Generically invariant set theory

John R. Steel

May 29, 2023

1 Introduction

Large cardinal hypotheses have been very successful at removing the incompleteness of ZFC in the realm of statements about concrete objects like natural numbers, real numbers, and simply definable sets of real numbers. On the other hand, nothing like them can decide the Continuum Hypothesis (CH), and many other natural questions about arbitrary sets of reals and objects of still higher type are provably beyond their reach. There is at present no generally agreed upon approach to deciding such questions.

This leads us to ask what it is about the large cardinal hypotheses, or ZFC itself for that matter, that justifies them, and whether there might be some other class of principles, justified similarly or perhaps on different grounds, that would have similar success in the realm of arbitrary sets of reals and beyond. And how would we recognize such principles if we had them in hand? What is it to be a solution to the Continuum Problem?

Penelope Maddy has written extensively on such questions. Indeed, she is one of the first to have written from a philosopher's point of view about the evidentiary standards employed by modern set theorists. Her approach is that of the Naturalist, informed by a careful analysis of the evolution of existing theory, and reluctant to use some pre-determined theory of evidence as a reason to reject what set theorists[1] regard as a very successful theory. It is probably not surprising that, as a set theorist, I am sympathetic to her approach. We have exchanged views many times over the years.

It would not be possible to engage here with the whole of Maddy's work in this area, but fortunately she and Toby Meadows have recently published [19], whose starting point is the semi-philosophical paper [30] that I wrote in 2014. My plan here is to make some replies to [19], connecting these to more general aspects of Maddy's work where appropriate. I shall also update the more technical side of [30], in a way that I hope will be accessible to most readers.

[1] "Practitioners".

The paper is organized as follows. §2 is devoted to general philosophical issues in the background. §3 discusses the twin heuristic principles *Maximize* and *Unify* underlying the search for new axioms in set theory. §4 reviews some definitions from [30], and §5 replies to some of the criticisms of its more philosophical parts made by Maddy and Meadows. §6 looks at the idea that the generic multiverse has a core, elaborating on a discussion in Meadows' paper [23] and responding to a criticism made there. §7 describes some recent technical progress related to the the idea that the generic multiverse has a core that resembles HOD in a model of the Axiom of Determinacy, and §8 makes some final remarks.

2 Some philosophy

Maddy has been heavily influenced by Quine, as was I. We have both come to disagree with him, in somewhat different ways. Some of those differences are relevant to the philosophical critique of [30] in [19].

2.1 Naturalism and Holism

Naturalist epistemology takes our existing best theory of everything as its starting point. It rejects the radical re-building projects of "First Philosophy", with its idealized knower who begins from nothing. Our theory and its language have been evolving for a very long time, and although individual humans may make important additions or revisions, no one can start from scratch. One beautiful statement of this attitude is Neurath's well-known boat metaphor:

> Wie Schiffer sind wir, die ihr Schiff auf offener See umbauen müssen, ohne es jemals in einem Dock zerlegen und aus besten Bestandteilen neu errichten zu können.

Holism adds the view that this theory does not have a distinguished starting point. The language in which we express it gains meaning, and the theory within that language that we adopt is confirmed, as a whole. Very well-developed parts of it (such as set theory) have the great virtue that we have organized them as axiomatic systems, but even then, the choice between equivalent axiomatizations is somewhat arbitrary.

One should hasten to add that "our existing best theory of everything" is itself an idealization. In even the fairly well-developed area of theoretical physics, what we have now are two very successful theories (General Relativity and Quantum Field Theory) that seem on the surface to be inconsistent with each other. Fortunately, our focus in this paper is mathematics, where (we claim) the superficially inconsistent alternative theories can be uni-

fied.[2] The holistic point is simply that our mathematical language gains meaning, and our mathematical theory is confirmed, by the way that it interacts with the rest of what we know.

Holism and naturalism lead to Quine's indispensability argument. To put it briefly: mathematics is essential to the rest of science, so let's not dispense with it. It might be hard to find anyone who would disagree so far; for Quine, this argument was a reason to reject his former nominalism, so it was accompanied by the claim that any consistent nominalist would have to dispense with mathematics. I agree with that, and it seems that perhaps Maddy no longer does[3], but it does not matter so much here. Burgess' paper [4] discusses the ontological issue thoroughly, and I agree pretty much completely with his point of view.

It's not so important here because our focus is the epistemology of mathematics, not its ontology. In this context we are dealing with mathematical theories, not mathematical objects; with set theories, not sets. Bringing sets into the discussion prematurely can lead one astray, and in particular into object-oriented formulations of *maximize* which, in my opinion, are mistaken.

Quine went on to say that we *should* dispense with the part of mathematics that is not essential to the rest of science. Quine drew his dividing line rather cavalierly, saying of the 'higher reaches of set theory' that

> We see them as meaningful because they are couched in the same grammar and vocabulary that generate the applied parts of mathematics. We are just sparing ourselves the unnatural gerrymandering of grammar that would be needed to exclude them.

The "we" here is apparently the subset of humanity sufficiently educated to have an opinion, and the passage implies that Quine is putting himself in that category. It is not clear whether he objects to putting up with nonsense in order to be spared the effort involved in avoiding it.[4] Others have tried to draw a dividing line with much more care[5], often motivated by constructivist scruples rather than nominalist ones. The line is always well below ZFC.

Needless to say, I disagree. Pure mathematics is inseparable from the rest of mathematics, and mathematics is inseparable from the rest of science. Gödel's 2nd incompleteness theorem shows that the highest reaches of set theory have consequences for the world we directly

[2]It seems that category theory has been successfully interpreted in set theory; see Maddy [17] and [18] for a recent discussion of the relationship between the two. The unification of superficially inconsistent alternative extensions of ZFC is the main topic of [30].

[3]See [16] Chapter IV, which seems to embrace a form of instrumentalism.

[4]See [26, p. 94]. It is ironic that Quine is using the notion of meaning here to make a philosophical claim, as he does when he claims that the "gratuitous flights of higher set theory" are "on a par rather with uninterpreted systems".([27, p. 788].) An inability to interpret higher set theory would seem to be a good reason to withhold judgement.

[5]See for example [5] and [6].

observe.[6] Those consequences do not affect the subject taught in Physics departments, but consequences like Projective Determinacy do have a bearing on the subject taught in Real Analysis courses. The subject taught in Real Analysis courses is important in Physics and many other areas. Higher set theory also interacts with the more abstract parts of Analysis, Algebra, and Topology.

The idea of dispensing with higher set theory, or "rounding it out"[7] in any old way, may seem to be hard-headed pragmatism, but it is not. Here as elsewhere, powerful unifying principles are preferable to "good enough for now", and it is both important and difficult to find the right ones. Long term pragmatism counsels that while there are basic unsolved problems in pure set theory, we should continue to develop it, regardless of applications. That can only make eventual applications more likely.

The more cogent objection to the work done on large cardinals and their consequences is that it is premature, that it deals with objects and notions so far removed from the rest of mathematics and science that long term pragmatism has no advice to give. The many mutually inconsistent ways of extending ZFC that we have discovered could be taken as evidence of this. What answers this objection is the "one road upward" phenomenon (cf. [29]): at the level of statements about the concrete (natural numbers, real numbers, and absolutely definable sets of real numbers) we have not found any divergences, but rather many different routes to the same theory.[8] The many different routes to it suggest that this theory will not wash away in the long term.

Maddy also rejects the dispensabilty sequel to the indispensability argument, but we may have some disagreement on its flaws. In [15, p. 95] she objects to confirmational holism:

> What's gone wrong with the Quinean picture is confirmational holism. This case suggests that we cannot regard a scientific theory as a homogeneous whole, confirmed as a unit, that a consistent naturalist must recognize different types of evidence and various different roles hypotheses can play in our theorizing.

The case under discussion is the evidence for the existence of atoms and molecules produced by Einstein's explanation of Brownian motion and Perrin's experiments supporting it. I would read the main moral of that case differently, as being that confirmation accumulates

[6] If there is an inaccessible cardinal, then in the next 25 years no contradiction in ZFC will be discovered. This is an experiment that will be performed. Of course, to decide how it turns out we need a proof-checking method; as always, what is confirmed is a theory as a whole.

[7] [27, p. 788].

[8] This "theory of the concrete" contains all $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ truths. (See §7.) There are various general theorems that capture aspects of this "one true hierarchy" phenomenon, for example Wadge's Lemma and the related prewellordering of jump operators (cf. [31]), and the Comparison Lemma of inner model theory. Montalban and Walsh have recently proved an interesting result in this direction at the level of Peano Arithmetic. See [24].

in stages, and that having evidence from different directions is important.[9] If anything, this seems to support the holistic view, since Physics and Chemistry were being asked to face the tribunal together.

Maddy continues

> The Second Philosopher sees the evidential relationships of modern science in its many branches as complex and varied, to be studied and assessed in their particular contexts of inquiry, not obviously subject to characterization like 'observation and the hypothetico-deductive method'.

I would put it differently. Observation and the hypothetico-deductive method is fine as a general characterization, it just doesn't take you very far. A careful assessment in the particular context of inquiry, in our case set theory and the foundations of mathematics, is necessary if you hope to say something useful.

I have brought out these perhaps minor differences because Maddy sometimes writes as if pure mathematics in general, and set theory in particular, had somehow become a domain of inquiry separate from the rest of mathematics and science, subject to its own standards of evidence.[10] Pushed too far, this makes set theory into an art form, with the artists and critics being one and the same.

Such a fictionalist view misses crucial connections. The theory of natural numbers, real numbers, and definable sets of real numbers that we get from large cardinal hypotheses is a key part of the justification for believing them. Those concrete objects show up everywhere in mathematics and its applications. The theory of them we get from large cardinal hypotheses is the only rational route we have toward going beyond ZFC in this realm of the concrete. This is not because we have only investigated a few different ways of extending ZFC, it is because each of the many ways we have investigated is either limiting, or leads to this road.

In this holist view, our mathematical beliefs are ultimately justified by by the central role they play in the whole of our conceptual system. In a sense, all the evidence for them is extrinsic. This may seem paradoxical; do we really believe 2+2=4 because of its consequences?

For the holist, this is an illformed question, because our belief that 2+2=4 cannot be isolated from our other beliefs. One must ask why we believe, or should believe, some reasonably large set of connected propositions. One cannot believe 2+2=4 except as part of some such complex.

---

[9] I would certainly agree that an epistemolgist should "recognize different types of evidence and various different roles hypotheses can play in our theorizing." In the epistemology of set theory, Maddy has gone well beyond Quine in that regard.

[10] For example: " my naturalist takes mathematics to be independent of both first philosophy and natural science (including naturalized philosophy that is continuous with science)–, in short, from any external standard. ([13, p. 184]). Also "she [the second philosopher] sees no opening for the familiar tools of that [scientific] perspective to provide supports, correctives, or supplements to the actual justificatory practices of set theory."[16, p. 55]. I may well be reading more into these passages than is intended.

Moving to that level, the question is still ambiguous. Are we asking the causal question or the justificatory one? And who is the "we"? If the question is the causal one, and the "we" refers to individual human beings alive today, then all sorts of mathematical beliefs are forced upon us as being true, by our biology, our teachers, or random factors. If "we" refers to humanity as a whole, there is at least 70,000 years of linguistic evolution[11], preceded by the evolution of neural structures over many millions of years, that have produced our core mathematical beliefs. They are indeed inevitable, at least for the foreseeable future. At this level, the causes seem more closely connected to the justification. Humans evolved this conceptual system because it does certain jobs well. But if we try to justify some extension of that system, we must say what those jobs are, and why the expanded system would do them better. This seems to be the realm of extrinsic evidence.

In the context of evaluating new axioms for set theory, the interesting question is "why should we adopt them?", the "we" in question is humanity as a whole, and the important evidence given for or against will be extrinsic. The immediate appeal of some axiom can only suggest it as a candidate for more consideration.

The naturalist idea of looking at the history of mathematical belief acquisition as a guide to what standards may be appropriate now does have merit. There is quite a lot to look at there, most of it belonging to prehistory. Fortunately, the part most relevant to evaluating new axioms for set theory belongs to the last 100-200 years. Moreover, the theories we are considering now can be rigorously formalized, so that our discussion of them can be informed by real proofs of various claims. Metamathematics is possible, and it is essential to the epistemology of set theory. One might think of the process of strengthening the current foundations of set theory as linguistic evolution gone self-conscious.

2.2 Meaning skepticism

One of the central criticisms of [30] by Maddy and Meadows has to do with its use of the notion of *meaning*. I shall make some general comments here, and more specific replies in §5.

Of course, philosophers have theorized about the meaning of mathematical statements a lot. The notion has been mis-used. In particular, the logical positivists' use of the claim that true mathematical statements are true solely in virtue of their meaning was very strongly criticized by Quine in his well known debate with Carnap, and in my opinion most of the criticism was just.

Nevertheless, it is pretty much impossible to discuss the epistemology of set theory without bringing the meaning of statements in the language of set theory into the picture. What we are evaluating are *interpreted* theories. How they are interpreted, and translated into one another, is important.

[11] See `https://en.wikipedia.org/wiki/Blombos_Cave`.

In the context of new axioms, it can be hard to distinguish sharpening the meaning of the language from discovering new truths. Perhaps there is some analog of the uncertainty principle that says this can never be fully done. But we do make translations, and they can be good or bad. We do it in daily life, in science, and in mathematics. We can define meaning operationally as that which is preserved by a good translation.

The reason that ZFC counts as a foundation for mathematics is that we can translate all mathematical language into the language of set theory, and prove the translated theorems from the axioms of ZFC. Taking over a term used by Enderton, Maddy calls this translation an "embedding", but "translation" or "interpretation" seem more accurate to me.[12]

Philosophers have written extensively on the notion of meaning, and used it in various contexts. It's in the job description of an *analytic* philosopher. Paraphrasing Hilbert, one might say that depriving an analytic philosopher of the notion of meaning is like depriving a boxer of the use of his fists. One can hope that something in this philosophical literature is relevant to the possibility of meaning indeterminacy in the language of set theory.[13]

But one need not have a theory of a notion in order to employ it. [30] is based on the idea that we do in real life identify and remove meaning indeterminacy, by either trimming back some piece of syntax, or fleshing out (further determining) the meaning of some existing syntax. There are episodes in the history of science that seem to fit this pattern.[14] The inclusion of the Axiom of Extensionality in ZFC can be seen as meaning clarification, in that it abandons the too-vague notion of property in favor of extensions. Several of the other axioms signal that one has abandoned the notion of set as division of all existing objects into members and non-members in favor of the iterative concept. Perhaps it will be useful to think of the resolution of CH as an analogous process. From an Olympian point of view, it is clear that the meaning humanity now assigns to the syntax of the language of set theory must have evolved over time, because at one time homo sapiens did not speak anything like the language of set theory, and its use of the language must have evolved in stages. We don't need a philosophical theory of meaning clarification to see it at work.

Let me contrast this view with the discussion in D. A. Martin's paper [21] of the possibility that what he calls *the concept of the sets* does not determine a truth value for CH. I think the main difference is that Martin's *concept of the sets* is not the same as what I am calling the meaning we currently assign to the syntax of $\mathcal{L}_\in$. Martin's *concept of the sets* does not seem to evolve over time, or depend on humans.[15] In Martin's view, understanding or "grasp"

[12] See [17, p. 290].

[13] Meaning holism seems closest to the notion of meaning that would be relevant as an underpinning for the philosophical parts of [30].

[14] For example, in Special Relativity, one abandons talk of an absolute time ordering, then fleshes out the meaning of "A happened before B" using the speed of light and the Minkowski metric.

[15] Martin says that he stands in between Gödel and Feferman on the nature of mathematical concepts, and that "it is more correct to say that they were discovered than that they were created by us". ([21, p.2]. So he does not rule out Feferman's view that they are social constructs at this point.

of the concept evolves, and perhaps that current understanding would correspond to what I am calling the meaning currently assigned to $\mathcal{L}_\in$. Perhaps what I will describe as removing a meaning indeterminacy in $\mathcal{L}_\in$ would be described by Martin as improving our grasp on the concept of the sets by discovering a particularly basic new truth.

Martin describes his sense of *the concept of the sets* in [20], where it functions as a boundary condition on universes of sets[16] that instantiate it. As Martin puts it,

> When we work on the mathematical subject of set theory, we can think of what we are doing as finding out what is implied by the concept, what has to be true of any instance of the concept.

Martin argues that if neither CH nor its negation is implied by the concept of the sets, then this concept is not instantiated. Nevertheless, set theory can go on, because it is really about what the sets "would be like" if the concept were instantiated. As fas as I can see, the counterfactual doesn't add anything (Martin also leans to that view), so the conclusion should be that instantiation is irrelevant, and that set theory is about what is implied by the concept of the sets.

This makes it hard to find a distinction between ϕ and "ϕ is implied by the concept of the sets". If there is only one *concept of the sets*, then it seems to be functioning as an ideal meaning for the language of set theory. For the naturalist epistemologist, actual humanity and the way it actually uses language are a more promising place to start.

3 Maximize and Unify

In [13] Maddy identifies two heuristic principles guiding the development of set theory, and calls them *Maximize* and *Unify*. Her statement of Maximize in [13] is

> The set theoretic arena in which mathematics is modelled should be as generous as possible; the set theoretic axioms from which mathematical theorems are proved should be as powerful and fruitful as possible.

I agree with the sentiment, and could certainly endorse everything after the semicolon. The more metaphorical first clause leads in the wrong direction, in my opinion. The metaphor is elaborated in her statement of *Generous Arena*, which seems to have supplanted *Maximize* in the discussion of [17]:

> Set theory's universe, V, provides the **Generous Arena** in which all this takes place, and that's why the 'final court' condition takes the form it does: to be a full participant in mathematical interaction, a so-and-so must appear along-side of the full range of its fellows, with all the tools of construction and interaction fully available.

[16] Alternatively, "concepts in the straightforward sense".

The more object-oriented parts of these formulations have led Maddy to attempt formalizations of *Maximize* based on the idea of a theory "providing an isomorphism type".[17] The attempted formalizations led to attempts to define *fair interpretation* formally: "ϕ is a fair interpretation of T in T' (where T extends ZFC) iff (i) T' shows ϕ is an inner model, and (ii) for all $\sigma \in T$, T' proves σ^{ϕ}." An inner model is taken to be a transitive class, perhaps cut off at some inaccessible cardinal. There are counterexamples to the attempts in [13] and [14] to define *Maximize* based on this notion of "fair interpretation".[18] In my opinion, the root of the problem is the idea that one should compare T with T' by considering the "sets they refer to", their "ontologies".

Talk of "the ontology of T" involves quantifying into a modal context. To compare "the ontology of T" to that of S, you need to translate the language of T into that of S. That's where the action is. Moreover, a good translation need not map referring expressions to referring expressions. In the set theory context, T might be interpreted by the user of S as the theory of what is forced in some partial order, for example.[19]

The idea of maximizing sets, rather than set theories, is sometimes given as a rationale for adopting forcing axioms like MM. I find it hard to see the power in such arguments. Why couldn't one argue that CH implies there is a wellorder of $\mathbb{R}$ of length ω_1, while MM implies there is no such thing, so CH implies there are "more sets of reals" than does MM? Moreover, if one takes the fact about the forcing relation that $M[G]$ has more sets than M as a guide to theory choice, isn't one led to "all sets are countable"? We end up with a theory (second order arithmetic) that is much weaker than ZFC in the sense that really matters, namely interpretative power.[20]

Woodin once argued for the axiom (*), now known to be a consequence of MM^{++}[21], on the basis that it maximizes the Π_2 theory of H_{ω_2} in a certain sense.[22] The maximization here is at the level of theories, not sets, but the order we are maximizing in is just inclusion. In other words, the translations have to map Π_2 sentences about H_{ω_2} to themselves. This is not the right notion of maximization, the one that is inspired by the role of set theory as a universal foundation.[23]

[17] Cf. [13, Ch. 6] and [14].

[18] See [13, p. 225]. There are further counterexamples in [10].

[19] The Bennaceraf problems arise from the demand that a translation that is well defined on sentences be well defined on formulae.

[20] Hamkins' paper [10] seems to advocate "all sets are countable" on precisely these grounds. He then attempts to recover logical strength via the instrumentalist dodge (cf. [29]). One symptom of the problem is his claim ([10, p.19]) that the map $t(\psi)$ = "there is a transitive model of $\mathsf{ZFC} + \psi$" is a translation. This is not true; $t(\psi \wedge \psi)$ is not equivalent to $t(\psi) \wedge t(\phi)$. You need to pick a transitive model that is independent of ψ, and there is no way to do that. The fundamental problem is that the confused notion of "more sets" is being used in a context where "more logical strength" is what is appropriate.

[21] See [1].

[22] See for example [38].

[23] Axiom (*) implies $\neg\mathsf{CH}$, but a theory that maximizes the Σ^2_1 theory of the reals in the inclusion order

The notion of fair interpretation that seems appropriate in a general formulation of *Maximize* is the the notion of a good translation. I doubt that this can be formally defined, but we do recognize good and bad translations in practice.

This leads to *maximize interpretative power* as the guiding heuristic for the discovery and evaluation of new axioms in set theory. Our foundational theory of sets should be such that all mathematical theories can be interpreted into it in a way that preserves their meaning. In this formulation, *Maximize* applies to set theories, rather than sets, and this seems more appropriate for an epistemological principle. *Maximize* is a principle of rationality, a broad rule of evidence, not a law of nature. It is parallel to *Unify* in that respect.[24]

To stay at the level of theories, rather than sets, is appropriate for a discussion of what theory one should believe, and of the extent to which believing one alternative conflicts with believing another. As a set theorist who believes some quite strong set existence principles, I am a realist. But set theory and the epistemology of set theory are two different subjects. The fundamental questions of epistemology have to do with meaning, evidence, and belief. In our case, the questions are "what would it be for humanity to add Axiom X to ZFC; how would we know that had happened?", and "what reasons could legitimately be given for or against doing that?". The first has to do with meaning: in an old slogan, the meaning is the use, and to adopt the axiom is to use it in a certain way. The second has to do with evidence. Of course, one can short-circuit a discussion of evidence by saying "adopt it if and only if it is true", but that goes nowhere. One hopes that it is possible to do better.

Maximizing interpretative power captures the central role that large cardinal hypotheses play in extending ZFC. They are our source of interpretative power. They are our most important tool for climbing the consistency strength hierarchy, or equivalently, generating new Π^0_1 truths. As we do that, we generate more complicated truths about the concrete in an orderly fashion. The apparent wellordering of natural consistency strengths is mirrored by the inclusion order on their canonical inner models, and natural models for all these natural theories can be built from the canonical inner models by forcing. This is how set theorists convince themselves that such theories are consistent.[25] [26]

Maddy states *Unify* in [13] (pp. 208-209) as follows:

> One methodological consequence of adopting the foundational goal is immediate: if your goal is to provide a single system in which all objects and structures of mathematics can be modelled or instantiated, then you must aim for a single

must include CH.

[24] Jeffrey Schatz's thesis [28] discusses and compares the two versions of *Maximize*.

[25] There are some qualifiers omitted here. The most important is that this picture has as of now been only partly verified, and there are many fundamental open problems to do with filling it in a lower levels and extending to higher ones.

[26] The consistency strength hierarchy and its higher reaches do not seem to show up naturally in the language of category theory. That gives the language of set theory an advantage as an ultimate foundational language, a language into which all other mathematical language can be interpreted.

> fundamental theory of sets. This methodological goal is just the flip side to one of the common objections to set theoretic foundations ...

At this point she quotes MacLane

> .. Cohen's method of forcing ...leads to the construction of many alternative models of set theory. Another result is the introduction of a considerable variety of axioms meant to supplement ZF ... for these reasons 'set' turns out to have many meanings, so that the purported foundation of all Mathematics upon set theory totters.([12, 358-9].)

Maddy then continues

> We arrive at the methodological goal UNIFY by running this argument in reverse. If set theorists were not motivated by a maxim of this sort, there would be no pressure to settle CH, to decide the questions of descriptive set theory, or to choose between alternative new axiom candidates; it would be enough to consider a multitude of alternative set theories.

Unify is not directly stated here. In [17] and [18] it seems to be replaced by the combination of *Shared Standard* and *Generous Arena*. She says for example ([18, p. 13])

> We need to bear in mind that the cash value of 'these things exist in V' is just 'the existence of (surrogates for) these things can be proved from the axioms of set theory'– a straightforward manifestation of set theory's role as a Shared Standard of proof. To say that 'the universe of sets is the ontology of mathematics' amounts to claiming that the axioms of set theory imply the existence of (surrogates for) all the entities of classical mathematics– a simple affirmation of set theory's role as Generous Arena.

Maddy goes on to warn against figurative ontological talk, but in my opinion the paragraph immediately above has not gone far enough in that direction. The notion of an "entity of classical mathematics" brings in the objects in a discussion that should be devoted to theories.

What MacLane and others miss is how interrelated the many different extensions of ZFC that we have investigated are. The common theory of the concrete that is generated by climbing the consistency strength hierarchy is a dramatic manifestation of this interconnectedness. MacLane, and perhaps Maddy at points, miss this because they do not regard the interpretations given by forcing and inner models to count as unification. But they are. They enable set theorists to use each other's work. The multiverse language and theory are an attempt to make this underlying unity more visible, but set theorists are already well aware of it, and they make use of it.[27]

[27]The paper [2] by Douglas Blue contains many examples of proofs that use the existence of generic extensions of V with some property to deduce conclusions about V itself.

Hamkins misses this unity in a different way. If one strips away the Platonist imagery, his advice to "embrace as real" all universes in his multiverse is equivalent to the formalist's advice to investigate all ("interesting") consistent theories, without any overall framework relating the results of one investigation to those of another. One might say that MacLane sees a mess, and shies away, whereas Hamkins sees the unity of the generic multiverse, but then obscures it with a mess. [28]

4 Generically invariant set theory

Large cardinal hypotheses are cofinal in the part of the interpretability hierarchy we know about. But, like ZFC itself, they are set-forcing-invariant, so they cannot decide CH and the many other statements that are not set-forcing-invariant.

This makes it natural to look for a sublanguage of $\mathcal{L}_\in$ in which the mathematics based on set-forcing-invariant principles can be carried out, and in which set-forcing-sensitive questions have no obvious formalization. This is done in [30], by introducing a *multiverse language* $\mathcal{L}_{MV}$ and an open-ended *multiverse theory* MV stated in that language. $\mathcal{L}_{MV}$ employs the usual syntax of $\mathcal{L}_\in$, but with two sorts, for the *worlds* and for the *sets*. Informally stated, the axioms of MV are:

Axioms of MV*:*

$(1)_\varphi$ φ^W, for every world W. (For each axiom φ of ZFC.)

(2) (a) Every world is a transitive proper class. An object is a set just in case it belongs to some world.

(b) If W is a world and $\mathbb{P} \in W$ is a poset, then there is a world of the form $W[G]$, where G is $\mathbb{P}$-generic over W.

(c) If U is a world, and $U = W[G]$, where G is $\mathbb{P}$-generic over W, then W is a world.

(d) (Amalgamation.) If U and W are worlds, then there are G, H set generic over them such that $W[G] = U[H]$.

A theorem of Laver shows that Axiom (2)(c) can be stated in $\mathcal{L}_{MV}$.[29]

The natural way to get a model of MV is as follows. Let M be a transitive model of ZFC, and let G be M-generic for $\mathrm{Col}(\omega, < \mathrm{OR}^M)$.[30] The worlds of the multiverse M^G are all those W such that

[28] Concerning the ontological question, if we are using $\mathcal{L}_\in$ in the standard way, the answer to "how many universes of sets are there?" was provided by Cantor and Russell: none. Set theorists occasionally talk about proper classes, but as I understand that talk, one can translate it into $\mathcal{L}_\in$ with its standard interpretation.

[29] See [11]. The result was re-discovered by Woodin.

[30] One might assume that M is countable, so that generics actually exist. We shall only be interested in various first order facts about M and M^G, so Lowenheim-Skolem lets us eliminate this assumption in various well-known ways.

$$W[H] = M[G \restriction \alpha],$$

for some H set generic over W, and some $\alpha \in \mathrm{OR}^M$.

Laver's theorem implies that the full first order theory of M^G is independent of G, and present in M, uniformly over all M. That is, there is a recursive translation function t such that whenever M is a model of ZFC and G is $\mathrm{Col}(\omega, < \mathrm{OR}^M)$-generic over M, then

$$M^G \models \varphi \Leftrightarrow M \models t(\varphi),$$

for all sentences φ of the multiverse language. $t(\varphi)$ just says "φ is true in some (equvalently all) multiverse(s) obtained from me".

If $\mathcal{W}$ is a model of MV, then for any world $M \in \mathcal{W}$, there is a G such that $\mathcal{W} = M^G$. Thus assuming MV indicates that we are using the multiverse language as a sublanguage of the standard one, in such a way that t is a correct translation. Also, it is clear that if φ is any sentence in the multiverse language, then MV proves

$$\varphi \Leftrightarrow \text{ for all worlds } M,\ t(\varphi)^M \ \Leftrightarrow \text{ for some world } M,\ t(\varphi)^M.$$

Thus everything that can be said in the multiverse language can be said using just one world-quantifier.

One can add large cardinal hypotheses that are preserved by small forcing to the theory MV as follows: given such a large cardinal hypothesis φ, we add "φ^W, for all worlds W" to MV. By adding large cardinal hypotheses to MV this way, we get as theorems "for all worlds W, φ^W", for any φ in the theory of the concrete they generate.

The $\mathcal{L}_\in$ sentence CH is not generically invariant, and hence not provably-in-ZFC equivalent to any sentence in $\mathrm{ran}(t)$. There is no obvious way to translate CH into the multiverse language.

One might call MV and its extensions in $\mathcal{L}_{MV}$ *generically invariant set theory*. One can find more thorough explanations and proofs of the elementary facts about it in [19].

$\mathcal{L}_{MV}$ is motivated by the idea that one might remove a meaning-indeterminacy in $\mathcal{L}_\in$ by "trimming back", that is, restricting it to a sublanguage. $\mathcal{L}_{MV}$ and its interpretation in $\mathcal{L}_\in$ via t are a convenient way to do that. It is somewhat unfortunate that *multiverse* has the connotations it does. $\mathcal{L}_{MV}$ is meant to be an artificial language, with no meaning except that given by the translation t. It does not go beyond $\mathcal{L}_\in$ in expressive power, it drops back within it. Whether that is a proper drop is the main question about it; perhaps we trimmed too much. "More universes" does not mean a stronger language or theory, it means a potentially weaker one.

5 Some replies to Maddy and Meadows

Let me address some specific questions and critiques raised in [19]. The first concerns the motivation for MV and its language.

> Why wouldn't it be sufficient to isolate a set of axioms that captures this central idea well enough to generate a mathematically successful theory, even if it wasn't complete for some natural collection of toy models? Without a satisfactory answer to this question, we have no reason to adopt the axiomatizabilty requirement, and we're left without a principled argument for Amalgamation.

The answer here is that my goal was to capture an existing successful theory, generically invariant set theory, not invent a new one. The existing theory has been developed in $\mathcal{L}_\in$, but $\mathcal{L}_{MV}$ and its translation t into $\mathcal{L}_\in$ seem useful in isolating it. Amalgamation is true under this translation. It records the intention to be translatable via t into $\mathcal{L}_\in$. Amalgamation is not meant to be an independent insight into the nature of multiverses. It clarifies that you want to be understood as speaking a certain sublanguage of $\mathcal{L}_\in$.

The Maddy-Meadows passage seems to say that there could have been some way to assign meaning to $\mathcal{L}_{MV}$ beyond translating it into $\mathcal{L}_\in$. I don't think that is true. I doubt anyone is going to find mathematical concepts that cannot be expressed in $\mathcal{L}_\in$ any time soon. I think that any candidate for a new concept should be regarded as suspect until it can be formulated in $\mathcal{L}_\in$, especially if there is no language community with an established agreement on how the new concept is used. I myself do not feel capable of going beyond what can be expressed in $\mathcal{L}_\in$.

This point answers some related objections:

> It seems to us that the gloss 'settled/unsettled by the current meaning' is a problematic fit for the role of defective/virtuous just described. First, assuming $\mathcal{L}_{MV}$ sentences enjoy this virtue, it isn't obvious that a translation would preserve it.

A bad translation might not preserve the virtue, of course, but t is guaranteed to be a perfect translation, because it is the only way $\mathcal{L}_{MV}$ sentences have been given any meaning at all.

> Whatever being 'settled by the current meaning' comes to, it seems that there might well be a sentence of $\mathcal{L}_\in$ that translates to a synonymous sentence of $\mathcal{L}_N$ (the language of arthmetic), where the former is settled by the current meaning of $\mathcal{L}_\in$ but the latter is unsettled by the current meaning of $\mathcal{L}_N$ (e.g. a strong consistency statement).

Here the authors seem to be identifying "current meaning" with what is provable in some formal theory. I definitely do not want to do that. I would say that, as of now, there is no meaning-indeterminacy visible in $\mathcal{L}_N$.

> Why should the the meaning currently assigned to $\mathcal{L}_{MV}$ do any better at settling all sentences of $\mathcal{L}_{MV}$ than the meaning currently assigned to $\mathcal{L}_\in$ does at settling all sentences of $\mathcal{L}_\in$?

The answer here is that it cannot do any worse, and if there is meaning indeterminacy remaining in $\mathcal{L}_{MV}$, then we should deal with it.

The central question about the multiverse language is whether our trimming has gone too far. Do we lose expressive power if we stay within it? There are two possible answers: yes and no. In [30], "no" is elaborated as

Weak relativist thesis:(WRT) Every proposition that can be expressed in the standard language $\mathcal{L}_\in$ can be expressed in the multiverse language.

Thinking of $\mathcal{L}_\in$ as being equivalent to $\mathcal{L}_{MV}$ with a constant symbol $\dot{V}$ for a reference universe, the "yes" answer was expanded as

Strong absolutist thesis:(SAT) "$\dot{V}$" makes sense, and that sense is not expressible in the multiverse language.

These two theses regarding the semantic completeness of $\mathcal{L}_{MV}$ belong squarely to epistemology. The word "proposition" has a long history in philosophy, where it is used for "sentence meaning". Perhaps it comes attached to theories of meaning that are not suitable for this context.[31] Another way to state WRT is: "everything that can be said in $\mathcal{L}_\in$ can be said in $\mathcal{L}_{MV}$". That is, $\mathcal{L}_{MV}$ has all the interpretative power present in $\mathcal{L}_\in$.

"Thesis" is a nod to the (perhaps illusory) parallel between the Weak Relativist Thesis and Church's Thesis. Neither is really capable of proof, and there are some similarities between their roles and the evidence for them. (Of course, the evidence for Church's Thesis is much stronger!)

WRT asserts that every sentence in $\mathcal{L}_\in$ that expresses a proposition is synonymous with one in $\mathrm{ran}(t)$. Feferman would say "definite proposition". That may be better, although definiteness here is not a property of the proposition, but of its relationship to the sentence. Maddy and Meadows propose replacing synonymy with provable equivalence in ZFC, or perhaps some extension T of ZFC. I think that this is probably consistent with the intent of WRT, so long as T has an axiomatization consisting of sentences in $\mathrm{ran}(t)$, and is a theory we believe.[32] If T has such an axiomatization, and $T \vdash \phi \leftrightarrow t(\psi)$, then letting θ be the conjunction of the axioms of T needed in the proof, $\theta \to \phi$ is logically equivalent to $\theta \to t(\psi)$, and the latter is logically equivalent to a sentence in $\mathrm{ran}(t)$. If we have accepted T, then adding ϕ is equivalent to adding $\theta \to \phi$. So the Maddy/Meadows version of WRT implies the version above because logical equivalence implies synonymy.

WRT does allow that one might decide sentences like CH that are not logically equivalent to sentences in $\mathrm{ran}(t)$ by asserting their equivalence with some ϕ in $\mathrm{ran}(t)$. From the Weak Relativist point of view, this assertion involves a clarification of their meaning.

Maddy and Meadows object to the use of the notion of meaning in WRT. They write

[31] In particular, it seems plausible that in this context one should treat theories, rather than sentences, as the primary bearers of meaning.

[32] If not, then T-provable equivalence does not capture the intent of WRT.

> It's fairly easy to explicate this type of concern [that the meaning we currently assign to $\mathcal{L}_\in$ does not decide CH] in a metaphysical theory like Hamkins's or Woodin's - there is an abstract ontology of worlds in some of which CH is true and others of which it is false ...

In my opinion, talk about ontologies of worlds does not explain anything, it just muddies the whole question. We are discussing what theory to adopt, and the extent to which adopting one conflicts with adopting another. The subject matter in this discussion is interpreted theories; set theories, not sets or worlds.

Maddy and Meadows continue

> ...but we've seen that Steel's thinking is strictly linguistic. In that context, it's not obvious how to characterize the potential problem without providing a substantive theory of meaning (no hint of which appears in [30].)

Here I would say that it would certainly be good if there were a substantive theory of meaning that would underpin and elaborate on WRT. Maddy and Meadows are skeptical that any exists or can be developed. People have certainly written a lot about the meaning of mathematical statements, so at least some have been more hopeful in this regard.

However, I think WRT has content that does not rest on such a theory. Its role is to identify a potential meaning indeterminacy in $\mathcal{L}_\in$, not to say what meaning indeterminacy in $\mathcal{L}_\in$ is. One way to do that in practice is to point to possible dis-ambiguations, and another is to trim away the allegedly meaningless part and claim that nothing has been lost. WRT makes a move in the latter direction. Moreover, when we trim back, the question as to whether the multiverse has a core immediately takes center stage, and this leads to what one might think of as a dis-ambiguation.

The strongest evidence for WRT is that the mathematical theory based on large cardinal hypotheses that we have produced to date can be naturally expressed in the multiverse sublanguage. Perhaps we lose something when we do that, some future mathematics built around an understanding of the symbol $\dot{V}$ that does not involve defining $\dot{V}$ in the multiverse language. But at the moment, it's hard to see what that is. WRT can be considered as a piece of advice: don't go looking for it. I don't think we can say yet with any confidence that this is good advice.

Maddy and Meadows have a different way of characterizing the virtue of generically invariant sentences of $\mathcal{L}_\in$, *impartiality*. Let me trace through their line of thought here. They begin

> We're assuming that our examination of the various candidates [for a foundational theory] shows them all to be on equal footing and that our best response is to trim the syntax of $\mathcal{L}_\in$.

To the extent this is meant to characterize my reason for isolating $\mathcal{L}_{MV}$, there is a subtle misunderstanding. It is based on something I wrote poorly:

> Before we try to decide whether some such theory is preferable to the others, let us try to find a neutral common ground on which to compare them. We seek a language in which all these theories can be unified, without bias toward any, in a way that exhibits their logical relationships, and shows clearly how they can be used together.

"Compare" here is misleading. "Unify" would be better. $\mathcal{L}_{MV}$ is not a device for comparing the merits of one of those theories with another, it is a way to make more visible their underlying unity.

Maddy and Meadows continue

> On those assumptions, consider the state of two imaginary set theorists, a universe theorist and a multiverse theorist. The universe theorist speaks $\mathcal{L}_\in$, embraces ZFC + LCs, and persists in trying to figure out the 'correct' way to extend it; under our current assumptions, this universe theorist is just wrong, making a mistake. In contrast, our multiverse theorist is aware that no candidate is preferable, speaks $\mathcal{L}_{MV}$, and embraces MV. This multiverse theorist thinks, with considerable justification on our assumptions, that the universe theorist is missing the fact that all the candidate foundational theories represented by worlds in the multiverse have equal standing. To put this another way, we might say that from the multiverse theorist's perspective, the universe theorist's $\mathcal{L}_\in$ sentences may reflect an improper bias, restricting attention to one world, while all $\mathcal{L}_{MV}$ sentences are suitably impartial.

Of course, I find the part about universes and multiverses, and theories represented by worlds, vague and at the wrong level. The distinction I was trying to draw was between views on the semantic completeness of $\mathcal{L}_{MV}$. Weak relativism and Strong Absolutism are views on the semantics of $\mathcal{L}_\in$. They are not expressible in $\mathcal{L}_\in$.

Our "multiverse theorist" could adopt the syntax of $\mathcal{L}_\in$, and just be careful to stay in the range of the translation function. At the level I was going for, nothing has changed, she is just expressing herself differently. I am not sure whether she still counts as a multiverse theorist in the Maddy/Meadows sense.

The sentences in $\mathrm{ran}(t)$ are, up to logical equivalence, precisely those that are generically invariant. So they are impartial in that sense. But the main claim in WRT is that *it is indeed suitable* to remain within the generically invariant. That is, that there is no mathematical meaning beyond that which can be captured by generically invariant principles. Or, to retreat to a more reasonable claim, that there is none on the horizon now.

Let me summarize my disagreements with Maddy and Meadows by replying to their own summary.

> The substance of Steel's thought can be formulated more effectively in philosophically innocent mathematical terms. By these means, we steer away from the

> vagaries of mathematical meaning, truth, and existence and toward the methodologically central questions: how exactly do we select our theories and by what right?

I would agree that truth and existence are out of place in a discussion of what set theory to adopt and why. Not because they are vague, but because bringing them in leads to question-begging. Meaning is at the center of it, because we select our theories based on how they are interpreted. In this set theoretic context, our framework theory should be such that all others can be translated into it. Maddy's focus on methodology is not in conflict with this idea; after all, the logical positivists' slogan was that the meaning is the method of verification. If we understand "method of verification" in a suitably broad sense, this seems ok to me.

6 Does the multiverse have a core?

Having retreated to $\mathcal{L}_{MV}$, one asks at once whether the absolutist's idea of a distinguished reference world is really gone. Perhaps there is an individual world that is definable in the multiverse language. An elementary forcing argument shows that if so, then there is a unique definable world, and this world is included in all the others.[33] In this case, we call this unique world included in all others the *core* of the multiverse. [34] This leads to what [30] calls the *Weak Absolutist Thesis*: The multiverse has a core.

It was a mistake to call this a thesis, and present it in parallel with the two semantic theses in the last section. "The multiverse has a core" is a statement in $\mathcal{L}_{\in}$, and in fact, in its sublanguage $\mathcal{L}_{MV}$. WRT and SAT are philosophical theses, and cannot be formulated in $\mathcal{L}_{\in}$. Whether the multiverse has a core is a question about sets, not set theories, and it can be formulated by the sentence $\exists U \forall W \forall x(x \in U \rightarrow x \in W)$ of $\mathcal{L}_{MV}$. Let us shorten this sentence to $\exists U(U = \mathfrak{C})$.

Whatever one thinks of the semantic completeness of the multiverse language, it does highlight $\exists U(U = \mathfrak{C})$ as a fundamental question. Because the multiverse language is a sublanguage of the standard one, this is a question for everyone. If the multiverse has a core, then surely it is important, whether it is the denotation of the absolutist's $\dot{V}$ or not.

Fuchs, Hamkins, and Reitz have shown that neither MV nor its extensions by large cardinal hypotheses up to the level of supercompact cardinals decides whether there is a core to the multiverse, or the basic theory of this core if it exists. (See [8].) But

Theorem 6.1 *(Usuba [35], [36]) If there is an extendible cardinal, then* $t(\exists U(U = \mathfrak{C}))$.

[33] This observation is due to Woodin.

[34] Fuchs, Hamkins, and Reitz, began the general study of such questions in what they call *set theoretic geology*. See [8]. Usuba's proof of their "downward directed grounds hypothesis" shows that W is the core iff W is what [8] calls a *bedrock*.

The Fuchs-Hamkins-Reitz work shows that nothing follows from extendible cardinals concerning the basic theory of the core.

Usuba's theorem is certainly evidence that there is a core, but there is some reason to be hesitant. First, the large cardinal hypothesis is "global", that is, Σ_3 rather than Σ_2, and that is essential.[35] Second, strong evidence that there is a core should be evidence that there is a core with well-determined properties. The fact that the existence of extendible cardinals decides very little about the theory of the core weakens the evidence provided by Usuba's proof.

Before we turn to the possibility of a well-determined core, let us consider the semantic consequences of adopting "there is a core", in either syntax. There is a discussion of this question in §3 of Meadows' paper [23], and I shall incorporate parts of it. I shall also reply to the "argument against the generic multiverse" made there.[36]

$\mathcal{L}_\in$ is equivalent to $\mathcal{L}^+_{MV}$, where $\mathcal{L}^+_{MV}$ is $\mathcal{L}_{MV}$ expanded by a constant symbol $\dot{V}$, in the following sense. If ϕ is a sentence of $\mathcal{L}_\in$, its translation in $\mathcal{L}^+_{MV}$ is just

$$u(\phi) = \phi^{\dot{V}}.$$

If ϕ is a sentence of $\mathcal{L}^+_{MV}$, then its translation in $\mathcal{L}_\in$ is

$$t^+(\phi) = \text{“}(V^G, \in, V) \models \phi\text{”},$$

where V^G is the (imaginary) "expanded" generic multiverse generated by V and V is treated as the interpretation of $\dot{V}$. The right hand side can be unpacked more fully using the definability of forcing. t^+ is well defined because the forcing is homogeneous. u actually acts on formulae, and is a standard relative interpretation. Meadows [23] shows using boolean ultrapowers that if V has a definable wellorder, then we can extend t^+ to a standard relative interpretation, but I don't see that this is important for our discussion. The extension does not preserve the meaning of formulae. Because the forcing is homogeneous, t^+ preserves $\wedge$ and $\neg$, moreover, letting MV^+ be the natural extension of MV to $\mathcal{L}^+_{MV}$,

(i) $\mathsf{ZFC} \vdash \phi \leftrightarrow t^+ \circ u(\phi)$, and

(ii) $\mathsf{MV}^+ \vdash \psi \leftrightarrow u \circ t^+(\psi)$,

for all sentences ϕ of $\mathcal{L}_\in$ and ψ of $\mathcal{L}^+_{MV}$. That is, ZFC proves that ϕ holds iff $\phi^{\dot{V}}$ holds in its expanded generic multiverse, and MV^+ proves that ψ holds iff $\dot{V} \models$ "ψ holds in my

[35] The Σ_2 sentences are those that are, provably in ZFC, equivalent to sentences of the form $\exists\alpha(V_\alpha \models \varphi)$. This is the sense in which Σ_2 sentences are local. "There is a supercompact cardinal" is also Σ_3 rather than Σ_2. Extendibles are strictly stronger than supercompacts. Usuba observed that $t(\exists U(U = \mathfrak{C}))$ does not follow from the existence of supercompacts. Goldberg [9] shows that nothing much weaker than the existence of extendibles suffice.

[36] Meadows attributes the argument to Woodin.

expanded generic multiverse". Adapting the terminology of [23], let us say that ZFC and MV^+ are *weakly sententially equivalent via* (u, t^+).

Now let's consider possible weak sentential equivalences between theories in $\mathcal{L}_\in$ extending ZFC and theories in $\mathcal{L}_{MV}$ extending MV. Composing with (u, t^+) above, we are looking for such equivalences between extensions S of MV^+ and extensions T of MV. Because MV^+ has its symbol $\dot{V}$, it is natural to add "there is a core", that is $\exists U(U = \mathfrak{C})$, to both S and T.[37] As Meadows notes, the resulting theories are still not weakly sententially equivalent, the reason being that we have yet to settle anything about what is true of $\dot{V}$. The simplest way to do that is to add $\dot{V} = \mathfrak{C}$ to MV^+.

Let us write "$V = \mathfrak{C}$" for $t^+(\dot{V} = \mathfrak{C})$, that is, for the $\mathcal{L}_\in$ version of "V is the core of its generic multiverse".

Proposition 6.1.1 *The theories*

(i) $T_0 = \mathsf{MV} + \exists U(U = \mathfrak{C})$,

(ii) $T_1 = \mathsf{MV}^+ + \dot{V} = \mathfrak{C}$, *and*

(iii) $T_2 = \mathsf{ZFC} + V = \mathfrak{C}$

are weakly sententially equivalent.

T_0 and T_1 are clearly definitionally equivalent. The equivalence between T_1 and T_2 is given by (t^+, u). In other words, the $\mathcal{L}_\in$ sentence ϕ is translated to the $\mathcal{L}_{MV}$ sentence "ϕ is true in $\mathfrak{C}$", and the $\mathcal{L}_{MV}$ sentence ψ is translated to the $\mathcal{L}_\in$ sentence "ψ is true in the multiverse generated by me".

I believe the proper conclusion is that $\mathsf{MV} + \exists U(U = \mathfrak{C})$ and $\mathsf{ZFC} + V = \mathfrak{C}$ are equivalent as foundations. From the point of view of the Weak Relativist, passing from $\mathsf{MV} + \exists U(U = \mathfrak{C})$ to $\mathsf{ZFC} + V = \mathfrak{C}$ amounts to making a definition. The real question was whether there is a core, and this is a question in the multiverse sublanguage.

It's worth noting that one could expand the list of weakly sententially equivalent theories given in the proposition. Let $\mathbb{P}$ be any homogeneous partial order in $\mathfrak{C}$ that is definable over $\mathfrak{C}$ (provably in ZFC); then the theory

$$T(\mathbb{P}) = \mathsf{MV}^+ + \text{"}\dot{V} \text{ is a } \mathbb{P}\text{-generic extension of } \mathfrak{C}\text{"}$$

is also weakly sententially equivalent to T_0. In contrast to T_1, $T(\mathbb{P})$ is not a definitional extension of T_0. It seems to me that $T(\mathbb{P})$ is an odd way of presenting the same foundation as that in T_0, T_1, and T_2.

Meadows presents the equivalence of T_0 with T_2 as a practical argument "against the generic multiverse". It is not clear to me how one can argue against an object. More

[37] There may be interesting equivalences that do not involve doing that.

importantly, it is not clear to me how the equivalence between T_0 and T_2 could be an argument against one and in favor of the other.

I think the idea is probably that T_2 is a simpler presentation of T_0, with a more familiar syntax, so in the end it would be better to work with it. That is certainly true for elementary set theory, but anyone who is able to understand and work with the hypothesis $V = \mathfrak{C}$ will understand the simple equivalence between T_2 and T_0, and be able to shift between $\mathcal{L}_\in$ and $\mathcal{L}_{MV}$ easily. In any case, convenience is not what motivates T_0 and its syntax. $\mathcal{L}_{MV}$ is useful in isolating generically invariant set theory, and that helps us to understand what might justify T_2, and the extension $V = \mathrm{Ult} - L$ of T_2 that we shall describe in the next section.

7 Absolute ordinal definability

Let us assume that the multiverse has a core. We still haven't gotten very far, because large cardinal hypotheses by themselves decide very little about the core. It seems reasonable to add that the sets in $\mathfrak{C}$ are all ordinal definable, but this still decides very little. The reason is that mild class forcings preserve the large cardinals, and can be used to construct models in which sets get into $\mathfrak{C}$ for reasons unrelated to the sets themselves, for example by being coded into the 2^α function at arbitrarily large α. The same construction shows that being ordinal definable, by itself, says very little about a set.[38] If we want to pin down $\mathfrak{C}$, we need an explanation for minimality and ordinal definability.

Inner model theory suggests a way to pin down $\mathfrak{C}$. The canonical inner model M_H for a large cardinal hypothesis H is its most concrete realization. Its construction yields a thorough *fine structure theory* for the model. We have constructed M_H for many H. Each M_H has a fine-grained hierarchy, and H is weaker than K if and only if M_H is a proper initial segment of the hierarchy of M_K. So the M_H we have constructed do fit together, in a hierarchy that seems to be the model-theoretic counterpart to the consistency strength hierarchy.[39]

At the moment, we only have a theory of the M_H at limited large cardinal levels. If they have a general form, one that is independent of H and goes as far as the large cardinal hypotheses do, then this would be a candidate for the structure of $\mathfrak{C}$. Hugh Woodin has suggested that this is the case, and that this form is that of the HODs in certain models of the Axiom of Determinacy.

To me, this suggestion for pinning down $\mathfrak{C}$ looks like meaning clarification. It seems similar to what happened when we adopted the Axiom of Extensionality, or the Axiom of

[38] In [8], Fuchs, Hamkins, and Reitz show this way that for any countable $M \models \mathsf{ZFC}$ and any ordinal α, M has a mild class forcing extension N such that $V_\alpha^M = V_\alpha^N$ and N satisfies "$\mathfrak{C}$ exists, and is equal to generic HOD". Here generic HOD consists of those sets that are ordinal definable in every set generic extension of V.

[39] Here I am assuming that M_H is pointwise definable, as it must be if it is truly minimal. If $\mathfrak{C}$ is an ultimate model along these lines, then the M_H would all be countable initial segments of it. See the introductory chapter of [33] for an overview of the inner model theory that is relevant in this section.

Regularity $V = WF$.

Let us give a brief explanation.

The sets in any M_H are ordinal definable in a certain generically absolute way.

Definition 7.1 *Let $A \subseteq \omega^\omega$; then A is* homogeneously Suslin (Hom_∞) *iff for all κ, there is a system $\langle M_s, i_{s,t} \mid s,t \in \omega^{<\omega}\rangle$ such that*

(1) $M_\emptyset = V$, and each M_s is closed under κ-sequences,

(2) for $s \subseteq t$, $i_{s,t}\colon M_s \to M_t$,

(3) if $s \subseteq t \subseteq u$, then $i_{s,u} = i_{t,u} \circ i_{s,t}$, and

(4) for all x, $x \in A$ iff $lim_n M_{x\restriction n}$ is wellfounded.

Martin showed in 1968 that all Hom_∞ sets are determined, and that if there are arbitrarily large measurable cardinals, then all Π^1_1 sets are Hom_∞. Stronger large cardinal hypotheses imply that more complicated sets are Hom_∞. Work of Martin-Solovay ([22]), Foreman-Magidor-Shelah ([7]), and various people in inner model theory led eventually to

Theorem 7.2 *(Martin, S., Woodin 1985) Assume there are arbitrarily large Woodin cardinals; then for any $A \in \mathrm{Hom}_\infty$, $L(A,\mathbb{R}) \models \mathsf{AD}^+$.*

AD^+ is a technical strengthening of the Axiom of Determinacy (AD), first isolated by Woodin.

Along with determinacy we get set-generic absoluteness.

Theorem 7.3 *(Woodin 1987?) If there are arbitarily large Woodin cardinals, then $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ truth is set-generically absolute.*

Recall that a set is *ordinal definable* (OD) iff it is definable over the universe of sets from ordinal parameters, and is *hereditarily ordinal definable* (HOD) just in case it and all members of its transitive closure are OD. The HOD of a determinacy model is close to the model itself; moreover these HOD's have Woodin cardinals.

Definition 7.4 *(Woodin) $V = \mathrm{Ult} - L$ is the statement: There are arbitrarily large Woodin cardinals, and for any Σ_2 sentence φ of LST: if if φ is true, then for some $A \in \mathrm{Hom}_\infty$, $HOD^{L(A,\mathbb{R})} \models \varphi$.*

This is read "V is ultimate L".

Theorem 7.5 *(Woodin) If $V = \mathrm{Ult} - L$, then*

(1) V is the core of its multiverse V^G.

(2) V is "generically absolute HOD".

See [37].

One can state $V = \mathrm{Ult} - L$ in the multiverse sublanguage.

The hope is that $V = \mathrm{Ult} - L$ is consistent with all the large cardinal hypotheses, so that adopting it does not restrict interpretative power. Whether it is consistent with hypotheses significantly stronger than the existence of many Woodin cardinals is a crucial open problem.

At the same time, one hopes that $V = \mathrm{Ult} - L$ will yield a detailed fine structure theory for V, removing the incompleteness that large cardinal hypotheses by themselves can never remove. It is known that $V = \mathrm{Ult} - L$ implies the CH, and many instances of the GCH. Whether it implies the full GCH is a crucial open problem.

Both problems have to do with the theory of HOD in models of AD^+. They have been recognized as central questions in descriptive set theory since the early 1980s, and in inner model theory since the mid 1990s. Many people have worked on them and obtained various encouraging results.[40] In the period after [30] appeared I made some further progress in this direction. Let me state one of the main new theorems, assuming as I do some significant technical background.

Definition 7.6 *(AD^+) A pointclass Γ is* long *iff there is an $A \in \Gamma$ such that A codes an (ω_1, ω_1) iteration strategy for a pure extender premouse with a long extender on its sequence. Otherwise Γ is* short*.*

Theorem 7.7 *(S. 2015-21, [34],[33]) Suppose there are arbitrarily large Woodin cardinals, and that there is a supercompact cardinal. Assume also that V* is uniquely iterable*; then*

(1) there is a short $\Gamma \subsetneq \mathrm{Hom}_\infty$ such that $L(\Gamma, \mathbb{R}) \models \mathsf{AD}_\mathbb{R}$ and $HOD^{L(\Gamma,\mathbb{R})} \models$ "there is a subcompact cardinal", and

(2) there is a long Γ in Hom_∞.

Theorem 7.8 *(S. 2015-21, [34],[33]) Assume $\mathsf{AD}^+ +$ "there is a long pointclass"; then for any short $\Gamma \subseteq P(\mathbb{R})$ such that $L(\Gamma, \mathbb{R}) \models \mathsf{AD}_\mathbb{R}$, $HOD^{L(\Gamma,\mathbb{R})}$ is a least branch premouse (so satisfies GCH, and has a fine structure).*

It should be possible to remove the iterability hypotheses in these theorems.[41] How to remove "V is uniquely iterable" from 7.7 is an instance of what has been the central problem in inner model theory since the mid-1980s. I suspect that removing the hypothesis "there

[40] There is a survey of work in this area in [32].

[41] In 7.8, the iterability hypothesis is "there is a long pointclass". We suspect that conclusion (1) of 7.7 also follows from $\mathsf{AD}^+ +$ "there is a long pointclass", but the claim made in [33, Theorem 1.5.1(2)] to have a proof of this was wrong.

is a long pointclass" from 7.8 is the most accessible of the big open problems to do with iterability.

Theorem 7.7 states that, granted its iterability hypothesis, $V = \mathrm{Ult} - L$ is consistent with subcompacts. This is a significant step beyond the large cardinal hypotheses that had been reached previously, and is therefore encouraging news. Subcompacts are near the upper limit of what can be modelled by mice with only short extenders. To reach larger cardinals like supercompacts, one would need a general comparison theorem for mice with long extenders. This is another long-open problem; there is a comparison theorem (modulo iterability) for mice in the lower reaches of the long extender realm, but no general one.[42]

If $V = \mathrm{Ult} - L$ truly pins down the theory of the core, then it must decide GCH. To prove that $V = \mathrm{Ult} - L$ implies a Π_2 sentence φ (like GCH), one must show that φ holds in $\mathrm{HOD}^{L(A,\mathbb{R})}$ for *all* $A \in \mathrm{Hom}_\infty$. One needs a general theory of these HOD's. Given what we know about "small" determinacy models, it seems likely that GCH is part of this theory. That leads to the following conjecture:

Conjecture 7.8.1 *Assume* $\mathsf{ZF} + \mathsf{AD}^+ + V = L(P(\mathbb{R}))$*; then HOD* $\models \mathsf{GCH}$.

In a sense this conjecture originates with the work of the Cabal in descriptive set theory in the late 1970s and early 1980s. Since then, we have seen that it belongs to inner model theory, and likely involves a general notion of mouse and iteration strategy, together with a general comparison lemma for these objects. Presumably these mice come equipped with a fine-grained hierarchy, the model-theoretic counterpart to the consistency strength hierarchy on the theories that hold in them or their generic extensions.

I think we are still pretty far from a proof of this conjecture. There are fragments of it that are more accessible, and would represent significant progress.

What we know from descriptive set theory, and from inner model theory where we have it, suggests that Conjecture 7.8.1 is true, and that its proof involves a general theory of mice and their iteration strategies. The mice must be capable of having long extenders, and probably supercompact cardinals and beyond. Although its less general approximations are much better targets at present, Conjecture 7.8.1 points to a long term future. The intricately structured world of inner model theory extends well beyond the part that we have discovered so far.[43]

[42] Those results are due to Woodin, Neeman, and the author. See [25].

[43] Here let me record a disagreement with [3]. The fate of inner model theory does not depend on the truth of the Ultimate-L conjecture. The Ultimate-L conjecture has to do with one approach to constructing canonical inner models. There is a fair amount of evidence that this approach does not always work in the way that the Ultimate-L conjecture requires, but there are other approaches. In particular, it is a different approach that leads to the positive results on Conjecture 7.8.1 that we have now.

8 Final remarks

What would it be to adopt $V = \mathrm{Ult} - L$ as a foundational axiom? Should we do that?

Adopting $V = \mathrm{Ult} - L$ would not, and should not, mean ending the further development of theories like the forcing axioms. What can be forced is of permanent interest in set theory. As the title of [2] puts it, the generic multiverse is not going away. The goal here is to unify those theories in a framework that lets us use them together properly, not to eliminate them.

The key questions are whether the HODs of determinacy models are ultimate, and whether they are L-like. Gödel taught us not to dream of final theories, so *Ultimate* should be taken with a few grains of salt. What we would like to show is that we can interpret (at the sentential level) the natural theories like those we know in extensions of $\mathsf{ZFC} + V = \mathrm{Ult} - L$. This seems to involve showing that the HODs of determinacy models can satisfy at least the local (Σ_2) forms of the very strong large cardinal hypotheses we know about. *L-like* means having a fine structure like that of L. At a minimum, we should be able to show that GCH holds. Experience suggests that if we do that, it will be the beginning of a thorough fine-structural analysis.

If all this works out, $V = \mathrm{Ult} - L$ would be a clarificatory axiom, like the Axiom of Extensionality, or the Axiom of Regularity. It lets people know how you are using the language of set theory. There is as of now no reason to believe that we lose interpretative power by using the language of set theory this way, but it is too early to have a definite opinion.

References

[1] D. Aspero and R. D. Schindler, Martin's Maximum++ implies Woodin's axiom (*). Annals of Mathematics, vol. 193 (2021), pp. 793-835.

[2] D. Blue, The generic multiverse is not going away. To appear.

[3] J. Bagaria, P. Koellner, and W. H. Woodin, Large cardinals beyond choice. Bulletin of Symbolic Logic, vol. 25 (2019), pp. 283-318.

[4] J. Burgess, Second philosophy: anti-nominalist reflections on Maddy's semi-nominalism, this volume.

[5] S. Feferman, Why a little bit goes a long way. Proceedings of the biennial meeting of the Philosophy of Science Association (1992), pp. 442-455.

[6] S. Feferman, In the light of Logic. Oxford University Press, New York, 1998.

[7] M. Foreman, M. Magidor, and S. Shelah, Martin's maximum, saturated ideals, and nonregular ultrafilters. Annals of Mathematics, vol. 127 (1998), pp. 1-47.

[8] G. Fuchs, J. D. Hamkins, and J. Reitz, Set-theoretic geology. Annals of Pure and Applied Logic, vol. 166 (2015), pp. 464-501.

[9] G. Goldberg, Usuba's extendible cardinal. Preprint, arXiv:2108.06903 [math.LO] (2021).

[10] J. D. Hamkins, A multiverse perspective on the Axiom of Constructibility. Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore. Infinity and Truth, (2014), pp. 25-45.

[11] R. Laver, Certain very large cardinals are not created in small forcing extensions. Annals of Pure and Applied Logic, vol. 149 (2007) 1–6.

[12] S. MacLane, Mathematics, form and function. Springer, New York (1986).

[13] P. Maddy, Naturalism in mathematics. Oxford University Press, New York, 1997.

[14] P. Maddy, $V = L$ and Maximize. In: Logic Colloquium '95, Proceedings of the European Assn. of Symbolic Logic held in Haifa, Israel, J. Makowsky et. al. eds, Lecture Notes in Logic vol. 11 (1998), pp. 134-152.

[15] P. Maddy, Second philosophy. Oxford University Press, New York, 2007.

[16] P. Maddy, Defending the axioms. Oxford University Press, New York, 2011.

[17] P. Maddy, Set-theoretic foundations, In: Set-theoretic foundations, A. Caicedo et. al. eds., Foundations of mathematics: essays in honor of W. Hugh Woodin's 60th birthday, Contemporary Mathematics, vol. 690 (2017), 289-320.

[18] P. Maddy, What do we want a foundation to do? Reflections on Foundations: Univalent Foundations, Set Theory, and General Thoughts. S. Centrone et. al eds, 36 pp. To appear.

[19] P. Maddy and T. Meadows, A reconstruction of Steel's multiverse project, Bulletin of Symbolic Logic, vol. 26 (2020), pp. 119-169.

[20] D. A. Martin, Gödel's conceptual realism. Bulletin of Symbolic Logic, vol. 11 (2005), pp. 207-224.

[21] D. A. Martin, Completeness and incompleteness of basic mathematical concepts, Preliminary draft Feb. 2018, 36pp., available at

[22] D. A. Martin and R. M. Solovay, A basis theorem for Σ^1_3 sets of reals. Annals of Mathematics, vol. 89 (1969), pp. 138-160.

[23] T. Meadows, Two arguments aganst the generic multiverse. Review of Symbolic Logic, vol. 14 (2021) pp. 347-379.

[24] A. Montalban and J. Walsh, On the inevitability of the consistency operator. Journal of Symbolic Logic, vol. 84 (2019), pp. 205-225.

[25] I. Neeman and J. R. Steel, Equiconsistencies at subcompact cardinals. Archive for Mathematical Logic, vol. 55, (2016) pp. 207-238.

[26] W. V. Quine, Pursuit of truth. Cambridge, Mass. Harvard University Press (1990).

[27] W. V. Quine, Review of Parsons. Journal of Philosophy, vol. 81 (1984), pp. 783-794.

[28] J. R. Schatz, Axiom selection by maximization: V = Ultimate L vs. forcing axioms. Ph.D. thesis, UC Irvine 2019.

[29] J. R. Steel, Mathematics needs new axioms. Bulletin of Symbolic Logic, vol. 6 (2000), pp. 422-433.

[30] J. R. Steel, Gödel's program. In: Interpreting Gödel, J. Kennedy ed., Cambridge Univ. Press (2014), pp. 153-179.

[31] J. R. Steel, A classification of jump operators. Journal of Symbolic Logic 47 (1982), p. 347–358.

[32] J. R. Steel, HOD in models of determinacy. In: Ordinal definability and recursion theory: the Cabal Seminar, volume III, A.S Kechris et. al. editors, volume 43 of Lecture Notes in Logic, ASL (2016) pp. 3-48.

[33] J. R. Steel, A comparison process for mouse pairs. To appear in Lecture Notes in Logic (ASL, Cambridge University Press).

[34] J. R.Steel, The comparison lemma. Annals of Pure and Applied Logic, to appear.

[35] T. Usuba, The downward directed grounds hypothesis and very large cardinals. Journal of Mathematical Logic, vol. 17 (2017).

[36] T. Usuba, Extendible cardinals and the mantle. Archive for Mathematical Logic, vol. 58 (2019), pp. 71-75.

[37] W. H. Woodin, In search of ultimate-L. The 19th Midrasha Mathematicae lectures. The Bulletin of Symbolic Logic, vol. 23 (2017), pp. 1-109.

[38] W. H. Woodin, The Continuum Hypothesis II. Notices of the AMS, vol. 48 (2001), pp. 681-690.