rules is allowed when the subderivations proceed first by eliminating η and then introducing it at the end.

$$(+\simeq\text{ I.})\ \dfrac{\begin{matrix}[+t\eta\hat{x}A] \\ \vdots \\ +t\eta\hat{x}B\end{matrix} \qquad \begin{matrix}[+t\eta\hat{x}B] \\ \vdots \\ +t\eta\hat{x}A\end{matrix}}{+\hat{x}A \simeq \hat{x}B}$$

if either the subderivations use no η-rules or they begin with an application of $(+\eta\text{E.})$ and $(-\eta\text{E.})$ to every premiss and end with an application of $(+\eta\text{I.})$ or $(-\eta\text{I.})$

With this weaker restriction in place, the above derivation of $+t\eta\hat{x}(x \not\eta x) \simeq t\eta\hat{x}(x \not\eta x \wedge x \not\eta x)$ goes through, as it perhaps should.

Clearly, however, if the motivation for weakening the restriction applies to the introduction rule for asserted class identity, it ought to apply whenever in MvFBL we had restrictions on the subderivations. Indeed, there is also a technical reason for this, namely that we are officially treating the introduction rule for asserted class identity as a derived rule, and in order to be able to derive the more permissive version of it, we must relax the restrictions on the subderivations on the (BET*) rule accordingly. This yields the following rule.

$$\text{(BET**)}\ \frac{\begin{matrix}[+A] & [-A] \\ \vdots & \vdots \\ \varphi & \varphi\end{matrix}}{\varphi}$$ if either the subderivations use no η-rules or they begin with an application of $(+\eta\text{E.})$ and $(-\eta\text{E.})$ to every premiss and end with an application of $(+\eta\text{I.})$ or $(-\eta\text{I.})$

A similar change must be made in the restrictions on the subderivations in the $(-\wedge\text{E.})$ and $(-\forall\text{E.})$ rules. Call (for reasons that will shortly become clear) the resulting theory MCBL. If we let a $\clubsuit_i$ denote one of the two force markers, MCBL can also be axiomatized by adding to MvFBL the following meta-rule of Classical Compositionality, which perhaps more directly captures the idea that classical reasoning is in order when the η-rules are only used to access the logical structure of the defining condition of the relevant class.[13]

$$\text{(CC)}\ \frac{\clubsuit_1 F_1 t_1, \ldots, \clubsuit_n F_n t_n \vdash_{\mathsf{MvFBL}^*} \clubsuit_o Gu}{\clubsuit_1 t_1 \eta \hat{x} F_1, \ldots, \clubsuit_n t_n \eta \hat{x} F_n \vdash_{\mathsf{MvFBL}^*} \clubsuit_o u \eta \hat{x} G}$$

We have arrived at MCBL via proof-theoretic considerations and by reflecting on the original motivation for the restrictions on hypothetical reasoning within a supervaluationist setting. What is remarkable is that MCBL is also a very natural theory from a model-theoretic point of view.

To explain and make precise the sense in which this is the case, we need to introduce some further model-theoretic terminology. I mentioned earlier that Kripke (1975) had already suggested the possibility of using a supervaluational scheme for handling truth-value gaps, and that the first scheme he considered was a straightforward adaption of van Fraassen's original notion of supervaluation to the case of truth. But in his paper, Kripke also considered another supervaluational scheme. The idea of this scheme is to take an expansion of an EA-pair to be admissible just in case it is maximally consistent. This gives rise to the following definition.

Definition 8.1 (mc-admissible expansion)**.** An EA-pair $\tau = \langle \tau^+, \tau^- \rangle$ is a *mc-admissible expansion* of an EA-pair $\sigma = \langle \sigma^+, \sigma^- \rangle$ iff (i) $\tau \sqsupseteq \sigma$, (ii) for no A, both $\tau \models A$ and $\tau \dashv A$, and (iiii) for every A, either $\tau \models A$ or $\tau \dashv A$.

Using this notion in the Supervaluational-jump template (Definition 6.1), we then obtain the maximally consistent jump J_{mc}, and replacing the Maddy jump with J_{mc} in the definition of the Maddy hierarchy, we obtain the *maximally consistent class hierarchy*. Once again, the construction reaches a fixed point σ^{mc}, and so we can consider σ^{mc} as providing the extension and the anti-extension of η according to the theory of classes obtained by adopting the maximally consistent supervaluational scheme within the context of a hierarchical theory of classes.

We are now ready to prove that there is a deep connection between MCBL and the maximally consistent class hierarchy. For, as I am now going to show, over TST, MCBL axiomatizes the fixed point of this hierarchy.[14] I first prove that MCBL is sound with respect to the consequence relation induced by σ^{mc}.

[13] The name is intended to be reminiscent of compositionality in the truth context, where two special cases of this rule have been shown by Incurvati and Schlöder (2023) to deliver material compositionality for the truth predicate. Indeed, the situation suggests to me that the problem of identity in the theory of classes is closely linked to the problem of compositionality in the theory of truth. I hope to explore these connections in future work.

[14] The result mirrors Incurvati and Schlöder's (2023) result that a suitable extension of their multilateral theory of truth axiomatizes the maximally consistent truth hierarchy.

Theorem 8.2. *Let Γ be a set of signed sentences and φ a signed sentence. Suppose that for all ψ in Γ, $\sigma^{\mathrm{mc}} \models A$ if $\psi = +A$, and $\sigma^{\mathrm{mc}} \dashv A$ if $\psi = -A$. Then, if $\Gamma \vdash_{\mathsf{MCBL}} \varphi$, $\sigma^{\mathrm{mc}} \models B$ if $\varphi = +B$ and $\sigma^{\mathrm{mc}} \dashv B$ if $\varphi = -B$.*

Proof. Recall that MCBL can be axiomatized by adding the Classical Compositionality metarule to MvFBL. Now the arguments from 6.3 carry over to the present case, except that for the soundness of ($+\eta$E.) we appeal to the fixed-point property. So it remains to check the soundness of (CC).

Let $\circ$ denote nothing if a given $\clubsuit_i$ is $+$ and denote negation if a given $\clubsuit_i$ is $-$. Then since $\vdash_{\mathsf{MvFBL}^*}$ denotes derivability in MvFBL without use of the η-rules, to prove that (CC) is sound, it suffices to show that if $\sigma^{\mathrm{mc}} \models \circ F_1 t_1 \wedge \ldots \wedge \circ F_n t_n \wedge \circ\neg Gu$, then $\sigma^{\mathrm{mc}} \models \circ t_1 \eta \hat{x} F_1 \wedge \ldots \wedge \circ t_n \eta \hat{x} F_n \wedge \circ u \not\eta \hat{x} G$. So suppose that $\models \circ F_1 t_1 \wedge \ldots \wedge \circ F_n t_n \wedge \circ\neg Gu$ and, for *reductio*, that $\sigma^{\mathrm{mc}} \not\models \circ t_1 \eta \hat{x} F_1 \wedge \ldots \wedge \circ t_n \eta \hat{x} F_n \wedge \circ u \not\eta \hat{x} G$. This means that there is an mc-admissible expansion τ of σ^{mc} such that $\tau \not\models \circ t_1 \eta \hat{x} F_1 \wedge \ldots \wedge \circ t_n \eta \hat{x} F_n \wedge \circ u \not\eta \hat{x} G$. It follows that $\tau \not\models \circ F_1 t_1 \wedge \ldots \wedge \circ F_n t_n \wedge \circ\neg Gu$. But since τ is maximally consistent, this means that $\tau \dashv \circ F_1 t_1 \wedge \ldots \wedge \circ F_n t_n \wedge \circ\neg Gu$. But since τ is an expansion of σ^{mc}, we also have that $\tau \models \circ F_1 t_1 \wedge \ldots \wedge \circ F_n t_n \wedge \circ\neg Gu$. Hence, τ is not classically consistent and hence not an mc-admissible expansion. □

Similarly to the cases of MBL and MvFBL, Theorem 8.2 yields that, over TST, MCBL is sound with respect to σ^{mc}.

Theorem 8.3. *For every sentence A, if $\mathsf{TST} \vdash_{\mathsf{MCBL}} +A$, then $\sigma^{\mathrm{mc}} \models A$, and if $\mathsf{TST} \vdash_{\mathsf{MCBL}} -A$, then $\sigma^{\mathrm{mc}} \dashv A$.*

Given our earlier results and proofs, it is then easy to see that, over TST, MCBL is also complete with respect to σ^{mc}.

Theorem 8.4. *For every sentence A, if $\sigma^{\mathrm{mc}} \models A$, then $\mathsf{TST} \vdash_{\mathsf{MCBL}} +A$, and if $\sigma^{\mathrm{mc}} \dashv A$, then $\mathsf{TST} \vdash_{\mathsf{MCBL}} -A$.*

Proof. The proofs of Lemma 6.5 and Theorem 6.6 go through as before. The same goes for the proof of Theorem 6.7, except that we now need to show that the τ defined in the course of the proof, besides satisfying condition (i) of the definition of an mc-admissible expansion of $\sigma^{\mathrm{mc}}_{\alpha}$ (that is, being an expansion of it), also satisfies conditions (ii) and and (iii) (that is, being consistent and maximal).

For consistency, suppose that $\tau \models A$ and $\tau \dashv A$ for some A. Then, by Theorem 6.6 adapted to the case of MCBL, it follows that $\Gamma \vdash_{\mathsf{MCBL}^*} +A$ and $\Gamma \vdash_{\mathsf{MCBL}^*} -A$, contradicting the assumption that Γ is b-consistent. The maximality of τ can be established in a similar manner. □

It is then straightforward to provide a general model theory for MCBL by requiring the models to be not only η-admissible but, in effect, to satisfy Classical Compositionality.

9. Conclusion

Far from being a failed research programme, the Maddian approach to classes is alive and well, once we pursue it along the supervaluational lines already mentioned by Kripke in 1975. The rampant indeterminacy of Maddy's original approach is not inherent to a hierarchical approach to classes based on a Kripke-style construction. Rather, it is due to the particular implementation using the scheme for handling truth-value gaps Kripke focused on, which gives rise to a Strong

Kleene Logic. By using a different, supervaluational scheme, we can address the basic problem of rampant indeterminacy. And by focusing on the specific supervaluational scheme based on maximally consistent expansions, we can give a natural solution to the longstanding problem of identity for a naïve theory of classes, meeting a more general challenge for naïve theories of classes issued by Restall. Looking at the problem from a proof-theoretic perspective, and providing bilateral and multilateral deductive systems for the theories of classes examined, played a key role in arriving at this solution, in that it made it clear exactly what shape a rule of identity among classes ought to have.

Many questions remain open. In closing, let me highlight two sets of questions that I deem especially pressing. The first is that of the conditional in MCML. Although the theory includes the unrestricted η-rules, it does not validate the comprehension schema where the schema is formulated using the material conditional of the theory. According to Field et al. (2017), any naïve theory of classes worth its name ought to validate the unrestricted comprehension schema. Field et al. (2017) go on to prove an impossibility result concerning extensionality in the setting of a naïve theory of classes: in the presence of what they deem to be very modest demands on extensionality and the conditional, it is not possible to consistently validate the unrestricted comprehension schema.

Now, Incurvati and Schlöder (2023, Ch. 8) have developed a theory of conditionals on which one of the principles Field et al. (2017) use in their proof (namely, Quasi-Substitutivity) appears to fail. For this reason, it might be profitable to investigate whether it is possible to validate a version of the comprehension schema formulated using the Incurvati-Schlöder indicative conditional:

Question 9.1. Is the theory obtained by adding the comprehension schema formulated using the Incurvati-Schlöder conditional to MCML consistent?

If the answer to this question is positive, a subsequent question would then be whether the Incurvati-Schlöder conditional suffices to sustain a decent amount of conditional reasoning within the theory.

This brings me to the second set of questions, which concern the mathematical strength of the maximally consistent theory of classes. Much is not known here at the time of writing. For instance, it is not known whether a fully satisfactory theory of Frege–Russell cardinal numbers can be developed within MCML. We know that MCML improves on Maddy's theory of classes by establishing basic cardinality facts. For instance, while Maddy's theory could not even prove that $\{\emptyset\}$ is not equinumerous with $\{\emptyset, \{\emptyset\}\}$, this holds in the maximally consistent theory of classes. The intuitive reason for this is that in any model over which we are supervaluating there is no class witnessing the one-to-one correspondence between $\{\emptyset\}$ and $\{\emptyset, \{\emptyset\}\}$, and so no such correspondence exists in the universe of the maximally consistent theory of classes. However, it seems plausible to require that MCML should be able to prove Hume's Principle, formulated as $\hat{x}(x \approx t) = \hat{x}(x \approx t') \leftrightarrow t \approx t'$.

Question 9.2. Is Hume's Principle provable in MCML?

Given the difficulties surrounding the development of a theory of Frege–Russell cardinal numbers within her theory of classes, Maddy (2000, 313–314) explores the prospects of developing a theory of cardinal numbers based on von Neumann ordinals. While the results are more encouraging in this case, the theory still falls short of deriving arithmetic. In particular, the

Induction Axiom is not provable because of the problem of rampant indeterminacy. One might therefore ask:

Question 9.3. Are the Peano Axioms provable in MCML given a definition of cardinality *à la* von Neumann?

In the theory of truth, much work has been devoted to establishing the proof-theoretic strength of various axiomatizations of the truth predicate. For instance, and relevantly for our purposes, it was established by Andrea Cantini (1990) that a theory of truth using a more demanding supervaluational scheme than the van Fraassen one but less demanding than the maximally consistent one has the same proof-theoretic strength as the theory of inductive definitions known as ID_1. For this reason, it is to be expected that the theory of classes developed here will have considerable proof-theoretic strength. I therefore ask:

Question 9.4. What is the proof-theoretic strength of MCML?

Acknowledgements. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101086295) within the project *Philosophical, Logical, and Experimental Routes to Substructurality*. For comments and discussion, I would like to thank Maria Beatrice Buonaguidi, Orestis Dimou Belegratis, Pablo Dopico Fernandez, Øystein Linnebo, Penelope Maddy, Simone Picenni, Graham Priest, Lorenzo Rossi, and a referee for this journal. Earlier versions of this material were presented at the Graduate Center of the City University of New York, at the University of Turin, at the University of Pisa, and at the University of Calcutta. I am grateful to the members of these audiences for their valuable feedback.

References

Aczel, P. (1980). Frege structures and the notions of proposition, truth and set. In Barwise, J., Keisler, J., and Kunen, K., editors, *The Kleene Symposium*, volume 101 of *Studies in Logic and the Foundations of Mathematics*, pages 31–59, North Holland, Amsterdam. Elsevier. DOI: https://doi.org/10.1016/S0049-237X(08)71252-7.

Barton, N. and Williams, K. J. (2024). Varieties of class-theoretic potentialism. *The Review of Symbolic Logic*, 17:272–304. DOI: https://doi.org/10.1017/S1755020323000126.

Bealer, G. (1982). *Quality and Concept*. Clarendon Press, Oxford. DOI: https://doi.org/10.1093/acprof:oso/9780198244288.001.0001.

Boolos, G. (1974). Reply to Charles Parsons' "Sets and classes". In Boolos, 1998, 30–36.

Boolos, G. (1985). Nominalist platonism. *Philosophical Review*, 94:327–344. Reprinted in Boolos, 1998, 73–87. DOI: https://doi.org/10.2307/2185003.

Boolos, G. (1998). *Logic, Logic, and Logic*. Harvard University Press, Cambridge, Massachusetts.

Button, T. and Trueman, R. (2024). A fictionalist theory of universals. In Fritz, P. and Jones, N. K., editors, *Higher-Order Metaphysics*, pages 245–290, Oxford. Oxford University Press. DOI: https://doi.org/10.1093/oso/9780192894885.003.0007.

Cantini, A. (1990). A theory of formal truth arithmetically equivalent to ID_1. *Journal of Symbolic Logic*, pages 244–59. DOI: https://doi.org/10.2307/2274965.

Chierchia, G. and Turner, R. (1988). Semantics and property theory. *Linguistics and Philosophy*, 11:261–302. DOI: https://doi.org/10.1007/BF00632905.

del Valle-Inclan, P. (2023). Harmony and normalisation in bilateral logic. *Bulletin of the Section of Logic*, 52:377–409. DOI: https://doi.org/10.18778/0138-0680.2023.14.

Ferrari, F. and Incurvati, L. (2022). The varieties of agnosticism. *The Philosophical Quarterly*, 72:365–380. DOI: https://doi.org/10.1093/pq/pqab038.

Field, H. (2008). *Saving Truth from Paradox*. Oxford University Press, New York. DOI: https://doi.org/10.1093/acprof:oso/9780199230747.001.0001.

Field, H. (2010). Replies to commentators on *Saving Truth from Paradox*. *Philosophical studies*, 147:457–470.

Field, H., Lederman, H., and Øgaard, T. F. (2017). Prospects for a naive theory of classes. *Notre Dame Journal of Formal Logic*, 58:461–506. DOI: https://doi.org/10.1215/00294527-2017-0010.

Fine, K. (1975). Vagueness, truth, and logic. *Synthese*, 30:265–300. DOI: https://doi.org/10.1007/BF00485047.

Florio, S. (2014). Unrestricted quantification. *Philosophy Compass*, 9:441–454. DOI: https://doi.org/10.1111/phc3.12127.

Frege, G. (1884). *Die Grundlagen der Arithmetik*. Wilhelm Koebner, Breslau. Translated as Frege, 1953.

Frege, G. (1892). On concept and object. *Vierteljahrsschrift für wisseschaftliche Philosophie*, 16:192–205. Translated in Geach and Black, 1952, 42–55.

Frege, G. (1953). *The Foundations of Arithmetic*. Basil Blackwell, Oxford, 2nd edition.

Fujimoto, K. (2019). Predicativism about classes. *The Journal of Philosophy*, 116:206–229. DOI: https://doi.org/10.5840/jphil2019116413.

Geach, P. and Black, M., editors (1952). *Translations from the Philosophical Writings of Gottlob Frege*. Basic Blackwell, Oxford.

Gilmore, P. C. (1974). The consistency of partial set theory without extensionality. In Jech, T. J., editor, *Axiomatic Set Theory II. Proceedings of Symposia in Pure Mathematics, 13*, pages 147–153, Providence, RI. American Mathematical Society.

Gitman, V., Hamkins, J. D., Holy, P., Schlicht, P., and Williams, K. J. (2020). The exact strength of the class forcing theorem. *The Journal of Symbolic Logic*, 85:869–905. DOI: https://doi.org/10.1017/jsl.2019.89.

Grišin, V. N. (1982). Predicate and set-theoretic calculi based on logic without contractions. *Mathematics of the USSR-Izvestiya*, 18:41. DOI: https://doi.org/10.1070/IM1982v018n01ABEH001382.

Hamkins, J. D., Kirmayer, G., and Perlmutter, N. L. (2012). Generalizations of the kunen inconsistency. *Annals of Pure and Applied Logic*, 163:1872–1890. DOI: https://doi.org/10.1016/j.apal.2012.06.001.

Hinnion, R. and Libert, T. (2003). Positive abstraction and extensionality. *The Journal of Symbolic Logic*, 68:828–836. DOI: https://doi.org/10.2178/jsl/1058448441.

Holy, P., Krapf, R., Lücke, P., Njegomir, A., and Schlicht, P. (2016). Class forcing, the forcing theorem and Boolean completions. *The Journal of Symbolic Logic*, 81:1500–1530. DOI: https://doi.org/10.1017/jsl.2016.4.

Horsten, L. and Linnebo, Ø. (2016). Term models for abstraction principles. *Journal of Philosophical Logic*, 45:1–23. DOI: https://doi.org/10.1007/s10992-015-9344-z.

Incurvati, L. (2020). *Conceptions of Set and the Foundations of Mathematics*. Cambridge University Press, Cambridge. DOI: https://doi.org/10.1017/9781108596961.

Incurvati, L. and Schlöder, J. J. (2017). Weak rejection. *Australasian Journal of Philosophy*, 95:741–60. DOI: https://doi.org/10.1080/00048402.2016.1277771.

Incurvati, L. and Schlöder, J. J. (2019). Weak assertion. *Philosophical Quarterly*, 69:741–770. DOI: https://doi.org/10.1093/pq/pqz016.
Incurvati, L. and Schlöder, J. J. (2022). Meta-inferences and supervaluationism. *Journal of Philosophical Logic*, 51:1549–1582. DOI: https://doi.org/10.1007/s10992-021-09618-4.
Incurvati, L. and Schlöder, J. J. (2023). Inferential deflationism. *Philosophical Review*, 132:529–578. DOI: https://doi.org/10.1215/00318108-10697531.
Incurvati, L. and Schlöder, J. J. (2023). *Reasoning With Attitude: Foundations and Applications of Inferential Expressivism*. Oxford University Press, New York. DOI: https://doi.org/10.1093/oso/9780197620984.001.0001.
Jones, N. K. (2016). A higher-order solution to the problem of the concept *horse*. *Ergo*, 3. DOI: https://doi.org/10.3998/ergo.12405314.0003.006.
Keefe, R. (2000). *Theories of Vagueness*. Cambridge University Press, Cambridge.
Kreisel, G. (1967). Informal rigour and completeness proofs. In Lakatos, I., editor, *Problems in the Philosophy of Mathematics*, pages 138–171, Amsterdam. North-Holland.
Kriener, J. (2014). The groundedness approach to class theory. *Inquiry*, 57:244–273. DOI: https://doi.org/10.1080/0020174X.2013.855657.
Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy*, 72:690–716. DOI: https://doi.org/10.2307/2024634.
Lear, J. (1977). Sets and semantics. *Journal of Philosophy*, 74:86–102. DOI: https://doi.org/10.2307/2025573.
Linnebo, Ø. (2006). Sets, properties, and unrestricted quantification. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 149–178. Oxford University Press, Oxford. DOI: https://doi.org/10.1093/oso/9780199276424.003.0006.
Linnebo, Ø. (2024). Maddy on classes. In Arbeiter, S. and Kennedy, J., editors, *The Philosophy of Penelope Maddy*, pages 65–80. Springer, Switzerland. DOI: https://doi.org/10.1007/978-3-031-58425-1_6.
Maddy, P. (1983). Proper classes. *Journal of Symbolic Logic*, 48:113–139. DOI: https://doi.org/10.2307/2273327.
Maddy, P. (1990). *Realism in Mathematics*. Clarendon Press, Oxford. DOI: https://doi.org/10.1093/019824035X.001.0001.
Maddy, P. (2000). A theory of sets and classes. In Sher, G. and Tieszen, R., editors, *Between Logic and Intuition. Essays in Honor of Charles Parsons*, pages 299–316, Cambridge. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511570681.016.
Maddy, P. (2024). Reply to Linnebo. In Arbeiter, S. and Kennedy, J., editors, *The Philosophy of Penelope Maddy*, pages 81–84. Springer, Switzerland. DOI: https://doi.org/10.1007/978-3-031-58425-1_7.
Meadows, T. (2015). Infinitary tableau for semantic truth. *Review of Symbolic Logic*, 8:207–35. DOI: https://doi.org/10.1017/S175502031500012X.
Parsons, C. (1974). Sets and classes. *Noûs*, 8:1–12. DOI: https://doi.org/10.2307/2214641.
Quine, W. V. (1951). Two dogmas of empiricism. *Philosophical Review*, 60:20–43. DOI: https://doi.org/10.2307/2181906.
Quine, W. V. O. (1948). On what there is. *The Review of Metaphysics*, 2:21–38.
Rayo, A. and Uzquiano, G. (1999). Toward a theory of second-order consequence. *Notre Dame Journal of Formal Logic*, 40:315–325. DOI: https://doi.org/10.1305/ndjfl/1022615612.
Rayo, A. and Williamson, T. (2003). A completeness theorem for unrestricted first-order languages. In Beall, J. C., editor, *Liar and Heaps: New Essays on Paradox*, pages 331–356, Oxford. Clarendon Press. DOI: https://doi.org/10.1093/oso/9780199264803.003.0016.

Restall, G. (2010). What are we to accept, and what are we to reject, while saving truth from paradox? *Philosophical Studies*, 147:433–443. DOI: https://doi.org/10.1007/s11098-009-9468-5.

Rumfitt, I. (2000). "Yes" and "no". *Mind*, 109:781–823. DOI: https://doi.org/10.1093/mind/109.436.781.

Russell, B. (1903). *Principles of Mathematics*. Allen & Unwin, London.

Schindler, T. (2019). Classes, why and how. *Philosophical Studies*, 176:407–435. DOI: https://doi.org/10.1007/s11098-017-1022-2.

Simonelli, R. (Forthcoming). "Yes", "no", neither, and both. *Synthese*.

Smiley, T. (1996). Rejection. *Analysis*, 56:1–9. DOI: https://doi.org/10.1111/j.0003-2638.1996.00001.x.

Standefer, S., French, R., and Restall, G. (2020). Proofs and models in naive property theory: A response to Hartry Field's 'Properties, propositions and conditionals'. *Australasian Philosophical Review*, 4:162–177. DOI: https://doi.org/10.1080/24740500.2021.1886690.

Trueman, R. (2021). *Properties and Propositions: The Metaphysics of Higher-Order Logic*. Cambridge University Press, Cambridge. DOI: https://doi.org/10.1017/9781108886123.

van Fraassen, B. C. (1966). Singular terms, truth-value gaps, and free logic. *The Journal of Philosophy*, 63:481–495. DOI: https://doi.org/10.2307/2024549.

White, R. B. (1979). The consistency of the axiom of comprehension in the infinite-valued predicate logic of Łukasiewicz. *Journal of Philosophical Logic*, 8:509–534. DOI: https://doi.org/10.1007/BF00258447.