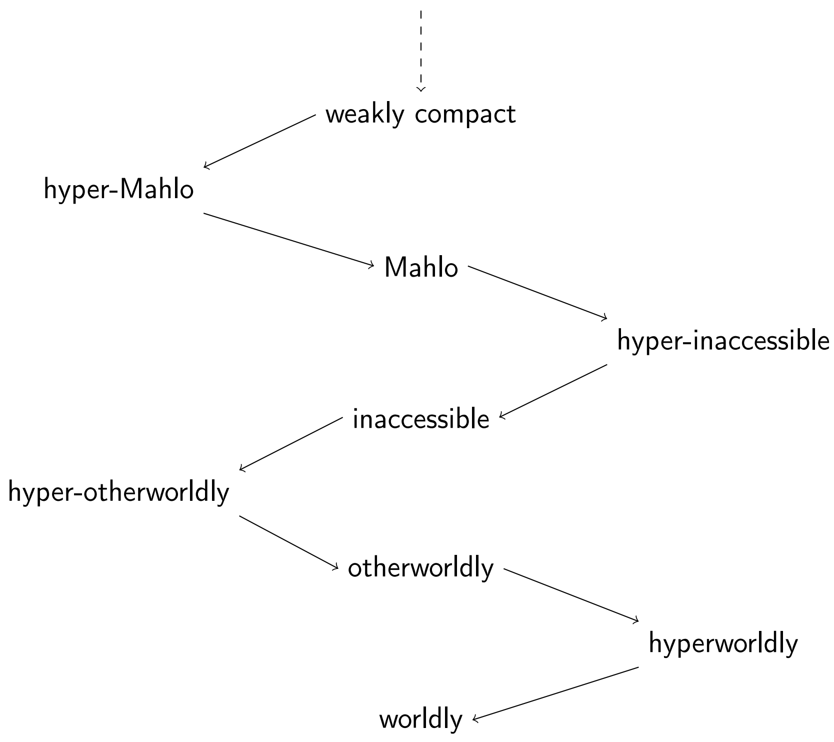
The hyperworldly and hyper-otherworldly cardinals and consistency strength linearity

Word count: *19,545*
Candidate number:
Candidate degree: *MMathPhil in Mathematics and Philosophy, Part C*
Submitted: *April 2022*



PART ∅: INTRODUCTION

This thesis concerns theories extending the 'Zermelo-Fraenkel with Choice', or ZFC, axioms. It will: introduce two such families of theories, which arise in the study of the small end of the large-cardinal hierarchy; place these families relative to each other and other such theories; and then consider more broadly the notion of 'placing' we have used. In particular, it will consider whether, given any two theories, it is the case that one will always place above the other (or at the same level).

In Part I, the two families of theories will be defined using the notions of worldly and otherworldly cardinals, which are large-cardinal concepts arising towards the bottom of any hierarchy commonly used to rank such concepts. These large-cardinal concepts can be extended in a straightforward way to obtain, first, the α-worldly and α-otherworldly cardinals (for α an ordinal), then the hyperworldly and hyper-otherworldly cardinals, then the α-hyperworldly and α-hyper-otherworldly cardinals, and so on.

In order to develop these concepts, we will require some set-theoretic and model-theoretic preliminaries. In addition we will briefly survey Gödel's incompleteness theorems and extend the ideas behind them to set theory. This will be the subject of Section 1. In Section 2, we will then give a development of the worldly and otherworldly cardinals and their hyper-extensions. We will prove some basic results, and relate them to each other and the inaccessible cardinals, the 'next largest' large-cardinal concept. Section 3 will then formally introduce the hierarchy of consistency strength, commonly used in ordering extensions of ZFC, and place our novel large-cardinal concepts within it. Other hierarchies used to order (some) theories extending ZFC will also be considered.

Part II examines the consistency strength hierarchy from a more philosophical perspective. It will primarily be concerned with the claim commonly made by set theorists that the hierarchy is linear in its natural theories. Specifically, that given two natural extensions of ZFC, we expect that one has consistency strength greater than or equal to the other. A central source in this part of the thesis will be Joel David Hamkins' paper 'Nonlinearity in the hierarchy of large-cardinal consistency strength', from which I will draw, expand on, and assess a number of ideas.

Section 4 explains why set theorists only consider 'natural' theories when arguing for consistency strength linearity. Using Gödelian ideas from Section 1, we can construct examples of incomparability in the hierarchy. With this restriction of natural theories noted, Section 5 investigates what we might hope to gain from such a linearity phenomenon, both philosophically and mathematically.

The thesis then shifts to critically consider this question of linearity. Section 6 gives the inductive argument for linearity, and notes that the cardinals we have presented here add to this case. Section 7 then considers two counter-arguments presented by Hamkins in his paper. I will expand these arguments from their original presentation, and examine how they impact linearity as well as our concerns from Section 5. In particular, I will conclude that on the basis of the arguments given, we do not currently have strong reason to believe that the hierarchy of consistency strength is linear in its natural theories.

PART I: THE HYPERWORLDLY AND HYPER-OTHERWORLDLY CARDINALS

1. Mathematical preliminaries

We begin with a number of preliminaries from set theory and logic. Familiarity with some set theory, such as the basic operations on sets, ordinal arithmetic, and basic cardinal arithmetic will be assumed. All the mathematical objects referred to will be sets unless otherwise specified (we know for example that On is not a set by Burali-Forti); we do not assume urelements (though doing so would change little). Throughout this essay proofs will be ended with $\dashv$, and definitions with $\lhd$.

1.1. Basic set theory.

The setting for this paper will be the *von Neumann hierarchy*, originally due to Zermelo in [Zer30]. This is defined as follows:

Definition 1.1. For $\alpha, \lambda \in \text{On}$, λ a limit ordinal (see Definition 1.9).

$$\begin{aligned} V_0 &\coloneqq \varnothing \\ V_{\alpha+1} &\coloneqq \mathcal{P}(V_\alpha) \\ V_\lambda &\coloneqq \bigcup_{\beta<\lambda} V_\beta. \end{aligned}$$

We then informally (since as above On is not a set) identify

$$V \coloneqq \bigcup_{\alpha\in\text{On}} V_\alpha.$$

$\lhd$

V is a proper class: for it to be a set it would have to be in some V_α, however then quickly $V \in V$, a contradiction. We also define the notion of rank:

Definition 1.2. Given a set x, the *rank* of x, written $\operatorname{rank} x$, is the minimal $\alpha \in \text{On}$ such that $x \subseteq V_\alpha$. $\lhd$

The following lemma will be useful in the proof of Theorem 1.10.

Lemma 1.3. *V_α is a transitive set for all $\alpha \in \text{On}$.*

Proof. Recall that a set x is transitive if and only if $z \in y \in x$ implies $z \in x$. We proceed by transfinite induction.

(i) $V_0 = \varnothing$ so this case is trivial.

(ii) If $z \in y \in V_{\alpha+1} = \mathcal{P}(V_\alpha)$, then $y \subseteq V_\alpha$; this means $z \in V_\alpha$. Since V_α is transitive this gives that $z \subseteq V_\alpha$, because any $w \in z$ must also be in V_α. Then we have that $z \in V_{\alpha+1} = \mathcal{P}(V_\alpha)$.

(iii) Let λ be a limit ordinal and suppose that for all $\alpha < \lambda$ we have that V_α is transitive. Then if $z \in y \in V_\lambda = \bigcup_{\alpha<\lambda} V_\alpha$ then $z \in y \in V_\beta$ for some $\beta < \lambda$. Then by hypothesis $z \in V_\beta \subsetneq V_\lambda$ (since V_λ is the union of all V_γ where $\gamma < \lambda$) so we are done. $\dashv$

We will also need this basic result.

Lemma 1.4. *If $\beta < \alpha$, then $x \in V_\beta$ implies that $x \in V_\alpha$.*

Proof. Suppose that $x \in V_\beta$ for some $\beta < \alpha$. We proceed by induction on α.

(i) There are no $\beta < 0$ so the base case is trivial.
(ii) If $\beta < \alpha + 1$ then either $\beta = \alpha$ or $\beta < \alpha$. By induction we may assume without loss of generality that $\beta = \alpha$ (if $\beta < \alpha$ then by hypothesis $x \in V_\alpha$). If $x \in V_\alpha$, then $x \subsetneq V_\alpha$ by transitivity (Lemma 1.3) and thus is an element of $V_{\alpha+1} = \mathcal{P}(V_\alpha)$.
(iii) Since, for λ a limit ordinal, $V_\lambda = \bigcup_{\beta<\lambda} V_\beta$, this case is immediate. ⊣

Finally, the following two results will be useful in the proof of Theorem 3.6.

Lemma 1.5. *For all $\alpha \in \mathrm{On}$, we have $\alpha \notin V_\alpha$.*

Proof. We proceed by transfinite induction.

(i) $V_0 = \varnothing$, so the base case is trivial.
(ii) Let $\alpha = \beta + 1$, where by induction $\beta \notin V_\beta$. Suppose for contradiction that $\alpha \in V_\alpha$, so $\beta + 1 = \beta \cup \{\beta\} \in V_{\beta+1} = \mathcal{P}(V_\beta)$. Then $\beta \cup \{\beta\} \subseteq V_\beta$, and thus $\{\beta\} \subseteq V_\beta$. This implies that $\beta \in V_\beta$, a contradiction.
(iii) Now suppose that λ is a limit ordinal, $\beta \notin V_\beta$ for all $\beta < \lambda$, and $\lambda \in V_\lambda$, for contradiction. Then since $V_\lambda = \bigcup_{\beta<\lambda} V_\beta$, we have that $\lambda \in V_\delta$ for some $\delta < \lambda$. Since $\delta \in \lambda$, by transitivity of V_δ (Lemma 1.3) we then get $\delta \in V_\delta$, a contradiction. ⊣

Lemma 1.6. *For all $\alpha \in \mathrm{On}$, $\alpha \subseteq V_\alpha$.*

Proof. We proceed by transfinite induction.

(i) $\varnothing \subseteq x$ for all $x \in V$, so the base case is trivial.
(ii) Let $\alpha = \beta + 1 = \beta \cup \{\beta\}$, where by induction $\beta \subseteq V_\beta$. Then $\beta \in V_{\beta+1} = \mathcal{P}(V_\beta)$, and thus $\{\beta\} \subseteq V_{\beta+1}$. Also by the transitivity of V_β, $\beta \subseteq V_{\beta+1}$. Therefore $\beta \cup \{\beta\} \subseteq V_{\beta+1}$ as required.
(iii) Let λ be a limit ordinal and suppose $\beta \subseteq V_\beta$ for all $\beta < \lambda$. Then if $\delta \in \lambda$, then $\delta \subseteq V_\delta$. Then since λ is a limit ordinal, $\delta + 1 < \lambda$, and so $\delta \in \mathcal{P}(V_\delta) = V_{\delta+1} \subseteq V_\lambda$. Thus $\lambda \subseteq V_\lambda$ as required. ⊣

Note that these are sufficient to show

Proposition 1.7. *For all $\alpha \in \mathrm{On}$, $\operatorname{rank} \alpha = \alpha$.*

Proof. Lemma 1.6 shows that $\operatorname{rank} \alpha \leqslant \alpha$. On the other hand, if $\operatorname{rank} \alpha < \alpha$, then $\alpha \subseteq V_\beta$ for some $\beta < \alpha$, and thus $\alpha \in V_{\beta+1} \subseteq V_\alpha$, which contradicts Lemma 1.5. ⊣

A brief note ought to be made on the use of ordinals and cardinals as indexing subscripts (for example V_κ for κ some large cardinal, as we will see below). On a modern view, we identify cardinalities with sets in a unique way via the *von Neumann cardinal assignment* (an excellent discussion of the historical context surrounding this is given in [Mos05, ch12]). This identifies the cardinality of a well-orderable set (which all sets are by AC) with the smallest ordinal equinumerous with it. For example we identify $\aleph_0 = \omega$, since the latter is the smallest infinite countable ordinal. In general the aleph numbers have the following definition:

Definition 1.8. Let α be any ordinal, and λ be a limit ordinal.

$$\begin{aligned}\aleph_0 &:= \omega \\ \aleph_{\alpha+1} &:= \aleph_\alpha^+ \\ \aleph_\lambda &:= \bigcup_{\beta<\lambda} \aleph_\beta.\end{aligned}$$

◁

In the presence of choice, all infinite cardinal numbers are of this form for some $\alpha \in \mathrm{On}$. Thus we see that all cardinals are also ordinals, and hence it is legitimate to subscript by the former. With this potential confusion between ordinals and cardinals in mind, we will use the term 'limit' in various ways in this essay; when not clear from context the different usages will be qualified.

Definition 1.9.

(i) An ordinal $\lambda > 0$ is a *limit* if and only if it is not an ordinal successor.
(ii) A cardinal $\kappa > 0$ is a *weak limit* if and only if it is not a cardinal successor. It is a *strong limit* if and only if for any $\delta < \kappa$, we have $2^\delta < \kappa$ (where 2^δ indicates cardinal exponentiation). If a cardinal is described as simply a 'limit', this will mean 'weak limit'. ◁

Note that in the presence of GCH (see Subsection 5.2), which says that $\kappa^+ = 2^\kappa$ for all $\kappa \geqslant \aleph_0$, the concepts of strong and weak limit coincide. Also note that all infinite cardinals are limit ordinals (since for any infinite ordinal α, we have a bijection $\alpha \to \alpha + 1$, so a cardinal can never be a successor ordinal). Another piece of terminology we will use which could be confused with the above is to say that κ is a limit of a collection of cardinals $\{\delta_i < \kappa \mid i \in I\}$. What this means is that for all $\gamma < \kappa$, there is a $j \in I$ with $\gamma < \delta_j < \kappa$. Of course in this case κ must be a (weak) limit cardinal (else take γ to be the cardinal immediately preceding κ).

A related cause of confusion could be the conflation of orderings of cardinals and ordinals: it is possible to have $\alpha < \beta$ as ordinals, however for α and β to have the same cardinality. Since we are using cardinals in the place of ordinals in some places, this could lead to ambiguity. It should almost always be obvious from context which ordering is meant, however where it isn't this will be clarified.

The following result will be useful; we also use it to establish the form in which we will assume the axioms of ZFC (we follow [Jec03, p3]).

Theorem 1.10.

(i) *For every $\alpha \in \mathrm{On}$, the structure $\langle V_\alpha, \in \rangle$ satisfies extensionality, foundation, union, choice, and separation.*
(ii) *For α any limit ordinal, $\langle V_\alpha, \in \rangle$ satisfies all the axioms in (i) as well as powerset and pairs.*
(iii) *For α a limit ordinal strictly greater than ω, $\langle V_\alpha, \in \rangle$ satisfies all the axioms in (i) and (ii), as well as infinity.*

Note that the relevant model-theoretic notions here are defined in Subsection 1.2. We do not consider replacement here, for reasons which will be explained in Subsection 1.3. Before we proceed with the proof, it is worth being clear about exactly what we are aiming to show and how. There are then two main methods of proof. The first method is to show that an axiom holds true of every element of the relevant level of the von Neumann hierarchy. For example for extensionality, we must show that it holds for every choice of $x, y \in V_\alpha$ where $V_\alpha \vDash x = y$. The second applies when an axiom posits the existence of a new set; we must then show that this set is in the relevant level of the hierarchy. For example to prove that choice holds at V_α, we must show that $\mathcal{C} \in V_\alpha$, where $\mathcal{C}$ is the choice set posited by the axiom.

Proof. We consider each of the axioms in turn.

Extensionality says that if x and y have the same elements, then $x = y$. Suppose that for some $\alpha \in \mathrm{On}$, V_α 'thinks' that x and y have the same elements. We show that in fact they do have the same elements, and thus $x = y$. This is almost immediate from transitivity of V_α, shown in Lemma 1.3: if $z \in x$, then since $x \in V_\alpha$ we have that $z \in V_\alpha$ and thus V_α 'thinks' that z is in y also, by hypothesis. Since our argument is symmetrical in x and y, we get $x = y$ as required.

Union says that for any x there exists a set $y = \bigcup x$, containing as its elements all and only elements of sets in x. We must show that if $x \in V_\alpha$ for some $\alpha \in \mathrm{On}$, then y, as given by union in V, is also in V_α. We proceed by induction.

(1) The base case is trivial as V_0 is empty.
(2) If $\alpha = \beta + 1$, then $x \in V_\alpha \to x \subseteq V_\beta$. In particular, any $z \in w \in x$ will be in V_{β_z} for $\beta_z < \beta$. Thus the collection of all such z – which is our y – must be a subset of $\sup_{z \in w \in x} V_{\beta_z} \subseteq V_\beta$, and thus an element of $\mathcal{P}(V_\beta) = V_\alpha$, as required.
(3) On the other hand if λ is a limit ordinal, then if $x \in V_\lambda$, then $x \in V_\beta$ for some $\beta < \lambda$, and then we are done by induction.

Foundation asserts that every non-empty set has an $\in$-minimal element; we require a transfinite induction to prove this.

(1) The base case is trivial as V_0 is empty.
(2) Suppose that the result holds for all $x \in V_\alpha$, then if $x \in V_{\alpha+1}$ were non-well-founded, then it would contain a non-well-founded element y witnessing this. However then $y \in V_\alpha$, which contradicts that V_α is well-founded. Note that this makes use of transitivity, since any element of y (and any element of an element of y, and so on) must also be in V_α, and thus V_α must see that y is non-well-founded.
(3) Now suppose that the result holds for all $\beta < \lambda$ for λ a limit ordinal. Any element of V_λ must be an element of V_α for some $\alpha < \lambda$, so the result is immediate.

Choice says that for any non-empty set x, there exists a 'choice set' $\mathcal{C}$ such that $|w \cap \mathcal{C}| = 1$ for all non-empty $w \in x$. We need to show that the choice set $\mathcal{C}$ for a given set $x \in V_\alpha$ is in V_α also. We proceed by induction.

(1) The base case is trivial as V_0 is empty.
(2) If $\alpha = \beta + 1$ and $x \in V_\alpha$, then as with union we note that elements z of sets in α will all come from V_β, and thus $\mathcal{C} \subseteq V_\beta$, so $\mathcal{C} \in V_\alpha$, as required.
(3) If λ is a limit and $x \in V_\lambda$, then $x \in V_\beta$ for some $\beta < \lambda$. The result is then immediate by induction.

Separation is an axiom schema which says for all $\varphi = \varphi(x_0, \ldots, x_n)$, a formula in $\mathcal{L}_{\mathrm{st}}$, and for any x and $p_0, \ldots, p_{n-1}$, there exists a set $y = \{u \in x \mid \varphi(u, p_0, \ldots, p_{n-1})\}$. We require that if $x \in V_\alpha$, then so is y. Clearly $y \subseteq x$; we show by induction that this implies that $y \in V_\alpha$.

(1) The base case is trivial as V_0 is empty.
(2) If $\alpha = \beta + 1$, then $x \subseteq V_\beta$, then since $\subseteq$ is a transitive relation, $y \subseteq V_\beta$. Thus $y \in V_\alpha$.
(3) If λ is a limit ordinal, then $x \in V_\lambda$ implies that $x \in V_\beta$ for some $\beta < \lambda$, then by induction $y \in V_\beta$ and we are done by Lemma 1.4.

Powerset says for any x there exists a set $y = \mathcal{P}(x)$, containing as its elements all and only the subsets of x. If α is not a limit ordinal (the case where $\alpha = 0$ is trivial), say $\alpha = \beta + 1$, then any $x \in V_{\beta+1} \setminus V_\beta$ will have $\mathcal{P}(x) \notin V_{\beta+1}$, else $x \in \mathcal{P}(x) \in V_{\beta+1} = \mathcal{P}(V_\beta)$ implies $x \in V_\beta$, which by hypothesis it isn't. Thus powerset doesn't hold at successor ordinals. On the other hand, if $x \in V_\lambda$ for λ a limit, we have that $x \in V_\alpha$ for some $\alpha < \lambda$. If $w \in \mathcal{P}(x)$, then $w \subseteq x$, so $w \subseteq V_\alpha$. Thus $w \in \mathcal{P}(V_\alpha) = V_{\alpha+1}$, and so $\mathcal{P}(x) \subseteq V_{\alpha+1}$. This gives $\mathcal{P}(x) \in V_{\alpha+2} \subseteq V_\lambda$ (since λ is a limit ordinal), so we are done.

Similarly for pairs, which asserts that for any a, b, there exists a set $\{a, b\}$ which contains exactly a and b: if $x, y \in V_{\beta+1} \setminus V_\beta$ (where as above we have let $\alpha = \beta + 1$), then if $\{x, y\} \in V_{\beta+1}$, we would have that $x, y \in V_\beta$, a contradiction. On the other hand if λ is a limit, $x, y \in V_\alpha$ for $\alpha < \lambda$, then $\{x, y\} \in \mathcal{P}(V_\alpha)$, and thus is an element of V_λ.

The axiom of infinity asserts that there exists a set Ω with the property that $\varnothing \in \Omega$, and for any $x \in \Omega$, we have $x \cup \{x\} \in \Omega$ (where $x \cup y := \bigcup\{x, y\}$). We recall that $\omega = \{0, 1, 2, \ldots\}$, where each n is identified with its von Neumann ordinal as standard, is the unique successor set contained (as a subset) in all other successor sets. Now we note that by construction V_ω is the set of all hereditarily finite sets (finite sets such that all of their elements are finite, and all of their elements' elements are finite, and so on): this follows quickly from an induction. In particular then, it can't contain ω, since this is an infinite set. On the other hand, $\omega \in V_{\omega+1}$, since $n \in V_\omega$ for all $n \in \omega$, thus the set $\{0, 1, \ldots\}$ must be in the powerset. Further, we then get by Lemma 1.4 that $\omega \in V_\alpha$ for $\alpha \geqslant \omega + 1$. ⊣

Note that at the lower levels, different formulations of the axioms can become non-equivalent: for example AC has another statement asserting the existence of a function $f : x \to \bigcup x$ such that for all $y \in x$, $f(y) \in y$. This however requires the construction of functions, which in turn uses the Cartesian product, which relies on the axiom of pairing, which we've established only works at limit ordinals.

A further definition which will be used in Subsection 2.3 is the following:

Definition 1.11. Given a partially ordered set $(x, \leqslant)$, we say that $y \subseteq x$ is *cofinal* in x if and only if for any $a \in x$, there is $b \in y$ with $a \leqslant b$. Noting that the ordinals are partially (in fact totally) ordered by $\subseteq$, we can then define the *cofinality* of an ordinal α, $\mathrm{cf}\,\alpha$, as the least ordinal such that there is a sequence of length $\mathrm{cf}\,\alpha$ which is cofinal in α. An infinite cardinal κ with $\mathrm{cf}\,\kappa = \kappa$ is called *regular*, else it is *singular*. ◁

Example 1.12.

(i) ω is regular, since for any $\alpha < \omega$ (i.e. α finite) a sequence of length α has a finite supremum.
(ii) $\mathrm{cf}\,(\omega + 1) = 1$, since $\{\omega\}$ is a sequence cofinal in $\omega + 1$. This result clearly generalises to all successor ordinals.

(iii) $\operatorname{cf} \aleph_\omega = \omega$, since the sequence $\{\aleph_n \mid n \in \omega\}$ is cofinal in $\aleph_\omega$, and could not be any shorter (for then it would be finite).

The following will be needed to define the Mahlo cardinals in Subsection 7.2.1:

Definition 1.13. Given an ordinal κ, a subset $C \subseteq \kappa$ is *closed unbounded*, or *club*, *in* κ if and only if (i) for every $\alpha \in \kappa$, there is $\beta \in C$ with $\alpha < \beta$ (unbounded), and (ii) if $B \subseteq C$ is such that every $\beta \in B$ is less than a uniform $\alpha \in \kappa$ (i.e. B is bounded in κ), then $\sup B \in \kappa$ (closed). $\lhd$

Definition 1.14. Given an ordinal κ, $S \subseteq \kappa$ is *stationary* if and only if $S \cap C \neq \varnothing$ for all club $C \subseteq \kappa$. $\lhd$

1.2. **Basic model theory.**

Definition 1.15. Given a first-order language $\mathcal{L}$ with predicate symbols $\{P_i\}_{i\in I}$, function symbols $\{f_j\}_{j\in J}$, and constants $\{c_k\}_{k\in K}$, an $\mathcal{L}$-*structure* is an object of the form

$$\mathcal{M} = \left\langle M, \left\{P_i^{\mathcal{M}}\right\}_{i\in I}, \left\{f_j^{\mathcal{M}}\right\}_{j\in J}, \left\{c_k^{\mathcal{M}}\right\}_{k\in K} \right\rangle,$$

where $P_i^{\mathcal{M}}$ is an assignment of the n-ary predicate symbol P_i to a subset of M^n, $f_j^{\mathcal{M}}$ is an assignment of the m-ary function symbol f_j to a function $M^m \to M$, and $c_k^{\mathcal{M}}$ is an assignment of c_k to any element of M. $\lhd$

We assume familiarity with the definitions of interpretations and assignments. Note that we will only be working with models of the form $\langle V_\kappa, \in^{V_\kappa}\rangle$; when unambiguous we will abbreviate these models to simply V_κ.

Given two models $\mathcal{M}$, $\mathcal{N}$, with universes M, N, respectively, we define the following:

Definition 1.16. An *embedding* $\mathcal{M} \to \mathcal{N}$ is an injective function $\pi\colon M \to N$ such that the interpretations of the predicate, function, and constant symbols are respected; i.e. for all n, m, for all n-ary predicate symbols P, for all m-ary function symbols f, and for all constant symbols c:

(i) $(a_0, \dots, a_{n-1}) \in P^{\mathcal{M}}$ if and only if $(\pi(a_0), \dots, \pi(a_{n-1})) \in P^{\mathcal{N}}$,
(ii) $\pi(f^{\mathcal{M}}(a_0, \dots, a_{m-1})) = f^{\mathcal{N}}(\pi(a_0), \dots, \pi(a_{m-1}))$,
(iii) $\pi(c^{\mathcal{M}}) = c^{\mathcal{N}}$. $\lhd$

In the above case, we abuse notation slightly and write $\pi\colon \mathcal{M} \to \mathcal{N}$. With a simple induction, one can show that an embedding in fact must respect all atomic $\mathcal{L}$-formulae. There are a number of special cases.

Definition 1.17. Given an embedding $\pi\colon \mathcal{M} \to \mathcal{N}$:

(i) if π is a bijection $M \to N$, then we say $\mathcal{M}$ and $\mathcal{N}$ are *isomorphic*, and write $\mathcal{M} \cong \mathcal{N}$;
(ii) if π is the inclusion map (i.e. if $M \subseteq N$), then we say that $\mathcal{M}$ is a *substructure* of $\mathcal{N}$, written $\mathcal{M} \leqslant \mathcal{N}$;
(iii) if π in fact preserves *all* $\mathcal{L}$-formulae, so that $\mathcal{M} \vDash \varphi(a_0, \dots, a_n)$ if and only if $\mathcal{N} \vDash \varphi(\pi(a_0), \dots, \pi(a_n))$, we call π an *elementary embedding.* If our elementary map is also an inclusion, then we say that $\mathcal{M}$ is an *elementary*

substructure of $\mathcal{N}$ and write $\mathcal{M} \preccurlyeq N$. If $\mathcal{M} \preccurlyeq N$ and $M \neq N$ we write $\mathcal{M} \prec N$. ◁

Example 1.18.

(i) $\langle \mathbb{Z}, < \rangle$ and $\langle 2\mathbb{Z}, < \rangle$ are both models of the theory of linear orderings with no endpoints. There is an isomorphism $\pi\colon \mathbb{Z} \to 2\mathbb{Z}$ given by 'fanning out from 0', $\pi(m) = 2m$, thus these models are isomorphic.

(ii) It is easy to see that $V_\kappa \leqslant V_\lambda$ for any $\kappa \leqslant \lambda$; this will not in general be *elementary*. For a simple example note that when $\varphi = \varphi(x, y)$ is the sentence asserting the existence of two distinct objects, i.e. $\exists x\, \exists y\, \neg x = y$, we have that $V_1 \nvDash \varphi$, whilst $V_2 \vDash \varphi$; thus whilst we have $V_1 \leqslant V_2$, we do not have $V_1 \preccurlyeq V_2$. We can extend this further: by using the same formulae expressing 'there exist 2^n distinct objects', we can readily see $V_m \not\preccurlyeq V_n$ for $m < n$ natural numbers. Moreover we can consider infinite cardinals: $V_\omega \not\preccurlyeq V_\kappa$ for any $\kappa > \omega$ (the 'greater than' here denoting the standard ordering of the ordinals), since $V_\omega \nvDash$ 'there is an infinite set', whilst $\omega \in V_\kappa$ for $\kappa > \omega$, so $V_\kappa \vDash$ 'there is an infinite set'. If $V_\kappa \prec V_\lambda$ then we must have that κ is a limit cardinal: if $\kappa = \delta^+$ then V_κ thinks that δ is the largest cardinal, whereas $\kappa \in V_\lambda$ for any $\lambda > \kappa$ (by Lemma 1.4), so V_λ will not agree. We will return to such concerns in Subsection 2.2.

The following classic theorem will also be useful (for proof see [CK90, ch2]).

Theorem 1.19 (Downwards Löwenheim-Skolem)**.** *If a set of first-order sentences Σ has a model of cardinality κ, then it has a model of size λ for all $|\mathcal{L}| \leqslant \lambda \leqslant \kappa$, where $|\mathcal{L}|$ indicates the size of the set of all the formulae of the language.*

Corollary 1.20 (Skolem's paradox)**.** *If* ZFC *has a model, then it has a countable model.*

Proof. $\mathcal{L}_{\text{st}}$ is countable, thus if ZFC has a model, by Theorem 1.19 we are done. ⊣

Skolem's result was called a paradox because the countable model obtained in Corollary 1.20 allegedly contains uncountably many sets (this is a theorem of ZFC). Really what's happening here is that the concept of 'uncountable' means different things to different models – the 'uncountable sets' in our countable model would appear 'from the outside' (i.e. from a larger superstructure) to be countable. What this shows is that the property of being countable is not *absolute* – it varies between models. We will take care below to note when we are using that a property is absolute.

The following result on absoluteness will be essential in several proofs below. The proof is beyond the author, however see [Kun14, II.4].

Proposition 1.21. *The notion '$V_\kappa \vDash \varphi$' is absolute between V and $V_\lambda \ni V_\kappa$ which satisfies* ZFC*.*

1.3. Gödel's Incompleteness Theorems.

Gödel's theorems are of central importance in almost all studies in the foundations of mathematics. In particular, in set theory they limit our ability to prove the consistency of our theories, and thus give rise to the consistency strength hierarchy which will be introduced in Section 3. The results as initially developed were about Peano Arithmetic (PA) and its extensions: Gödel noticed that, given a certain amount of arithmetic, one could 'translate' the formal language of arithmetic into statements about natural numbers, a process called *arithmetisation*. In particular, we are able to arithmetise the notions of provability and consistency, and use this to place limits on what we are able to show within PA. We will give here a sketch of how these ideas may be extended and applied in set theory. For a full presentation for PA see [Smi13].

Perhaps the key idea of arithmetisation is the use of *Gödel numbering*, which is a process by which one can uniquely (and thus reversibly) associate each string in the language of arithmetic with a natural number. It is well known that set theory may be used as a foundation for all known mathematics, in that every assertion of mathematics has a translation into the language of set theory, $\mathcal{L}_{\text{st}}$. With this in mind, a development of Gödel's ideas many be given for ZFC, where instead of natural numbers, we associate formulae in the language of set theory with sets, in a unique way. This allows us to extend Gödel's results to ZFC. The analogue of Gödel numbers are then *Gödel sets*, which are defined as follows. The presentation follows [Dra74, p90].

Definition 1.22. Given a formula φ, its Gödel set will be denoted $\ulcorner\varphi\urcorner$, and is defined recursively. For atomic formulae with variables v_i, v_j:

(i) $\ulcorner v_i = v_j \urcorner = \langle 0, i, j\rangle$.
(ii) $\ulcorner v_i \in v_j \urcorner = \langle 1, i, j\rangle$.

For non-atomic formulae:

(i) $\ulcorner \varphi \vee \psi \urcorner = \langle 2, \ulcorner\varphi\urcorner, \ulcorner\psi\urcorner\rangle$.
(ii) $\ulcorner \neg\varphi \urcorner = \langle 3, \ulcorner\psi\urcorner\rangle$.
(iii) $\ulcorner \exists v_i \varphi \urcorner = \langle 4, i, \ulcorner\varphi\urcorner\rangle$. ◁

Where as usual the ordered pair $\langle x, y\rangle$ is defined to be the Kuratowski pair $\{\{x\}, \{x, y\}\}$, and then inductively $\langle x, y, z\rangle = \langle x, \langle y, z\rangle\rangle$. By defining $\ulcorner\#\urcorner$ to be $\{\{\varnothing\}\}$, we may then extend this definition to include sequences of formulae:

Definition 1.23. If $\varphi_0, \dots, \varphi_n$ are formulae in $\mathcal{L}_{\text{st}}$, then the Gödel set representing the sequence $(\varphi_0, \dots, \varphi_n)$ is $\langle\ulcorner\#\urcorner, \ulcorner\varphi_0\urcorner, \ulcorner\#\urcorner, \dots, \ulcorner\#\urcorner, \ulcorner\varphi_n\urcorner, \ulcorner\#\urcorner\rangle$. ◁

Note we must distinguish between sets as referred to in the metalanguage, and as referred to in $\mathcal{L}_{\text{st}}$; the latter will be disambiguated by writing them with an overline: $\overline{x}$. Thus if a formula $F(v_0)$ takes as an argument the set represented in $\mathcal{L}_{\text{st}}$ by the Gödel set $\ulcorner x\urcorner$, this will be written $F(\overline{\ulcorner x\urcorner})$.

Once we have this duality between a certain subcollection of sets and formulae of $\mathcal{L}_{\text{st}}$, we are able to make $\mathcal{L}_{\text{st}}$ 'talk about itself' by writing down a formula $\varphi(\overline{x})$ which is true in V if and only if x is the Gödel set of some formula. With this φ, we can then write down a formula $\psi_S(\overline{x})$ which is true if and only if the formula with Gödel set x is a logical axiom of the system S. Whilst we will not do either of these constructions rigorously, it is intuitive that it should be possible, as both our axiom system and language were defined recursively. With these, we can write down a formula $\text{proof}_S(\overline{x}, \overline{y})$, which asserts that x is the Gödel set representing a proof of the formula with Gödel set y from axioms S, and then a formula $\text{Pr}_S(\overline{y}) = \exists x\, \text{proof}_S(x, \overline{y})$, which asserts that there is a proof of the formula with Gödel set y from axioms S. As above we will not go through these constructions in full detail, however it should be fairly intuitive from the above remarks that it is possible. For a more complete presentation (for PA), see [Kni21, pp8–12].

A key result is then the *Diagonal Lemma*:

Theorem 1.24. *For any formula $F(v_0)$ in the language of set theory, there is a formula C such that*

$$\mathsf{ZFC} \vdash C \leftrightarrow F\left(\ulcorner C \urcorner\right).$$

Which has as an important corollary Tarski's theorem on the undefinability of truth:

Theorem 1.25. *There is no formula* True *in $\mathcal{L}_{\mathrm{st}}$ such that $V \vDash \mathrm{True}\left(\ulcorner \varphi \urcorner\right)$ if and only if $V \vDash \varphi$.*

Proof. Let $F(v_0) = \neg\mathrm{True}(v_0)$. Then by the Diagonal Lemma there is a formula C in the language of set theory such that $\mathsf{ZFC} \vdash C \leftrightarrow \neg\mathrm{True}\left(\ulcorner C \urcorner\right)$. But then $V \vDash C$ if and only if $V \vDash \neg\mathrm{True}\left(\ulcorner C \urcorner\right)$ if and only if $V \vDash \neg C$, a contradiction. ⊣

Finally note we may define the formula $\mathrm{Con}\, S$ to be $\neg\,\mathrm{Pr}_S\left(\ulcorner 0 = 1 \urcorner\right)$. We are now in a position to state Gödel's incompleteness theorems. For simplicity (to avoid having to introduce the notions of n- and ω-consistency), and since it will be useful to us later, we will state the first theorem as generalised by Rosser in 1936.

Theorem 1.26. *If T is a theory extending* ZFC*, then there is a sentence R_T such that if T is consistent, then $T \nvdash \mathrm{R}_T$, and $T \nvdash \neg\mathrm{R}_T$.*

The sentence R_T is known as the *Rosser sentence* for T, and asserts that if R_T itself is provable, then there is a shorter proof of $\neg\mathrm{R}_T$. Here 'shorter' means coming earlier in the lexicographical ordering defined on the Gödel sets, where we must extend the usual ordering $\leqslant$ on On and define $\ulcorner \# \urcorner \leqslant i$ for all $i \in \omega$. This order defines $\langle \alpha_0, \ldots, \alpha_n \rangle \leqslant_{\mathrm{lex}} \langle \beta_0, \ldots, \beta_m \rangle$ if and only if $(\alpha_0 \leqslant \beta_0)$, or $(\alpha_0 = \beta_0$ and $\alpha_1 \leqslant \beta_1)$, and so on.

Following on from the first, Gödel's second incompleteness theorem relates to proofs of consistency.

Theorem 1.27. *If T is a consistent theory extending* ZFC*, then $T \nvdash \mathrm{Con}\, T$.*

Corollary 1.28. *If* ZFC *is consistent, then* $\mathsf{ZFC} \nvdash \mathrm{Con}\,\mathsf{ZFC}$*.*

This result is crucial to the properties of large cardinals, to be examined below. With this result stated, we can now see why we excluded the axiom of replacement in Theorem 1.10.

Theorem 1.29. *If consistent,* ZFC *does not prove the existence of an $\alpha \in \mathrm{On}$ such that $V_\alpha \vDash \mathsf{ZFC}$.*

Proof. Let $\alpha > \omega$ be such that $V_\alpha \vDash \mathsf{ZFC}$. The existence of a model of ZFC implies $\mathrm{Con}\,\mathsf{ZFC}$, which we know by Corollary 1.28 that we can't prove from within ZFC. Thus the existence of such an α is not provable in ZFC. ⊣

2. Large cardinals

The notion of a 'large cardinal' is difficult, and very possibly impossible, to make rigorous. The core idea however is that a cardinal κ is 'large' if its existence is (a) not known to be inconsistent with ZFC, and (b) can't be proven from ZFC (i.e. ZFC is consistent with the statement expressing 'κ exists'). (This said, some large-cardinal concepts such as the Reinhardt cardinals, introduced in Example 7.6(v), are known to be inconsistent with ZFC, however may be consistent with ZF.) A typical first introduction to large cardinals (such as [Kan08]) usually begins with the weakly and strongly inaccessible cardinals (see Subsection 2.3), leading sometimes to the erroneous view that these are the 'smallest' of the large-cardinal concepts. This is not the case, and in fact I would argue a more natural place to start is with the more recent notion of a 'worldly' cardinal, introduced by J.D. Hamkins in lectures at the CUNY Graduate Centre and NYU.

Whilst the focus of this thesis will be theories extending ZFC more broadly, much of the below will focus only on large cardinals. We do this because there is a close relationship between the two ideas: any large-cardinal concept A gives rise to a family of associated theories: ZFC + 'there is a cardinal of type A', ZFC + 'there are κ cardinals of type A' (for some cardinal κ), ZFC + 'there is a proper class of cardinals of type A', and so on. On the other hand it is speculated that the duality goes both ways: that every theory can, in some sense, be characterised in terms of large cardinals. The sense intended here is equiconsistency, which will be defined in Subsection 3.1. This will be returned to in Subsection 5.2.

2.1. **Worldly and hyperworldly cardinals.**

Definition 2.1. A cardinal κ is called *worldly* if and only if $V_\kappa \models \mathsf{ZFC}$. ◁

By Theorem 1.29, we note that the existence of such cardinals can't be proven within ZFC. We may give a generalisation of this concept:

Definition 2.2. A cardinal κ is *α-worldly* if and only if it is worldly and for all $\beta < \alpha$, κ is a limit of β-worldly cardinals. We call κ *hyperworldly* if and only if it is κ-worldly. ◁

Note that on this definition, a worldly cardinal is exactly a 0-worldly cardinal. Before we show that this notion of hyperworldly is the correct one, we note that α-worldly cardinals satisfy the following desirable property:

Proposition 2.3. *If κ is α-worldly, then it is β-worldly for all $\beta < \alpha$.*

Proof. Almost immediate: let κ be α-worldly and let $\beta < \alpha$, then since every $\gamma < \beta$ is also less than α, we see that κ is a limit of γ-worldly cardinals for all $\gamma < \beta$, thus is β-worldly. ⊣

Proposition 2.4. *There is no κ which is $(\kappa + 1)$-worldly, so more generally there is no κ which is λ-worldly for $\lambda > \kappa$.*

Proof. Suppose, for contradiction, that a cardinal κ can in fact be $(\kappa + 1)$-worldly, and let δ be the minimal such cardinal (which must exist by the well-ordering theorem); then δ is a limit of δ-worldly cardinals. Let $\gamma < \delta$ be one of these. Since δ is a limit, we know that $\gamma + 1 < \delta$, and in particular, since γ is δ-worldly, by Proposition 2.3 it is $(\gamma + 1)$-worldly, which contradicts the minimality of δ.

For the second part note that if κ could be λ-worldly for $\lambda > \kappa$, then by Proposition 2.3 it would be $(\kappa + 1)$-worldly, a contradiction. ⊣

This theorem shows that our notion of 'hyperworldly' is a good one: when κ is κ-worldly, this is the furthest we can iterate our construction. Of course it isn't the furthest we can go overall, since we can simply define:

Definition 2.5. κ is *α-hyperworldly* if and only if it is hyperworldly and a limit of β-hyperworldly cardinals for all $\beta < \alpha$. κ is *hyper-hyperworldly* if and only if it is κ-hyperworldly. $\triangleleft$

We can keep proceeding with this in the obvious way by calling a hyper-hyperworldly cardinal a hyper^2-worldly cardinal and extending to the hyper$^\beta$-worldly cardinals, then by taking limits obtain the α-hyper$^\beta$-worldly cardinals. In more detail (style following [Car15, p.14]):

Definition 2.6. A cardinal κ is α-hyper$^\beta$-worldly if and only if

(i) κ is worldly;
(ii) for all $\eta < \beta$, κ is κ-hyper$^\eta$-worldly;
(iii) for all $\gamma < \alpha$, κ is a limit of γ-hyper$^\beta$-worldly cardinals. $\triangleleft$

We get a result parallel to Proposition 2.4 which gives the 'limit' of this hierarchy:

Proposition 2.7. *There is no κ which is 1-hyper$^\kappa$-worldly.*

Proof. We may essentially copy the proof of Proposition 2.4: suppose that δ is the minimal 1-hyper$^\delta$-worldly cardinal, hence δ is a limit of (0-)hyper$^\delta$-worldly cardinals. Let $\gamma < \delta$ be one of these. Then by definition γ is γ-hyper$^\beta$-worldly for all $\beta < \delta$, and thus in particular since $\gamma < \delta$, γ is γ-hyper$^\gamma$-worldly.This implies that γ is 1-hyper$^\gamma$-worldly, which contradicts the minimality of δ. $\dashv$

We may however proceed as before and simply define new terminology to get around this (again we follow [Car15, p.16]): call a hyper$^\kappa$-worldly cardinal *richly* worldly. Then we may define the α-richly worldly cardinals, the hyper-richly worldly cardinals, the hyper$^\alpha$-richly worldly cardinals, the hyper$^\alpha$-richly$^\beta$ worldly cardinals, and so on: we may keep doing this as long as we can invent new language/notation to show what we are doing.

It should be noted that the property of being a worldly cardinal is not, in general, absolute:

Theorem 2.8. *It is consistent with* ZFC *that there is a cardinal κ such that V believes that κ is worldly, however there is a model of* ZFC*, $M \leqslant V$, with the same ordinals, such that M does not believe that κ is worldly.*

The proof of this involves forcing and thus is above the level of this essay, however see [Ham17].

2.2. Otherworldly and hyper-otherworldly cardinals.

Definition 2.9. A cardinal κ is called *otherworldly* if and only if there exists a $\lambda > \kappa$ such that $V_\kappa \prec V_\lambda$. $\triangleleft$

Example 1.18 gives a sense of 'how large' such cardinals must be. In fact as a preliminary result we have:

Theorem 2.10. *If κ is an otherworldly cardinal then it is worldly, so $V_\kappa \models \mathsf{ZFC}$.*

Proof. Suppose that $\lambda > \kappa$ is such that $V_\kappa \prec V_\lambda$; we first show that κ is a limit greater than ω, and thus by Theorem 1.10 must model ZFC without replacement. Suppose, for contradiction, that $\kappa = \alpha + 1$ for some α. Then we have that $V_\kappa \models$ 'α is a maximal element'. On the other hand, in V_λ, α is not maximal (since $\lambda > \kappa$, V_λ must at least contain $\alpha+1 = \kappa$), thus $V_\lambda \not\models$ 'α is a maximal element'; this contradicts our assumption that V_κ and V_λ agree on all formulae of $\mathcal{L}_{\mathrm{st}}$. Similarly to see that $\kappa > \omega$ note that if $\kappa = \omega$ then $\lambda > \omega$ and thus $V_\kappa \not\models$ infinity, whilst $V_\lambda \models$ infinity.

We now turn to replacement: let $A \in V_\kappa$ and suppose that for each $a \in A$, there is a unique $b \in V_\kappa$ such that $\varphi(a, b, z_0, \dots, z_n)$, for some formula φ in $\mathcal{L}_{\mathrm{st}}$ and where the z_i are arbitrary parameters. Then if B is the collection of all such b, we have that $B \subseteq V_\kappa$; thus since λ is strictly greater than κ, so in particular is at least $\kappa + 1$, we have that $B \in V_\lambda$. In particular, V_λ thinks that B – the replacement set – is a set; since this is expressible in $\mathcal{L}_{\mathrm{st}}$, V_κ must also think that the replacement set is a set, thus replacement is satisfied. $\dashv$

This result gives us a sense of the size of the otherworldly cardinals, and shows that they are genuinely large cardinals, insofar as we can define what it means to be a large cardinal (since, being worldly, their existence can't be proven in ZFC). We note that we also see that the smallest worldly cardinal is not otherworldly, so the concepts genuinely are distinct: let κ be the smallest worldly cardinal, so that $V_\kappa \not\models$ 'there is a worldly cardinal'. On the other hand for any $\lambda > \kappa$, $V_\lambda \models$ 'there is a worldly cardinal', so V_κ can never be an elementary substructure of a larger V_λ.

We may sharpen this result on the size of the otherworldly cardinals to further emphasise how these concepts diverge.

Theorem 2.11. *If κ is otherworldly, it is hyperworldly, and hyper-hyperworldly, and in fact α-hyper$^\beta$-worldly for all α, β for which Definition 2.6 makes sense.*

Proof. Let $\lambda > \kappa$ be such that $V_\kappa \prec V_\lambda$. We will first show that κ is hyperworldly, i.e. that it is α-worldly for all $\alpha \leqslant \kappa$. We induct on α: the base case follows from Theorem 2.10, since by definition 0-worldly just means worldly. Now suppose that κ is α-worldly; then it is a witness to the sentence asserting that for all $\gamma < \kappa$ there is a β such that β is α-worldly and $\beta > \gamma$ in V_λ. Therefore since by elementarity V_κ and V_λ must agree on all $\mathcal{L}_{\mathrm{st}}$-formulae, there must be such a witness in V_κ (which must then be less than κ, since $\kappa \notin V_\kappa$ by Lemma 1.5). Thus κ is a limit of α-worldly cardinals so is $(\alpha+1)$-worldly. The limit case is immediate: if κ is β-worldly for all $\beta < \lambda$ a limit, then for any $\gamma < \lambda$, κ is a limit of γ-worldly cardinals (since it is also $(\gamma+1)$-worldly), so is λ-worldly.

To generalise this result, we note that at no point in the proof did we use any properties of the α-worldly cardinals other than the limit requirement in the definition; since this is maintained as we ascend the hyper-hierarchy, the proof will still go through for these levels too (the base case just shifts, however will always follow from the 'previous' induction – we have shown an otherworldly cardinal is hyperworldly, and thus 0-hyperworldly, and so on). $\dashv$

Thus we see that the otherworldly cardinals exceed the entire worldly hierarchy with respect to direct implication, as defined in Subsection 3.3. As with the worldly cardinals, we may similarly extend the otherworldly cardinals.

Definition 2.12. A cardinal κ is *α-otherworldly* if and only if it is worldly and for all $\beta < \alpha$ there is a $\lambda > \kappa$ such that $V_\kappa \prec V_\lambda$ and λ is β-otherworldly. κ is *hyper-otherworldly* if and only if it is α-otherworldly for all $\alpha \in \mathrm{On}$. $\lhd$

Note that on this definition 0-otherworldly means worldly, and 1-otherworldly means otherworldly. Further, this definition certainly doesn't give us that an α-otherworldly cardinal gives rise to an elementary chain of length α – for example being ω-otherworldly only gives that we can get chains arbitrarily close to ω in length, rather than a chain of length ω itself. In fact we have the two following results.

Theorem 2.13. *If there is an elementary chain of length ω*

$$V_{\kappa_0} \prec V_{\kappa_1} \prec \cdots \prec V_{\kappa_n} \prec \cdots$$

then each κ_i, $i \in \omega$ is hyper-otherworldly.

Proof. We prove by strong induction on α that κ_i is α-otherworldly for each $\alpha \in \mathrm{On}$, for each $i \in \omega$. Being 0-otherworldly by above simply means being worldly, however this is clearly true since each κ_i is otherworldly. Now suppose that for all $\beta < \alpha$ we have that every κ_i is β-otherworldly. Then clearly any V_{κ_j} can be extended by a $V_{\kappa_{j'}}$ where $j' > j$ and $\kappa_{j'}$ is β-otherworldly; thus κ_j is α-otherworldly and we are done. $\dashv$

Theorem 2.14. *If κ is otherworldly, then there is a $\lambda > \kappa$ such that $V_\kappa \prec V_\lambda$ and $V_\lambda \not\prec V_{\lambda'}$ for all $\lambda' > \lambda$.*

Proof. Suppose not for a contradiction: then given any otherworldly κ, we get an elementary chain

$$V_\kappa \prec V_\lambda \prec V_{\lambda'} \prec \cdots$$

unbounded in V. But since we can define a truth predicate True_{V_α} in each of the V_α, each only being a set model, where $\alpha \in I = \{\kappa, \lambda, \lambda', \ldots\}$, this would give us a means of defining a truth predicate in $V = \bigcup_{\alpha \in I} V_\alpha$, by setting

$$\mathrm{True}_V\left(\ulcorner\varphi\urcorner\right) = \mathrm{True}_{V_\alpha}\left(\ulcorner\varphi\urcorner\right),$$

for some $\alpha \in I$ (where the choice of α is unimportant by elementarity). However the construction of a truth predicate in V contradicts Theorem 1.25. $\dashv$

We easily obtain a theorem paralleling Proposition 2.3, however note that we get no parallel of Proposition 2.4 (see the comment below Corollary 2.22 for more on this).

Proposition 2.15. *If κ is α-otherworldly, then it is β-otherworldly for all $\beta < \alpha$.*

We may then extend the otherworldly cardinal hierarchy in the natural way.

Definition 2.16. A cardinal κ is α-hyper$^\beta$-otherworldly if and only if

(i) κ is otherworldly;
(ii) For all δ, for all $\eta < \beta$, κ is δ-hyper$^\eta$-otherworldly;
(iii) For all $\gamma < \alpha$, κ is a limit of γ-hyper$^\beta$-otherworldly cardinals. $\lhd$

Note as with worldliness, the property of being otherworldly is not in general absolute:

Theorem 2.17. *If κ is otherworldly in V, then we may find a model $M \leqslant V$ which does not believe that κ is otherworldly.*

Proof. Let $\lambda > \kappa$ be the minimal cardinal such that $V_\kappa \prec V_\lambda$. V_λ does not believe that κ is otherworldly, else there would be a $\lambda' > \kappa$ in V_λ such that $V_\kappa \prec V_\lambda$. However, this contradicts minimality of λ. ⊣

We will now move to relate these new concepts to another more well-studied class of large cardinals.

2.3. **Inaccessibility.**

Definition 2.18. A cardinal κ is *inaccessible* if and only if it is uncountable, regular ($\operatorname{cf}\kappa = \kappa$), and a strong limit cardinal ($\delta < \kappa$ implies $2^\delta < \kappa$). ◁

Note we may also define the *weakly* inaccessible cardinals by exchanging 'strong limit' for 'weak limit' in the above. We will show, among other things, that inaccessible cardinals are worldly, for which we will need the following (proofs following [Kan08, ch1.1]).

Lemma 2.19. *For $x \subseteq V_\kappa$ where κ is inaccessible, we have that $x \in V_\kappa$ if and only if $|x| < \kappa$.*

Proof. [$\rightarrow$] Suppose that $x \in V_\kappa$, then for some $\alpha < \kappa$, $|x| \leqslant |V_\alpha| < |V_\kappa|$, since x is a subset of V_κ. Thus it suffices to show that $|V_\alpha| < \kappa$ for all $\alpha < \kappa$. This follows from a quick induction: the base case is obvious; if $|V_\alpha| < \kappa$, then $|V_{\alpha+1}| = 2^{|V_\kappa|} < \kappa$ since κ is a strong limit; if $|V_\beta| < \kappa$ for all $\beta < \lambda$, then note that $\{|V_\beta| \mid \beta < \lambda\}$ is a sequence bounded in κ of length less than κ; since κ is regular, this implies that $|V_\lambda| = \sup_{\beta<\lambda} |V_\beta| < \kappa$.

[$\leftarrow$] Recall $\operatorname{rank} z$ is the minimal α such that $z \subseteq V_\alpha$. Now let $|x| < \kappa$, then (similarly to above) the set $R = \{\operatorname{rank} y \mid y \in x\}$ consists of a sequence bounded in κ of length less than κ, thus $R \subseteq \alpha$ for some $\alpha < \kappa$; then $x \in V_{\alpha+1} \subsetneq V_\kappa$. ⊣

We may now show:

Theorem 2.20. *If κ is inaccessible, then $V_\kappa \vDash \mathsf{ZFC}$, so κ is worldly.*

Proof. Since κ is an infinite cardinal and thus a limit ordinal by definition, and since by definition it is greater than $\aleph_0$, we simply need to show that $V_\kappa \vDash$ replacement; the rest follows immediately from Theorem 1.10. Let $x \in V_\kappa$ and suppose that φ is such that for all $y \in x$ and $v_0, \ldots, v_n$ variables, there is a unique $z \in V_\kappa$ with $\varphi(y, z, v_0, \ldots, v_n)$. Then if $Z \subseteq V_\kappa$ is the collection of all such z, note that $|Z| \leqslant |x| < \kappa$; then by Lemma 2.19, $Z \in V_\kappa$, so replacement holds. ⊣

Thus we see that the inaccessible cardinals exceed the entire worldly hierarchy, also in the sense of direct implication, as defined in Subsection 3.3. We have a related, however different result for the otherworldly cardinals, Proposition 3.9, for which we will need the following.

Proposition 2.21. *If κ is inaccessible, then there are arbitrarily large $\lambda < \kappa$ such that $V_\lambda \prec V_\kappa$.*

Proof. Let $\{\varphi_0, \varphi_1, \ldots\}$ be an enumeration of the formulae of $\mathcal{L}_{\mathrm{st}}$; let $\alpha_0 < \kappa$. Our choice of α_0 determines our λ: we may choose $\alpha_0 < \kappa$ arbitrarily large and the proof still goes through. For each $n, m \in \omega$ and $(a_0, \ldots, a_m) = \mathbf{a} \in V_{\alpha_0}^{m+1}$ consider whether $V_\kappa \vDash \exists x\, \varphi_n(x, \mathbf{a})$ ($V_{\alpha_0}^{m+1}$ denotes the $(m+1)$-fold Cartesian product of V_{α_0}). If this is the case, then define $\beta_{n,\mathbf{a}}$ to be the minimal ordinal less than κ such that there is a witness for $\exists x\, \varphi_n(x, \mathbf{a})$ in $V_{\beta_{n,\mathbf{a}}}$.

Now let $\alpha_1 = \sup_{n,m,\mathbf{a}} \beta_{n,\mathbf{a}}$; since κ is regular, and we are considering the union of a set with fewer elements than κ, we must have $\alpha_1 < \kappa$. Continuing this inductively we may define $\lambda = \sup_n \alpha_n$ which must by similar regularity considerations also be less than κ.

Since we have constructed our λ such that V_λ contains witnesses for all the true existentials in V_κ, by the Tarski-Vaught criterion and the fact that $V_\lambda < V_\kappa$ (since $\lambda < \kappa$), we have that $V_\lambda \prec V_\kappa$. $\dashv$

Corollary 2.22. *If κ is inaccessible, then there is a continuous elementary chain*

$$V_{\gamma_0} \prec V_{\gamma_1} \prec \cdots \prec \bigcup_{\alpha<\kappa} V_{\gamma_\alpha} = V_\kappa.$$

Note by 'continuous' here we mean that at limit ordinals $\lambda \leqslant \kappa$ we have $\gamma_\lambda = \bigcup_{\beta<\lambda} \gamma_\beta$.

Proof. We can define such a sequence by transfinite recursion as follows: (i) let γ_0 be any λ as given by Proposition 2.21; (ii) let $\gamma_{\alpha+1}$ be any λ' such that $\gamma_\alpha < \lambda' < \kappa$, which we also know we can construct by Proposition 2.21; (iii) finally define γ_δ for δ a limit ordinal to be $\bigcup_{\beta<\delta} \gamma_\beta$, which is less than κ by regularity. We show that this sequence has the desired properties.

To see that for all α, $V_{\gamma_\alpha} \prec V_\kappa$, note that in case (i) it is immediate. In case (ii) it follows since $V_{\gamma_\alpha} < V_{\gamma_{\alpha+1}}$ and $V_{\gamma_\alpha}, V_{\gamma_{\alpha+1}} \prec V_\kappa$, thus $V_{\gamma_\alpha} \prec V_{\gamma_{\alpha+1}}$ (if not then there would be a formula they disagree on, but this would contradict their elementarity in V_κ). Case (iii) follows immediately since the union of an elementary chain is an elementary superstructure of each of its elements, and thus in particular an elementary substructure of V_κ, since $\gamma_\delta \leqslant \kappa$. Continuity is immediate from the definition. $\dashv$

Note that by definition the γ_αs are κ-otherworldly, since they each have a chain of length κ extending them; this shows that a cardinal γ may be λ-otherworldly for $\lambda > \gamma$, in contrast to the α-worldly cardinals.

Theorem 2.23. *If κ is inaccessible, then it is hyperworldly, and hyper-hyperworldly, and...*

Proof. Let κ be an inaccessible cardinal and consider the chain $\{\gamma_\alpha \mid \alpha < \kappa\}$ given by Corollary 2.22.

First note that γ_α is worldly for all $\alpha < \kappa$: by construction γ_α is otherworldly; then apply Theorem 2.10. Then we have that γ_ω, which by continuity is $\bigcup_{n\in\omega} \gamma_n$, is a limit of worldly cardinals, and thus is 1-worldly. Similarly we conclude that $\gamma_{\omega\cdot 2} = \bigcup_{n\in\omega} \gamma_{\omega+n}$ is 1-worldly, and so on for $n \in \omega$. Thus $\gamma_{\omega^2} = \bigcup_{n\in\omega} \gamma_{\omega\cdot n}$ is a limit of 1-worldly cardinals, so is 2-worldly.

We can inductively continue this construction, which must remain bounded in κ by regularity, to get α-worldly γ_α for all $\alpha < \kappa$. In particular, we can iterate this construction at any ordinal below κ to obtain an α-worldly γ_α, $\alpha < \kappa$, which again by regularity must be less than κ. By construction κ must be a limit of these γ_α, so is thus a limit of α-worldly cardinals for all $\alpha < \kappa$; this is the definition of κ being hyperworldly.

As with previous results to generalise we note that all the above proofs also go through *mutatis mutandis* (where we must apply Theorem 2.11 to ensure the base case goes through). $\dashv$

Whilst we will see in Subsection 3.2 that the inaccessible cardinals exceed the otherworldly cardinals in terms of consistency strength, the following result might make us nervous:

Theorem 2.24. *There is an inaccessible cardinal κ which is not otherworldly.*

Proof. Let κ be the smallest inaccessible cardinal (which must exist by well-ordering). Then whilst $V_\kappa \not\models$ 'there is an inaccessible cardinal', for any $\lambda > \kappa$ we have $\kappa \in V_\lambda$ and thus $V_\lambda \models$ 'there is an inaccessible cardinal' (since the property of being an inaccessible cardinal is absolute between V and the von Neumann hierarchy). Thus we can't have $V_\kappa \prec V_\lambda$ for any $\lambda > \kappa$. $\dashv$

We do however have:

Theorem 2.25. *Every inaccessible cardinal is a limit of otherworldly cardinals.*

Proof. Note that in Theorem 2.22, κ is a limit of the γ_αs, which as noted above are all κ-otherworldly, and in particular by Proposition 2.15, otherworldly. $\dashv$

We may strengthen this result by noting that in fact each of the γ_αs must be *hyper*-otherworldly in V_κ, since they are κ-otherworldly (outside of V_κ) and hence α-otherworldly for all $\alpha < \kappa$. V_κ will attest this for each $\alpha < \kappa$, thus we are done. Compare this with Theorem 2.13. We can continue this to show that in fact V_κ proves that there is a class of hyper^2-otherworldly cardinals, and so on as above. This result allows us to prove that the inaccessible cardinals 'exceed' the otherworldly cardinals in another sense – minimal occurrence – which will be defined in Subsection 3.3.

We now move on to discuss the consistency strength hierarchy, in which we will place our new cardinals.

3. The consistency strength hierarchy

Several different hierarchies have been developed in the study of set theory; more discussion around the philosophy surrounding the one we study here will be given in Part II. Here I will introduce the consistency strength hierarchy as defined by Steel in [Ste12], and locate within it the cardinals we have considered above.

3.1. Steel's hierarchy.

In [Ste12, p3], John Steel introduces the notion of the *consistency strength hierarchy*, which is intended to provide a 'ranking' of theories extending ZFC; let T, U be two such theories.

Definition 3.1. We say $T \leqslant_{\text{Con}} U$, or that U is of greater-than or equal consistency strength to T, if and only if $\mathsf{ZFC} \vdash \operatorname{Con} U \to \operatorname{Con} T$. If we additionally have that $\mathsf{ZFC} \nvdash \operatorname{Con} T \to \operatorname{Con} U$ then we write $T <_{\text{Con}} U$. If $T \leqslant_{\text{Con}} U \leqslant_{\text{Con}} T$, then we write $T \equiv_{\text{Con}} U$ and say T and U have the same consistency strength, or are *equiconsistent.* ◁

Note by Example 3.3(i) we see that $\leqslant_{\text{Con}}$ is not antisymmetric and thus not a partial order: we have $\mathsf{ZFC} \leqslant_{\text{Con}} \mathsf{ZFC} + \mathsf{CH} \leqslant_{\text{Con}} \mathsf{ZFC}$, however $\mathsf{ZFC} \neq \mathsf{ZFC} + \mathsf{CH}$ (see Corollary 5.1). We do however get reflexivity and transitivity:

Proposition 3.2. *If $\mathfrak{T}$ is the space of theories extending* ZFC*, then $\leqslant_{\text{Con}}$ forms a reflexive and transitive order on $\mathfrak{T}$.*

Proof. Reflexivity: clearly $\mathsf{ZFC} \vdash \operatorname{Con} T \to \operatorname{Con} T$ for any T, and thus $T \leqslant_{\text{Con}} T$ for any T.

Transitivity: if $S \leqslant_{\text{Con}} T \leqslant_{\text{Con}} U$, then $\mathsf{ZFC} \vdash \operatorname{Con} U \to \operatorname{Con} T$ and $\mathsf{ZFC} \vdash \operatorname{Con} T \to \operatorname{Con} S$. Then by hypothetical syllogism, $\mathsf{ZFC} \vdash \operatorname{Con} U \to \operatorname{Con} T$. ⊣

In the standard way we can then turn $\leqslant_{\text{Con}}$ into a partial order by considering theories which are equivalent up to equiconsistency.

Example 3.3.

(i) It is well-known from the work of Gödel and Cohen that $\mathsf{ZFC} \equiv_{\text{Con}} \mathsf{ZFC} + \mathsf{CH} \equiv_{\text{Con}} \mathsf{ZFC} + \neg\mathsf{CH}$. See Subsection 5.2 for more detail.
(ii) Trivially, the inconsistent theory is a maximal element under $\leqslant_{\text{Con}}$.
(iii) Clearly $\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC} \vdash \operatorname{Con} \mathsf{ZFC}$, thus by Proposition 3.4 $\mathsf{ZFC} <_{\text{Con}} \mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}$. More generally, we note that no theory (which is not known to be inconsistent) has the greatest consistency strength, since we can always add to a theory T the statement $\operatorname{Con} T$, which by Gödel will be strictly stronger.

In [Ham21, p3], Joel David Hamkins notes a more convenient sufficient condition for $T <_{\text{Con}} U$:

Proposition 3.4. *If T, U are consistent extensions of* ZFC *and $U \vdash \operatorname{Con} T$, then $T <_{\text{Con}} U$.*

Proof. First we show that $\mathsf{ZFC} \vdash \operatorname{Con} U \to \operatorname{Con} T$, i.e. that $T \leqslant_{\text{Con}} U$. Suppose that $U \vdash \operatorname{Con} T$ and let $\mathcal{M}$ be any model of $\mathsf{ZFC} + \operatorname{Con} U$; we must have that $\mathcal{M} \vDash \operatorname{Con}(U + \operatorname{Con} T)$. For suppose not, so $\mathcal{M} \vDash \neg \operatorname{Con}(U + \operatorname{Con} T)$. Then $\mathcal{M}$ believes there is a proof of a contradiction from $U + \operatorname{Con} T$. Since $U \vdash \operatorname{Con} T$, this proof can be recast purely in terms of axioms of U, and thus we would have that $\mathcal{M} \vDash \neg \operatorname{Con} U$, which contradicts the definition of $\mathcal{M}$.

We must then have $\mathcal{M} \vDash \operatorname{Con} T$, since if $M \vDash \neg \operatorname{Con} T$, then $\mathcal{M}$ believes there is a proof of this, and thus that there is a proof of this from $U + \operatorname{Con} T$, since adding axioms doesn't affect proofs we already have. Clearly $\mathcal{M}$ believes that there is a proof of $\operatorname{Con} T$ from $U + \operatorname{Con} T$, thus $\mathcal{M}$ believes that $U + \operatorname{Con} T$ is inconsistent, since it proves $\varphi, \neg\varphi$ for some φ. This contradicts that $\mathcal{M} \vDash \operatorname{Con}(U + \operatorname{Con} T)$. Thus we must have $\mathcal{M} \vDash \operatorname{Con} T$. In particular, since $\mathcal{M}$ was arbitrary, we have shown that $\mathsf{ZFC} + \operatorname{Con} U \vdash \operatorname{Con} T$, and thus $\mathsf{ZFC} \vdash \operatorname{Con} U \to \operatorname{Con} T$, as required.

On the other hand, to see that $\mathsf{ZFC} \nvdash \operatorname{Con} T \to \operatorname{Con} U$, note that if the implication were provable, then since $U \vdash \operatorname{Con} T$ by hypothesis, and U extends ZFC and thus also $U \vdash \operatorname{Con} T \to \operatorname{Con} U$, we would have by *modus ponens* $U \vdash \operatorname{Con} U$. This contradicts Theorem 1.27.

Therefore we have $T <_{\mathrm{Con}} U$, as required. ⊣

It should be noted that the condition in Proposition 3.4, whilst sufficient, is not necessary: there exist theories T, U such that $T <_{\mathrm{Con}} U$ although $U \nvdash \operatorname{Con} T$. For example, adapting Theorem 4 of [Ham21], we may find a U such that $\mathsf{ZFC} <_{\mathrm{Con}} U <_{\mathrm{Con}} \mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}$. Following Hamkins' notation, such a U will be given by the theory

$$(\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}) \vee (\mathsf{ZFC} + \mathrm{R}_S),$$

where R_S is the Rosser sentence of the theory

$$S = \mathsf{ZFC} + \operatorname{Con}(\mathsf{ZFC}) + \neg \operatorname{Con}(\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}).$$

This is the theory with sentences of the form $\varphi \vee (\psi \wedge \mathrm{R}_S)$ for φ a theorem of $\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}$, and ψ a theorem of ZFC. Note that such a U entails all the theorems of ZFC, since any theorem of ZFC is derivable from both $\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}$ and $\mathsf{ZFC} + \mathrm{R}_S$. For further details on the construction see Hamkins' paper.

Proposition 3.5. *With U as above, it is the case that (i) $\mathsf{ZFC} <_{\mathrm{Con}} U$ and (ii) $U \nvdash \operatorname{Con} \mathsf{ZFC}$.*

Proof. (i) is immediate from the definition of U and Proposition 3.4. For (ii), suppose for contradiction that $U \vdash \operatorname{Con} \mathsf{ZFC}$. Now as noted above, by the construction of U we also have that $U \vdash \mathsf{ZFC}$, so we have $U \vdash \mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}$. Thus any proof of a contradiction from $\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}$ would also be a proof a contradiction from U, thus we have (noting as usual ZFC is our background theory)

$$\mathsf{ZFC} \vdash \neg \operatorname{Con}(\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}) \to \neg \operatorname{Con} U,$$

thus

$$\mathsf{ZFC} \vdash \operatorname{Con} U \to \operatorname{Con}(\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}).$$

However by hypothesis, since $U <_{\mathrm{Con}} \mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}$, we have

$$\mathsf{ZFC} \nvdash \operatorname{Con} U \to \operatorname{Con}(\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}),$$

which gives our contradiction. ⊣

3.2. **Placing the worldly and otherworldly cardinals in the hierarchy.** Figure 1 at the end of this section illustrates where the cardinals developed lie in the consistency strength hierarchy. We will show a selection of corresponding results (those stated but not proven follow in a similar manner). For typographical ease, we introduce the notation $\alpha\mathsf{W}_\beta$ and $\alpha\mathsf{O}_\beta$ to mean 'there are β many α-worldly/otherworldly cardinals', respectively, with the additional stipulations that we write simply W_β and O_β for 'there are β worldly/otherworldly cardinals', and we write $\beta = \infty$ (i.e. $\alpha\mathsf{W}_\infty, \alpha\mathsf{O}_\infty$) to mean 'there is a proper class of α-worldly/otherworldly cardinals'. The same style but with HW and HO indicates the same statements about hyperworldly/hyper-otherworldly cardinals (we will not prove results that require notation beyond this). For example $\alpha\mathsf{HW}_\beta$ will mean 'there are β many α-hyperworldly cardinals'. We will also write I for the statement 'there is an inaccessible cardinal'.

Theorem 3.6.

(i) $\mathsf{ZFC}, \mathsf{ZFC} + \mathrm{Con}\,\mathsf{ZFC}, \ldots <_{\mathrm{Con}} \mathsf{ZFC} + \mathsf{W}_1$.
(ii) $\mathsf{ZFC} + \mathsf{W}_1 <_{\mathrm{Con}} \mathsf{ZFC} + \mathsf{W}_2$.
(iii) $\mathsf{ZFC} + \mathsf{W}_\infty <_{\mathrm{Con}} \mathsf{ZFC} + 1\mathsf{W}_1$.
(iv) $\mathsf{ZFC} + 1\mathsf{W}_\infty <_{\mathrm{Con}} \mathsf{ZFC} + 2\mathsf{W}_1$.
(v) $\mathsf{ZFC} + \alpha\mathsf{W}_\infty <_{\mathrm{Con}} \mathsf{ZFC} + \mathsf{HW}_1$, *where α is any ordinal less than the hyperworldly cardinal stipulated.*
(vi) $\mathsf{ZFC} + \mathsf{HW}_1 <_{\mathrm{Con}} \mathsf{ZFC} + \mathsf{O}_1$.
(vii) $\mathsf{ZFC} + \mathsf{O}_1 <_{\mathrm{Con}} \mathsf{ZFC} + 2\mathsf{O}_1$
(viii) $\mathsf{ZFC} + \mathsf{HO}_1, \mathsf{ZFC} + \mathsf{HO}_2, \ldots <_{\mathrm{Con}} \mathsf{ZFC} + \mathsf{I}$.

Proof. We use the sufficient condition on of $T <_{\mathrm{Con}} U$ from Proposition 3.4 and in each case show that $U \vdash \mathrm{Con}\,T$ (for relevant T, U). To show that $\mathrm{Con}\,T$ follows from U, we give a model of T, given that U holds; this suffices to show that $U \vdash \mathrm{Con}\,T$ since by soundness, the existence of a model of a theory implies that theory's consistency.

(i) Define by recursion the following sequence of theories:

$$S_0 = \mathsf{ZFC}$$
$$S_{n+1} = S_n + \mathrm{Con}\,S_n.$$

The required result by Proposition 3.4 is then that $\mathsf{ZFC} + \mathsf{W}_1 \vdash \mathrm{Con}\,S_n$ for all $n \in \omega$. Proceed by induction. Note that if κ is the worldly cardinal given by $\mathsf{ZFC} + \mathsf{W}_1$, then by soundness it suffices to show that $V_\kappa \vDash S_n$ for all $n \in \omega$.

The base case follows because by hypothesis $V_\kappa \vDash \mathsf{ZFC} = S_0$.

Suppose that $V_\kappa \vDash S_n$. Then V believes that S_n is consistent, i.e. believes $\mathrm{Con}\,S_n$ (since S_n has a model). Now V_κ is transitive by Lemma 1.3, and thus by a result beyond the author (which will be mentioned again in Subsection 7.2.2), must agree with V on arithmetic truths; see [Kun14, Lemma II.4.14] for more details. Consistency statements are, by construction, arithmetic truths, and thus $V_\kappa \vDash \mathrm{Con}\,S_n$. Therefore we have $V_\kappa \vDash S_n + \mathrm{Con}\,S_n = S_{n+1}$, as required.

(ii) Let $\kappa_1 < \kappa_2$ be worldly cardinals. Since $\kappa_1 < \kappa_2$, we must have that $V_{\kappa_1} \in V_{\kappa_2}$. By worldiness, the hypotheses of Proposition 1.21 are satisfied, and thus V_{κ_2} believes that $V_{\kappa_1} \vDash \mathsf{ZFC}$. Thus V_{κ_2} believes that there is a worldly cardinal (i.e. κ_1).

(iii) Let κ be 1-worldly, so it is a limit of worldly cardinals, so there are worldly $\kappa_\alpha < \kappa$ ($\alpha < \beta$ for some $\beta \in \mathrm{On}$) such that $\kappa = \sup\{\kappa_\alpha \mid \alpha < \beta\}$. By similar remarks as in (ii), κ must believe all of the κ_i are worldly. I claim that this collection is a proper class in V_κ, i.e. $\{\kappa_\alpha \mid \alpha < \beta\} \notin V_\kappa$ and $\{\kappa_\alpha \mid \alpha < \beta\} \subseteq V_\kappa$. For the latter note that $\kappa \subseteq V_\kappa$ by Lemma 1.6, and thus since each $\kappa_\alpha \in \kappa$, their collection is a subset of V_κ. For the former suppose not for contradiction: then since $V_\kappa \vDash \mathsf{ZFC}$ by Proposition 2.3, and in particular the axiom of union, $\bigcup\{\kappa_\alpha \mid \alpha < \beta\} = \sup\{\kappa_\alpha \mid \alpha < \beta\} = \kappa \in V_\kappa$. This contradicts Lemma 1.5. In particular, we have shown V_κ believes that there is a proper class of worldly cardinals, which suffices.

(iv) This proof is identical to (iii), however with '2-worldly' replacing '1-worldly', and '1-worldly' replacing 'worldly'.

(v) Let κ be a hyperworldly cardinal, so κ is worldly and a limit of β worldly cardinals for all $\beta < \kappa$. Fix any $\alpha < \kappa$, and let κ be a limit of the α-worldly cardinals $\{\kappa_\delta \mid \delta < \gamma\}$ for some $\gamma \in \mathrm{On}$. Similarly to (iii), we must have that $\{\kappa_\delta \mid \delta < \gamma\} \notin V_\kappa$ and $\{\kappa_\delta \mid \delta < \gamma\} \subseteq V_\kappa$, so as in (iii) we are done. Since $\alpha < \kappa$ was arbitrary this completes the proof.

(vi) Suppose that κ is otherworldly, so there is $\lambda > \kappa$ with $V_\kappa \prec V_\lambda$. Then by the proof of Theorem 2.11, V_λ believes that κ is hyperworldly, so we have a model of $\mathsf{ZFC} + \mathsf{HW}_1$, as required.

(vii) Let κ be 2-otherworldly, so there are $\kappa < \lambda < \lambda'$ with $V_\kappa \prec V_\lambda \prec V_{\lambda'}$. Then in particular, in $V_{\lambda'}$ we have that $V_\kappa \prec V_\lambda$, and thus $V_{\lambda'}$ believes there is an otherworldly cardinal. (Note the elementary equivalence holds according to $V_{\lambda'}$ by Proposition 1.21, since $V_{\lambda'}$ is transitive and models ZFC).

(viii) Let κ be inaccessible, so there is by Theorem 2.22 an elementary chain $V_{\kappa_0} \prec V_{\kappa_1} \prec \cdots \prec V_\kappa$. In particular, V_κ believes that there is an elementary ω-chain extending V_{κ_0}, and thus by Theorem 2.13 it believes that κ_0 is hyper-otherworldly. Indeed V_κ will believe that each of the κ-many κ_i are hyper-otherworldly by identical logic. Thus we have our result. ⊣

3.3. **Alternative hierarchies.** A number of alternative hierarchies for classifying large-cardinal concepts have been developed. I will briefly detail two here, and relate them to each other and $\leqslant_{\mathrm{Con}}$.

Definition 3.7. Given two large-cardinal concepts A and B, we say that $\mathsf{A} \leqslant_{\leftarrow} \mathsf{B}$ if and only if whenever ZFC proves that a cardinal κ is of type B, ZFC also proves that κ is of type A This gives the *direct implication* hierarchy. ◁

Definition 3.8. Given two large-cardinal concepts A and B, we say that $\mathsf{A} \leqslant_{\min} \mathsf{B}$ if and only if the least cardinal of type A is less than or equal to the least cardinal of type B. This gives the *minimal instance* hierarchy. ◁

Note that in talking here about $\leqslant_{\leftarrow}$, $\leqslant_{\min}$, and $\leqslant_{\mathrm{Con}}$, there is a slight elision between theories and large-cardinal concepts, the latter of which more properly refers to definitions. For example ZFC + 'there is a worldly cardinal' is a theory, whereas the large-cardinal concept 'worldly' refers to the definition of a worldly cardinal. Properly speaking, $\leqslant_{\mathrm{Con}}$ is an order on theories (not just about large cardinals), and $\leqslant_{\min}, \leqslant_{\leftarrow}$ are orders on large-cardinal concepts. In the below therefore we abuse notation slightly and pretend $\leqslant_{\mathrm{Con}}$ is an order on large-cardinal concepts, simply by defining $\mathsf{A} \leqslant_{\mathrm{Con}} \mathsf{B}$ if and only if ZFC + 'there is a cardinal of type A' $\leqslant_{\mathrm{Con}}$ ZFC + 'there is a cardinal of type B'.

We may now formalise the remark made at the end of Subsection 2.3.

Proposition 3.9. $\mathsf{O} \leqslant_{\min} \mathsf{I}$.

Proof. Recall as in Corollary 2.22 that any inaccessible cardinal has many otherworldly cardinals below it. Thus if κ is the smallest inaccessible cardinal, then the smallest otherworldly cardinal must be smaller than it. ⊣

Theorem 3.10. *Considered as classes, we have (i) $\leqslant_{\leftarrow} \subsetneq \leqslant_{\min}$, (ii) $\leqslant_{\leftarrow} \subsetneq \leqslant_{\mathrm{Con}}$, (iii) $\leqslant_{\mathrm{Con}} \nsubseteq \leqslant_{\min}$ and (iv) $\leqslant_{\min} \nsubseteq \leqslant_{\mathrm{Con}}$.*

Proof. In order to show $\leqslant \subseteq \preccurlyeq$ for two large-cardinal orders $\leqslant, \preccurlyeq$, we need to show that for any large-cardinal concepts A, B, we have $\mathsf{A} \leqslant \mathsf{B}$ implies that $\mathsf{A} \preccurlyeq \mathsf{B}$. Similarly to show $\leqslant \nsubseteq \preccurlyeq$, we need to give examples of large-cardinal concepts A, B such that $\mathsf{A} \leqslant \mathsf{B}$, but $\mathsf{A} \not\preccurlyeq \mathsf{B}$.

(i) Suppose that $\mathsf{A} \leqslant_{\leftarrow} \mathsf{B}$, for two large-cardinal concepts A and B. Then if V believes that κ is of type B, V believes that κ is of type A. Thus the smallest instance of a cardinal of type B will also be a cardinal of type A, and thus the smallest instance of a cardinal of type A will be at most this size, thus $\mathsf{A} \leqslant_{\min} \mathsf{B}$. To see that the containment is proper, note that by 2.25, the smallest otherworldly cardinal will be smaller than the smallest inaccessible cardinal, since the latter must be a limit of otherworldly cardinals; thus $\mathsf{O} \leqslant_{\min} \mathsf{I}$ (where we have adapted the notation from Subsection 3.2 in the obvious way). On the other hand, Theorem 2.24 gives us that there is an inaccessible cardinal which is not otherworldly, and thus $\mathsf{O} \not\leqslant_{\leftarrow} \mathsf{I}$.

(ii) Suppose that $\mathsf{A} \leqslant_{\leftarrow} \mathsf{B}$. Then if $\mathsf{ZFC} \vdash \neg\,\mathrm{Con}\,(\mathsf{ZFC} + \mathsf{A})$, since ZFC proves that every A-cardinal is also a B-cardinal, this means we must also have $\mathsf{ZFC} \vdash \neg\,\mathrm{Con}\,(\mathsf{ZFC} + \mathsf{B})$. Thus by contraposition $\mathsf{ZFC} \vdash \mathrm{Con}\,(\mathsf{ZFC} + \mathsf{B}) \to \mathrm{Con}\,(\mathsf{ZFC} + \mathsf{A})$, so $\mathsf{A} \leqslant_{\mathrm{Con}} \mathsf{B}$, as required. On the other hand, by Theorem 3.6, $\mathsf{O} \leqslant_{\mathrm{Con}} \mathsf{I}$, however as above, we have that $\mathsf{O} \not\leqslant_{\leftarrow} \mathsf{I}$, so our containment is strict.

(iii), (iv) Both can be tackled in the same way, by noting that the existence of strong cardinals has lower consistency strength than the existence of superstrong cardinals (both defined in Example 7.6), yet the least superstrong cardinal is less than the least strong cardinal. The details of these proofs may be found in [Kan08, pp358–365]. If we let Strong and $\mathsf{Superstrong}$ denote the strong and superstrong large-cardinal concepts, respectively, then we have $\mathsf{Strong} \leqslant_{\min} \mathsf{Superstrong}$, and $\mathsf{Superstrong} \leqslant_{\mathrm{Con}} \mathsf{Strong}$, showing that the two orders $\leqslant_{\min}$, $\leqslant_{\mathrm{Con}}$, are $\subseteq$-incomparable. $\dashv$

ZFC + ‘there is an inaccessible cardinal’

⋮

ZFC + ‘there is a richly otherworldly cardinal’

⋮

ZFC + ‘there is a hyper-hyper-otherworldly cardinal’

⋮

ZFC + ‘there is a hyper-otherworldly cardinal’

⋮

ZFC + ‘there is a 2-otherworldly cardinal’

⋮

ZFC + ‘there are two 1-otherworldly cardinals’
ZFC + ‘there is a 1-otherworldly cardinal’

⋮

ZFC + ‘there are two otherworldly cardinals’
ZFC + ‘there is an otherworldly cardinal’

⋮

ZFC + ‘there is a richly worldly cardinal’

⋮

ZFC + ‘there is a hyper-hyperworldly cardinal’

⋮

ZFC + ‘there is a hyperworldly cardinal’

⋮

ZFC + ‘there is a 2-worldly cardinal’

⋮

ZFC + ‘there are two 1-worldly cardinals’
ZFC + ‘there is a 1-worldly cardinal’

⋮

ZFC + ‘there is a proper class of worldly cardinals’

⋮

ZFC + ‘there are two worldly cardinals’
ZFC + ‘there is a worldly cardinal’

⋮

ZFC + Con ZFC + Con (ZFC + Con ZFC)
ZFC + Con ZFC
ZFC

Figure 1. The new cardinals fit linearly into the consistency strength hierarchy. The vertical dots indicate in general a large leap in consistency strength.

PART II: IS THE CONSISTENCY STRENGTH HIERARCHY LINEAR?

A common – indeed almost unanimous – view among modern set theorists is that the consistency strength hierarchy exhibits linearity in its 'natural' theories. In his paper on the subject, Hamkins quotes five prominent researchers in the field giving this opinion [Ham21, pp1–2]. For example Stephen Simpson [Sim09] says

> [i]t is striking that a great many foundational theories are linearly ordered by $<$. Of course it is possible to construct pairs of artificial theories which are incomparable under $<$. However, this is not the case for the 'natural' or non-artificial theories which are usually regarded as significant in the foundations of mathematics. The problem of explaining this observed regularity is a challenge for future foundational research.

(Simpson's relation '$<$' is defined similarly enough to our $\leqslant_{\text{Con}}$ that the difference will not be relevant.) In this part of the essay, I will attempt to assess the extent to which this linearity phenomenon is genuine, and if it is, what we may conclude from this. I will begin by explaining why we do not have linearity outright and thus must restrict our attention to a particular class of theories. I will then discuss what we might hope to gain from such a linearity phenomenon: to that end I will introduce Gödel's program and the universe view of set theory, as well as considering linearity's implications within mathematics. I'll then move to a discussion of whether the hierarchy is indeed linear in 'natural' theories.

4. The consistency strength hierarchy is not linear outright

When discussing linearity of the consistency strength hierarchy what we really mean is linearity of the order $\leqslant_{\text{Con}}$: that for any theories T, U which extend ZFC, we have at least one of $T \leqslant_{\text{Con}} U$ and $U \leqslant_{\text{Con}} T$. The stronger claim is in fact sometimes made that this order is a well-order (see [Ste12, p5]), so that every descending chain of theories has a $\leqslant_{\text{Con}}$-minimal element; this slightly stronger condition will not concern us in this paper. However note that a proposed counter-example is given in Section 4 of [Ham21], using the 'cautious enumeration', ZFC°, of ZFC. The enumeration works by listing the theorems of ZFC until a proof of a contradiction is found, at which point it halts. We then have:

Proposition 4.1. $\mathsf{ZFC}^\circ <_{\text{Con}} \mathsf{ZFC}$.

Proof. Clearly we have $\mathsf{ZFC}^\circ \subseteq \mathsf{ZFC}$, and thus $\mathsf{ZFC} \vdash \operatorname{Con}\mathsf{ZFC} \to \operatorname{Con}\mathsf{ZFC}^\circ$. On the other hand, suppose that $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC}$ is consistent; then we can find a model of $\mathsf{ZFC} + \operatorname{Con}(\mathsf{ZFC}) + \neg\operatorname{Con}(\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC})$ by Gödel's second incompleteness theorem. In this model, any model of ZFC must have $\neg\operatorname{Con}\mathsf{ZFC}$ (since $\mathcal{M} \vDash \neg\operatorname{Con}(\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC})$), thus the model believes that there is a proof from ZFC that ZFC is not consistent. So the model believes that the enumeration of ZFC° will halt after finitely many iterations. It is well-known that $\mathsf{ZFC} \vdash \operatorname{Con}\Gamma$ for any finite $\Gamma \subseteq \mathsf{ZFC}$, the idea being we can conjunct the elements of Γ to obtain a single formula, and then Lévy reflect to find a model of this formula, which suffices for consistency; for details see [Kun14, pp131–132]. These considerations give us that the model believes $\mathsf{ZFC} \vdash \operatorname{Con}\mathsf{ZFC}^\circ$. By construction this model also believes that $\mathsf{ZFC} + \neg\operatorname{Con}\mathsf{ZFC}$ is consistent, and thus with this and the above, we may construct a model inside our original of $\mathsf{ZFC} + \neg\operatorname{Con}(\mathsf{ZFC}) + \operatorname{Con}\mathsf{ZFC}^\circ$. Thus we arrive at the conclusion $\mathsf{ZFC}^\circ <_{\text{Con}} \mathsf{ZFC}$. ⊣

Note that we may then iterate this construction to get $\mathsf{ZFC}^{\circ\circ} <_{\text{Con}} \mathsf{ZFC}^{\circ} <_{\text{Con}} \mathsf{ZFC}$, where $\mathsf{ZFC}^{\circ\circ}$ is obtained by enumerating ZFC until either a proof of a contradiction, or a proof that there is such a proof, is found. This continues to obtain an ill-founded chain of theories. It would take us beyond the scope of this paper to examine whether this example suffices to demonstrate natural ill-foundedness, though this certainly seems an interesting area for future study.

For $\leqslant_{\text{Con}}$ to be linear outright would mean that for any theories T, U extending ZFC, we have at least one of $T \leqslant_{\text{Con}} U$ and $U \leqslant_{\text{Con}} T$. That is, any two theories have comparable consistency strengths. This is the most simple, and therefore arguably would be the most desirable, kind of linearity for the hierarchy to have. Due to the work of Gödel and those following, however, it is now well-known that this is not the case, and we can easily produce examples of non-comparability.

The following result is sufficient to see that the consistency strength hierarchy is not linear *simpliciter*. The proof is adapted from Theorem 2 of [Ham21]: it was originally stated for PA, so I have extended it to ZFC, and expanded the explanation significantly.

Theorem 4.2. *There is a sentence* R *in the language of set theory such that* $\mathsf{ZFC} + \mathrm{R} \nleqslant_{\text{Con}} \mathsf{ZFC} + \neg\mathrm{R}$ *and* $\mathsf{ZFC} + \neg\mathrm{R} \nleqslant_{\text{Con}} \mathsf{ZFC} + \mathrm{R}$*.*

Proof. As in Theorem 1.26 let R be the Rosser sentence for $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC}$, so that R asserts that for any proof of R from $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC}$, there is a shorter proof of $\neg$R from $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC}$. We know that if $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC}$ is consistent, then (i) $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC} \nvdash \mathrm{R}$, and (ii) $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC} \nvdash \neg\mathrm{R}$.

For (i) this is because if R were provable in this theory, then by its definition, this would show that we can prove $\neg$R also (and that this proof will be shorter), which would show that $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC}$ is inconsistent (since it proves both a statement and its negation). For (ii) we reason similarly: if $\neg$R were provable in this theory, then there would be a proof of R (with no shorter proof of $\neg$R) also, so we similarly contradict our assumption of consistency.

We will now use these facts to give a model $\mathcal{M}$ of

$$\mathsf{ZFC} + \operatorname{Con}(\mathsf{ZFC} + \neg\mathrm{R}) + \neg\operatorname{Con}(\mathsf{ZFC} + \mathrm{R}),$$

and a model $\mathcal{N}$ of

$$\mathsf{ZFC} + \operatorname{Con}(\mathsf{ZFC} + \mathrm{R}) \ + \neg\operatorname{Con}(\mathsf{ZFC} + \neg\mathrm{R}).$$

This suffices for our conclusion, since this shows

$$\begin{aligned} \mathsf{ZFC} &\nvdash \operatorname{Con}(\mathsf{ZFC} + \neg\mathrm{R}) \rightarrow \operatorname{Con}(\mathsf{ZFC} + \mathrm{R}) \\ \mathsf{ZFC} &\nvdash \operatorname{Con}(\mathsf{ZFC} + \mathrm{R}) \ \rightarrow \operatorname{Con}(\mathsf{ZFC} + \neg\mathrm{R}) \end{aligned}$$

by the semantics of $\rightarrow$.

(i) gives us that there is a model $\mathcal{M}$ of $\mathsf{ZFC} + \operatorname{Con}(\mathsf{ZFC}) + \neg\mathrm{R}$: if every model of $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC}$ believed R, then by completeness we would have that $\mathsf{ZFC} + \operatorname{Con}\mathsf{ZFC} \vdash \mathrm{R}$, which we know is not the case. Thus inside $\mathcal{M}$, by the definition of R, there is a proof of R such that there is no shorter proof of $\neg$R. Thus this model believes that $\mathsf{ZFC} \vdash \neg\mathrm{R}$, and thus $\operatorname{Con}(\mathsf{ZFC} + \neg\mathrm{R})$, since $\mathcal{M}$ believes that ZFC is consistent. (It is a basic fact of logic that if T is consistent and $T \vdash \varphi$, then $T + \varphi$ is consistent.)

Further, we must have that $\mathcal{M}$ believes $\neg \operatorname{Con}(\mathsf{ZFC} + \mathrm{R})$: from before, $\mathcal{M}$ believes that $\mathsf{ZFC} \vdash \neg\mathrm{R}$, and thus must also believe $\mathsf{ZFC} + \mathrm{R} \vdash \neg\mathrm{R}$. However clearly $\mathsf{ZFC} + \mathrm{R} \vdash \mathrm{R}$, thus this theory must be inconsistent according to $\mathcal{M}$.

Thus we have a model of $\operatorname{Con}(\mathsf{ZFC} + \neg\mathrm{R}) + \neg \operatorname{Con}(\mathsf{ZFC} + \mathrm{R})$, as required.

Now turning to (ii). By Theorem 1.27, the theory $\mathsf{ZFC} + \operatorname{Con}(\mathsf{ZFC}) + \mathrm{R}$ can't prove its own consistency, and thus there is a model $\mathcal{N}$ which believes that this theory is inconsistent. Note that in general if a model believes that a theory $T + \varphi$ is inconsistent, then it thinks that any model of T must have $\neg\varphi$ true (since there is no model of an inconsistent theory). Then by completeness, as above, we get that this model must believe that $T \vdash \neg\varphi$. Applying this here with $T = \mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC}$ and $\varphi = \mathrm{R}$, we have that $\mathcal{N}$ must believe that $\mathsf{ZFC} + \operatorname{Con} \mathsf{ZFC} \vdash \neg\mathrm{R}$.

Since R is true in $\mathcal{N}$, this means that the model believes that there is a proof of $\neg\mathrm{R}$ such that there is no smaller proof of R (this is just the definition of R restated, given that we have a proof of $\neg\mathrm{R}$). Then $\mathcal{N}$ believes that $\mathsf{ZFC} \vdash \mathrm{R}$, so as above believes $\operatorname{Con}(\mathsf{ZFC} + \mathrm{R})$. But since $\mathcal{N}$ believes ZFC is consistent, inside $\mathcal{N}$ we must then have $\mathsf{ZFC} \nvdash \neg\mathrm{R}$ by identical logic to the case for (i). Thus by the same logic as above, we have a model of $\operatorname{Con}(\mathsf{ZFC} + \mathrm{R}) + \neg \operatorname{Con}(\mathsf{ZFC} + \neg\mathrm{R})$.

We have given our two models, and so the proof is complete. $\dashv$

Thus we have seen that the consistency strength hierarchy does indeed contain some non-linearity. In light of this I will now discuss what linearity in the class of 'natural' theories might give us, both philosophically and mathematically.

5. Why might we want consistency strength hierarchy linearity?

Before we examine the question of linearity in natural theories, some consideration must be given to what such a result might hope to achieve. Central to this are Gödel's program and the universe view of set theory, which I will now briefly introduce.

5.1. **The universe view.** The 'universe view' of set theory [Ham20b, p286] claims that there is a unique set-theoretic reality, within which all set-theoretic questions have determinate answers (though we may not know these answers – see Subsection 5.2). Since all the mathematics we know may be interpreted in set theory, this could be argued to extend to the claim that within such a universe, all mathematical questions have determinate answers. Gödel is well-known to have been an advocate of the universe view (though he did not call it this): see Subsection 5.2 for more details on this. The universist typically believes that there is a correct set of axioms extending ZFC with which we should do set theory, and therefore that set-theoretic principles which conflict with these axioms should not be accepted, even if we can study their consequences. For example, if our correct set of axioms proves CH (see Subsection 5.2), then the universist would argue that the forcing extensions of Cohen in which $\neg$CH holds are simply misguided or else just a formalism, and have no set-theoretic reality underlying them.

In opposition to the universe view, there is the 'multiverse view' [Ham12, p1], which claims that there are many valid concepts of set, each of which gives rise to its own set-theoretic universe. It claims that modern work on forcing and inner model theory has given us large amounts of experience in these different universes, allowing us to choose which set-theoretic principles we want to hold: there are universes with CH, universes without, universes with inaccessible cardinals, and so on.

Each of these positions raise a host of ontological questions concerning the nature of the objects to which we are referring: are they of the same sort as objects in the physical world? how do we have access to these objects? and so on. Interesting though these questions are, this essay will be more concerned with how results about the consistency strength hierarchy can be brought to bear on the issue of universism more broadly, and thus will sidestep these concerns.

5.2. **Gödel's program.** In [Ste12, p1], Steel describes Gödel's program as the task of 'decid[ing] mathematically interesting questions independent of ZFC in well-justified extensions of ZFC'. It is worth making a few clarifications about exactly what I will take this to mean.

Gödel's philosophy of maths was not always clear – his account of intuition in [Göd83] has been described as 'one of the most difficult and obscure passage[s] in [his] finished philosophical writings' [Par95, p67] – though it is clear that he held some form of realism to be true. This is to say that he believed that there is a 'well-determined' [Göd83, p476] mathematical reality, which exists independently of our experience of it. The crucial corollary of this is that Gödel believed that all mathematical questions ought to have determinate answers – in his case he was considering the continuum hypothesis, which he suspected would eventually be answered in the negative [Göd83, p480].

The continuum hypothesis, CH, first investigated by Cantor, is the claim that the cardinality of the set of real numbers is the first uncountable cardinal, $\aleph_1$ (see Definition 1.8). We can prove that the cardinality of the reals is $2^{\aleph_0}$, and thus the continuum hypothesis equivalently asserts that $2^{\aleph_0} = \aleph_1$. Since by definition $\aleph_1$ is the cardinal successor of $\aleph_0$, this leads to a natural generalisation of CH, the *generalised* continuum hypothesis, GCH, which asserts that the cardinal successor operation is always the power set operation, so we have $\aleph_\alpha^+ := \aleph_{\alpha+1} = 2^{\aleph_\alpha}$ for all $\alpha \in \mathrm{On}$. It is now well known that CH and GCH are independent of ZFC: in the late 1930s Gödel showed in [Göd38] and [Göd39] using his constructible universe L that if ZFC is consistent, then so is $\mathsf{ZFC} + (\mathsf{G})\mathsf{CH}$. Later, in 1963 Paul Cohen published a result to the contrary, [Coh63], showing that if ZFC is consistent, then so is $\mathsf{ZFC} + \neg\mathsf{CH}$. These results taken together show us that the continuum hypothesis must be independent of ZFC:

Corollary 5.1. *If* ZFC *is consistent, then* $\mathsf{ZFC} \nvdash \mathsf{CH}, \neg\mathsf{CH}$.

Proof. I will show $\mathsf{ZFC} \nvdash \mathsf{CH}$. The corresponding proof for $\neg\mathsf{CH}$ is essentially identical, however uses Gödel's result rather than Cohen's.

Suppose for contradiction that $\mathsf{ZFC} \vdash \mathsf{CH}$. Then we also have $\mathsf{ZFC}+\neg\mathsf{CH} \vdash \mathsf{CH}$, since a proof in ZFC also counts as a proof in $\mathsf{ZFC}+\neg\mathsf{CH}$. We also have $\mathsf{ZFC}+\neg\mathsf{CH} \vdash \neg\mathsf{CH}$, thus $\mathsf{ZFC} + \neg\mathsf{CH}$ is inconsistent, and thus by Cohen's result so is ZFC, which contradicts our assumption. $\dashv$

To this day, the status of CH is not settled in the mathematical community: see [Ham15] for more on this. Gödel himself believed that it would would eventually be proven false by addition of further axioms to ZFC. An interpretation of some of Gödel's philosophical beliefs, including around CH, is given in [Mad00, II.1, III.2].

Before I move to explain Gödel's program, a brief note must be made on ontology and epistemology. We must be careful not to conflate the ontological claims of realism with any claims about knowability – it is possible that one could believe there is a unique set-theoretic reality in which all mathematical questions are settled, without also believing that we can hope to know the answers to these questions ourselves (even in principle). Questions whose answers we may never know (even if they do have answers) have been termed 'absolutely undecidable', and are considered for example in [Koe06], where Koellner concludes that we don't at present have good reason to believe that we know of any such questions. Gödel eventually moved towards the view that not only did all mathematical questions have determinate answers, but also we could expect to know such answers. He was open to the idea that there could be a a 'generalized completeness theorem' [Koe06, p10], 'which would say that every proposition expressible in set theory is decidable from the present axioms plus some true assertions about the largeness of the universe of all sets' [Göd90, p151].

With this clarified, we arrive at Steel's characterisation of Gödel's program above. What we take from Gödel is his belief that all set-theoretic, and thus all mathematical questions, have determinate, objective answers, and his hope that we could come to know these answers. Thus stated, these views are a strengthening of the universe perspective, adding to it a positive epistemological claim. For an example: though we can construct models where CH is true and models where CH is false, one collection of these models is misguided, and does not accurately represent mathematical reality, and we might hope to find some principle which tells us which collection this is.

I will briefly pause on a potential objection to Gödel's program, since if it held true, then it seems the program would be misguided, and our questions about linearity perhaps moot.

5.2.1. *Epistemological issues.* The objection pushes back on the idea underlying Gödel's program that, if we are able to climb the hierarchy in a unique way, this supports the epistemological claim that this will answer all our mathematical questions. This issue is considered in part in Koellner's [Koe06], as mentioned above: if climbing arbitrarily high in the ladder of consistency strength failed to decide a particular question, then such a question would be a strong candidate for an absolutely undecidable statement. It is outside the scope of the paper to detail Koellner's arguments here. However I will note that there is a quantity of evidence suggesting that climbing the consistency strength hierarchy – typically by assuming larger large-cardinal axioms – allows us to answer more questions. For example, if there are infinitely many Woodin cardinals, then the axiom of projective determinacy (which is about two-player infinite games) holds (see [MS89]).To the contrary, the continuum hypothesis is not settled by any known large-cardinal axiom (provided that we do not take $V = L$ to be a large-cardinal axiom): this is a result of Lévy and Solovay, generalising the one given in [LS67]. We have that (following its statement in [Ham15, p136]):

Theorem 5.2. *The set-theoretic universe V has forcing extensions (i) $V[G] \models \neg\mathsf{CH}$, where $V[G]$ collapses no cardinals (ii) $V[H] \models \mathsf{CH}$, where $V[H]$ has no new reals.*

It is beyond this paper to detail exactly what this theorem means. However the key takeaway is that given any large-cardinal axiom we want (since these only ever affect what is inside V), we can always 'force' CH or $\neg\mathsf{CH}$, to our liking. Thus, large-cardinal axioms will never settle CH.

On balance, however, I believe that were the consistency strength hierarchy to be linear, this would support that Gödel was right in his suspicions about generalised completeness. Whilst CH is yet unsettled by a sufficiently tall rung in the consistency strength ladder, this stands out against a background of many problems which *have* been decided in this way: we have a quantity of evidence suggesting that climbing the hierarchy does provide answers to all our questions, and our only evidence against is that as yet we are unsure about some propositions. Lévy-Solovay is a powerful result and certainly provides some credibility to the view that CH might not be settled by height in the hierarchy. However new set-theoretic principles rich in consequences which do not take the form of large cardinals, such as determinacy axioms, are being studied increasingly more. On this basis I would argue that Gödel's program is still a worthwhile endeavour.

5.3. **Philosophical implications of linearity.**

5.3.1. *Linearity builds the road.* Central to Gödel's program is the idea that we can ascribe a 'strength' to a particular extension of ZFC: for Gödel this manifested in his belief that the use of increasingly strong 'axioms of infinity' (referred to today as 'large-cardinal axioms') would answer an increasing number of mathematical questions. For the modern advocate of Gödel's program, such as John Steel, this is seen in talk of the 'one road upward' [Ste13, Slide 5], and the idea that we can 'climb' the consistency strength hierarchy [Ste12, p5, p6, p7, p8]. (Strictly speaking, Steel's 'one road upward' remark was just in reference to a particular instance of linearity at the level of sentences about real numbers, though it should be clear that this view applies elsewhere.) In both cases, what is important is that there is a *unique* way to extend a particular theory extending ZFC: if our extension is not unique, so there are two or more of possible extensions, then our hierarchy may branch off in many directions. In this case it is not clear how we are to assign 'strength' to any particular extension, or indeed choose between different branches when we try and 'climb'. Linearity gives us this uniqueness, since for each node in our hierarchy, there can be only one directly above it.

It is noted by Steel [Ste13, Slide 6], we don't need full linearity to arrive at the 'one road', merely *directedness*, where we recall:

Definition 5.3. A partially ordered set $(\mathbb{P}, \leqslant)$ is *directed* if and only if for any $x, y \in \mathbb{P}$, there exists a $z \in \mathbb{P}$ with $x, y \leqslant z$. $\lhd$

The reason for this is that directedness allows us to keep climbing in a unique way by simply choosing an upper bound for any given two theories. It might be objected that, absent linearity, the upper bound for two nodes in a directed poset need not be unique. In particular then, it is not the case in general that if any two theories have an upper bound, that we recover the 'one road upwards', since this upper bound might sit amongst many, all of which may be incomparable. Consider Figure 2 for an example of such a situation. This objection does not damage Steel's point however, since we can simply iterate, and find an upper bound for these incomparable upper bounds, and so on. We are still able to climb in a unique way, since we can always find a theory which bounds any two given theories. These remarks on directedness will concern us in Subsection 7.1.

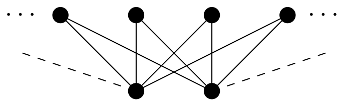


Figure 2. Two theories (the two nodes along the bottom) may have many upper bounds (the nodes along the top).

In any case, linearity would give us a unique way of climbing the consistency strength hierarchy, and thus would provide the advocate of Gödel's program with the notion of 'strength' or 'height' which they require for their project to make sense.

5.3.2. *The universe view.* Though the universe view underpins Gödel's program, we note that consistency strength linearity could also be used as an argument in favour of this position, and against multiversism. In this way it is doubly useful to an advocate of the program.

More specifically, if our theories do sit in a linear hierarchy of consistency strength, then this seems to suggest that there is a determinate set-theoretic reality. The 'determinate' aspect of this characterisation ought to be emphasised more strongly here than the 'reality' part – the existence of a linear hierarchy of consistency strength presumably tells us nothing about ontology. Whatever the ontology of this universe, and despite any epistemological issues we may have (*à la* Koellner), it doesn't seem like the freedom required for multiversism, which relies on maintaining a plurality of possible theories, is possible if we have such a hierarchy.

5.4. **Linearity and mathematics.** On the more mathematical side, linearity would surely be a desirable result, if we can attain it. Not only would it be a beautiful phenomenon in its own right – one described by H. Friedman as one of the 'great mysteries' [Fri98] of the foundations of mathematics – but it would also provide an elegant classification framework for theories extending ZFC: we simply aim to find a rung on the ladder with which the theory is equiconsistent. A similar classification problem arises even if the hierarchy does not have such a structure, however this is certainly a less pleasing phenomenon. (It should be noted that these are purely mathematical considerations – a theory being pleasing is not here being taken as evidence for its truth, as surely it would be 'pleasing' if powerful theories could prove their own consistencies.)

Having considered why we may want linearity to hold for the consistency strength hierarchy, I will now move to consider the phenomenon in more depth, before I consider whether, given what I will conclude below, any of these goals succeed.

6. Evidence for linearity in 'natural' theories

As Hamkins notes [Ham21, p8], 'nobody likes' the example of non-linearity given in Theorem 4.2, the reason being that the theories considered are not 'natural'. Whilst the concept of 'naturality' presumed by most mathematicians working in this field has been called into question [Ham21, Section 9], for present purposes note that *prima facie* there is a genuine concept being identified here (all five of the mathematicians quoted by Hamkins mention naturality or something intended as a synonym). Some possible glosses could be 'not constructed specifically to demonstrate non-linearity', or else 'occurring in the normal studies of a set theorist interested in extensions of ZFC'. Making for now the (potentially quite large) assumption that we have a reliable way of distinguishing the 'natural' theories from the 'unnatural', I will consider here evidence for consistency strength linearity in the natural theories.

As noted above, linearity in natural theories seems widely accepted among set theorists; in his [Ste12, p5] Steel conjectures that 'if T and U are natural extensions of ZFC, then either $T \leqslant_{\text{Con}} U$ or $U \leqslant_{\text{Con}} T$.' Indeed most appear to agree that no counter-example has yet been found; in conjunction with the many instances of linearity we have proved, this appears to be good inductive evidence for his claim.

A particular case study of this is seen in the cardinals we defined in Part I: as proved by Theorem 3.6, the worldly and otherworldly cardinals, as well as all their hyper-extensions, fit linearly below the inaccessible cardinals. In defence of the 'naturality' of these concepts (despite our avoidance of defining exactly what 'naturality' means), we might note that the worldly cardinals arise as the smallest cardinals which provide models of ZFC, surely of interest to set theorists, whilst the otherworldly cardinals have been described as occurring 'frequently' in set-theoretic studies by Joel David Hamkins (a set-theorist) on his blog [Ham20a]. Moreover Hamkins claims he was introduced to the idea by a PhD student in set theory by a tweet here [Che20], so these cardinals must belong to the class of those 'studied by set theorists'. As for the hyper-extensions, these were defined in line with typical hyper-extensions of large-cardinal concepts. In particular, the hyperworldly cardinals and their extensions essentially copied their definition from those of the hyper-inaccessible cardinals from [Car15, ch2].

Definition 6.1. A cardinal κ is α-inaccessible if and only if it is inaccessible and for every $\beta < \alpha$, κ is a limit of β-otherworldly cardinals. ◁

Definition 6.2. A cardinal κ is α-hyper$^\beta$-inaccessible if and only if

(i) κ is inaccessible;
(ii) for all $\eta < \beta$, κ is κ-hyper$^\eta$-inaccessible;
(iii) for all $\gamma < \alpha$, κ is a limit of γ-hyper$^\beta$-inaccessible cardinals. ◁

Compare these with Definitions 2.2 and 2.6, respectively.

7. Questioning linearity in 'natural' theories

I will now turn to consider how we might call into question the orthodoxy surrounding natural linearity in the consistency strength hierarchy. There are a number of ways we might do this: we could offer examples of non-linearity which are natural, we could argue that our current evidence base in fact does not support this conclusion, or we could reject that the notion of 'naturality' has a sufficiently precise, non-circular definition which is able to support the conclusion. In the final case, this is not strictly speaking evidence against linearity in natural theories, but is instead to be taken as a rejection of the question more broadly. We will not have sufficient space in this essay to consider this last option, however it is certainly an area worth further study. A survey of the concept of naturality (or 'naturalness') in mathematics is given in [MV15]; in his [Bag04], Joan Bagaria suggests some criteria for axioms to count as natural: maximality, fairness, consistency, and success. For more see the works cited.

Note that Hamkins' paper only considers theories which posit the existence of certain large cardinals – these being by far the most common extensions of ZFC. It has been suggested that for any natural extension T of ZFC we can find a large-cardinal hypothesis H with $T \equiv_{\text{Con}} H$, see for example [Ste12, p5], though we will not pursue this claim any further. For our purposes, it suffices to note that if we seek to demonstrate non-linearity, then an example from the realm of large cardinals will suffice. On the other hand when seeking to evidence linearity, we must note that the majority of theories extending ZFC currently studied by set theorists begin with some large-cardinal assumption, and thus evidence for linearity among theories defined via large cardinals constitutes good inductive evidence for linearity among all theories.

7.1. **Natural examples of non-linearity.** Perhaps the easiest way to reject the view that the consistency strength hierarchy is linear in natural theories would be to provide natural theories which witness this. Joel David Hamkins suggests in [Ham21, Sec2] that there is an abundance of such examples. Hamkins' result is:

Theorem 7.1. *There is a computable function f such that for $n \neq m$, we have that* $\mathsf{ZFC} \nvdash \text{Con}(\mathsf{ZFC} + \mathsf{I}_n) \to \text{Con}(\mathsf{ZFC} + \mathsf{I}_m)$*, where I_k asserts that there are $f(k)$ inaccessible cardinals.*

Stated using $\leqslant_{\text{Con}}$, this theorem gives us countably many instances of $\leqslant_{\text{Con}}$-incomparability – whenever $n \neq m$ it is neither the case that $\mathsf{ZFC} + \mathsf{I}_n \leqslant_{\text{Con}} \mathsf{ZFC} + \mathsf{I}_m$ nor that $\mathsf{ZFC} + \mathsf{I}_m \leqslant_{\text{Con}} \mathsf{ZFC} + \mathsf{I}_n$.

The proof of this statement uses ideas from computability theory, in particular the universal function (which is a computable function which can replicate the behaviour of all other computable functions). The details are beyond the knowledge of the author, however will not be relevant to the conclusions we draw from this result.

As Hamkins notes ([Ham21, p10]), this example not only gives an instance of non-linearity at a specific level of the consistency strength hierarchy, but in fact at many levels – the result clearly goes through if we replace 'inaccessible' with any (consistent) large-cardinal concept we like.

There are a number of possible replies to Hamkins' result. The first could be to question in what sense, if any, these theories count as 'natural'. Another reply, which accepts the naturality of the construction, notes that whilst this does then count as non-linearity in the hierarchy, this doesn't mean the same philosophical conclusions can't be drawn as from full linearity. In particular, it seems to be the case that above any collection of instances of non-linearity, we can find a stronger principle which implies them all. For example if we define a cardinal to be 1-inaccessible if and only if it is inaccessible and a limit of inaccessible cardinals, as in Definition 6.1, then ZFC + 'there is a 1-inaccessible cardinal' has consistency strength greater than any $\mathsf{ZFC} + \mathsf{I}_n$ for $n \in \omega$. More formally:

Proposition 7.2. $\mathsf{ZFC} + \mathsf{I}_n <_{\text{Con}} \mathsf{ZFC}$ + 'there is a 1-inaccessible cardinal' *for any $n \in \omega$.*

Proof. Since $f(n)$ is (if defined) a natural number, it suffices to show that ZFC + 'there is a 1-inaccessible cardinal' $\vdash \text{Con}(\mathsf{ZFC}$ + 'there are n inaccessible cardinals') for any $n \in \omega$. Suppose that there is a 1-inaccessible cardinal κ, then κ is a limit of inaccessible cardinals; since κ is also inaccessible and thus regular, we have that there must be κ inaccessible cardinals less than κ. In particular since $\kappa > \aleph_0$, there are more than n inaccessible cardinals for any $n \in \omega$. Since this will hold in any model where there is a 1-inaccessible cardinal, we are done by completeness. $\dashv$

This result gives us that even if Hamkins' construction is natural, we could still hope for directedness in the consistency strength hierarchy, where we recall Definition 5.3. Indeed, we may have as strong evidence for directedness as in Section 6, which may well be sufficient to draw the philosophical conclusions we want. Reformulating the central claim of interest to say that for any T, U which are natural extensions of ZFC there is a natural extension S such that $T, U \leqslant_{\mathrm{Con}} S$, we still arrive at John Steel's 'one road upward': since we are always able to 'unify' any two theories generated as in Theorem 7.1 with a stronger one above, we still have a useful notion of 'height'. If we note that Hamkins' construction only enables us to posit finite numbers of particular types of cardinals, we can see we will still obtain directedness, absent examples of non-linearity of a different type. Since we can always posit the theory which asserts that there is a proper class of cardinals of a particular type, and this theory will be strictly stronger than any theory positing only finitely many such cardinals, we can still get directedness.

Thus even if we take Theorem 7.1 to be a genuine example of natural non-linearity, we needn't abandon any of the philosophical goals of Subsection 5.3: it seems we are just as able to climb the hierarchy in a unique way as we thought before. The only case in which we lose this uniqueness is if we wanted to stop at one of the $(\mathsf{ZFC} + \mathsf{I}_k)$-theories, though I am not aware of any arguments that this would ever be the case, rather than continuing up the hierarchy to the next type of large cardinal. Indeed there is a school of thought among set theorists which argues that we should always seek principles which 'maximise', see for example [Mad88, II.2]. The epistemological goal of Gödel's program is also unaffected by this result. In particular, if we believe that climbing the hierarchy will offer answers to more and more questions, since we are still able to climb the hierarchy in a unique way, we should keep this belief. I will now turn to consider arguments which call into question whether our current evidence base supports natural linearity.

7.2. **Does our current evidence support natural linearity?** Two main arguments are raised in [Ham21] which call into question whether our current evidence for natural linearity is as strong as commonly assumed. The first says that since our large-cardinal concepts are constructed in similar ways – frequently as critical points of elementary embeddings of the universe V into a transitive class M – in many cases it is entirely unsurprising that we get linearity. Furthermore, there is also the suggestion that if we were to move beyond this construction of large cardinals, we would then lose this linearity phenomenon. The second argues that we suffer from confirmation bias in asserting linearity, since our methods for constructing models (forcing and inner model theory) preserve arithmetic truth, and thus preserve consistency statements. I will consider these arguments in turn, critically assessing them and drawing some new conclusions.

7.2.1. *Similarity of construction.* The first argument notes that many of our large-cardinal concepts were created in ways which were intended as strengthenings of concepts we already had: this has been seen in this paper with the construction of both the hyper-hierarchies, from whose definitions we were easily able to prove linearity. As noted in Subsection 6, this type of 'hyper'-construction occurs elsewhere in the study of large cardinals: as noted above [Car15] greatly generalises the notion of hyper-inaccessible to higher and higher levels. The same sort of construction arises if we consider Mahlo cardinals, which are defined as follows. Recall the definitions of club and stationary sets, 1.13, 1.14, respectively.

Definition 7.3. A cardinal κ is *Mahlo* if and only if the set $\{\alpha < \kappa \mid \alpha \text{ is inaccessible}\}$ is stationary in κ. $\lhd$

We may then define, following [Car15, p28].

Definition 7.4. A cardinal κ is *α-Mahlo* if and only if it is Mahlo, and for all $\beta < \alpha$, the set of β-Mahlo cardinals less than κ is stationary in κ. κ is *hyper-Mahlo* if and only if it is κ-Mahlo. $\lhd$

Away in consistency strength from the smaller large cardinals, many large-cardinal concepts are now defined via non-trivial elementary embeddings $j\colon M \to N$, where M, N are transitive classes (this allows $M, N = V$). We have the following result.

Proposition 7.5. *If $j\colon M \to N$ is a non-trivial elementary embedding and $N \subseteq M$, then there is an ordinal κ with $j(\kappa) \neq \kappa$; at this κ we in fact have $j(\kappa) > \kappa$.*

Proof. By hypothesis there is some $x \in M$ of minimal rank such that $j(x) \neq x$. If $y \in x$, then y has lower rank than x and thus $j(y) = y$ by the minimality of x. Further, by elementarity $j(y) \in j(x)$, and thus $y \in j(x)$. Therefore $x \subsetneq j(x)$, and in particular then $\operatorname{rank} x \leqslant \operatorname{rank} j(x)$.

Let $z \in j(x) \setminus x$, and suppose for contradiction that $\operatorname{rank} j(x) = \operatorname{rank} x = \kappa$. By assumption z has lower rank than $j(x)$, thus by our assumption lower rank than x. Therefore we must have $z = j(z)$, and so $j(z) \in j(x)$, which by elementarity implies that $z \in x$, which contradicts the definition of z. This gives us that $\operatorname{rank} x < \operatorname{rank} j(x)$.

Finally note that j is elementary, and so for any formula φ and $x, y \in M$, we have $M \vDash \varphi(x, y)$ if and only if $N \vDash \varphi(j(x), j(y))$. By definition $M \vDash \operatorname{rank} x = \kappa$, hence $N \vDash \operatorname{rank} j(x) = j(\kappa) = j(\operatorname{rank} x)$, so combining with the above, and the fact that rank is absolute between transitive models ([Kan08, pp5–6]) we get that $\kappa < j(\kappa)$, as required. $\dashv$

The minimal such κ, which must exist by the well-ordering of On, is called the *critical point* of j. We can define various large cardinals as the critical points of elementary embeddings $j\colon M \to N$, where we posit various conditions on j, M, and N to vary the result.

Example 7.6.

(i) A cardinal κ is *weakly compact* if and only if for every transitive set M with cardinality κ such that $\kappa \in M$, there is a transitive set N and an elementary embedding $j\colon M \to N$ with critical point κ. (Note the existence of this critical point does not follow from 7.5 since we don't necessarily have $N \subseteq M$; see [Kan08, Proposition 5.1(b)] for a proof that we still have a critical point in this case, as long as $M \vDash \mathsf{AC}$.)

(ii) A cardinal κ is *γ-strong* if and only if if is the critical point of a non-trivial elementary embedding $j\colon V \to M$, $\gamma < j(\gamma)$, and $V_{\kappa+\gamma} \subseteq M$. A cardinal is *strong* if and only if it is γ-strong for arbitrarily large $\gamma \in \mathrm{On}$.

(iii) A cardinal κ is *superstrong* if and only if it is the critical point of a non-trivial elementary embedding $j\colon V \to M$ and $V_{j(\kappa)} \subseteq M$.

(iv) A cardinal is *θ-supercompact* if and only if it is the critical point of a non-trivial elementary embedding $j\colon V \to M$ and $M^\theta \subseteq M$, where M^θ denotes the set of all functions $\theta \to M$. A cardinal is *supercompact* if and only if it is θ-supercompact for all $\theta \in \mathrm{On}$.

(v) A cardinal is *Reinhardt* if and only if it is the critical point of a non-trivial elementary embedding $j\colon V \to V$. It was shown by Kunen in [Kun71] that the existence of such a cardinal is inconsistent with ZFC; the corresponding result for ZF is still an open problem. For more see [BKW19].

Before we detail Hamkins' arguments more fully, it is worth clarifying that he is not suggesting that we can explain away *all* the linearity phenomena we have observed by claiming that all the large-cardinal concepts we are interested in are all simple variants on a central theme. With this said, it will be worth briefly dwelling on this line of argument.

First we may note this is patently not the case: there exist large cardinals, such as the worldly cardinals, for which no characterisation in terms of elementary embeddings is currently known. With this said, as the area has matured, an increasing number of large cardinals have been given elementary embedding characterisations. Quite recently at the 'small' end of the hierarchy, Victoria Gitman [Git20, Slide 6] has given a necessary condition for inaccessibility in terms of elementary embeddings. However even if all known large-cardinal concepts could be characterised in this way, I would argue that this would still not explain away the phenomenon of natural linearity. For one, for many cardinals their elementary embedding characterisation came after they were initially defined. A good example of this is the measurable cardinals, which have the following equivalent definitions (among others: see [Va]). A brief history follows below.

We recall the definition of a measure.

Definition 7.7. Given an infinite set S, a *non-trivial, σ-additive measure* on S is a function $m\colon \mathcal{P}(S) \to [0, \infty]$ such that:

(i) $m(\varnothing) = 0$, $m(S) > 0$;
(ii) if $X \subseteq Y \subseteq S$, then $m(X) \leqslant m(Y)$;
(iii) if $X_i \subseteq S, i \in \omega$ are such that $X_i \cap X_j \neq \varnothing$ for $i < j$, then

$$m\left(\bigcup_{n\in\omega} X_n\right) = \sum_{n=0}^{\infty} m(X_n).$$

If $\operatorname{im} m \subseteq \{0, 1\}$, then m is called *2-valued*. Note cases where the summands are infinite are defined as in [Tao11, pp xi–xii]. ◁

We then have:

Definition 7.8. A cardinal κ is measurable if and only if there is a 2-valued, σ-additive measure definable on κ. ◁

We also equivalently have the following definition:

Definition 7.9. A cardinal κ is measurable if and only if there is a transitive M and a non-trivial elementary embedding $j\colon V \to M$ with critical point κ. ◁

It is worth noting that because Definition 7.9 imposes no other conditions on j other than its domain being V, in order to characterise large cardinals smaller than the measurables with respect to direct implication, we will need to take a domain strictly smaller than V: this partly explains the definition of weakly compact cardinals given in Example 7.6(i), as these have smaller implication and consistency strength than the measurable cardinals.

The first of these definitions was investigated by Stanisław Ulam among others in the 1920s and 1930s, whilst the characterisation in terms of elementary embeddings had to wait until the work of Scott in the 1950s. For more of the historical background see [Kan08, pp22–27, p40]. The relevant point here is that whilst it is true that measurable cardinals admit an elegant characterisation in terms of elementary embeddings (indeed arguably the simplest such characterisation), this was a result in itself, and far from obvious.

Further to this, we can note that there are many different ways we vary our elementary embeddings to arrive at our various large cardinals, as in Example 7.6. I would argue it is far from obvious that these should then lead to linearity in consistency strength. More subtly than just the observation that the two closure conditions are not immediately related to each other, I would note that consistency strength does not always line up with the more intuitive notions of implication strength, and 'least instance' strength as mentioned in Subsection 3.3. As noted there, it can be the case that the other two hierarchies – whose configuration we might guess more easily from different elementary-embeddings definitions – can differ from consistency strength. In particular, it may be easy to see by comparing different definitions using elementary embeddings whether the smallest instance of one cardinal will be bigger than the smallest instance of another, though, given this result about strong and superstrong cardinals, it does not follow that we can assume much about their consistency strengths from this.

It should be conceded here that we do have the partial result that if we have direct implication, then by Theorem 3.10 we also have consistency strength implication. For an example of this, it was noted above that any non-trivial elementary embedding with domain V will have a critical point which is measurable; thus for any large cardinal defined in this way, we will also have that it has consistency strength of at least a measurable cardinal.

Hamkins is instead making the point that our stock of genuinely surprising instances of linearity are 'simply many fewer than one might have expected' [Ham21, p26]. Moreover, he adds that there are numerous instances where we have thus far failed to prove linearity, for example not much is known about how the strongly compact cardinals fit into the hierarchy (see [Ham21, p26]). Given the above remarks, though the unexpectedness of such instances is subjective, Hamkins clearly has a point: in any case where it is clear that we have direct implication between two large cardinals, as is often the case with those given elementary embedding characterisations, we know that we will also then have the corresponding result about consistency strength.

It is not clear to me however that these concerns present too much of an issue for the defender of natural linearity. The fact still remains that the hierarchy is, for all we know, linear in its natural theories. Even if many of these instances of linearity have simple explanations, this does nothing to counter the phenomenon as a whole. Moreover, as in the end of Section 6, given that we have a large stock of cases where we do have linearity, the fact that there are cases where we are unsure needn't necessarily damage our hope that we will prove linearity for these too. What we would need instead is a reason to believe that these cases *won't* turn out to be linear, which as of now I don't believe we do. With this said, Hamkins makes a further case which damages this point.

7.2.2. *Confirmation bias.* Hamkins raises a subtler argument against our current stock of evidence for linearity on page 27 of his paper: he claims that since our two main methods of constructing models for our theories preserve arithmetic truth, and consistency statements are arithmetic truths, it is entirely unsurprising that we will never see consistency strength non-linearity with these methods. I will now consider this in more detail, expanding on the presentation in [Ham21].

Our two main techniques in investigating models of extensions of $\mathsf{ZF(C)}$ are the method of forcing and the construction of transitive inner models. The former of these takes a transitive model M of ZFC and adjoins to it a new set G, which is constructed using a partial order $\mathbb{P}$. This method gives us very precise control over what is true in the resulting model, which is termed $M[G]$. For an introduction to forcing, see [Ung14]. On the other hand, to construct inner models we consider transitive subclasses of the von Neumann universe V satisfying ZFC, or some fragment thereof. Perhaps the canonical example of this is Gödel's constructible universe L, with which he demonstrated the consistency of both the axiom of choice and the generalised continuum hypothesis with ZF. There will not be sufficient space to discuss this here, however see [Sua21] for an exposition. Inner model theory is applied to the study of the consistency strength hierarchy as follows: if it can be shown that any model of a theory U has an inner model satisfying another theory T, then it follows that the consistency of U implies the consistency of T (since the existence of models implies consistency), and thus $T \leqslant_{\text{Con}} U$.

Importantly, when doing forcing and inner model theory we are always considering transitive models, and thus we have that both of these methods preserve the truths of arithmetic (the proof of this is beyond the author; as above see [Kun14, Lemma II.4.14]), which will, in particular, include the codings of (the proofs of) any provable consistency statements. More precisely, if we start with a transitive model $\mathcal{M}$ of a theory T, and then construct using it a transitive model $\mathcal{N}$ of a theory U which extends T, it is immediate that U believes all the consistency statements that T does, and thus $T \leqslant_{\text{Con}} U$. Put another way, using these methods we could never construct a model of U which doesn't believe $\operatorname{Con} T$ – thus we could never demonstrate the incomparability of U and T, which requires models of $\operatorname{Con} T + \neg \operatorname{Con} U$ and of $\neg \operatorname{Con} T + \operatorname{Con} U$.

Hamkins' conclusion is then that a large proportion of our evidence for natural linearity suffers from confirmation bias. In particular, since our methods can only demonstrate linearity, the lack of evidence of non-linearity should not be taken as a particularly strong signal that it doesn't occur. For an analogy, suppose that we were searching the night sky for a body emitting radiation at a certain wavelength; if the instrument we were using to search for the object was unable to detect radiation of that particular wavelength, we would not be any more justified in concluding that the body we were searching for was less likely to be there than if we had no evidence at all. In Subsection 7.2.1, I argued that despite various cases where we are not sure whether linearity holds, our stock of examples where it does gives us reason to believe these unknown cases will be settled in the positive. Hamkins' response is that of the theories we know about, we have only been able to properly investigate and classify those which are linearly related, and thus I am unjustified in making this inductive claim.

This conclusion doesn't deny that there are instances of linearity, or that these instances are genuinely surprising. It simply suggests that based on our current available evidence we are not justified in believing that linearity in natural theories holds across the board.

Given these considerations, the question then arises of how we might gather genuine evidence for linearity in the consistency strength hierarchy. The first, and perhaps easiest, would be to significantly reduce the stock of open cases: even if we only use methods which incorporate transitive models (and thus can only demonstrate linearity), if these methods are widely applicable across our class of theories, then I would argue this still gives us inductive evidence for linearity. To see this we note that in the limit case, where all known theories have been demonstrated to be linear, confirmation bias would not concern us, since there are no potential counter-instances which we may have missed. Thus in the (admittedly very contrived) 'limit case but one', where all but one theory have been shown to fit linearly, and thus where our methods have proved to be applicable in almost all circumstances, this would seem to give us inductive evidence that our methods ought to be applicable here too – and thus that we should expect consistency strength linearity. The reason that these arguments do not apply to our current state is that there are a great number of open cases remaining, indeed enough that it would be an overstatement to class our methods of forcing and inner models as 'widely applicable' in a way which gives us inductive support for their applicability to cases about which we are yet unsure.

The other approach would be to develop methods which do not, by default, preserve arithmetic truths, such as (indeed necessarily) the study of non-transitive models. If using these methods we continued to observe exclusively linearity in our theories, then this would constitute confirmation-bias-free evidence. This certainly seems an area worthy of further investigation. On the other hand, were it argued that transitivity of models was somehow inherent to our study of the concept of set, then we would appear to arrive at linearity, albeit now simply by the way we are selecting our models.

7.3. **Where does this leave us?** We have seen three arguments against the natural linearity phenomenon. The first argument attempts to give an example of natural non-linearity, and thus refute linearity wholesale. The second two arguments take a more philosophical approach by suggesting that our evidence base does not support the claim of natural linearity as strongly as we think.

It is difficult to appropriately assess the success of the first argument without a clear concept of 'naturality' in mind. It is beyond doubt that Hamkins has given an example of non-linearity, yet it remains to be argued that such an example is natural in a way that the spurned Gödelian cases are not. Although no attempt will be made here to define 'naturality', it could be argued that Hamkins' construction fails the two vague criteria given in Section 6 (not being designed specifically to reach non-linearity and occurring in the normal work of a set theorist). For the first, without presuming too much about the history of the work, I would argue that it strongly seems that the theories were developed with the intention of arriving at non-linearity, given that this is the only place such theories arise. The second is less clear: Hamkins is by trade a set theorist, and is certainly interested in extensions of ZFC. On the other hand he is also a philosopher. The theory he uses to demonstrate non-linearity seems to me to have come up with this latter purpose in mind, rather than as part of a broader mathematical endeavour. Put another way, were a set theorist unconcerned with more philosophical questions surrounding linearity of the consistency strength hierarchy, it seems unlikely they would have studied such a theory. With this said, the question of linearity is as much a mathematical one as a philosophical one, and in any case, especially as far as set theory is concerned, philosophy and maths have a fuzzy boundary between them.

In any case, it is far from clear that these two criteria are necessary or sufficient for naturality; whether this is the case could only hope to be answered by a serious philosophical investigation of the concept. My conclusion on this first argument against natural linearity is that since it preserves directedness, it still allows for the philosophical goals as outlined in Section 5 (which is not to suggest Hamkins argued otherwise). We are still able to climb the hierarchy and have a coherent notion of the strength of a theory, and thus pursue Gödel's program. Likewise, since we can always bound our theories above, the arguments for the universe view also go through in the same way.

The second two arguments are less definitive in nature. I would argue, however, they should be more troubling to those who believe that we do have consistency strength linearity, who can respond to the above by simply rejecting naturality. The confirmation bias argument shows that the evidence we currently have for linearity supports this significantly less strongly than the prevailing consensus (as at the beginning of Part II) suggests. Further, the argument from similarity of construction gives us that this evidence base is smaller than previously believed. In the face of these two objections, and absent any further study which bypasses them, it seems to me that, despite the consensus among set theorists, we don't currently have good reason to believe that the hierarchy is linear in its natural theories.

PART III: CONCLUSION

This thesis began with a mathematical investigation into the properties of two recently developed large-cardinal concepts, the worldly and otherworldly cardinals, as well as their extensions via various 'hyper' operators. In particular, we demonstrated that in terms of consistency strength, the worldly cardinals, whilst small (and indeed perhaps as small as large cardinals can be), exceed the entire hierarchy of consistency statements of the form $\mathsf{ZFC}, \mathsf{ZFC} + \mathrm{Con}\,\mathsf{ZFC}, \ldots$ In turn, the otherworldly cardinals exceed the worldly cardinals and their entire 'hyper' hierarchy. Finally, both these concepts and their extensions are exceeded by the inaccessible cardinals, which are still considered 'small' as large cardinals.

Further, we gave a brief survey of other hierarchies which have been developed to compare large cardinals. I would highlight these hierarchies and others like them (perhaps which apply to all theories extending ZFC, rather than just large-cardinal theories), and in particular the relationships between them, as an interesting area for further study. No space has been given in this essay to justifying why it is linearity in consistency strength which concerns us, rather than linearity (say) in implication strength; answering such a question and others surrounding it could be the task of a future research.

We then moved to discuss the question of linearity of the consistency strength hierarchy, which we noted could only occur for a subclass of the class of all theories, since using Gödelian ideas we can readily construct formal instances of non-linearity. We noted that, philosophically, linearity points us towards the universe view and gives us a means of pursuing Gödel's program by allowing us to meaningfully talk of 'climbing' the hierarchy. We further noted that linearity would be an elegant mathematical result. The argument for linearity was then given, which is empirical and notes that our current evidence base contains only instances of linearity for natural theories (or instances about which we are unsure). Our original results in Part I of this thesis, which demonstrated that the worldly and otherworldly cardinals as well as their hyper-extensions fit into the hierarchy linearly, contributes to this evidence base.

I then considered arguments which call natural linearity into question. First, I examined Hamkins' purported example of non-linearity and concluded that whilst it does constitute a genuine instance of non-linearity in the hierarchy, it is not clear whether it counts as 'natural'. Further, I noted that even if it is natural, it does not prevent the philosopher from pursuing the goals as mentioned in the previous paragraph, since it does not prevent directedness. The next two arguments showed us that our evidence base for linearity is smaller than expected; but also, more seriously, that they may not be able to demonstrate linearity at all, since the evidence base suffers from confirmation bias throughout. We took these arguments as severely damaging to the current orthodoxy, which maintains that we do have natural linearity, or at least that our available evidence supports that it is the case. We noted that, in order to properly demonstrate linearity, we would have to begin by using methods that employ non-transitive models of set theory.

As mentioned, no detailed consideration of the concept of naturality has been undertaken in this essay, though this is a challenge given at the end of Hamkins' paper on the topic. Further research into this area is clearly needed before we can properly answer the question of natural linearity. With this said, as things stand, I argue that based on what we have discussed in this thesis, we ought to reject the consensus that we are justified in believing that the hierarchy is linear in its natural theories.

References

[Bag04] Joan Bagaria. Natural axioms of set theory and the continuum problem. Available at `https://www.researchgate.net/publication/228371092_Natural_axioms_of_set_theory_and_the_Continuum_Problem`, 2004.

[BKW19] Joan Bagaria, Peter Koellner, and W. Hugh Woodin. 'Large cardinals beyond choice'. *Bulletin of Symbolic Logic*, 25(3):283–318, 2019.

[Car15] Erin Carmody. *Force to change large cardinal strength.* PhD thesis, City University of New York, 2015.

[Che20] Jason Chen, 2020. Tweet available here: `shorturl.at/tILRS`.

[CK90] Chen C. Chang and Howard J. Keisler. *Model Theory.* Elsevier, 1990.

[Coh63] Paul J. Cohen. 'The independence of the continuum hypothesis'. *Proceedings of the National Academy of Sciences of the United States of America*, 50(6):1143, 1963.

[Dra74] Frank R. Drake. *Set theory, An Introduction to Large Cardinals.* North-Holland Publishing Company, 1974.

[Fri98] Harvey Friedman. 'Proof-theoretic degrees', 1998. Foundations of Mathematics Mailing list, Post 20. Available at `https://cs.nyu.edu/pipermail/fom/1998-August/001929.html`.

[Git20] Victoria Gitman. 'Elementary embeddings and smaller large cardinals', 2020. Slides for a talk given at the Oxford Set Theory Seminar, University of Oxford. Available at `https://victoriagitman.github.io/talks/2020/05/06/elementary-embeddings-and-smaller-large-cardinals.html`.

[Göd38] Kurt Gödel. 'The consistency of the axiom of choice and of the generalized continuum-hypothesis'. *Proceedings of the National Academy of Sciences of the United States of America*, 24(12):556, 1938.

[Göd39] Kurt Gödel. 'Consistency-proof for the generalized continuum-hypothesis'. *Proceedings of the National Academy of Sciences of the United States of America*, 25(4):220–224, 1939.

[Göd83] Kurt Gödel. 'What is Cantor's continuum problem?'. In Paul Benacerraf and Hilary Putnam, editors, *Philosophy of Mathematics, Selected Readings*, pages 470–485. Cambridge University Press, 1983.

[Göd90] Kurt Gödel. Remarks before the Princeton bicentennial conference on problems in mathematics, 1946. In Paul Benacerraf and Hilary Putnam, editors, *Collected Works, Volume II: Publications 1938–1974*, pages 150–153. Oxford University Press, 1990.

[Ham12] Joel David Hamkins. 'The set-theoretic multiverse'. *The Review of Symbolic Logic*, 5(3):416–449, 2012.

[Ham15] Joel David Hamkins. 'Is the dream solution of the continuum hypothesis attainable?'. *Notre Dame Journal of Formal Logic*, 56(1):135–145, 2015.

[Ham17] Joel David Hamkins. 'Worldly cardinals are not always downwards absolute'., 2017. Blog post available at `http://jdh.hamkins.org/worldly-cardinals-are-not-always-downwards-absolute/`.

[Ham20a] Joel David Hamkins, 2020. Blog post 'The Otherworldly Cardinals'. Available at `http://jdh.hamkins.org/otherwordly-cardinals/`.

[Ham20b] Joel David Hamkins. *Lectures on the Philosophy of Mathematics.* MIT Press, 2020.

[Ham21] Joel David Hamkins. 'Nonlinearity in the hierarchy of large cardinal consistency strength'. 2021.

[Jec03] Thomas Jech. *Set Theory, 3rd Edition.* Springer, 2003.

[Kan08] Akihiro Kanamori. *The Higher Infinite: Large Cardinals in Set Theory From Their Beginnings.* Springer Science & Business Media, 2008.

[Kni21] Robin Knight. Lecture notes on the Gödel incompleteness theorems, 2021. Available at `https://courses-archive.maths.ox.ac.uk/node/view_material/53786`.

[Koe06] Peter Koellner. 'On the question of absolute undecidability'. *Philosophia Mathematica*, 14(2):153–188, 2006.

[Kun71] Kenneth Kunen. 'Elementary embeddings and infinitary combinatorics'. *The Journal of Symbolic Logic*, 36(3):407–413, 1971.

[Kun14] Kenneth Kunen. *Set Theory: an Introduction to Independence Proofs.* Elsevier, 2014.

[LS67] Azriel Lévy and Robert M. Solovay. 'Measurable cardinals and the continuum hypothesis'. *Israel Journal of Mathematics*, 5(4):234–248, 1967.

[Mad88] Penelope Maddy. 'Believing the axioms I'. *The Journal of Symbolic Logic*, 53(2):481–511, 1988.

[Mad00] Penelope Maddy. *Naturalism in Mathematics.* Oxford University Press, 2000.

[Mos05] Yiannis Moschovakis. *Notes on Set Theory.* Springer, 2005.

[MS89] Donald A. Martin and John R. Steel. 'A proof of projective determinacy'. *Journal of the American Mathematical Society*, 2(1):71–125, 1989.

[MV15] Luca San Mauro and Giorgio Venturi. 'Naturalness in mathematics'. In *From Logic to Practice*, pages 277–313. Springer, 2015.

[Par95] Charles Parsons. 'Platonism and mathematical intuition in Kurt Gödel's thought'. *Bulletin of Symbolic Logic*, 1(1):44–74, 1995.

[Sim09] Stephen G. Simpson. 'The Gödel hierarchy and reverse mathematics'. Manuscript on the author's webpage: `http://www.personal.psu.edu/t20/papers/gh/`, 2009.

[Smi13] Peter Smith. *An Introduction to Gödel's Theorems*. Cambridge University Press, 2013.

[Ste12] John R. Steel. 'Gödel's program'. Available at `https://math.berkeley.edu/~steel/papers/godelsprogrampub.pdf`, 2012.

[Ste13] John R. Steel. 'Gödel's program', 2013. Slides for a talk given at the CSLI meeting, Stanford University. Available at `https://math.berkeley.edu/~steel/talks/stanford2013copy.pdf`.

[Sua21] Rolf Suabedissen. Axiomatic set theory lecture notes, 2021. Available at `https://courses-archive.maths.ox.ac.uk/node/view_material/56507`.

[Tao11] Terence Tao. *An Introduction to Measure Theory*, volume 126. American Mathematical Society Providence, 2011.

[Ung14] Spencer Unger. Forcing summer school lecture notes, 2014. Available at `http://homepages.math.uic.edu/~shac/forcing/forcing2014.pdf`.

[Va] Various authors. 'Measurable cardinal'. Available at `http://cantorsattic.info/Measurable`. Accessed 28th March 2022.

[Zer30] Ernst Zermelo. 'Über grenzzahlen und mengenbereiche: neue untersuchungen über die grundlagen der mengenlehre'. *Fundamenta mathematicae*, 16, 1930.