Gödel's program

John R. Steel

October 29, 2012

Abstract

Set theorists have discovered many mutually incompatible natural theories extending ZFC. It is possible that these incompatibilities will be resolved by interpreting all such theories in a useful common framework theory.

1 Introduction

In the 19th and early 20th centuries, it was shown that all mathematical language of the time could be translated into the language of set theory (LST), and all mathematical theorems of the time could be proved in ZFC. A century later, mathematicians have yet to develop any mathematics that cannot be expressed in LST, and there are probably few who believe that this will happen any time soon. However, the remarkable work of Kurt Gödel in the 1930's, and Paul Cohen and his successors in the period from 1963 to the present, has shown that ZFC is incomplete in significant ways. There are very concrete statements about natural numbers that it fails to decide. There are conceptually central questions about real numbers and sets of real numbers that it fails to decide. Although mathematicians can say everything they have to say in LST, they cannot decide all the questions they would like to decide using only the axioms of ZFC.

Gödel anticipated that Cohen's theorem on the independence of the Continuum Hypothesis would eventually be proved. In his 1947 paper "What is Cantor's Continuum Problem" ([3]), he suggested a program of research now known as Gödel's program:

Gödel's Program: Decide mathematically interesting questions independent of ZFC in well-justified extensions of ZFC.

The mathematical question Gödel had foremost in mind was the Continuum Hypothesis (CH), and the well-justified extensions of ZFC he had in mind were those that begin with

$\mathsf{ZFC} + \mathrm{Con}(\mathsf{ZFC})$, and make their way up through the hierarchy of strong axioms of infinity.[1]

Gödel's program is inevitable, if you believe, as Gödel did, that the axioms of ZFC describe a "well-determined reality", or in less colorful language, that the meaning we have assigned to sentences in the syntax of LST suffices to determine a truth value for CH. [2] The qualification *well-justified* in the statement of Gödel's program also introduces a philosophical component. How does one justify statements in LST? General philosophical questions concerning the nature of meaning, evidence, and belief rear their ugly heads. Of course, the same questions come up vis-a-vis whatever fragment of LST and ZFC one might choose as a safe retreat. Moreover, for those who accept some reasonable fragment of ZFC, but believe that the truth value of CH is not determined by the meaning we currently assign to the syntax of LST, the Continuum Problem does not disappear. Certainly we don't want to employ a syntax which encourages us to ask psuedo-questions, and the problem then becomes how to flesh out the current meaning, or trim back the current syntax, so that we can stop asking psuedo-questions.

For these reasons, work in Gödel's program provides an interesting case study for the epistemology of mathematics, and perhaps epistemology will be able to give something back, for it seems likely that any solution to the Continuum Problem must be accompanied by a better understanding of what it is to be a solution to the Continuum Problem.

In the author's opinion, the key methodological maxim that epistemology can contribute to the search for a stronger foundation for mathematics is: *maximize interpretative power*. As Gallileo put it in his well-known dictum, mathematics is the language of science. Our foundational language and theory should enable us to say as much as possible, as efficiently as possible. A set theorist cannot help but be aware that there are limits to the usefulness of such maxims, and that proofs usually count for more than plausibility arguments.[3] But the idea that set theorists ought to seek a language and theory[4] that maximizes interpretative power seems to carry us a long way. We shall discuss just how far it goes in this paper.[5]

[1]Gödel did not formulate his program in the broad terms we have used. But [3] is the earliest paper we know of advocating this response to the incompleteness of ZFC with respect to the CH, and the program as we have stated it is an obvious extrapolation. See also [2].

[2]I am using LST to stand for the usual syntax, together with the meaning we currently assign to it. This is why it makes sense to talk about translating pre-set-theoretic language into LST. What the translation preserved was meaning; meaning is that which is preserved by a good translation.

[3]In Einstein's words, "A scientist ...must appear to the systematic epistemologist as a type of unscrupulous opportunist."

[4]We believe the two evolve together.

[5]*Maximize interpretative power* has a lot in common with the old idea that in mathematics, consistency guarantees existence. The trouble with the old formulation is that consistency is a property of theories, not objects. If we try to use the old formulation to guide us to a foundational theory, we face the seemingly hopeless task of making useful sense of "objects that T is about".

2 The consistency strength hierarchy

One thing set theorists have understood much better in the years since Gödel [3] is the family of possible extensions of ZFC. At one level, this family is rich indeed. There is a plethora of seemingly natural, basic questions about uncountable sets which ZFC does not decide. But underlying this great variety of consistent extensions of ZFC, and the corresponding wealth of models of ZFC, there is a good deal more order than might at first be apparent. To explain this, we must introduce the consistency strength hierarchy.

Definition 2.1 *Let T and U be axiomatized theories extending* ZFC*; then $T \leq_{\mathrm{Con}} U$ iff* ZFC *proves* $\mathrm{Con}(U) \Rightarrow \mathrm{Con}(T)$*. If $T \leq_{\mathrm{Con}} U$ and $U \leq_{\mathrm{Con}} T$, then we write $T \equiv_{\mathrm{Con}} U$, and say that T and U have the same consistency strength, or are* equiconsistent*.*

There is an intensional aspect here, in that the order really is on presentations of theories, rather than theories, but we shall ignore that detail here.

Of course, it was Gödel who discovered that there is no largest consistency strength. One can increase the consistency strength of T by adding to it $\mathrm{Con}(T)$, or equivalently, that there is a model of T. One can take larger steps upward by requiring that this model be arithmetically correct, or wellfounded, or an initial segment of the cumulative hierarchy. Still stronger reflection principles lead into the large cardinal hierarchy.

A terminological digression: Some people would rather use "reflection principle" in such a way that large cardinal hypotheses at the level of measurables and beyond do not qualify (cf. [8]). I prefer the more liberal usage. It is the hierarchy of principles that starts with Con(ZFC), or Con(PRA) for that matter, and goes on through the existence of rank-to-rank embeddings that is the important natural kind. *Strong axioms of infinity* is a traditional term, and perhaps the best, but it is a mouthful. I shall tend to use *large cardinal hypothesis* below, although I mean to include here some statements that do not assert the existence of a cardinal number, such as "$0^\sharp$ exists".

Large cardinal hypotheses play a very special role in our understanding of the consistency of theories extending ZFC. Many natural extensions T of ZFC have been shown to be consistent relative to some large cardinal hypothesis H, via the method of forcing. This method is so powerful that, at the moment, we know of no interesting T extending ZFC which seems unlikely to be provably consistent relative to some large cardinal hypothesis via forcing. Thus the extensions of ZFC via large cardinal hypotheses seem to be cofinal in the part of the consistency strength order on extensions of ZFC which we know about. These days, the way a set theorist convinces people that T is consistent is to show by forcing that $T \leq_{\mathrm{Con}} H$ for some large cardinal hypothesis H. He probably had heuristic reasons to believe in advance that T is consistent, but this was the proof. Of course, such a proof only carries weight if there is evidence that H itself is consistent.

We do have pretty good evidence that even quite strong large cardinal hypotheses like the existence of rank-to-rank embeddings are consistent with ZFC. For supercompacts, the evidence is stronger, and when we move down to large cardinal hypotheses like the existence of Woodin cardinals, for which we have an inner model theory, it is much stronger still. In all cases, the evidence is basically the existence of a coherent theory in which the hypothesis plays a central role, a theory that extends in a natural way the theory we obtain from weaker hypotheses.[6]

The evidence from inner model theory is especially strong. If H has an inner model theory, then by assuming H we can give a systematic, detailed description of what a minimal model satisfying H might look like. The prototypical such model is Gödel's universe L of constructible sets, which satisfies ZFC, and various weak large cardinal hypotheses. Larger canonical models satisfy hypotheses like "there is a measurable cardinal" and "there is a Woodin cardinal". Canonical inner models admit a systematic, detailed, "fine structure theory" much like Jensen's theory of L. Such a thorough and detailed description of what a universe satisfying H might look like provides evidence that H is indeed consistent, for a voluble witness with an inconsistent story is more likely to contradict himself than a reticent one. Moreover, inner model theory lets us prove consistency strength lower bounds $\mathrm{Con}(T) \Rightarrow \mathrm{Con}(H)$ for various T which we may have some independent reason to believe consistent, and this provides further evidence that H is consistent. For example, Δ^1_2-determinacy, the existence of a saturated ideal on ω_1, and the existence of one Woodin cardinal are all equiconsistent over ZFC. The consistency of all of them follows from the consistency of the Proper Forcing Axiom. None of these statements has much to do with any of the others; there are different intuitions, motivations, and connections that led set theorists to them. This state of affairs makes it more likely that $\mathsf{ZFC}+$ "there is a Woodin cardinal" is consistent. [7]

Often, a consistency strength upper bound for T obtained by forcing can, with additional work, be made optimal. That is, we can find a large cardinal hypothesis H such that $T \equiv_{\mathrm{Con}} H$. The proof that $H \leq_{\mathrm{Con}} T$, involves constructing inside any model of T a canonical inner model of H. Unfortunately, although the basic theory of forcing is well-understood, inner model theory lags well behind. At the moment, it cannot produce any nontrivial consistency strength lower bounds of the form $H \leq_{\mathrm{Con}} T$, where H is significantly stronger than "There is a Woodin limit of Woodin cardinals", a hypothesis of only middling strength.

The large cardinal hypotheses are themselves wellordered by consistency strength.[8] The pattern described in the last two paragraphs then leads to the following vague conjecture.

[6] With regard to rank-to-rank embeddings, I am thinking of Woodin's theory of $L(V_{\lambda+1})$. With regard to supercompact cardinals, there are the many uses of them in forcing, by many people over a period of decades.

[7] We do not as yet have a full inner model theory for large cardinal hypotheses at the level of superstrong or supercompact cardinals. The author believes that one day we will.

[8] The ones we know.

Natural consistency strengths wellordered: If T is a natural extension of ZFC, then there is an extension H axiomatized by large cardinal hypotheses such that $T \equiv_{\text{Con}} H$. Moreover, $\leq_{\text{Con}}$ is a prewellorder of the natural extensions of ZFC. In particular, if T and U are natural extensions of ZFC, then either $T \leq_{\text{Con}} U$ or $U \leq_{\text{Con}} T$.

It is difficult to see how one could make the conjecture more precise. One can construct unnatural extensions (using self-referential sentences, for example) that are of incomparable consistency strengths. By "natural" we mean considered by set theorists, because they had some set-theoretic idea behind them. Here the standards are very liberal, as the many thousands of pages published by set theorists will testify. There is some vagueness in "large cardinal hypothesis" too, but it is less important. Practically speaking, we may as well take "large cardinal hypothesis" to mean "informative marker of consistency strength, used to compare consistency strengths". Perhaps the main thrust of this vague conjecture at the moment is programmatic: understand better, and develop further, our methods for comparing consistency strengths. At present, this devolves at once into: understand better, and develop further, the theory of canonical inner models satisfying large cardinal hypotheses. One very ambitious conjecture here is that Con(PFA) $\Rightarrow$ Con(there is a supercompact cardinal). This has been a target of inner model theory for about 30 years. [9]

3 A theory of the concrete

A set theory T is consistent just in case all its Π^0_1 consequences are true. As emphasized by Kreisel [7], this is the best way to understand Hilbert's program: a consistency proof for classical mathematics would justify the use of its ideal machinery in deriving "real", that is Π^0_1, statements. The consistency proof would yield a general elimination of classical mathematics from the proofs of Π^0_1 statements, in favor of whatever was used in the consistency proof.

Of course, Hilbert's dream that mathematicians of his time had found a final theory went up in smoke, and it is Gödel's program that rises from the ashes. But perhaps Hilbert's view that Π^0_1 sentences have a special status has something to it. They are, after all, the logically simplest, most concrete sentences that can be true, but not provable in ZFC. Remarkably, climbing the consistency strength hierarchy in any natural way seems to decide uniquely not just Π^0_1 sentences, but more complicated sentences about the concrete as well. *Concrete* refers here to natural numbers, real numbers, and certain sets of real numbers.

[9] There are various ways to attach ordinals to set theories that correspond to the consistency strength order in the case of natural theories. One can look at the provably recursive ordinals, or the minimal ordinal height of a transitive model, for example.

3.1 A theory of natural numbers

Definition 3.1 *Let* Γ *be a set of sentences in the syntax of* LST*, and* T *a set theory; then* $\Gamma_T = \{\varphi \mid \varphi \in \Gamma \wedge T \vdash \varphi\}$.

We shall use this notation with $\Gamma = \Pi^0_1$, the set of Π^0_1 sentences, as well as $\Gamma = \Pi^0_\omega$ and $\Gamma = \Pi^1_\omega$, the sentences in the languages of first and second order arithmetic, respectively.

It is not quite true that if $T \leq_{\text{Con}} U$, then $(\Pi^0_1)_T \subseteq (\Pi^0_1)_U$. (We get a counterexample by letting T be U together with a Rosser sentence for U.) But it is true for natural theories[10], and in fact we observe the following.

Phenomenon: If T and U are natural extensions of ZFC, then

$$\begin{aligned} T \leq_{\text{Con}} U \quad &\Leftrightarrow \quad (\Pi^0_1)_T \subseteq (\Pi^0_1)_U \\ &\Leftrightarrow \quad (\Pi^0_\omega)_T \subseteq (\Pi^0_\omega)_U \end{aligned}$$

Thus the wellordering of natural consistency strengths corresponds to a wellordering by inclusion of theories of the natural numbers. There is no divergence at the arithmetic level, if one climbs the consistency strength hierarchy in any natural way we know of. Even more, the theory of the natural numbers generated by T is a monotonically increasing function of the consistency strength of T.[11]

3.2 A theory of the reals

Natural ways of climbing the consistency strength hierarchy do not diverge in their consequences for the reals.

Phenomenon: Let T, U be natural theories of consistency strength at least that of "there are infinitely many Woodin cardinals"; then either $(\Pi^1_\omega)_T \subseteq (\Pi^1_\omega)_U$, or $(\Pi^1_\omega)_U \subseteq (\Pi^1_\omega)_T$.

In other words, the second-order arithmetic generated by natural theories is an eventually monotonically increasing function of their consistency strengths.

[10] Let $T \leq_{\text{Con}^*} U$ iff for every finite $F \subset T$, U proves $\text{Con}(F)$. By the reflection schema, for any S extending ZFC, $(\Pi^0_1)_S$ is axiomatized by $\{\text{Con}(G) \mid G \subseteq S \wedge G \text{ is finite }\}$. Using this, it ie easy to see that if T and U extend ZFC, then $T \leq_{\text{Con}^*} U$ iff $(\Pi^0_1)_T \subseteq (\Pi^0_1)_U$. Finally, for natural theories extending ZFC, $T \leq_{\text{Con}} U$ iff $T \leq_{\text{Con}^*} U$.

[11] There are counterexamples to eventual monotonicity on the margins. For example, let T be $\mathsf{ZFC}+$ "there is an inaccessible cardinal", and U be $\mathsf{ZFC} + \text{Con}(T)$; then $T \leq_{\text{Con}} U$, but $(\Pi^0_\omega)_T \not\subseteq (\Pi^0_\omega)_U$. One could argue about whether U is natural. (Its extra axiom is about theories, not sets. It is an instrumentalist's theory.) In any case, $(\Pi^0_\omega)_T \cup (\Pi^0_\omega)_U \subseteq^(\Pi^0_\omega)_S$, where S asserts that there are two inaccessibles. We do not know of an example of divergence, or even non-directedness, with any claim to involve natural theories, in the realm of arithmetic.

This phenomenon extends to statements about sets of reals generated by reasonably simple means. As an example of the latter, we consider sets of reals in $L(\mathbb{R})$, the minimal model of ZF containing all reals and ordinals. Let us write $(\mathrm{Th}^{L(\mathbb{R})})_T$ for the set of consequences of T of the form "$L(\mathbb{R}) \models \varphi$".

Phenomenon: Let T, U be natural theories of consistency strength at least that of $\mathsf{ZFC}+$ "there are infinitely many Woodin cardinals with a measurable cardinal above them all"; then either $(\mathrm{Th}^{L(\mathbb{R})})_T \subseteq (\mathrm{Th}^{L(\mathbb{R})})_U$, or $(\mathrm{Th}^{L(\mathbb{R})})_U \subseteq (\mathrm{Th}^{L(\mathbb{R})})_T$.

So there is no divergence at the level of statements about $L(\mathbb{R})$, if one climbs sufficiently high in the consistency strength hierarchy in any natural way we know of. Even more, the theory of $L(\mathbb{R})$ generated by T is an eventually monotonically increasing function of the consistency strength of T.

The eventual monotonicity phenomenon extends somewhat beyond statements about $L(\mathbb{R})$. In general, we classify the complexity of statements by the type of objects they quantify over, and then by the number of alternations of such quantifiers. For example, statements of the form $V_\omega \models \varphi$ would be Σ^0_n, for some n. They quantify over type 0 objects (elements of V_ω, or up to simple coding, natural numbers). Statements of the form $V_{\omega+1} \models \varphi$ would be Σ^1_n, for some n. They quantify over type 1 objects (up to simple coding, real numbers). Statements of the form $V_{\omega+2} \models \varphi$ would be Σ^2_n, for some n. They quantify over type 2 objects (sets of reals). In between the Σ^1_n and Σ^2_n statements are statements about the reals with infinitely many alternations of quantifier. One can make precise sense of this by interpreting such statements using infinite games; see for example [13]. For example, if there is a measurable cardinal, statements of the form

$$L(\mathbb{R}) \models \varphi$$

can be expressed using the real game quantifier of length ω in the form

$$\exists x_0 \forall x_1 \exists x_2 \forall x_3 \ldots .(V_{\omega+1}, \in) \models \varphi^*[\vec{x}].$$

(The quantifier string is interpreted as saying that player I has a winning strategy in the game in which he plays the x_{2i} and his opponent plays the x_{2i+1}, and his goal is to insure that $(V_{\omega+1}, \in) \models \varphi^*[\vec{x}]$.) One can obtain still larger classes of statements by considering games on the reals of length $> \omega$, and the eventual monotonicity phenomenon extends to some of these larger classes.

3.3 Completeness, correctness, and generic absoluteness

In addition to the question of divergence in the realm of the concrete, there is the question of completeness in this realm. Of course, no axiomatizable theory is literally complete, even in the realm of Π^0_1 statements, but we can ask for completeness with respect to mathematically natural statements. Here we find that, roughly speaking, if we climb to a consistency strength sufficient to insure monotonicity in the realm of Γ-statements, then our theory decides the natural statements in the realm of Γ statements.

There is a partial explanation of the phenomena of non-divergence, eventual monotonicity, and practical completeness in the realm of the concrete, for theories of sufficiently high consistency strength. It lies in the way we obtain independence theorems, by interpreting one theory in another. Recall that there are two interpretation methods, one producing generic extensions, and one producing canonical inner models.

For natural T, U extending ZFC, if $T \leq_{\mathrm{Con}} U$, then for every finite $F \subseteq T$, U proves "there is an ω-model M_F of F". This implies at once that $(\Pi^0_\omega)_T \subseteq (\Pi^0_\omega)_U$: if φ is an arithmetic consequence of T, then it is a consequence of some finite $F \subseteq T$, so it is true in M_F, so it is true. This is a partial explanation of arithmetic monotonicity, and it extends to Σ^1_2 monotonicity because in practice, M_F will be a transitive model containing ω_1, and hence Σ^1_2-correct. If T and U are strictly stronger than "there are infintely many Woodin cardinals", then M_F will be $L(\mathbb{R})$-correct. In general, our model-producing methods lead to eventual Γ-monotonicity because in order to produce a model for a theory T that is sufficiently strong with respect to Γ, we must produce a Γ-correct model.

Γ-non-divergence provides also partial explanation of practical Γ-completeness, for natural T that are of sufficient consistency strength. For if φ is a Γ statement, and we obtained (by forcing or inner model theory) models M and N of T from models P and Q of the natural theories S and U, then for all we know, P and Q agreed on φ, and therefore M and N agreed on φ, because M and N were Γ-correct from the point of view of P and Q respectively.

Generic absoluteness theorems help explain why forcing yields Γ- correct models. Here are two such theorems at the level of statements about $L(\mathbb{R})$.

Theorem 3.2 (Woodin 1985) *Suppose there are arbitrarily large Woodin cardinals, and let M and N be set-generic extensions of V; then M and N satisfy the same statements about $L(\mathbb{R})$.*

Theorem 3.3 (Woodin 1988) *Suppose there is an iterable proper class model with infinitely many Woodin cardinals, and let M and N be set-generic extensions of V; then M and N satisfy the same statements about $L(\mathbb{R})$.*

The hypothesis of the second theorem follows from that of the first, so the second theorem properly contains the first. The second theorem also helps explain why canonical inner models of sufficiently strong large cardinal hypotheses are correct about the theory of $L(\mathbb{R})$.

The currrent large cardinal hypotheses are themselves generically absolute. If κ is measurable in V, then it is measurable in $V[G]$, whenever G is generic over V for a poset of size $< \kappa$. The same goes for Woodinness, superstrongness, supercompactness, etc. Large cardinal hypotheses resembling the current ones in very basic ways will be *preserved by small forcing*. If our hypothesis says there are arbitrarily large cardinals of some type, it will be preserved by set forcing. These are results of Levy and Solovay from the 60s.

Thus all the consequences of "there are arbitrarily large Φ-cardinals" are preserved by set forcing, for any of the current large cardinal properties Φ. Theorem 3.2 gives an indication of how extensive those consequences are, in one case.

Finally, we should point out one very important reason natural theories of sufficient consistency strength are complete in the realm of Γ statements: they imply Γ-determinacy. Γ determinacy in turn yields a practically complete theory in the Γ-realm. This theory coheres well with the theory of Borel and analytic sets developed by classical descriptive set theorists under ZFC, and in particular it bears out their intuition that definable sets of reals are free of the pathologies generated by a wellorder of the reals. Decades of work by modern descriptive set theorists have shown this. Since this key point has been made so often before, we shall not dwell on it here. See [10], [18], and [13]. To give a concrete example: if there are arbitrarily large Woodin cardinals, then AD holds in $L(\mathbb{R})$, and this in turn is the basis for a thorough, detailed theory of the internal structure of $L(\mathbb{R})$.

4 A boundary

If we regard Theorem 3.2 as a key indicator of the completeness of large cardinal axioms for sentences about $L(\mathbb{R})$, then it is natural to ask whether this kind of completeness extends further. We can get near the boundary of what is known using the real game quantifier corresponding to games of length ω_1 which are *closed*-Π^1_1, in the sense that there is a Π^1_1 set A of (codes for) countable sequences of reals such that I wins a run $\vec{x}$ just in case for all $\alpha < \omega_1$, $\vec{x} \restriction \alpha \in A$. Let $\Game^{\omega_1}(\text{closed} - \Pi^1_1)$ be the class of all statements of the form: I wins such a game.

Theorem 4.1 (Woodin 1985) *Suppose there are arbitrarily large measurable Woodin cardinals, and let M and N be set-generic extensions of V; then M and N satisfy the same $\Game^{\omega_1}(closed - \Pi^1_1)$ statements.*

Theorem 4.2 (Woodin, Steel) *Suppose there is an iterable inner model satisfying "there is a measurable Woodin cardinal", and let M and N be set-generic extensions of V; M and N satisfy the same $\Game^{\omega_1}(closed - \Pi^1_1)$ statements.*

Whilst proving 4.1, Woodin also showed that if there is a measurable Woodin cardinal, then all length-ω games on ω with $\Game^{\omega_1}(\text{closed} - \Pi^1_1)$ payoff are determined. Recently Itay Neeman has shown that the determinacy of the length-ω_1 closed-Π^1_1 games themselves follows from the existence of iterable inner model satisfying "there is a measurable Woodin cardinal".

(See [15].) The determinacy of these long games implies at once all length-ω games on ω with $\Game^{\omega_1}(\text{closed} - \Pi^1_1)$ payoff are determined. Neeman's theorem also gives a new proof of 4.2. In contrast to the situation lower down, we do not know yet that any large cardinal hypothesis implies that there is an iterable inner model satisfying "there is a measurable Woodin cardinal". Presumably, the existence of arbitrarily large measurable Woodins does so, and thus Neeman's determinacy theorem extends Woodin's, and yields proofs of both 4.2 and 4.1. But we have as yet no proof, and finding such a proof is a manifestation of the fundamental iterability problem in inner model theory.

What happens if we move from long strings of quantifiers over reals to quantifiers over sets of reals? Then we immediately encounter CH, which is a Σ^2_1 statement. None of our current large cardinal axioms decide CH, because they are preserved by small forcing, whilst CH can both be made true and made false by small forcing.[12] Because CH is provably not generically absolute, it cannot be decided by large cardinal hypotheses that are themselves generically absolute.

Theorem 4.3 (Levy, Solovay) *Let A be one of the current large cardinal axioms, and suppose $V \models A$; then there are set generic extensions M and N of V which satisfy $A + \mathsf{CH}$ and $A + \neg\mathsf{CH}$ repectively.*

There may be consistency strengths beyond those we have reached to date that cannot be reached in any natural way without deciding CH. But the consistency strengths of which we currently have some understanding can be reached by the current large cardinal axioms, and since these are generically absolute, they do not decide CH. On the other hand, the statement of CH involves just one quantifier over arbitrary sets of reals.

It is worth remarking that CH plays a fundamental role among Σ^2_1 sentences, in that it implies that every Σ^2_1 sentence can be translated into a $\Game^{\omega_1}(\text{closed} - \Pi^1_1)$ sentence. Thus theorem 3.3 gives us at once the following *conditional* generic absoluteness theorem for Σ^2_1 sentences

Theorem 4.4 (Woodin) *Suppose $V \models$ "There are arbitrarily large measurable Woodin cardinals". Let M and N be set-generic extensions of V satisfying CH; then M and N are Σ^2_1-equivalent.*

It is open whether if $V \models$ "There are arbitrarily large supercompact cardinals", and M and N are set-generic extensions of V satisfying $\diamondsuit$, then M and N are Σ^2_2-equivalent. If so, this would indicate that in the presence of large cardinal hypotheses, $\diamondsuit$ yields a complete theory at the Σ^2_2 level, just as CH does at the Σ^2_1 level. Such *conditional generic absoluteness* theorems may go much further.

[12]To vary CH, we need a poset of size at most $(2^{\aleph_0})^+$, so small with respect to any large cardinal.

5 The multiverse language

What does this picture of what is possible suggest as to what we should believe, or give preferred development, as a framework theory?

We have good evidence that the consistency hierarchy is not a mirage, that the theories in it we have identified are indeed consistent. This is especially true at the lower levels, where we already have canonical inner models, and equiconsistencies with fragments of definable determinacy. This argues for developing the theories in this hierarchy. All their Π^0_1 consequences are true, and we know of no other way to produce new Π^0_1 truths.

Developing one natural theory develops them all, via the boolean-valued interpretations. At the level of statements about the concrete (including most of what non-set-theorists say), all the natural theories agree. This might suggest that we need no further framework: why not simply develop all the natural theories in our hierarchy as tools for generating true statements about the concrete? Let 1000 flowers bloom! This is Hilbertism without the consistency proof, and with perhaps an enlarged class of "real" statements.

The problem with this watered-down Hilbertism is that we don't want everyone to have his own private mathematics. We want one framework theory, to be used by all, so that we can use each other's work. It's better for all our flowers to bloom in the same garden. If truly distinct frameworks emerged, the first order of business would be to unify them.

In fact, the different natural theories we have found in our hierarchy are not independent of one another. Their common theory of the concrete stems from logical relationships that go deeper, and are brought out in our relative consistency proofs. These logical relationships may suggest a unifying framework.

The central role of the theories axiomatized by large cardinal hypotheses argues for adding such hypotheses to our framework. The goal of our framework theory is to *maximize interpretative power*, to provide a language and theory in which all mathematics, of today, and of the future so far as we can anticipate it today, can be developed. Maximizing interpretative power entails maximizing consistency strength, but it requires more, in that we want to be able to translate other theories/languages into our framework theory/language in such a way that we preserve their meaning. The way we interpret set theories today is to think of them as theories of inner models of generic extensions of models satisfying some large cardinal hypothesis, and this method has had amazing success. We don't seem to lose any meaning this way. It is natural then to build on this approach.

Nevertheless, large cardinal hypotheses like our current ones cannot decide CH, and so our theory of the concrete still has many different possible theoretical superstructures, some with CH, some with $\diamondsuit$, some with MM, some with $2^{\aleph_0}$ being real-vaulued measurable, and so on: all the behaviors that can hold in set-generic extensions of V, no matter what large cardinals exist. 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. That is, we want one neat package they all fit into.

To this end, we describe a *multiverse language*, and an open-ended *multiverse theory*, in an informal way. It is routine to formalize completely.

Multiverse language: usual syntax of set theory, with two sorts, for the *worlds* and for the *sets*.

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 $W[G] = U[H]$.

It is a theorem of Laver and Woodin (cf. [22]) that there is a formula ψ of LST such that if $N = W[H]$ where H is N-generic for a set forcing $\mathbb{P}$, then ψ defines W over N from $\mathbb{P}$ and H. This result must be used to formalize (2)(c) precisely; W must be defined via ψ inside U in the statement of (2)(c).

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 $\mathrm{Col}(\omega, < \mathrm{OR}^M)$. The worlds of the multiverse M^G are all those W such that

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

for some H set generic over W, and some $\alpha \in \mathrm{OR}^M$. It follows from the result of Laver and Woodin 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 for any sentence φ of the multiverse language 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".

Notice also that 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 then that we are using the multiverse language as a sublanguage of the standard one, in the way described above. 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 MV as follows: given such a large cardinal hypothesis φ, we add "φ^W, for all worlds W" to MV. For example, we might add "(there is a Woodin cardinal)^W, for all worlds W". By standard preservation theorems for large cardinals, this implies that every world has arbitrarily large Woodin cardinals. The same goes for superstrongs, supercompacts, etc. 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.

Why have we included the worlds in M^G; why not define it so as to have only the sets? Then we have only a model of "every set is countable"! We have lost important information, the information-level of the worlds. Set theory cannot be formalized in the language of this structure.

Why haven't we put more into the multiverse? Why not declare definable inner models of worlds to be worlds, or even declare the sets to be worlds in their own right? The answer is that they are already there, we can talk about them in the multiverse language already, and if we go too far in the direction of obscuring what we are now calling the worlds with other objects, we may lose information.[13] Our multiverse is an equivalence class of worlds under "has the same information". Definable inner models and sets may lose information, and we do not wish to obscure the original information level.

For the same reason, our multiverse does not include class-generic extensions of the worlds. There seems to be no way to do this without losing track of the information in what we are now regarding as the multiverse, no expanded multiverse whose theory might serve as a foundation.[14] We seem to lose interpretative power.

The multiverse language is a sublanguage of the standard one, under the translation indicated above. It is sufficiently expressive to state versions of the axioms of ZFC, and of the large cardinal hypotheses preserved by set forcing: we replace φ by "for all worlds W, φ^W". Clearly we cannot state the CH this way. The same goes for the many other statements about the uncountable which are sensitive to set forcing, no matter what large cardinals there may be. Whether there are traces of CH and these other sentences in the multiverse language is the issue we consider next.

One can think of the standard language as the multiverse language, together with a constant symbol $\dot{V}$ for a reference universe. Statements like CH are intended as statements about the reference universe. To what extent is this constant symbol meaningful? Does

[13]An extreme here is Hamkins' multiverse in [4], whose first order theory, if it were formalized so as to have one, would probably be an arithmetic set.

[14]Amalgamation will fail if we start counting sets, definable inner models, or class-generic extensions as worlds.

one lose anything by retreating to the superficially less expressive multiverse language? We distinguish three answers to this question:

Weak relativist thesis: Every proposition that can be expressed in the standard language LST can be expressed in the multiverse language.

It follows from the weak relativist thesis that the symbol $\dot{V}$, which produces the standard language when we add it to the multiverse language, only makes sense if we can define it in the multiverse language. The weak relativist thesis is silent on whether one can do that.

At the other pole, we have

Strong absolutist thesis: "$\dot{V}$" makes sense, and that sense is not expressible in the multiverse language.[15]

Finally, perhaps weak relativism and the absolutist's idea of a distinguished reference world can be combined, in that that there is an individual world that is definable in the multiverse language. An elementary forcing argument shows that if the multiverse has a definable world, then it has a unique definable world, and this world is included in all the others.[16] In this case, we call this unique world included in all others the *core* of the multiverse.

Weak absolutist thesis: There are individual worlds that are definable in the multiverse language; that is, the multiverse has a core.

The strongest evidence for the weak relativist thesis is that the mathematical theory based on large cardinal hypotheses that we have produced to date can be naturally expressed in the multiverse language. 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.[17]

One argument for strong absolutism is that the multiverse language is parasitic on the standard one. We have only been able to understand it by giving the translation function t.[18] By itself, this is not a very strong objection, as one could think of what we are doing

[15] The strong absolutist thesis is not, strictly speaking, the negation of the weak relativist one.

[16] The author is grateful to Hugh Woodin for this observation.

[17] There is a passage in [22] in which, making some translations from Woodin's language to ours that may not be valid, the weak relativist is challenged to come up with some mathematical truth that cannot be formalized in the multiverse language.(See [22][page 26].) This is odd indeed, since the weak relativist thesis is just that, if we stay within the language of set theory, this cannot be done.

[18] Set theorists sometimes over-estimate the value of models of set theory as a tool for fixing meaning. What we do in the formal semantics (model theory) of set theory is make translations. You don't get a model of ZFC unless you start with one. Our description of M^G showed how to translate the multiverse

as isolating the meaningful part of the standard language, the range of t, while trimming away the meaningless, in order to avoid psuedo-questions. After climbing our ladder, we throw it away, and from now on, MV can serve as our foundation. After all, people learning how to use LST are shown diagrams in the shape of the letter "V", and told stories about random collection being iterated over transfinite time, and supplied with other metaphors with features that will produce psuedo-questions if taken too seriously. Perhaps one ought to discard more of this than is at first apparent. The true test is whether we thereby lose the ability to formalize any mathematics.

Whatever one thinks of the semantic completeness of the multiverse language, it does bring the weak absolutist thesis to the fore, 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! Indeed, if there is an inclusion-least world in the multiverse, why don't we use $\dot{H}$ to denote it[19], and agree to retire $\dot{V}$ until we need it? The question as to whether the multiverse has a core is an important question for everyone, relativist or absolutist.

Neither MV nor its extensions by large cardinal hypotheses of the sort we currently understand decides whether there is a core to the multiverse, or the basic theory of this core if it exists. (See [16], [5],[6].) So what we have here is another basic question, like the CH, that large cardinals do not decide. But it is a different question, and its role in our search for a universal framework theory seems crucial.

There is some reason to hope for a positive answer. Hugh Woodin has recently proposed an axiom which

(a) implies the multiverse has a core,

(b) suggests an approach toward developing a detailed, systematic "fine structure theory" for this core, and

(c) may be consistent with all our large cardinal hypotheses.

The new mathematics needed in order to turn (b) and (c) from promise into reality is formidable, but there is some reason for optimism. In the next two sections I shall describe this axiom, where it comes from, and why one might hope that it pins down our multiverse, without restricting it.[20]

language into the standard one, but it did not explain the meaning of the multiverse language in any other way. The meaning of a foundational language is determined by its use, and in particular the theory we develop within it. The multiverse sublanguage has been in use a long time, and all of set theory has been developed within it.

[19] As we shall see below, there is a reason for the letter H.

[20] The author wishes to thank Hugh Woodin for his permission to discuss his axiom, and his unpublished results regarding it.

There are other ways the multiverse may have important global structure, as indicated by conditional generic absoluteness theorems. Let us say that an axiomatizable theory T in LST is Γ-*complete* iff T is true in some world of the multiverse, and whenever M and N are worlds satisfying T, then they satisfy the same Γ sentences.[21] If the weak absolutist thesis is true, then "I am the core of my multiverse" is Γ-complete, where Γ is the set of all sentences in LST whatsoever. So too is "I am a generic extension of the core of my multiverse via the poset adding ω_2 Cohen reals". But there are other interesting examples of Γ completeness. For example, CH is Σ^2_1-complete, by theorem 4.1. There may be axiomatizable T that are Σ^2_n-complete for all n, and yet do not imply the weak absolutist thesis.

Historical note. The author has thought off-and-on about the various theses above for a long time, and published some remarks on the role of the generic multiverse at the end of [20]. Recently, Hamkins ([5]) and Woodin ([22]) have published papers on the multiverse idea. Hamkins' realization of the idea is quite far from ours, and as far as I can tell, does not lead to any candidate for a framework language and theory. Woodin's generic multiverse V_M is much closer, but it is not actually a model of MV, and it is not at all clear what its theory would be.[22] Neither Hamkins nor Woodin presented a language and a first order theory in that language, both of which seem necessary for a true foundation. If one takes Woodin's statement of what he calls the *generic multiverse position*, namely that "truth is equal to truth in all universes in V_M"[23], and then tries to find a language that expresses all truths and all falsehoods and nothing else, one will probably arrive at the multiverse language and MV. The generic multiverse position seems to correspond to what I am calling the weak relativist thesis. Woodin's paper makes some arguments against the generic multiverse position, based on the logical complexity of certain truth predicates, but those arguments do not seem valid to me.[24] In particular, the weak relativist thesis does not seem to have much in common with formalist doctrines.

[21] If we require this also for M and N being rank initial segments of worlds, we get completeness in the sense of Woodin's $\models_\Omega$ relation. Woodin has conjectured that no T is Σ^2_3-complete in this stronger sense. See [22].

[22] V_M does not satisfy amalgamation.

[23] Where $M = V$?

[24] For those familiar with the arguments: the problem is that the decision to stay within the multiverse language does not commit one to a view as to what the multiverse looks like. The "multiverse laws" do not follow from the weak relativist thesis. The argument that they do is based on truncating the worlds at their least Woodin cardinals. However, this leaves one with nothing, an unstructured collection of sets with no theory. The Ω-conjecture does not imply a paradoxical reduction of MV or its language to something simpler, because there is no simpler language or theory describing a "reduced multiverse" consisting of truncated worlds. There is no "rejection of the transfinite beyong $H(\delta_0^+)$" implicit in staying within the multiverse sublanguage; indeed, saying that every world has a Woodin cardinal automatically entails that every world has arbitrarily large Woodin cardinals.

6 V looks like the HOD of a model of AD

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. Gödel first isolated HOD in the 1940s; see [2]. Myhill and Scott showed in [14] that if $M \models \mathsf{ZF}$, then $\mathrm{HOD}^M \models \mathsf{ZFC}$. Woodin's axiom says that V looks like HOD^M, for models M of the axiom of determinacy.

Here is a precise statement. It's not the strongest form of the axiom one could put down, but is simple, and gives the idea.

Axiom H. For any sentence φ of LST: if if φ is true, then for some $M \models \mathsf{AD}^+ + V = L(P(\mathbb{R}))$ such that $\mathbb{R} \cup \mathrm{OR} \subseteq M$, $(\mathrm{HOD} \cap V_\Theta)^M \models \varphi$.

Here AD^+ is a slight strengthening of AD, introduced by Woodin for technical reasons that need not concern us here. Axiom H is a schema, but one could restrict it to Σ_2 sentences without much loss. The schema is stated above in the standard language, but Woodin has shown that it implies that V is the core of its own multiverse. So one could state Axiom H in the multiverse language: the multiverse has a core, and it satisfies Axiom H.

Axiom H holds in $(\mathrm{HOD} \cap V_\Theta)^M$[25], if M is a model of $\mathsf{AD}_\mathbb{R}$ plus "Θ is regular". It is thus consistent relative to fairly weak large cardinal hypotheses. (See [17].) The hope is that Axiom H is consistent with all the large cardinal hypotheses, so that adopting it does not restrict interpretative power, whilst at the same time it yields a detailed fine structure theory for V, removing the incompleteness that large cardinal hypotheses by themselves can never remove.[26] It is known that Axiom H implies the CH, and many instances of the GCH. Whether it implies the full GCH is a crucial open problem.

Once again, Axiom H can be stated in the multiverse language. The strong absolutist who believes that V does not satisfy CH must still face the question whether the multiverse has a core satisfying Axiom H. If he agrees that it does, then the argument between him and someone who accepts Axiom H as a strong absolutist seems to have little practical importance.

7 Homogeneously Suslin sets of reals

In this section, I shall say something about where Axiom H came from, and why one might entertain such hopes for it. The technical prerequisites for following this material are significantly greater than what was needed to follow the paper up to this point. For the reader who is not a set theorist, it would make sense to skip this section.

[25] Θ is the least ordinal not the surjective image of the reals.

[26] If they are preserved by small forcing.

What makes a set of reals amenable to the techniques of descriptive set theory? [27] The key property is that of being ∞-*homogeneously Suslin*, or Hom_∞ for short. We shall give the technical definition now, but the reader will not lose much by reading Hom_∞ as "well-behaved" in the sequel.

Definition 7.1 *A set $A \subseteq \omega^\omega$ is Hom_∞ iff for any κ, there is a continuous function associating to reals x direct limit systems $\langle (M_n^x, i_{n,m}^x) \mid n, m < \omega \rangle$ on ω^ω such that for all x,*

(a) $M_0^x = V$, each M_n^x is a transitive model of ZFC closed under κ-sequences, and

(b) $x \in A \Leftrightarrow \lim_n M_n^x$ is wellfounded.

The concept was abstracted in the early 70s by Kechris and Martin, from Martin's 1968 proof of $\underset{\sim}{\Pi}^1_1$-determinacy. Hom_∞ sets are determined, Lebesgue measurable, have the Baire property, and so on. The definition seems to capture what it is about sets of reals that makes them "well-behaved". [28] In general, a given large cardinal hypothesis will show that sets of reals in a certain logical complexity class are well-behaved by showing that the sets in question are Hom_∞.

A *pointclass* is a collection of sets of reals closed under complements and continuous preimages. One can think of a pointclass as a logical complexity class for sets of reals. If there are arbitrarily large Woodin cardinals, then Hom_∞ is a pointclass, and closed under fairly complicated types of definability. In fact,

Theorem 7.2 (Martin, S., Woodin 1985) *If there are arbitrarily large Woodin cardinals, then for any pointclass Γ properly contained in Hom_∞, every set of reals in $L(\Gamma, \mathbb{R})$ is in Hom_∞, and thus $L(\Gamma, \mathbb{R}) \models \mathsf{AD}^+$.*

Here AD^+ is a slight strengthening of the axiom of determinacy AD. The fact that $L(\Gamma, \mathbb{R}) \models \mathsf{AD}^+$ leads to a thorough understanding of this model. One of the most basic features of such a model is given by

Theorem 7.3 (Wadge, Martin 1970) *Assume AD; then the pointclasses are wellordered by inclusion.*

[27] For us, a *real* is usually an infinite sequence of natural numbers, considered as an element of the Baire space ω^ω.

[28] *Universally Baire* is an intensionally weaker property, discovered in the late 80s by Feng, Magidor, and Woodin ([1]). All Hom_∞ sets are universally Baire, and if there are arbitrarily large Woodin cardinals, then all universally Baire sets are Hom_∞. See [19].

By theorem 7.3, the sets of reals in Hom_∞ fall into a well-defined, wellordered hierarchy based on their logical complexity. As we climb this hierarchy, new facts as to the existence of well-behaved sets of reals with various properties may be verified. More precisely, new $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ statements may be verified, where a $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ *statement* is one of the form: $\exists A \in \mathrm{Hom}_\infty(V_{\omega+1}, \in, A) \models \varphi$.

Definition 7.4 *0^Ω is the set of all true $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ statements. For any theory T, put $0^\Omega_T = \{\varphi \mid \varphi$ is $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ and $T \vdash \varphi\}$.*

Phenomenon: Let T, U be natural theories extending ZFC + "there are arbitrarily large Woodin cardinals"; then either $0^\Omega_T \subseteq 0^\Omega_U$ or $0^\Omega_U \subseteq 0^\Omega_T$.

There is a version of this phenomenon that applies to theories that are consistencywise as strong as "there are arbitrarily large Woodin cardinals", rather than outright extend that theory, but it would take us too far afield to try to explain it.

7.1 Generic absoluteness revisited

The central idea of descriptive set theory is that the *definable* sets of reals are well-behaved (Lebesgue measurable, determined, etc.) in ways that arbitrary sets of reals are not. How is definability related to homogeneity? In brief, a large cardinal hypothesis H will prove that the pointclass Γ of sets of reals definable in a certain way consists of well-behaved sets by showing $\Gamma \subseteq \mathrm{Hom}_\infty$.

In a similar vein, the generic absoluteness of a class of statements can be proved by reducing them to $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ statements. The following theorem is an abstract statement of part of the method. (See [19].)

Theorem 7.5 (Woodin) *If there are arbitrarily large Woodin cardinals, then $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ statements are absolute for set forcing.*

In order to prove a class Γ of statements is generically absolute under the large cardinal hypothesis H this way, one produces a recursive translation function t, and uses H to show that for all φ,

$$\varphi \in \Gamma \Leftrightarrow t(\varphi) \in 0^\Omega.$$

One may need an H stronger than "there are arbitrarily large Woodin cardinals" to do this. For example, if Γ is the class of $\Game^{\omega_1}($ closed-$\Pi^1_1)$ statements, then H must be "there is an iterable model with a measurable Woodin cardinal". It is natural to view the proof of 4.2 this way.

The proofs of generic absoluteness that proceed directly from large cardinals, such as that of 4.1, need not produce reductions of Γ-truth to $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$-truth. However, it is reasonable to believe that one day the iterability problem will be solved, and that Γ-generic absoluteness will be shown to follow from reductions of Γ-truth to $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$-truth, for very complicated Γ.

How complicated? It is easy to see that every $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ sentence φ is (up to provable-in-ZFC equivalence) a Π_2 sentence of LST.[29] Assuming there are arbitraily large Woodin cardinals, φ is therefore equivalent to the sentence in the multiverse sublanguage: "for all worlds W, φ^W". This puts an upper bound on the complexity of sentences whose generic absoluteness can be guaranteed by a reduction to 0^Ω.

Woodin's Ω-*conjecture* (cf.[22]) implies that this upper bound is optimal. It states that any true sentence of the multiverse sublanguage of the form "for all worlds W, φ^W", where φ is Π_2, has a Hom_∞ *proof*, in a certain precisely defined sense. If the Ω-conjecture is true, then generic absoluteness for Π_2 statements is always guaranteed by Hom_∞ proofs, in a way that resembles abstractly the proofs of 3.3 and 4.2. So up to provable equivalence in ZFC+"there are arbitarily large Woodin cardinals", the class of $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ statements coincides with the class of statements of the form "for all worlds W, φ^W", where φ is Π_2.

It is worth pointing out that the large cardinal hypotheses are Σ_2, and sometimes Σ_3. (The "local" ones like the existence of measurables, Woodins, superstrongs, or huges are Σ_2. The existence of strong or supercompact cardinals is a Σ_3 sentence.) "For all worlds W, (there is a measurable)^W" is a true statement of the multiverse language. This makes it hard to see how the weak relativist thesis implies an identification of truth with Ω-provability, as [22] seems to argue.[30]

Here are two natural test questions regarding the extent of generic absoluteness:

Open questions: Does any large cardinal hypothesis (e.g. the existence of arbitarily large supercompact cardinals) imply

(a) that statements of the form $\forall x \in \mathbb{R} \exists A \in \mathrm{Hom}_\infty (V_{\omega+1}, \in, A) \models \varphi[x]$ are absolute for set forcing?

(b) that $L(\mathrm{Hom}_\infty, \mathbb{R}) \models \mathsf{AD}$?

The canonical inner models for such a large cardinal hypothesis would have to be different in basic ways from those we know. It is unlikely that superstrong cardinals would suffice.

[29]The best way to think of Π_2 sentences is to think of them as Π_1 relative to the power set function. In other words, they are just sentences of the form "for all α, $V_\alpha \models \psi$".

[30]Gödel sentences like "I have no Hom_∞ proof from T" and "there is no Hom_∞ proof of $0 = 1$ in T" also show that truth for sentences of the form "for all W,φ^W", with φ being Σ_2, exceeds Hom_∞-provability from T. Here we assume T is an axiomatizable theory extending ZFC+"there are arbitrarily large Woodin cardinals, and T is true in V^W_α, for some world W and some ordinal α.

Axiom H plus "there are arbitrarily large Woodin cardinals" implies the negations of both (a) and (b).

7.2 Extender models and iteration strategies

The natural attempt to complete set theory in a way that is compatible with all the large cardinal hypotheses is try to model those hypotheses in some minimal way. This leads us into the theory of canonical inner models for large cardinal hypotheses. Gödel's L is the prototypical such model, but it is too small to satisfy even moderately strong large cardinal hypotheses. In the period 1966–1990, set theorists generalized Gödel's construction in such a way that it could produce models with superstrong cardinals. Let us call the resulting models *short-extender models.* [31] It is not known how to generalize this framework so that it produces inner models with supercompact cardinals and beyond. The author believes this can be done. Let us call the as-yet-undiscovered models *long-extender models.*

What makes a countable short-extender model M canonical is a Hom_∞ set of reals called an *iteration strategy for M*. Let us call such M *iterable.* If M is iterable, then every real in M is ordinal definable, and in fact, $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$-definable from a countable ordinal. Moreover, if φ is a $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ statement, and there is an iterable short-extender model M with arbitrarily large Woodin cardinals such that $M \models \varphi$, then φ is true. The converse is true for φ that are not too complicated. For example, true $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ statements of the form $(V_{\omega+1}, \in, \mathbb{R}^\sharp) \models \varphi$ are certified by iterable short-extender models this way, and this is the basis for the proof of 3.3. More generally, the Hom_∞ proofs of Ω-logic, in the region where we understand them very well, are Hom_∞ iteration strategies for short-extender models.

It is natural to guess that the as-yet-undiscovered long-extender models are also certified by Hom_∞ sets, and that a $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ statement is true just in case it is true in some iterable extender model. That is, the Hom_∞ proofs of Ω-logic can always be taken to be an iteration strategy for a canonical extender model reaching some large cardinal hypothesis or other. If this is not the case, then Axiom H loses plausibility, for the determinacy models in which those HOD's are formed may just not go far enough. We are still some considerable distance from proving this general *mouse set conjecture* (see [21]). We do not know that there are iteration strategies for short-extender models with superstrongs, and we do not know what the long-extender models look like. ([23] has made progress on the latter problem.)

7.3 HOD in models of determinacy

There is another family of canonical models of ZFC. Let $M = L(\Gamma, \mathbb{R})$, where Γ is a pointclass properly included in Hom_∞. Put

$$\mathcal{H}^M = \mathrm{HOD}^M \cap V_\theta,$$

[31] The basics of this framework are due to W. Mitchell, in [11].

where $\theta = \Theta^M$. Note that $\mathcal{H}^M$ really only depends on the Wadge ordinal of Γ, and therefore it is contained in the HOD of V. Indeed, all its reals are $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ in a countable ordinal.

To simplify the discussion a bit, let us assume M is a model $\mathsf{AD}_\mathbb{R}$+"Θ is regular". By results of G. Sargsyan [17], the existence of a Woodin limit of Woodin cardinals implies that there are such M. Hugh Woodin has shown that in this case

$$\mathcal{H}^M \models \mathsf{ZFC} + \text{ "there are arbitrarily large Woodin cardinals"},$$

and that every $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ truth with witness in Γ is true in $\mathcal{H}^M$. (See [9] for the first statement, and [19] for a proof of the second.) So $\mathcal{H}^M$ keeps pace with M with respect to $(\Sigma^2_1)^{\mathrm{Hom}_\infty}$ truth, and it has at least moderately large cardinals. Moreover, Woodin has shown that

$$\mathcal{H}^M \models \text{ Axiom } \mathsf{H}.$$

(For the experts: $\mathcal{H}^M$ can be elementarily embedded into the HOD of its own derived model. But this derived model is a proper Wadge initial segment of $j((\mathrm{Hom}_\infty)^{\mathcal{H}^M})$, for some stationary tower embedding j. So $\mathcal{H}^M$ is elementarily equivalent to the HOD of some proper initial segment of $j((\mathrm{Hom}_\infty)^{\mathcal{H}^M})$, and now we can pull this fact back using j.)

Can $\mathcal{H}^M$ be analyzed? It is known to satisfy CH, and many instances of GCH[32], but in general we don't know. However, G. Sargsyan has shown that for M_0 the minimal Wadge initial segment of Hom_∞ satisfying $\mathsf{AD}_\mathbb{R}$+"Θ is regular", a full analysis is possible. Sargsyan's analysis actually applies to the HOD of $L(\Gamma, \mathbb{R})$, for any Wadge initial segment of Hom_∞ up to, and somewhat past M_0.[33] For the M to which Sargsyan's work applies, HOD^M is a short-extender model, expanded with information as what its own iteration strategy is.[34] It is natural to suppose that this analysis will go all the way, and that one day we will prove

Conjecture. If Γ is a pointclass properly contained in Hom_∞, then GCH holds in $\mathrm{HOD}^{L(\Gamma,\mathbb{R})}$.

The conjecture is a $(\Pi^2_1)^{\mathrm{Hom}_\infty}$ statement, so it is generically absolute. We would guess that it is provable in ZFC plus "there are arbitrarily large Woodin cardinals.

The other important question here is whether $\mathcal{H}^M$ can satisfy very strong large cardinal hypotheses. We don't know yet whether even a Woodin limit of Woodins is possible.[35] If $M \models \mathsf{AD}_\mathbb{R}$, then $\mathcal{H}^M$ has no strong cardinals, by a theorem of Woodin. However, the local, Σ_2 large cardinal axioms may hold in $\mathcal{H}^M$ for sufficiently large $M \models \mathsf{AD}_\mathbb{R}$, and the global ones , like the existence of supercompacts, may hold in $\mathcal{H}^M$ for M that do not satisfy $\mathsf{AD}_\mathbb{R}$.[36]

[32] Woodin's proof that Σ^2_1 has the scale property (cf. [19]) implies that for $\Gamma = \Sigma^2_1$, the results of [12] apply.

[33] Sargsyan's work builds on work of the author and Hugh Woodin. See [21].

[34] The idea of expanding extender models this way is due to Hugh Woodin.

[35] G. Sargsyan contributed to Woodin's discovery of Axiom H by persisting in the idea that this might be possible.

[36] Which of these two kinds of HOD is a model for the core of the multiverse seems to me to be a matter of convention, like the decision between Kelley-Morse and ZFC. Indeed, the universe above a strong cardinal is something like the level of proper classes in Kelley-Morse.

8 Gödel's program

What are our prospects today for reaching Gödel's original goal, deciding the CH?

The work of Gödel, Cohen, and their successors has shown us just how important the metamathematics of set theory is in this endeavor. An understanding of the possible set theories is useful in finding the true one. Our metamathematical work has also shown us just how many other natural questions are in the same boat as the CH, and thereby broadened our focus. It seems highly unlikely that the solution to CH will be a "one-off".[37] One idea motivating Gödel's work on L, the first substantial work in the metamathematics of set theory, was that one cannot count the reals until one has specified in a more thorough way what it is one is counting. This seems more likely to carry the day. Most likely, the Continuum Problem will not be solved without a significant, far-reaching clarification of the notion of set. Our metamathematical work is a necessary prelude to that.

Our current understanding of the possibilities for maximizing interpretative power has led us to to one theory of the concrete, and a family of theoretical superstructures for it, each containing all the large cardinal hypotheses. These different theories are logically related in a way that enables us to use them all together. Whatever the strong absolutist may believe about V, it is surely an important fact about the global structure of V that it has the generic extensions it does have.

The logical relationships between the different theories extending ZFC plus large cardinal hypotheses we have discovered are brought out clearly by formalizing them in the multiverse language. This language is a sublanguage of the standard one, and in it we can formalize naturally all the mathematics that set theorists have done. Remaining within this sublanguage has the additional virtue that our attention is directed away from CH, which has no obvious formalization within it, and toward the global question as to whether the multiverse has a core.

Remarkably, we can see now the outlines of a positive answer to this question, a way in which the multiverse may indeed have a core, and this core may admit a detailed fine-structural analysis that resembles that of Gödel's L. There are formidable technical mathematical problems that need to be answered in a certain way to realize this promise: we must show that there are models M of AD^+ such that HOD^M satisfies "there are supercompact cardinals", for example, and we must produce a fine structure theory for HOD^M. Although these are difficult questions, large cardinal hypotheses should settle them.

Perhaps the mathematics will turn out some other way. Perhaps the multiverse has no core, but some other, more subtle structure. There are many basic open questions at the foundations of set theory: the extent of generic absoluteness, the existence of iterable structures, the Ω-conjecture, the form of canonical inner models with supercompacts, and

[37] Freiling's "darts argument" and some passages in Gödel's [3] are examples of attempted "one-offs". The Banach-Tarski paradox ought to make clear how little weight can be attached to appeals to intuition in this context.

the properties of HOD in models of determinacy, to give my own partial list. Our path toward a stronger foundation will be lit by the answers to such questions.

References

[1] Qi Feng, Menachem Magidor, and W. Hugh Woodin, *Universally Baire sets of reals*, in Set theory of the Continuum, H. Judah, W. Just, and W.H. Woodin eds., MSRI publications 26 (1992), Springer-Verlag.

[2] Kurt Gödel, Remarks before the Princeton bicentennial conference on problems in mathematics, 1-4 (1946), first published in The undecidable: Basic papers on undecidable propositions, unsolvable problems and computable functions, M. Davis ed., Hewlitt, N.Y.: Raven Press (1965).

[3] Kurt Gödel, What is Cantor's Continuum Problem?, American Mathematical Monthly 54 (1947), 515-525.

[4] Victoria Gitman and Joel Hamkins, A natural model of the multiverse axioms, Notre Dame J. Form. Log., vol. 51, iss. 4, pp. 475-484, 2010.

[5] Joel Hamkins, The set-theoretic multiverse : a natural context for set theory(mathematical logic and its applications), Annals of the Japan Association for Philosophy of Science, vol. 19, pp. 37-55, 2011.

[6] Joel Hamkins, Jonas Reitz, and W. Hugh Woodin, The Ground Axiom is consistent with V=HOD, Proc. AMS. I, (2007), 6 pp.

[7] Georg Kreisel, Hilbert's programme, in Philosophy of Mathematics, Paul Benacerraf and Hilary Putnam, eds., Cambridge: Cambridge University Press, 207-238 (1983), 2nd ed.

[8] Peter Koellner, On Reflection Principles, Annals of Pure and Applied Logic, Vol. 157, Nos. 2-3, 2009, pp. 206-219.

[9] Peter Koellner and W. Hugh Woodin, Large cardinals from determinacy, in Handbook of Set Theory, vol. 3, M. Foreman and A. Kanamori eds., Springer 2010, pp. 1951–2119.

[10] Donald Martin, Mathematical evidence, in Truth in Mathematics, H.G. Dales and G. Oliveri (eds.), Oxford: Clarendon Press (1998), 215-231.

[11] William Mitchell, Hypermeasurable Cardinals, in Maurice Boffa, Dirk van Dalen, Kenneth McAloon eds., Logic Colloquium '78, Proceedings of the Colloquium held in Mons, August 1978, Amsterdam 1979, 303-316.

[12] Yiannis Moschovakis, Ordinal games, in Cabal Seminar 77-79, A.S. Kechris, D.A. Martin, and Y.N. Moscovakis eds., Lecture Notes in Math, vol. 839 (1981), Springer-Verlag, Berlin, 169-201.

[13] Yiannis Moschovakis, Descriptive set theory, second edition, American Mathematical Society, 2009.

[14] John Myhill and Dana Scott, Ordinal definability, in Axiomatic set theory, Proceedings of symposia in pure mathematics vol. 13, part 1, 271-278, D. Scott ed. (Providence, R.I., American Mathematical Society).

[15] Itay Neeman, Games of length ω_1, Journal of Mathematical Logic, vol. 7 (2007), pp. 83124.

[16] Jonas Reitz, The Ground Axiom, The Journal of Symbolic Logic, 72(4):1299-1317, 2007.

[17] Grigor Sargsyan, A tale of hybrid mice, Ph.D. thesis, Berkeley 2009.

[18] John Steel, Mathematics needs new axioms, Bulletin of Symbolic Logic,**6** (Dec. 2000), no. 4, pp. 422-433.

[19] John Steel, The derived model theorem, in Logic Colloquium 2006, Cooper et. al. eds., Cambridge University Press, 2009.

[20] John Steel, Generic absoluteness and the Continuum Problem, Slides for talk given at the Laguna workshop on philosophy and the Continuum Problem (P. Maddy and D. Malament organizers), March 2004, http://math.berkeley.edu/steel/talks/laguna.ps

[21] John Steel and W. Hugh Woodin, HOD as a core model, to appear in: Ordinal definability and recursion theory: the Cabal seminar, A. Kechris, B. Loewe, and J. Steel eds., Cambridge Univ. Press.

[22] W. Hugh Woodin, The Continuum Hypothesis, the generic multiverse of sets, and the Ω conjecture, in: Set theory, arithmetic, and foundations of mathematics, J. Kennedy and R. Kossak eds., ASL lecture notes in Logic, Cambridge Univ. Press (2011), 13-42.

[23] W. Hugh Woodin, Suitable Extender Models I, Journal of Math Logic, 1-2:101341, 2010.