Our Knowledge of the Mathematical World
Part II: Mathematics

Peter Koellner

(Harvard University)

Introduction

Now that we have discussed the epistemological framework we will discuss the case for new axioms by focusing on specific examples and couching each case in the terms of the epistemological framework.

I will begin by discussing the limitations of one standard approach to making an *intrinsic* (or *a priori*) case for new axioms, namely, the traditional approach based on set-theoretic reflection principles.

I will then turn to *extrinsic* (or *a posteriori*) approaches.

Finally, I will discuss the prospect of bifurcation.

Intrinsic

Recall Gödel's distinction between *intrinsic* and *extrinsic* justifications and his description of the former:

> *[the] axioms of set theory [ZFC] by no means form a system closed in itself, but, quite on the contrary, the very concept of set on which they are based suggests their extension by new axioms which assert the existence of still further iterations of the operation "set of"...*

He mentions large cardinal axioms such as those asserting the existence of inaccessible and Mahlo cardinals.

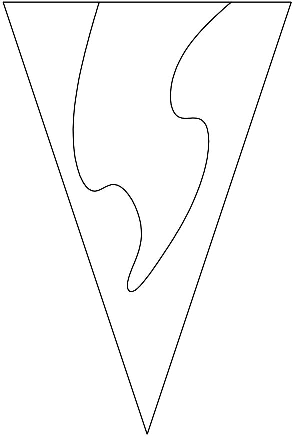
> *... [t]hese axioms show clearly, not only that the axiomatic system of set theory as used today is incomplete, but also that it can be supplemented without arbitrariness by new axioms which only* **unfold the content of the concept of set.** *(Gödel 1964, 260–261).*

Reflection Principles

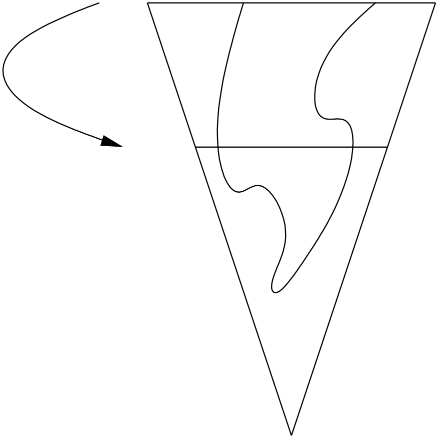
Reflection principles aim to articulate the idea (inherent in the iterative conception of set) that the universe of sets is “absolutely” infinite and hence cannot be “described from below”.

Some have thought that these are the *only* principles that are intrinsically justified on the basis of the iterative conception of set.

> *On the [iterative] conception, the “intrinsic necessity” of an axiom arises from the fact that it expresses that some property possessed by the totality of ordinals is possessed by some ordinal. (Tait, 2001)*



$$V \models \varphi(A)$$



$V \models \varphi(A)$

$V_\alpha \models \varphi^\alpha(A^\alpha)$

Schematically, a reflection principle has the form

$$V \models \varphi(A) \;\rightarrow\; \exists\alpha\; V_\alpha \models \varphi^\alpha(A^\alpha)$$

where $\varphi^\alpha(\,\cdot\,)$ is the result of relativizing the quantifiers of $\varphi(\,\cdot\,)$ to V_α and A^α is the result of relativizing an arbitrary parameter A to V_α.

Tait found the most general classification of such principles:

$\Gamma^{(2)}_1$*-reflection*, ..., $\Gamma^{(2)}_n$*-reflection*, ...
$\Gamma^{(3)}_1$*-reflection*, ..., $\Gamma^{(3)}_n$*-reflection*, ...
...

Theorem (K.)

Assume $\kappa = \kappa(\omega)$ *exists. Then there is a* $\delta < \kappa$ *such that* V_δ *satisfies* $\Gamma_n^{(2)}$*-reflection for all* $n < \omega$.

Theorem (K.)

$\Gamma_1^{(3)}$*-reflection is inconsistent.*

Theorem (K.)

The classification is sharp.

SUMMARY:

Under a very fine calibration generalized reflection principles are either weak (consistent relative to $\kappa(\omega)$) or inconsistent.

Insofar as intrinsic justifications are exhausted by reflection principles, such an approach cannot yield a significant reduction in incompleteness.

Extrinsic

From the above (and related) results it appears that the prospect of finding (for each interpretability degree) statements of high intrinsic plausibility is rather dim. For as we climb the hierarchy it appears that the degree of intrinsic plausibility of the most intrinsically plausible statement diminishes. That was our puzzle.

Is there a way to pick up the slack?

The hope is that we can pick up the slack by leveraging mathematical results to pick. In outline the way this works is as follows: If a statement (which may have a low degree of intrinsic plausibility) implies a great variety of statements with high intrinsic plausibility then that is evidence for that statement. This is the nature of *a posteriori* justification.

There are other ways of accumulating evidence. For example, if an intrinsically plausible statement is shown to imply another statement then the latter statement inherits the evidence of the former. More generally, if at a certain stage of inquiry a statement has a certain degree of evidence, then if that statement is shown to imply another statement then the latter statement inherits the evidence of the former.

Moreover, if we have two statements, each with a degree of evidence, and they are shown to imply one another through a deep series of results, then in addition to the transference of evidence of the one statement to the other statement, the connection itself counts as additional, higher-order evidence.

The case we shall present will have all of these features and more.

Large Cardinals

Let us begin with a preliminary discussion of large cardinals and their inner models.

Large cardinals are a form of *principle of pure strength* which aim to cash out the idea that the universe of sets is "absolutely infinite".

Reflection principles provide us with a template for formulating *small* large cardinal axioms. And elementary embeddings provide us with a template for formulating *large* large cardinal axioms.

How might one justify a large cardinal axioms or, more generally, a principle of pure strength?

To approach this question let us consider the first non-trivial principle of pure strength—the totality of exponentiation.

Suppose you slip up and find yourself face to face with the oracle. You ask: "Is it consistent that exponentiation is total?" The oracle responds: "Yes." And then, after a pause, continues: "However, although it is *consistent* that it is total it is not *in fact* total."

Does this response make sense? Of course, it is consistent. The question is whether it is coherent or tenable.

I don't think that it is. The only thing standing in the way of the *truth* of the totality of exponentiation is its possible inconsistency.

More generally, in the case of principles of pure strength "consistency implies truth".

Of course the principle that "consistency implies truth" cannot apply across the board since in general there are statements which are incompatible each of which is consistent. But the principles of pure strength that we know of are *well-ordered* and there is a cluster of results (on jump operators) which provides good evidence that this situation will persist as long as we focus on genuine principles of pure strength.

If one grants this principle then the case for large cardinal axioms reduces to the problem of providing evidence for their consistency. The case for this is, I believe, one of *a posteriori* justification and involves inner model theory and structure theory. Some of this evidence will appear in the next section.

Definable Determinacy

The primary case for new axioms that I want to consider is the case for *axioms of definable determinacy*. These axioms appear in *descriptive set theory*, a subject that concerns *definable sets of reals*. There is a hierarchy of definability and for each level in this hierarchy there is an associated axiom of definable determinacy. I will focus on three levels in this hierarchy—the Borel sets, the projective sets, and the sets of reals in $L(\mathbb{R})$.

The interest of axioms of definable determinacy is that they have virtually *no* intrinsic plausibility. The case for these axioms is entirely *a posteriori*.

"Verifiable" Consequences

Recall that Gödel spoke of an axiom having 'verifiable consequences'. The analogy is clearly with physics. But what is meant by 'verifiable consequence' in our present setting?

There are two forms of verifiable consequences that I want to consider. The first is where 'verifiable' means 'provable in an accepted theory' (like ZFC)'. The second is where 'verifiable' means 'has a high degree of intrinsic plausibility'.

There are numerous instances of the former. For example:

(1) Martin originally tried to refute definable determinacy by refuting some of its consequences. For example, definable determinacy implies the cone theorem for definable sets. Martin ended up verifying these consequences by proving them in ZFC. This constituted evidence for definable determinacy.
(2) Shortening of proofs: Martin's proof in 1970 of $\utilde{\Delta}^1_1$ determinacy from a measurable is "considerably simpler and easier to discover" than his subsequent proof in 1975 in ZFC.

This was the initial evidence for definable determinacy. But the really convincing case for these axioms lay further down the road.

Structure Theory

Theorem (ZFC)

(1) $\undertilde{\Sigma}^1_1$ *sets are Lebesgue measurable (Luzin, 1917),*
(2) $\undertilde{\Sigma}^1_1$ *sets have the property of Baire (Luzin, 1917), and*
(3) *Every* $\undertilde{\Sigma}^1_2$ *subset of the plane can be uniformized by a* $\undertilde{\Sigma}^1_2$ *set (Kondô, 1937).*

The generalizations of these statements from the Borel sets to sets of reals in $L(\mathbb{R})$ have a certain degree of intrinsically plausibility. However, the statement of definable determinacy for $L(\mathbb{R})$ is not intrinsically plausible at all.

The evidence for the latter lies in the fact that it implies the former statements, and many others that are also intrinsically plausible.

Theorem

Assume $\mathsf{AD}^{L(\mathbb{R})}$*. Then*

(1) *Every set of reals in* $L(\mathbb{R})$ *is Lebesgue measurable.*

(2) *Every set of reals in* $L(\mathbb{R})$ *has the property of Baire.*

(3) *(Martin-Steel)* Σ^2_1*-uniformization holds in* $L(\mathbb{R})$*.*

Concern: There might be other (incompatible) theories sharing this fruitful consequence.

Theorem (Woodin)

Assume

(1) *every set of reals in* $L(\mathbb{R})$ *is Lebesgue measurable,*

(2) *every set of reals in* $L(\mathbb{R})$ *has the property of Baire, and*

(3) Σ^2_1*-uniformization holds in* $L(\mathbb{R})$.

Then $\mathrm{AD}^{L(\mathbb{R})}$ *holds.*

SUMMARY: Evidence for $\mathrm{AD}^{L(\mathbb{R})}$ lies in the fact that it implies many statements that are intrinsically plausible.

The worry arises: Perhaps there are other statements which are incompatible with $\mathrm{AD}^{L(\mathbb{R})}$ which also have these consequences. But this possibility is closed off: One can *recover* $\mathrm{AD}^{L(\mathbb{R})}$ from its consequences. This shows that the evidence (through intrinsic plausibility) for the consequences transfers to $\mathrm{AD}^{L(\mathbb{R})}$.

Moreover, the connections in the recovery result provide additional, higher-order evidence.

Large Cardinals

Large cardinal axioms have a certain degree of intrinsic plausibility and through results in inner model theory there is evidence of their consistency, which, by our earlier considerations, provides further evidence for them. This accumulated evidence transfers to definable determinacy:

Theorem (Martin-Steel)

Assume there are infinitely many Woodin cardinals. Then PD.

Theorem (Woodin)

Assume there are infinitely many Woodin cardinals and a measurable cardinal above them all. Then $\mathrm{AD}^{L(\mathbb{R})}$.

Conversely, definable determinacy implies (inner models) of large cardinals:

Theorem (Woodin)

Assume $\mathrm{AD}^{L(\mathbb{R})}$. *Then there is an inner model* N *of* ZFC + *there are* ω*-many Woodin cardinals.*

SUMMARY: The evidence for inner models of large cardinal axioms transfers to evidence for definable determinacy and the evidence for definable determinacy transfers to evidence for inner models of large cardinal axioms.

Moreover, the connection between the two series of axioms is so deep and involved, that the connection constitutes a form of higher-order evidence.

Generic Absoluteness

Theorem (Shoenfield)

Suppose φ is a Σ^1_2 sentence, $\mathbb{P}$ is a partial order and $G \subseteq \mathbb{P}$ is V-generic. Then

$$V \models \varphi \text{ iff } V[G] \models \varphi.$$

Theorem (Martin-Solovay)

Assume there is a proper class of measurable cardinals. Suppose φ is a Σ^1_3 sentence, $\mathbb{P}$ is a partial order and $G \subseteq \mathbb{P}$ is V-generic. Then

$$V \models \varphi \text{ iff } V[G] \models \varphi.$$

Theorem (Woodin)

Assume there is a proper class of Woodin cardinals. Suppose φ is a sentence, $\mathbb{P}$ is a partial order and $G \subseteq \mathbb{P}$ is V-generic. Then

$$L(\mathbb{R}) \models \varphi \text{ iff } L(\mathbb{R})^{V[G]} \models \varphi.$$

Call a set *absolutely* Δ^2_1 if there are Σ^2_1 formulas which define complementary sets of reals in all generic extensions. Woodin showed that if there is a proper class of measurable Woodin cardinals then all absolutely Δ^2_1 sets of reals are universally Baire (Γ^∞).

Theorem (Woodin)

Suppose there is a proper class of Woodin cardinals and let φ be a sentence of the form

$$\exists A \in \Gamma^{\infty}\ (H(\omega_1), \in, A) \models \psi.$$

Suppose $G \subseteq \mathbb{P}$ is V-generic. Then

$$V \models \varphi \ \text{ iff } \ V[G] \models \varphi.$$

Thus we have generic absoluteness for a large chunk of the existential quantifier in Σ^2_1.

In this sense, in choosing CH as a test case for his program, Gödel put his finger in precisely the point at which it breaks down.

Conversely:

Theorem (Woodin)

Suppose there is a proper class of strongly inaccessible cardinals. Suppose that the theory of $L(\mathbb{R})$ is generically absolute. Then $\mathrm{AD}^{L(\mathbb{R})}$.

SUMMARY: There is a 'good' theory (one that freezes the theory of $L(\mathbb{R})$) and all good theories imply $\mathsf{AD}^{L(\mathbb{R})}$.

Inner Model Theory and the Overlapping Consensus

Definable determinacy is implicated in an even *more* dramatic fashion:

Theorem (Harrington, Martin)

The following are equivalent:

(1) $\underset{\sim}{\Pi}^1_1$*-determinacy.*

(2) *For all* $x \in \mathbb{R}$, $x^{\#}$ *exists.*

Theorem (Woodin)

The following are equivalent:

(1) PD.

(2) *For each $n < \omega$, there is a transitive ω_1-iterable model M such that*

$$M \models \text{“ZFC} + \textit{there exist } n \textit{ Woodin cardinals”}.$$

This generalizes.

The machinery for proving the above "recovery" results (the core model induction) can be used to show that virtually *every* natural mathematical theory of sufficiently strong consistency strength actually implies $\mathrm{AD}^{L(\mathbb{R})}$.

Here are two representative examples:

Theorem (Woodin)

Assume ZFC + *there is an* ω_1 *dense ideal on* ω_1*. Then* $\mathrm{AD}^{L(\mathbb{R})}$*.*

Theorem (Steel)

Assume ZFC + PFA*. Then* $\mathrm{AD}^{L(\mathbb{R})}$*.*

Point: These two theories are incompatible and yet both imply $\mathrm{AD}^{L(\mathbb{R})}$.

SUMMARY: $AD^{L(\mathbb{R})}$ lies in the overlapping consensus of all sufficiently strong natural theories. It inherits whatever evidence has accumulated for these theories. Everything flows toward $\mathrm{AD}^{L(\mathbb{R})}$. Moreover, there is a metamathematical reason for this which provides evidence that it is prevalent throughout the space of possible theories. Furthermore, the very existence of this phenomenon is itself additional, higher-order evidence.

Summary:

(1) ("Verifiable" Consequences). Cone theorem. Shorter and more illuminating proofs.

(2) (Structure Theory) There is a good theory (one that lifts the structure theory that to $L(\mathbb{R})$) and all good theories imply $\mathsf{AD}^{L(\mathbb{R})}$.

(3) (Reflective Equilibrium/Convergence) $\mathsf{AD}^{L(\mathbb{R})}$ is implied by large cardinals and so inherits the considerations in their favour. Conversely, $\mathsf{AD}^{L(\mathbb{R})}$ implies the existence of inner models of large cardinals. Ultimately, $\mathsf{AD}^{L(\mathbb{R})}$ is equivalent to the existence of certain inner models of large cardinals.

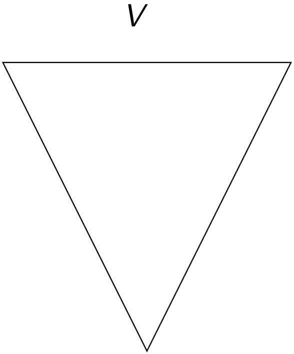
(4) (FREEZING) There is a good theory (one that freezes the theory of $L(\mathbb{R})$) and all good theories imply $\mathsf{AD}^{L(\mathbb{R})}$.

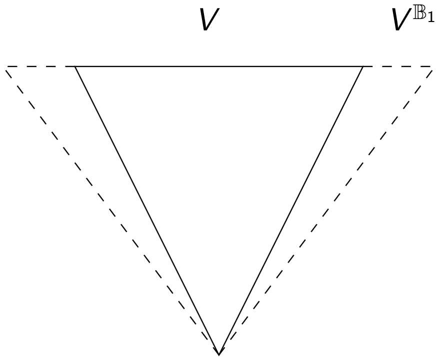
(5) (INEVITABILITY/OVERLAPPING CONSENSUS) $\mathsf{AD}^{L(\mathbb{R})}$ is inevitable in that it lies in the overlapping consensus of all sufficient strong, natural theories. This includes incompatible theories from radically distinct domains.

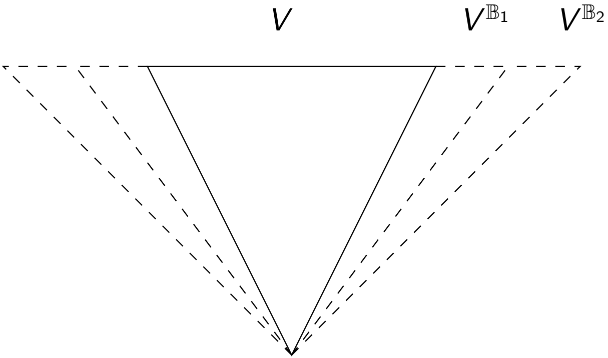
All of this amounts to a compelling extrinsic case for $\mathsf{AD}^{L(\mathbb{R})}$ and a similar case holds for higher forms of definable determinacy.

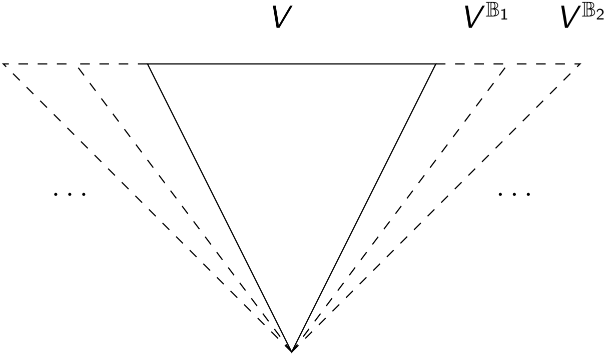
The Prospect of Bifurcation—Ω Complete Pictures

The success of the above case for $\mathrm{AD}^{L(\mathbb{R})}$ can be quantified in terms of a strong logic—Ω-logic.


V


V
$V^{\mathbb{B}_1}$


V
$V^{\mathbb{B}_1}$
$V^{\mathbb{B}_2}$


V
$V^{\mathbb{B}_1}$
$V^{\mathbb{B}_2}$
. . .
. . .

Definition (Woodin)

Suppose that T is a countable theory in the language of set theory and φ is a sentence. Then

$$T \models_\Omega \varphi$$

if for all complete Boolean algebras $\mathbb{B}$ and for all ordinals α, if $V_\alpha^{\mathbb{B}} \models T$ then $V_\alpha^{\mathbb{B}} \models \varphi$.

This logic is "robust":

Theorem (Woodin)

Suppose that T is a countable theory in the language of set theory and φ is a sentence. Suppose that there is a proper class of Woodin cardinals. Then for all complete Boolean algebras $\mathbb{B}$,

$$T \models_\Omega \varphi \text{ iff } V^{\mathbb{B}} \models \text{“}T \models_\Omega \varphi.\text{”}$$

There is a flexible characterization of *generically invariant logic* which "scales up" as one increases the background assumptions to subsume β-logic, Σ^1_2-logic, etc.

Theorem (K.)

Ω-logic is the strongest generically invariant logic.

There is a corresponding "proof relation" (denoted $\vdash_\Omega$) which is also robust in the above sense. The analogue of the Soundness Theorem is known to hold:

Theorem (Woodin)

Suppose T is a countable theory in the language of set theory and φ is a sentence. If $T \vdash_\Omega \varphi$ then $T \models_\Omega \varphi$.

The key open question is the analogue of the Completeness Theorem:

Definition (Ω Conjecture)

Assume there is a proper class of Woodin cardinals. Then for each sentence φ,

$$\varnothing \models_\Omega \varphi$$

if and only if

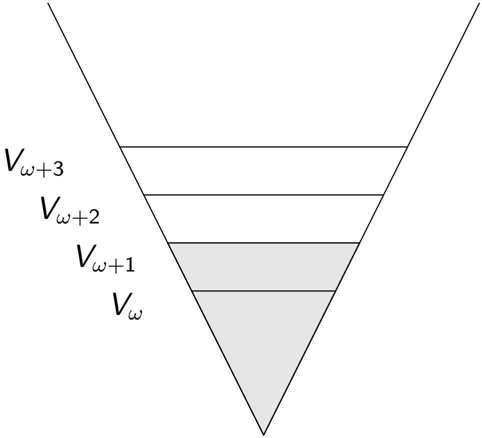
$$\varnothing \vdash_\Omega \varphi.$$

We say that a theory T is Ω-*complete* for a collection of sentences Γ if for each $\varphi \in \Gamma$, either $T \models_\Omega \varphi$ or $T \models_\Omega \neg\varphi$.

In this language we can reformulate the above result as follows:

Theorem (Woodin)

Assume ZFC *and that there is a proper class of Woodin cardinals. Then* ZFC *is* Ω*-complete for statements about second-order arithmetic (and, in fact,* $L(\mathbb{R})$*).*


$V_{\omega+3}$
$V_{\omega+2}$
$V_{\omega+1}$
V_{ω}

Can large cardinals provide a larger Ω-complete picture?

Yes, all the way "up to" CH.

What about "at" the level of CH, namely, Σ^2_1?

Theorem (Levy-Solovay . . .)

Assume L is a standard large cardinal axiom. Then $\mathrm{ZFC} + L$ is not Ω-complete for Σ^2_1.

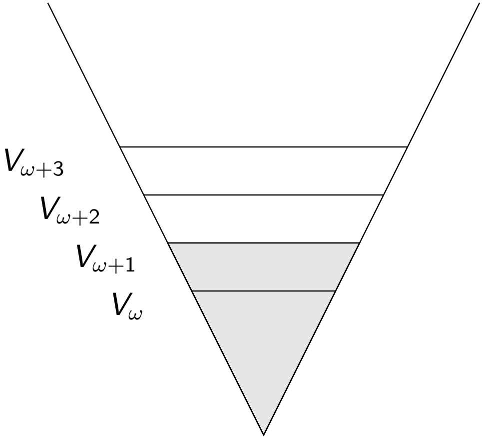
But what if one supplements large cardinal axioms?

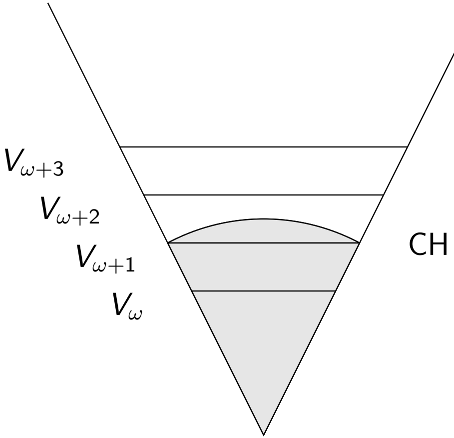
In fact, there is a special sentence: CH.

Theorem (Woodin)

Assume ZFC *and that there is a proper class of measurable Woodin cardinals. Then* $\mathrm{ZFC} + \mathrm{CH}$ *is* Ω*-complete for* Σ^2_1*.*

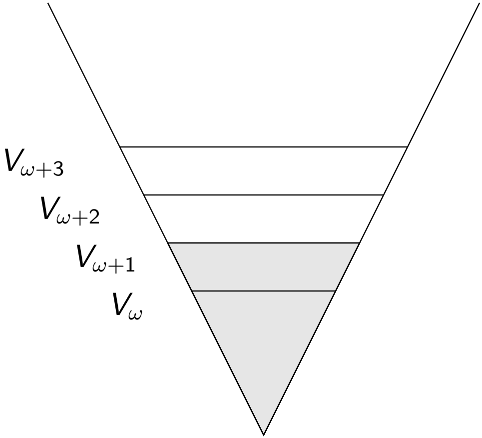
Moreover, up to Ω-equivalence, CH is the unique Σ^2_1-statement that is Ω-complete for Σ^2_1.

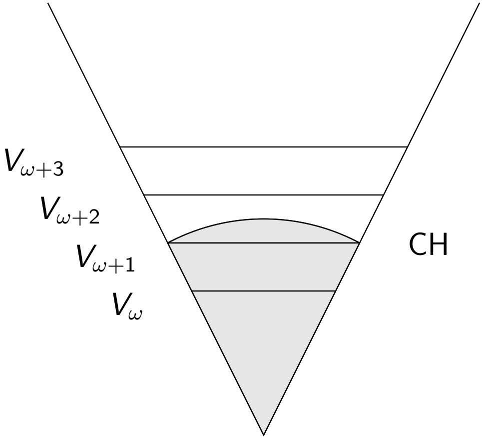

$V_{\omega+3}$
$V_{\omega+2}$
$V_{\omega+1}$
V_{ω}

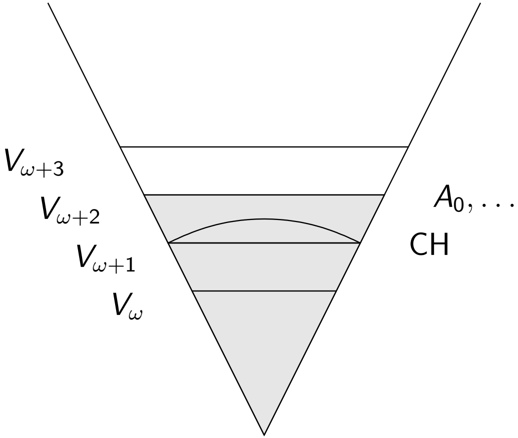

$V_{\omega+3}$
$V_{\omega+2}$
$V_{\omega+1}$
V_{ω}
CH

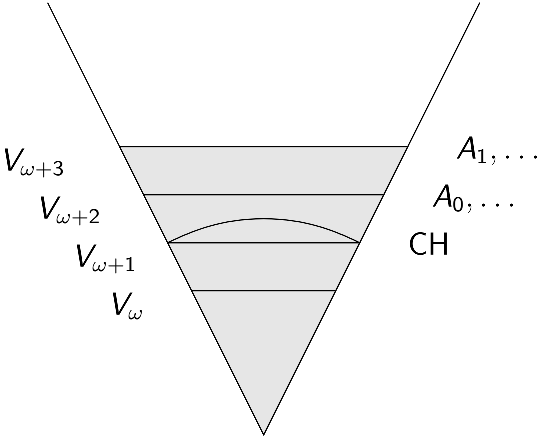
Steel proposed the following scenario:

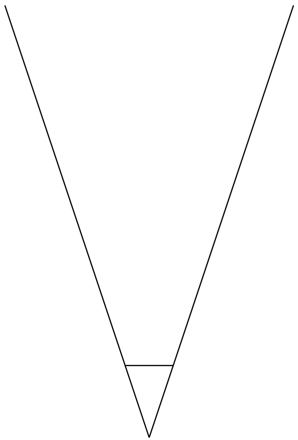
For each specifiable fragment V_λ of the universe there is a recursive set of axioms $\vec{A}$ such that $\mathrm{ZFC} + L + \vec{A}$ is Ω-complete for the theory of V_λ. Moreover, if $\mathrm{ZFC} + L + \vec{B}$ is any sequence with this feature then $\mathrm{ZFC} + \mathrm{LCA} + \vec{A}$ and $\mathrm{ZFC} + L + \vec{B}$ provide the same Ω-picture of V_λ.

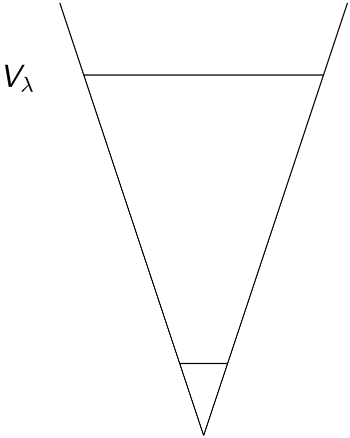

$V_{\omega+3}$
$V_{\omega+2}$
$V_{\omega+1}$
V_ω

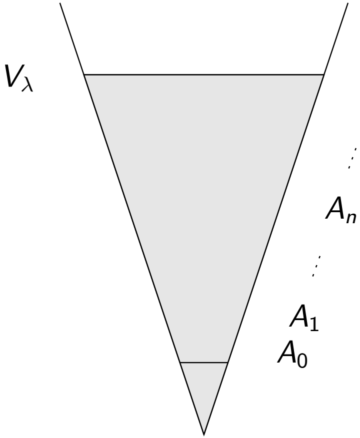

$V_{\omega+3}$
$V_{\omega+2}$
$V_{\omega+1}$
V_{ω}
CH


$V_{\omega+3}$
$V_{\omega+2}$
$V_{\omega+1}$
V_{ω}
$A_0, \ldots$
CH


$V_{\omega+3}$
$V_{\omega+2}$
$V_{\omega+1}$
V_{ω}
$A_1, \ldots$
$A_0, \ldots$
CH




V_λ

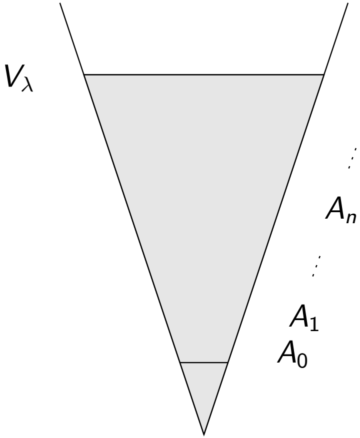

V_λ
A_n
A_1
A_0

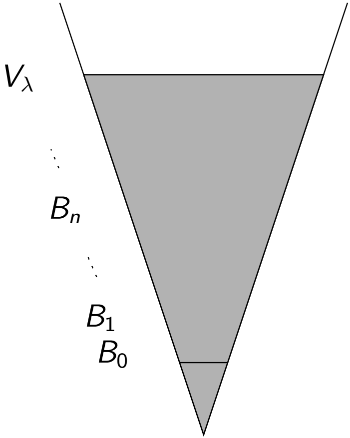
This scenario would lead to a strong case for new axioms, for there would be a *unique* Ω-complete picture of the universe of sets.

Unfortunately, this scenario fails! In particular, *uniqueness* must fail.

Theorem (K. and Woodin)

Assume that there is a proper class of Woodin cardinals. Let V_λ *be a specifiable fragment of the universe. Suppose* $\mathrm{ZFC} + L + \vec{A}$ *is* Ω*-complete for the theory of* V_λ*. Then there exists* $\vec{B}$ *such that* $\mathrm{ZFC} + L + \vec{B}$ *is also* Ω*-complete for the theory of* V_λ *and yet the two theories differ on* CH*.*


V_λ
A_n
A_1
A_0


V_λ
B_n
B_1
B_0

So we are actually led to a *bifurcation* scenario.

To my knowledge this is the most plausible scenario in which one could argue that CH is *absolutely undecidable*.

Virtue: This scenario is sensitive to developments in mathematics.

Key Question: (Strong) Ω Conjecture.

Ultimate-L

In his attempt to prove the (Strong) Ω Conjecture Woodin made a surprising discovery.

There is evidence of an *ultimate inner model*—Ultimate-L—which can accomodate all standard large cardinal axioms and is *close to* V.

If this model exists it would provide a "tidy" picture of the universe. The open-ended theory "$\mathrm{ZFC} + \mathrm{LCA} + V = \text{Ultimate-}L$" would be "effectively complete". In particular, it would imply CH.

Side Structure

But perhaps this picture of the universe is too tidy. Perhaps there is structure "off to the side".

Three Principles:

(1) (Forcing Axioms) $\mathbb{P}_{\mathrm{max}}$.

(2) (Large Cardinal Axioms) There exists a non-trivial elementary embedding

$$j : \mathrm{HOD}_{V_{\lambda+1}} \to \mathrm{HOD}_{V_{\lambda+1}}$$

with critical point less than λ.

(3) (Definable Determinacy) (Steel). Every game on $\mathrm{OD}_{\mathbb{R}}$ wich ends as soon as the play is not $\mathrm{OD}_{\mathbb{R}}$ which has an $\mathrm{OD}_{\mathbb{R}}$ payoff condition, is determined.

FACT: Each of these principles is *incompatible* with Ultimate-L.

Key Question: HOD Conjecture.

Summary

(1) The nature of justification in mathematics is a quite subtle affair—it differs dramatically from the traditional conception.

(2) There is a strong case for axioms that settle questions of complexity "below" that of CH.

(3) We now have precise scenarios for how the future might unfold. These scenarios turn on mathematical conjectures.

So ... Resolve the conjectures!