Are Large Cardinal Axioms Restrictive?

Neil Barton*

04 May 2023†

Abstract

The independence phenomenon in set theory, while pervasive, can be partially addressed through the use of large cardinal axioms. A commonly assumed idea is that large cardinal axioms are species of *maximality principles*. In this paper, I argue that this claim is questionable. I point to two ways in which the term 'maximality' is used in the philosophy of set theory, namely *maximality in interpretive power* and *maximality in set existence*. I argue that there is a conception of capturing set existence maximality through *absoluteness*, on which large cardinal axioms come out as *restrictive* relative to maximality in interpretive power. Despite this, I argue that within this framework large cardinals are still important axioms of set theory and can play many of their usual foundational roles.

Introduction

Large cardinal axioms are widely viewed as some of the best candidates for new axioms of set theory. They are (apparently) linearly

*IFIKK, Universitetet i Oslo. E-mail: n.a.barton@ifikk.uio.no.

†I would like to thank David Asperó, David Fernández-Bretón, Monroe Eskew, Sy-David Friedman, Victoria Gitman, Luca Incurvati, Penelope Maddy, Toby Meadows, Sandra Müller, Michael Potter, Chris Scambler, Giorgio Venturi, Matteo Viale, Kameryn Williams, and audiences in Cambridge, New York, Konstanz, and São Paulo for helpful discussion. Two anonymous referees also provided helpful comments, and I am grateful for their input. I am also very grateful for the generous support of the FWF (Austrian Science Fund) through Project P 28420 (*The Hyperuniverse Programme*), the VolkswagenStiftung through the project *Forcing: Conceptual Change in the Foundations of Mathematics*, and Norges forskningsrådet (the Research Council of Norway) for their support via the project *Infinity and Intensionality: Towards A New Synthesis* (no. 314435).

ordered by consistency strength, have substantial mathematical consequences for questions independent from ZFC (such as consistency statements and Projective Determinacy[1]), and appear natural to the working set theorist, providing fine-grained information about different properties of transfinite sets. They are considered mathematically interesting and central for the study of set theory and its philosophy.

In this paper, I do not deny any of the above points. I will, however, argue that the status of large cardinal axioms as *maximality* principles is questionable. In particular, I will argue that there is a way of trying to capture maximality through *absoluteness* in set theory on which large cardinal axioms can appear *restrictive*. Despite this, we'll see that they are nonetheless able to carry out many of their key foundational roles in this context.

Here's the plan: First (§1) I'll provide some background, identify the two notions of maximality we'll consider (*maximality in set-existence* and *maximality in interpretive power*), and introduce the idea that large cardinals are maximality principles. In §2 I'll explain the absoluteness principle we will consider (the Class-Generic Inner Model Hypothesis or CIMH). In §3 I argue that an appealing formulation of restrictiveness in interpretive power due to Maddy delivers the result that many formulations of ZFC with large cardinals added are restrictive in interpretive power when compared to theories utilising the CIMH. Next (§4) I'll compare the justification of the CIMH with that of bounded forcing axioms, arguing that they rest on similar set-existence maximality motivations, though they calibrate them in very different ways. I'll then (§5) show that theories using the CIMH also Maddy-maximise over some bounded forcing axioms. Next (§6) I argue that large cardinals can still fulfil many of their required foundational roles. Finally (§7) I'll conclude with some possible ramifications for the study of the philosophy of set theory and some open questions.

1 Large cardinals and notions of maximality

Let's start with some background on large cardinals and how they have been viewed philosophically. Given a set theory capable of axiomatising a reasonable fragment of arithmetic (i.e. able to support the coding of the relevant syntactic notions), we start our discussion with the following celebrated theorem:

Theorem 1. [Gödel, 1931] (Second Incompleteness Theorem). No ω-consistent recursive theory T capable of axiomatising primitive recur-

[1] See [Schindler, 2014] for a textbook treatment of large cardinals and determinacy.

sive arithmetic can prove its own consistency sentence $Con(\mathsf{T})$.

Given then some appropriately strong set theory T, we can then obtain a *strictly stronger* theory by adding $Con(\mathsf{T})$ to T. So, if we accept 'ZFC' then, ZFC + $Con(\mathsf{ZFC})$ is a strictly stronger theory, and ZFC + $Con(\mathsf{ZFC} + Con(\mathsf{ZFC}))$ is strictly stronger still. More generally:

Definition 2. A theory T has *greater consistency strength* than S if we can prove $Con(\mathsf{S})$ from $Con(\mathsf{T})$, but cannot prove $Con(\mathsf{T})$ from $Con(\mathsf{S})$. They are called *equiconsistent* iff we can both prove $Con(\mathsf{T})$ from $Con(\mathsf{S})$ and $Con(\mathsf{S})$ from $Con(\mathsf{T})$.[2]

The interesting fact for current purposes is that in set theory we are not limited to increasing consistency strength solely through adding Gödel-style diagonal sentences. The axiom which asserts the existence of a transitive model of ZFC is stronger still (such an axiom implies the consistency of theories with transfinite iterations of the consistency sentence for ZFC). As it turns out, by postulating the existence of certain kinds of models, embeddings, and varieties of sets, we discover theories with greater consistency strength. For example:

Definition 3. (ZFC) A cardinal κ is *strongly inaccessible* iff it is uncountable, regular[3], and a strong limit[4].

Such an axiom provides a model for *second-order* ZFC_2, namely $(V_\kappa, \in, V_{\kappa+1})$. These cardinals represent the first steps on an enormous hierarchy of logically and combinatorially characterised objects.[5] More generally, we have the following rough idea: A large cardinal axiom is a principle that serves as a natural stepping stone in the indexing of consistency strength.

In the case of inaccessibles, many of the logical properties attaching to the cardinal appear to derive from its brute size. For example, because such a κ cannot be reached 'from below' by either of the axioms of Replacement or Powerset, we can show that $(V_\kappa, \in, V_{\kappa+1})$

[2] A subtlety here concerns what base theory we should use to prove these equiconsistency claims. Number theory will do (since consistency statements are number-theoretic facts), but we will keep discussion mostly at the level of a suitable set theory (e.g. ZFC).

[3] A cardinal κ is *regular* iff there is no function from a smaller cardinal unbounded in κ.

[4] A *strong limit cardinal* is a cardinal κ such that $|x| < \kappa$ then $|\mathcal{P}(x)| < \kappa$.

[5] Often, combinatorial and logical characterisations go hand in hand, such as in the case of measurable cardinals. However, sometimes it is not clear how to get one characterisation from another. Recently, cardinals often thought of as having only combinatorial characterisations have been found to have embedding characterisations. See [Holy et al., 2019] for details.

satisfies ZFC_2. Similar considerations apply to other kinds of cardinal and other consistency implications. A *Mahlo cardinal*, for example, is a strongly inaccessible cardinal κ beneath which there is a stationary set[6] of inaccessible cardinals. The fact that such a cardinal has higher consistency strength than that of strong inaccessibles (and mild strengthenings thereof) is simply because it contains many models of these axioms below it.

It is not the case, however, that consistency strength is inextricably tied to size. For example, the notion of a *strong*[7] cardinal has lower consistency strength than that of *superstrong*[8] cardinal, but the least strong cardinal is larger than the least superstrong cardinal.[9] The key point is that despite the fact that the least superstrong is not as *big* as the least strong cardinal, one can always *build* a model of a strong cardinal from the existence of a superstrong cardinal (but not vice versa). Thus, despite the fact that a superstrong cardinal can be 'smaller', it still witnesses the *consistency* of the existence of a strong cardinal.

Before we move on, we note some foundational uses for large cardinals that make them especially attractive objects of study. First:

(Linearity) The 'natural' large cardinal principles appear to be linearly ordered by consistency strength.

One can gerrymander principles (via metamathematical coding) that would produce only a partial order of consistency strengths[10], however it is an empirical fact that the large cardinal axioms that set theorists have naturally come up with and view as interesting *are* linearly ordered.[11] This empirical fact has resulted in the following feature of mathematical practice:

(Indices) Large cardinals serve as the the natural indices of consistency strength in mathematics.

In particular, if consistency concerns are raised about a new branch of mathematics, the usual way to assess our confidence in the consistency of the practice is to provide a model for the relevant theory with

[6] $S \subseteq \kappa$ is *stationary in* κ iff S intersects every closed and unbounded subset of κ.

[7] A cardinal κ is *strong* iff for all ordinals λ, there is a non-trivial elementary embedding (to be discussed later) $j : V \longrightarrow \mathfrak{M}$, with critical point κ, and in which $V_\lambda \subseteq \mathfrak{M}$.

[8] A cardinal κ is *superstrong* iff it is the critical point of a non-trivial elementary embedding $j : V \longrightarrow \mathfrak{M}$ such that $V_{j(\kappa)} \subseteq \mathfrak{M}$.

[9] See [Kanamori, 2009], p. 360.

[10] See [Koellner, 2011] for discussion.

[11] There are some open questions to be tied up, for example around strongly compact cardinals and around Jónsson cardinals.

sets, possibly using large cardinals.[12] For example, worries of consistency were raised during the emergence of category theory, and were assuaged by providing a set-theoretic interpretation, which then freed mathematicians to use the category-theoretic language with security. Grothendieck postulated the existence of universes (equivalent to the existence of inaccessible cardinals), and Mac Lane is very careful to use universes in his expository textbook for the working mathematician.[13] These later found application in interpreting some of the cohomological notions used in the original Wiles-Taylor proof of Fermat's Last Theorem (see [McLarty, 2010]). Of course now category theory is a well-established discipline in its own right, and quite possibly stands free of set-theoretic foundations. Nonetheless, set theory was useful providing an upper bound for the consistency strength of the emerging mathematical field. More recently, several category-theoretic principles (even some studied in the 1960s) have been calibrated to have substantial large cardinal strength.[14]

Related to indexing consistency strength is the use of large cardinals in building inner models, which has become a focal point of study in its own right. For many large cardinal axioms we can (using the relevant large cardinal axioms) build a model that takes itself to contain a cardinal of a particular kind. Many of these models are L-like and satisfy properties such as condensation, revealing a good deal of information about the properties of the sets they contain (relative to the model). Again, the details are rather technical, so we omit them.[15] The point is the following: Often in set theory we have very little information about the properties of certain sets, as exhibited by the independence phenomenon. This is not so for large cardinals with L-like inner models, where (whilst there are open questions) there is a large amount of highly tractable information concerning the objects. The

[12]See here, for example, Steel:

> "The central role of the theories axiomatized by large cardinal hypotheses argues for adding such hypotheses to our framework. The goal of our framework theory is to maximize interpretive power, to provide a language and theory in which all mathematics, of today, and of the future so far as we can anticipate it today, can be developed." ([Steel, 2014], p. 11)

[13]See [Mac Lane, 1971], Ch.1, §6. Also interesting here is [McLarty, 1992], Ch. 12.

[14]See [Bagaria and Brooke-Taylor, 2013] for details. The consistency strength is really quite high; many category-theoretic statements turn out to be equivalent to Vopěnka's Principle.

[15]For the state of the art concerning inner model theory and the challenges faced, see [Sargsyan, 2013] and [Woodin, 2017]. For an overview of what makes a model L-like, see [Steel, 2010], especially §5 and Theorems 5.1 and 5.4.

building of inner models for large cardinals thus represents an important and technically sophisticated area of study, and many of the major open questions in set theory concern their construction.

This observation concerning the role of large cardinals in contemporary mathematics point to a central desideratum for their use:

(Maximality in Interpretive Power) Large cardinals are required to *maximise interpretive power*: We want our theory of sets to facilitate a unified foundational theory in which all mathematics can be 'appropriately' interpreted.

This idea is strongly emphasised in [Steel, 2014]. Of course, work has to be done in saying what we mean by an 'appropriate' interpretation. We will discuss this in detail in §3, when we consider Penelope Maddy's notion of restrictiveness. For now we can simply note that such inner models provide 'nice' interpretations; they are transitive, well-founded, and contain all the ordinals.

Maximising interpretive power entails maximising consistency strength; we want a theory that is able to incorporate as much consistent mathematics as is possible whilst preserving a sense of intended interpretation, and hence (assuming the actual consistency of the relevant cardinals) require the consistency strength of our framework theory to be very high. However, maximality in interpretive power is not the only kind of maximality in play. Often large cardinals seen as species of *set* maximality principles. For example, Gödel famously wrote (concerning small large cardinals like inaccessibles, Mahlos etc.):

> "...the axioms of set theory by no means form a system closed in itself, but, quite on the contrary, the very concept of set on which they are based suggests their extension by new axioms which assert the existence of still further iterations of the operation "set of". These axioms can also be formulated as propositions asserting the existence of very great cardinal numbers or (which is the same) of sets having these cardinal numbers." ([Gödel, 1947], p. 181)

Here, we see Gödel argue that the postulation of small large cardinals serves as a good way of asserting that the universe contains many different kinds of large sets.

This is indicative of a second kind of maximality:

(Maximality in Set-Existence) The more sets the axiom asserts to exist, the better.

Several authors discuss this idea, and are sensitive to its pitfalls.[16] In particular, this formulation is taken directly from [Bagaria, 2005], and we shall discuss it in detail in §4, examining how Bagaria makes it precise and how it relates to discussion of large cardinals.

Of course, one might think that the two notions are related. If we maximise the number and variety of sets available, one might think we thereby maximise the possibilities for interpreting mathematical theories. Part of this paper will provide an example to show that this link is perhaps not as clear as one might think.

One final point that we shall make about large cardinals is their use in proving axioms of definable determinacy. The full details will be familiar to specialists and obscure to non-specialists, so we omit them here.[17] Nonetheless, a coarse description will be helpful in stating our arguments. Roughly put, axioms of definable determinacy assert the existence of winning strategies for games played with natural numbers.[18] Importantly, some authors have argued that these axioms have various pleasant consequences we would like to capture.[19] One salient fact is that Projective Determinacy yields high degree of completeness for the hereditarily countable sets (i.e. there are no known statements apart from Gödel style diagonal sentences independent from the theory ZFC - Powerset + $V = H(\omega_1)$ + PD).[20] Moreover, whilst it is a theorem of ZFC that not all games are determined, certain restricted forms can be proved from large cardinals. For example:

Fact 4. Borel Determinacy is provable in ZFC, but any proof requires ω_1-many applications of the Powerset Axiom.

Fact 5. Analytic Determinacy is provable in ZFC + "There exists a measurable cardinal", but is independent from ZFC.

Fact 6. Projective Determinacy is provable in ZFC + "There are ω-many Woodin cardinals", but is independent from ZFC + "There exists a measurable cardinal".

[16] See, for example, [Hauser, 2001], (p. 257) [Incurvati, 2017] (p. 162), [Maddy, 2011] (pp. 125–126), [Drake, 1974] (p. 186).

[17] The interested reader is directed to [Schindler, 2014] for a recent presentation of the technical details, and [Koellner, 2006], [Maddy, 2011], and [Koellner, 2014] for a philosophical discussion.

[18] $\mathsf{AD}^{L(\mathbb{R})}$ is one axiom of definable determinacy that goes beyond second-order arithmetic. There are also versions of determinacy for real-valued games, or games of longer length. We put aside these issues here.

[19] See, for example, [Maddy, 2011] and [Welch, 2017].

[20] [Koellner, 2014] provides a detailed survey of the literature here, and is quick to point out that axioms of definable determinacy seem to be the consequence of any strong 'natural' theory extending ZFC (e.g. ZFC + PFA). Given the focus of this paper, we shall concern ourselves only with the argument from large cardinals.

Again, we will not go through the definitions of Borel, Analytic, or Projective here. Suffice to say, each admits progressively more sets of reals with a more permissive notion of definability, and each is resolved by strictly stronger large cardinal axioms. Some authors have pointed out that it may well be that our 'best' theory of sets uses axioms of definable determinacy.[21]

We can now summarise the following points from our discussion of large cardinals:

(1.) Large cardinals appear to be linearly ordered by consistency strength and are the standard indicies of consistency strength.

(2.) They are used in various technical model-building constructions in inner model theory.

(3.) They can be used to prove axioms of definable determinacy.

(4.) They are often regarded as species of 'maximality' axiom, in particular with respect to *maximality in interpretive power* and *maximality in set-existence*.

In the rest of the paper, I will argue that large cardinals can appear restrictive from a maximality in interpretive power perspective. The issue, we shall see, depends upon how we calibrate maximality in set existence. I thus conclude that the status of large cardinal axioms as maximality principles is questionable. We will, however, see that their important foundational roles as outlined by (1.) and (2.) are unaffected, and that the question of their role in (3.) is still open.

2 The Class-Generic Inner Model Hypothesis

The principle we will consider (the Class-Generic Inner Model Hypothesis) stems from *absoluteness* considerations; if something is satisfied in an extension of the universe then it is already satisfied in the universe (subject to terms and conditions, and in certain contexts). We will discuss this absoluteness idea in detail in §4, for now we focus on defining the principle in order to move forward with our restrictiveness arguments. Before we get going, however, it is useful to provide some set up. We will work under *von Neumann-Bernays-Gödel class theory* (NBG) which has a second sort of variables $X, Y, Z, X_0, ..., X_n, ...$ (intended to range over classes), and has as axioms all first-order axioms of ZFC, with the axiom schemas of Replacement and Separation

[21] See, for example, [Woodin, 2001].

replaced by single second-order axioms, and an axiom of predicative comprehension for classes.[22] We will work with the version that does *not* include the axiom of Global Choice, but not much hangs on this for our proofs. We start with a couple of definition we'll need for the set up. For now we'll characterise them as claims about *models*, the syntactic formulation will be made clear in due course:

Definition 7. (NBG) An *inner model* in a universe $\mathcal{V} = (V, \in, \mathcal{C}^{\mathcal{V}}) \models \mathsf{NBG}$ (where V is the domain of sets of V, and $\mathcal{C}^{\mathcal{V}}$ is a domain of classes) is a $\mathcal{V}$-transitive proper class $X \in \mathcal{C}^V$ satisfying ZF. Further:

(i) We say that an inner model X is *proper* when $X \subsetneq V$.

(ii) We say that an inner model X is *first-order definable in the parameters* $\vec{a}$ when there is some first-order formula $\phi(\vec{a}, x)$ with one free variable in the language of ZFC (with $\vec{a}$ added) such that $\phi(\vec{a}, x)$ iff $x \in X$.

(iii) We say that an inner model X is *first-order definable* when there is some first-order formula $\phi(x)$ with one free variable in the language of ZFC such that $\phi(x)$ iff $x \in X$.

Definition 8. (NBG) A *width extension* of a universe $\mathcal{V}$ is a universe $\mathcal{V}'$ such that $\mathcal{V}$ is an inner model of $\mathcal{V}'$. A width extension $\mathcal{V}'$ is *proper* when $\mathcal{V}$ is a proper inner model of $\mathcal{V}'$.

We can then consider what inner models V (*our* universe) must contain relative to its extensions. In particular we can formulate:

Definition 9. [Friedman, 2006] Let ϕ be a parameter-free first-order sentence. The *Inner Model Hypothesis* (or IMH) states that if ϕ is true in an inner model of a width extension of V, then ϕ is already true in an inner model of V.

A core feature of the IMH is that it depends upon quantifying over *arbitrary* extensions of a universe. Initially, it is thus unclear what theory we should use to formalise it. It can be formulated as about countable transitive models[23] or (using well-founded top-extensions of V)

[22] A brief note on nomenclature: In set theory is usual to refer to theories that do not have class variables as first-order, and those that do as second-order. This is despite the fact that, strictly speaking, NBG and its cousins are two-sorted first-order theories, even if they could be given a second-order formulation in which we quantify into predicate position.

[23] See here [Arrigoni and Friedman, 2013].

as about infinitary logics[24], or can be coded using a variant of Morse-Kelley class theory[25]. Since we want to consider an axiom that is easily formalisable across a range of possible perspectives, we shall consider a modified version of the IMH that is formalisable in NBG:

Definition 10. (NBG) Let $(V, \in, \mathcal{C})$ be a NBG structure. The *Class-Generic Inner Model Hypothesis* (or CIMH) is the claim that if a (first-order, parameter-free) sentence ϕ holds in an inner model of a tame class forcing extension $(V[G], \in, \mathcal{C}[G])$ (where where $V[G]$ consists of the interpretations of set-names in V using G, and $\mathcal{C}[G]$ consists of the interpretations of class-names in $\mathcal{C}$ using G), then ϕ holds in an inner model of V.

In other words, if there is a $(V[G], \in, \mathcal{C}[G])$ and an inner model $X \in \mathcal{C}[G]$ such that $X \models$ ZF and $X \models \phi$, then there is an inner model $Y \in \mathcal{C}$ such that $Y \models$ ZF and $Y \models \phi$.

The requirement of *tameness* is somewhat technical to state, but is equivalent to the forcing preserving NBG.[26] This is needed in the present context, since there are non-tame-forcings that are not NBG-preserving, and if such forcings are allowed we would easily obtain a contradiction (without some other further modification). For example, if we allow a Powerset-violating non-tame forcing, we would obtain a quick contradiction by finding an inner model violating the Powerset Axiom (whilst also—per impossibile—satisfying ZF).

The restriction to tame forcings also aids with formalisation. Since forcing relations are definable for tame class forcings, the following way of expressing the CIMH is equivalent:

Definition 11. (NBG) $(V, \in, \mathcal{C})$ satisfies the *Absolute Class-Generic Inner Model Hypothesis* (or $\mathsf{CIMH}^{\Vdash}$) iff whenever $\mathbb{P} \subset V$ is a tame class forcing, and ϕ is a parameter-free first-order sentence, then if there is a $p \in \mathbb{P}$ such that $p \Vdash_{\mathbb{P}}$ "ϕ is true in an inner model" then ϕ is true in an inner model of V.

In this way, the Class-Generic Inner Model Hypothesis can be formalised in NBG without quantifying directly over extensions. Even a believer in just one maximal universe of sets, for example, could consider the $\mathsf{CIMH}^{\Vdash}$ as a possible axiom candidate, since presumably they accept the use of NBG class theory. After all, for any model

[24]See here [Antos et al., 2015], [Barton and Friedman, 2017].

[25]Morse-Kelley class theory (or MK) has an *impredicative comprehension scheme* instead of NBG's predicative axiom. In fact, a variant of NBG + Σ^1_1-Comprehension is enough to formalise the IMH, see [Antos et al., 2021].

[26]For the details of tameness (and pretameness) see [Friedman, 2000].

$\mathcal{M} = (M, \in) \models \mathsf{ZFC}$, NBG is satisfied by $(M, \in, Def(M))$, where $Def(M)$ are the definable classes of $\mathcal{M}$ (at least the version of NBG formulated without Global Choice). A believer in one universe of sets would thus have to reject the use of definable classes in rejecting NBG, and would thereby give up on a large amount of standard set theory. The question then of whether the $\mathsf{CIMH}^{\Vdash}$ is truth-evaluable is thus not dependent upon ontological perspective.[27] Since they are formally equivalent, we will drop the distinction between the CIMH and $\mathsf{CIMH}^{\Vdash}$ from here on out.

3 Interpretive maximality and large cardinals

As we will now see, the CIMH can be used to obtain theories that suggest that some large cardinal axioms are restrictive with respect to interpretive maximality (we will discuss set-existence maximality in §4). The core points we'll see are:

(1.) The CIMH implies the negation of large cardinal axioms, even some of the weakest such principles.

(2.) The CIMH nonetheless validates the *consistency in inner models* of large cardinals up to the level of many measurable cardinals.

(3.) ZFC-based set theories obtained from the CIMH can only be interpreted in 'impoverished' contexts using theories incorporating large cardinals.

We deal with these points in turn. (1.) Anti-large cardinal properties of the IMH were noticed early on. Many results using the full IMH can be incorporated to the current context, since they only require tame class forcings. For instance, we can immediately identify:

Theorem 12. [Friedman, 2006] (NBG) If the Class-Generic Inner Model Hypothesis holds, there are no inaccessible cardinals in V.[28]

Given acceptance of the CIMH, this would mean that there could be no (significant) large cardinals in V. However, the existence of large cardinals in inner models is positively *implied*:

[27] Of course, one might worry that the *natural* reading of the CIMH is in terms of quantifying over extensions, whatever the coding possibilities. See [Barton, 2020] for discussion of this point.

[28] See here [Friedman, 2006], Theorem 15. The proof is formulated for the full IMH, but since it uses only tame class forcings, we can import it directly to the current context.

Theorem 13. [Friedman et al., 2008] (NBG) The Class-Generic Inner Model Hypothesis implies that there is an inner model with measurable cardinals of arbitrarily large Mitchell order.[29]

Thus, while the Inner Model Hypothesis does not permit the existence of large cardinals in V, it *does* vindicate their existence in inner models. By contrast, we can *prove* the consistency of the IMH (and hence the CIMH) from large cardinals:

Theorem 14. [Friedman et al., 2008] (ZFC) Assuming the consistency of the existence of a Woodin cardinal with an inaccessible above, the Inner Model Hypothesis is consistent.

A remark about the proof will be useful for motivating our discussion. For a real x, let M_x be the least transitive model of ZFC containing x. By collapsing the Woodin cardinal to ω, and using Σ^1_2-Determinacy and the (preserved) inaccessible in the forcing extension, one has that for every real x in a Turing cone $(M_x, \in, Def(M_x))$ satisfies the IMH. The core point to retain is the following: The IMH-satisfying structure we find in the proof is small (in that it is countable).

We thus have a rough guide as to the consistency strength of the Class-Generic Inner Model Hypothesis (somewhere between many measurables and a Woodin with an inaccessible above). But now there is something of a stand off between the friend of large cardinals and the supporter of the CIMH. The friend of large cardinals looks at the supporter of the CIMH and thinks that her theory is true in small countable transitive models, and certainly does not hold in the universe. The supporter of the CIMH, on the other hand, looks at the friend of large cardinals and thinks that his theory can only be true when we leave out some subsets that destroy the inaccessibility of particular cardinals in V. Is there any way to resolve this stand off?

There is at least one technically precise sense in which we can say that the CIMH-theorist is in better shape with respect to *interpretive* maximality. We will use [Maddy, 1998]'s notion of theories *maximizing* over one another and (and some being *restrictive* on these grounds). Her idea is that one set theory T_1 maximises over another T_2 (and hence shows it to be restrictive) when one can use T_1 to provably find

[29]The Mitchell ordering is a way of ordering the normal measures on a measurable cardinal. Roughly, it corresponds to the strength of the measure, where a measure U_1 is below another U_2 in the Mitchell order if U_1 belongs to the ultrapower obtained through U_2. See [Jech, 2002] Ch. 19. Again, only tame class forcings are used in proving the existence of an inner model with measurable cardinals of arbitrarily large Mitchell order from the full IMH.

an interpretation of T_2 in an 'appropriate'[30] context, but not vice versa, and the two theories are jointly inconsistent with one another. More precisely, Maddy begins with the following definition:

Definition 15. [Maddy, 1998] A theory T extending ZFC *shows ϕ is an internal model* iff ϕ is a formula in one free variable such that:

(i) For all σ in ZFC, $\mathsf{T} \vdash \sigma^\phi$ (i.e. σ holds relative to the ϕ-satisfiers).

(ii) $\mathsf{T} \vdash \forall\alpha\phi(\alpha)$ or $\mathsf{T} \vdash \exists\kappa((\text{“}\kappa \text{ is inaccessible”}) \wedge \forall\alpha(\alpha < \kappa \rightarrow \phi(\alpha)))$ (i.e. the ϕ-satisfiers either contain all ordinals, or all ordinals up to some inaccessible), and

(iii) $\mathsf{T} \vdash \forall x\forall y((x \in y \wedge \phi(y)) \rightarrow \phi(x))$ (i.e. ϕ defines a transitive interpretation).

This definition serves to specify the 'appropriate' interpretations we are interested in; proper class inner models and truncations thereof at inaccessibles. Maddy uses the term 'inner model' instead of 'internal model', but we opt for the latter in order to avoid confusion with the notion of 'inner model' being employed in versions of the CIMH. We can then define:

Definition 16. [Maddy, 1998] ϕ is a *fair interpretation of* T_1 *in* T_2 iff:

(i) T_2 shows ϕ is an internal model, and

(ii) For all σ in T_1, $\mathsf{T}_2 \vdash \sigma^\phi$.

i.e. a fair interpretation of one theory T_1 in another T_2 is provided by finding some ϕ defining an inner model (or truncation thereof) in T_2 that satisfies T_1.

Maddy then goes on to define what it means for a theory to maximise over another. First, she thinks that there should be new isomorphism types outside the interpretation, which, in the presence of Foundation, amounts to the existence of sets not satisfying ϕ (i.e. the internal model defined by ϕ is proper):

Definition 17. [Maddy, 1998] T_2 *maximizes* over T_1 iff there is a ϕ such that:

(i) ϕ is a fair interpretation of T_1 in T_2, and

[30] [Maddy, 1998] uses the terminology 'preservation' and 'fair interpretation', instead of 'appropriate'. We use 'appropriate' in order to keep the technically-loaded notions of 'fair interpretation' and 'preservation' separate from the informal notion.

(ii) $\mathsf{T}_2 \vdash \exists x \neg \phi(x)$.

With this idea of maximisation in play, she next sets up some additional definitions to make sure that weak but unrestrictive theories, whilst not maximising, do not count as restrictive. This is dealt with by the following definitions.

Definition 18. [Maddy, 1998] T_2 *properly maximizes* over T_1 iff T_2 maximizes over T_1 but not vice versa.

Definition 19. [Maddy, 1998] T_2 *inconsistently maximizes* over T_1 iff T_2 properly maximises over T_1 and T_2 is inconsistent with T_1.

Definition 20. [Maddy, 1998] T_2 *strongly maximizes* over T_1 iff T_2 inconsistently maximizes over T_1, and there is no consistent T_3 extending T_1 that properly maximizes over T_2.

Thus we have a picture on which one theory T_2 (strongly) maximises over T_1 when we can prove in T_2 that a certain formula ϕ defines a proper inner model (or truncation thereof), satisfying T_1, and such that we cannot extend T_1 to a theory capable of finding such an interpretation for T_2. If there is a theory T_2 strongly maximising over T_1, then we say that T_1 is *Maddy-restrictive*[31]. A natural example here is contrasting the theories ZFC + $V = L$ and ZFC + "There exists a measurable cardinal". The latter strongly maximises over the former, since we can always build L to find a model of ZFC + $V = L$, but there are no fair interpretations with measurable cardinals under $V = L$ (though they can exist in other kinds of model, e.g. countable).

One difficult issue in this context concerns the use of parameters. Let U_0 be an ultrafilter on the least measurable cardinal κ_0. We would like our definition of $L[U_0]$ to provide a fair interpretation of ZFC + "$V = L[U_0]$" using the theory ZFC + "There are two measurable cardinals", and on this basis show that the axiom $V = L[U_0]$ is also a restrictive axiom much like $V = L$ (in particular, $V = L[U_0]$ implies that there is *exactly* one measurable cardinal). This situation is in fact considered by Maddy:

> $V = L$, as we know, is the claim that the universe... is identical with the smallest proper class inner model, L. Given the development of 'wider' proper class inner models that include various large cardinals, beginning with measurables, the next test case would be to ask what the criterion

[31] We use the term 'Maddy-restrictive' as it is a substantial open question whether or not Maddy-restrictiveness and restrictiveness are coextensive.

> has to say about various theories of the form ZFC + 'V = the canonical inner model with such-and-such large cardinals'. A theory like this seems restrictive because it rules out the existence of additional large cardinals; the expectation is that it would be strongly maximized over by a theory of the form ZFC + 'the next larger large cardinal exists'. In fact, this pattern is satisfied for a considerable distance. So, for example, ZFC + 'there are two measurable cardinals' (ZFC + 2MC) provides a fair interpretation of ZFC + '$V =$ the canonical inner model with one measurable cardinal' (ZFC + $V = L[U]$). Furthermore, ZFC + $V = L[U]$ implies that there is no inner model of ZFC + 2MC, so neither it nor any extension of it can fairly interpret ZFC + 2MC. Thus, ZFC + 2MC strongly maximizes over ZFC + $V = L[U]$, and the latter is restrictive, as seems right. [Maddy, 1997, p. 225]

The use of parameters, however, poses an additional complication. The issue is that a parameter is normally taken to be the 'name' for a particular set; an essentially model-theoretic notion. Maddy's definition of restrictiveness is meant to be purely 'syntactic' in character—it concerns theory-to-theory interpretation *within the same language*. Especially in the context where we are meant to be adjudicating between foundational theory, it's thus unclear how to handle parameters.

In the specific case Maddy considers, however, we can do away with the parameters. By a theorem of [Kunen, 1971][32] one can show that if U_1 and U_2 are normal ultrafilters witnessing the measurability of κ, then $L[U_1] = L[U_2]$. It's thus possible to quantify away their use abbreviating "$x \in L[U_0]$" (where U_0 is a normal ultrafilter on the least measurable cardinal κ_0) with the following formula (that I'll informally paraphrase for readability):

$x \in L[U_0]$ iff "there exists a normal measure U on the least measurable cardinal κ such that there is an α with $x \in L_\alpha[U]$"

Similar remarks apply to any case where we have a parameter that is definable if it exists and whose existence is implied by a theory. For example, if T implies that "$0\sharp$ exists", we will not need $0\sharp$ as a parameter in order to define $L[0\sharp]$ using T. This kind of strategy will be used repeatedly to eliminate parameters in what follows.[33]

[32] See [Kanamori, 2009, p. 267], Theorem 20.10.

[33] I thank an anonymous referee for pushing me to be more precise on the use of parameters, and Sandra Müller and Kameryn J. Williams for additional discussion here.

Maddy's definitions are not without their problems (notably some false negatives and positives), a fact which Maddy herself is admirably transparent about.[34] Subsequent developments of the notion have been considered by Löwe and Incurvati.[35] Our point here is not that Maddy's definitions provide *the* definitive word on restrictiveness relative to maximality in interpretive power, but rather that they provide an interesting perspective on which the rough ideas sketched earlier (concerning the stand-off) could be made precise, if one so desired.

First, the CIMH. The CIMH is formulated in NBG, and since Maddy's formulation is intended to apply only to first-order set theories, we require a further modification. One option would be to change her definition to apply to second-order theories. However, the presence of second-order quantification would perhaps suggest different notions of *fair interpretation* (since we could directly quantify over inner models), and so we will not pursue this strategy here (despite its interest). Rather, we will look at the *first-order fragment* of NBG + CIMH. We can then prove:

Proposition 21. (NBG) Let $\mathsf{ZFC}^{\mathsf{CIMH}}$ be the parameter-free first-order part of NBG + CIMH (i.e. take NBG + CIMH and remove all sentences and formulas containing at least one set parameter, class variable, or class parameter). Let $\mu(\alpha)$ be a parameter-free ZFC-sentence defining some ordinal $\alpha > 0$, such that $\exists\alpha\mu(\alpha)$ is provable in ZFC alone. Then $\mathsf{ZFC}^{\mathsf{CIMH}}$ strongly maximises over ZFC + "There exist α-many measurables".

Proof. Note that by assumption, ZFC + "There exist α-many measurables" can be expressed in a parameter-free way by using the formula $\mu(\alpha)$. We first need to show that $\mathsf{ZFC}^{\mathsf{CIMH}}$ shows that some ϕ is an internal model with α-many measurables. Theorem 2 of [Friedman et al., 2008] establishes that $\mathsf{NBG}+\mathsf{CIMH}$ proves that there is a definable inner model with measurable cardinals of arbitrarily large Mitchell order.[36] The model in question is Mitchell's core model K for sequences of measures, and this version of K is definable using a parameter-free formula.[37] Thus, by going high enough in the Mitchell

[34] In the original [Maddy, 1998] and [Maddy, 1997].

[35] See here [Löwe, 2001], [Löwe, 2003], and [Incurvati and Löwe, 2016] (which responds to some criticisms of [Hamkins, 2014]).

[36] Note: Friedman, Welch, and Woodin are explicit about the fact that whilst their theorems are formulated about the IMH, none of their theorems depend on arbitrary outer models, but rather could be formulated in terms of the CIMH. See [Friedman et al., 2008] pp. 391–392.

[37] See [Mitchell, 2010] for details. Thanks to Sy-David Friedman for some discussion of the definition of K.

order, $\mathsf{ZFC}^{\mathsf{CIMH}}$ provides a fair interpretation of ZFC + "There exist α-many measurables".

Moreover $\mathsf{ZFC}^{\mathsf{CIMH}}$ also *maximises* over ZFC + "There exist α-many measurables", since there are always sets outside this interpretation. In particular, since $\mathsf{ZFC}^{\mathsf{CIMH}}$ implies that there are no inaccessible cardinals, for any particular β that is measurable in K, K misses out the sets witnessing the accessibility of β. Clearly, the two theories are also inconsistent with one another.

It just remains to show that ZFC + "There exist α-many measurables" does not maximise over $\mathsf{ZFC}^{\mathsf{CIMH}}$ (for inconsistent maximisation), nor can any consistent extension (for strong maximisation). These are established by the following claim:

Claim 22. No consistent extension of ZFC + "There exist α-many measurables" can provide a fair interpretation of $\mathsf{ZFC}^{\mathsf{CIMH}}$.

To show this, we need to show that under any extension of ZFC + "There exist α-many measurables", none of (i) there is an inner model of $\mathsf{ZFC}^{\mathsf{CIMH}}$, (ii) there is a truncation at an inaccessible with $\mathsf{ZFC}^{\mathsf{CIMH}}$, or (iii) there is a truncation at an inaccessible of an inner model with $\mathsf{ZFC}^{\mathsf{CIMH}}$, are possible. For (i) it suffices to note that being accessible is upwards absolute. Since all cardinals are accessible under $\mathsf{ZFC}^{\mathsf{CIMH}}$ (a parameter-free claim) if $\mathsf{ZFC}^{\mathsf{CIMH}}$ holds in an inner model, then all cardinals are accessible, ruling out (i). For (ii) and (iii) we first note that no truncation V_κ for κ above the least inaccessible β can satisfy $\mathsf{ZFC}^{\mathsf{CIMH}}$, since then V_κ would see the inaccessibility of β. Nor can such a V_κ have an inner model satisfying $\mathsf{ZFC}^{\mathsf{CIMH}}$ since inaccessibility is downwards absolute and so the inner model would still see the inaccessibility of β. The only possible case is thus when κ is the least inaccessible cardinal. If this holds, then $(V_\kappa, \in, \mathcal{P}(V_\kappa))$ contains a proper class of worldly cardinals.[38] However, the CIMH implies that there is a definable inner model of the form $L[r]$, where r is a real, with no worldly cardinals (see Theorem 15 of [Friedman, 2006]). This implies that "There is a real r such that for every β, $L_\beta[r] \not\models \mathsf{ZFC}$" is in $\mathsf{ZFC}^{\mathsf{CIMH}}$ (we have effectively quantified away the parameter r). We argue for a contradiction by showing (via an argument due to Joel-David Hamkins) that worldliness is downwards absolute to models of the form $L[r]$, for r a real. Let $\beta < \kappa$ be a worldly cardinal in V_κ. Since $r \in V_\beta$ and $V_\beta \models \mathsf{ZFC}$, it is a standard theorem of relative constructibility[39] that for any r,

[38] β is *worldly* iff $V_\beta \models \mathsf{ZFC}$. We have such a proper class in V_κ since $(V_\kappa, \in, \mathcal{P}(V_\kappa))$ satisfies MK (where MK is Morse-Kelley class theory—NBG augmented with the *impredicative* comprehension scheme), which in turn allows us to prove the existence of such a class.

[39] See Theorem 13.22 on p. 192 of [Jech, 2002].

$(L_\beta[r])^{V_\beta} \models \mathsf{ZFC}$. We now just need to check that $(L_\beta[r])^{V_\beta} = (V_\beta)^{L_\kappa[r]}$, but this follows from the fact that β is a $\beth$-fixed point (a consequence of the worldliness of β) in both V_κ and hence $L_\kappa[r]$. Thus $(V_\beta)^{L_\kappa[r]} \models \mathsf{ZFC}$ (for any r) contradicting the claim (in $\mathsf{ZFC}^{\mathsf{CIMH}}$) that there is an r such that $L_\kappa[r]$ contains no worldly cardinals, and so if V_κ is inaccessible then V_κ cannot satisfy $\mathsf{ZFC}^{\mathsf{CIMH}}$. This deals with (ii). For (iii) note that *any* inner model $\mathfrak{M}$ of V_κ such that $\mathfrak{M} \models \mathsf{ZFC}^{\mathsf{CIMH}}$ will have to contain a model of the form $L_\kappa[r]$ containing no worldly cardinals. Since any r is also in V_κ if it is in $\mathfrak{M}$, we can again build $L_\kappa[r]$ to obtain a $L_\kappa[r] \subseteq V_\kappa$ with no worldly cardinals; a contradiction. This proves Claim 22 and hence Proposition 21.[40] □

Remark 23. The downward absoluteness of worldly cardinals to inner models of the form $L[r]$ for r a real is especially interesting for two related reasons. First, it shows that the CIMH prohibits not just the existence of inaccessible cardinals in V, but worldly cardinals too. Secondly, it shows that $\mathsf{ZFC}^{\mathsf{CIMH}}$ will strongly maximise over any natural parameter-free extension of ZFC of weaker consistency strength (witnessed by a fair interpretation in $\mathsf{ZFC}^{\mathsf{CIMH}}$) that proves "There exists a worldly cardinal".

We can thus see that the CIMH has maximising properties with respect to large cardinals, in particular how it shows them to be restrictive in the sense of Maddy-restrictiveness (our current way of getting a grip on maximality in interpretive power). Of course, for stronger large cardinals that are capable of proving the CIMH consistent (e.g. anything stronger than the existence of a Woodin cardinal with an inaccessible above), it is not possible to provide a fair interpretation of those large cardinals within $\mathsf{ZFC}^{\mathsf{CIMH}}$ alone, and so neither strongly maximizes over the other. However, if we are able to augment our theory of NBG + CIMH (albeit somewhat artificially) with a a claim that allows us to define an inner model of ZFC with the relevant large cardinals in a parameter-free way, then parallel reasoning yields the same restrictiveness result. One can do this, for example, by asserting the existence of mice; small structures that allow us to construct inner models for large cardinals by iterated ultrapowers. If we have a mouse $\mathfrak{N}$ whose iterated ultrapower generates an inner model of $\mathfrak{M} \models \phi$ for some large cardinal ϕ, the first-order part of the theory NBG + CIMH + "$\mathfrak{N}$ exists" will strongly maximise over ZFC + ϕ as before, assuming that the existence of the relevant mouse is consistent with the CIMH,

[40] I am grateful to Kameryn Williams and Victoria Gitman for some useful discussions concerning this proof, and to an anonymous reviewer for pressing me on the details with parameters.

and the model can be defined without using parameters. There are some limitations here since; (i) the CIMH implies that the reals are not closed under $\sharp$ and PD is false,[41] and (ii) the existence of mice generating n-many Woodin cardinals for every n is equivalent to PD. We thus cannot go as far as ω-many Woodin cardinals using this tactic. These complications aside, for many large cardinals the restrictiveness result does hold, and we'll return to these issues in §6.

4 The CIMH and set-existence maximality

We have reached a point where:

(1.) We have seen that there are axioms (e.g. CIMH) that have anti-large cardinal properties.

(2.) There is an apparent standoff: From the perspective of the advocate of large cardinals the CIMH appears to consider only very small transitive models, and from the perspective of the supporter of the CIMH, the truth of large cardinal axioms requires missing out subsets that witness accessibility and/or non-worldliness.

(3.) If we analyse this debate in terms of Maddy's notion of restrictiveness, it is the large cardinal axioms, at least up to the level of many measurable cardinals, that appear restrictive.

We are thus at a point where some large cardinal axioms are viewed as restrictive relative to maximality in interpretive power given theories based on the CIMH. However, it is one thing to provide an axiom for which the restrictiveness results hold, and another to argue that said axiom is a reasonable one. Maddy herself is aware that her notion of restrictiveness delivers far too many false-positives when 'dud' theories are considered. So, is the CIMH (and the first-order theory it generates) a 'dud'?

In this section I argue that the CIMH can be motivated along similar lines to bounded forcing axioms via considerations of maximality in set-existence. In seeing this we shall use the idea of *absoluteness* (in width) which is appealed to by both the friend of bounded forcing axioms and the supporter of the CIMH. We'll see though that these are calibrated in very different ways. On this basis, I'll argue that the CIMH should count as a natural axiom in the sense of [Bagaria, 2005].

One can formulate a general template for a width-absoluteness principle as follows:

[41] This also holds in virtue of Theorem 15 in [Friedman, 2006].

Width Absoluteness Principles. Let Γ be a class of sentences in some appropriate logic. If $\phi \in \Gamma$ is true in some appropriate extension of V with the same ordinals (i.e. a *width* extension) then ϕ is already realised in some appropriate structure contained in V.

Clearly the idea of a width absoluteness principle is schematically formulated, and the content a width absoluteness principle has will be relative to the logical resources, extensions, and internal structures allowed. Some precedents exist for justification of axioms by this means. *Bounded forcing axioms* are a good example here. To facilitate understanding of the ideas later in this section, we first provide a very coarse and intuitive sketch of the forcing technique.

Forcing, loosely speaking, is a way of adding subsets of sets to certain kinds of model. For some model $\mathfrak{M}$ and atomless partial order $\mathbb{P} \in \mathfrak{M}$, we (via ways of naming possible sets and evaluating these names) add a set G that intersects every dense set of $\mathbb{P}$ in $\mathfrak{M}$.[42] The resulting model (often denoted by '$\mathfrak{M}[G]$'), can be thought of as the smallest object one gets when one adds G to $\mathfrak{M}$ and closes under the operations definable in $\mathfrak{M}$.

A *forcing axiom* expresses the claim that the universe has been saturated under forcing for certain kinds of partial order and families of dense sets. For example we have the following axiom:

Definition 24. Let κ be an infinite cardinal. $\mathsf{MA}(\kappa)$ is the statement that for any forcing poset $\mathbb{P}$ in which all antichains are countable (i.e. $\mathbb{P}$ has the countable chain condition), and any family of dense sets $\mathcal{D}$ such that $|\mathcal{D}| \leq \kappa$, there is a filter G on $\mathbb{P}$ such that if $D \in \mathcal{D}$ is a dense subset of $\mathbb{P}$, then $G \cap D \neq \emptyset$.

Definition 25. *Martin's Axiom* (or just MA) is the statement that for every κ smaller than the cardinality of the continuum, $\mathsf{MA}(\kappa)$ holds.

One can think of Martin's axiom in the following way: The universe has been saturated under forcing for all posets with a certain chain condition and less-than-continuum-sized families of dense sets.

There are several kinds of forcing axiom, each corresponding to different permissions on the kind of forcing poset allowed (the countable chain condition is quite a restrictive requirement). Many of these have interesting consequences for the study of independence, notably many

[42] A subtle philosophical and technical question is exactly which models are extendible in width (e.g. must the model be countable?) and how we should understand the metamathematics of this practice, given different ontological outlooks. See [Barton, 2020] for discussion.

(e.g. the Proper Forcing Axiom) imply that CH is false and that in fact $2^{\aleph_0} = \aleph_2$.

It is, however, unclear exactly why we should accept forcing axioms. As it stands, though they seem to correspond to some rough idea of 'saturating' under forcing, they are nonetheless combinatorially characterised principles, and it is not clear if this idea can be cashed out in more foundational terms.[43]

One idea is to capture some of the content of forcing axioms by assimilating them under principles of width absoluteness. This project has been developed by Bagaria who provides the following characterisations of bounded forcing axioms:

Definition 26. [Bagaria, 1997] (ZFC) *Absolute*-MA. We say that V satisfies *Absolute*-MA iff whenever $V[G]$ is a generic extension of V by a partial order $\mathbb{P}$ with the countable chain condition in V, and $\phi(x)$ is a $\Sigma_1(\mathcal{P}(\omega_1))$ formula (i.e. a first-order formula containing only parameters from $\mathcal{P}(\omega_1)$), if $V[G] \models \exists x \phi(x)$ then there is a y in V such that $\phi(y)$.

and we can characterise the Bounded Proper Forcing Axiom (BPFA) as follows:

Definition 27. [Bagaria, 2000] (ZFC) *Absolute*-BPFA. We say that V satisfies *Absolute*-BPFA iff whenever ϕ is a Σ_1 sentence with parameters from $H(\omega_2)$, if ϕ holds a forcing extension $V[G]$ obtained by proper forcing, then ϕ holds in V.

and Bounded Martin's Maximum (BMM):

[43]There are those that think that forcing axioms are well-justified just on the basis of the saturation idea. Magidor, for example, argues:

> Forcing axioms like Martin's Axiom (MA), the Proper Forcing Axiom (PFA), Martin's Maximum (MM) and other variations were very successful in settling many independent problems. The intuitive motivation for all of them is that the universe of sets is as rich as possible, or at the slogan level: A set [whose] existence is possible and there is no clear obstruction to its existence [exists]...
>
> ...What do we mean by "possible"? I think that a good approximation is "can be forced to [exist]"... I consider forcing axioms as an attempt to try and get a consistent approximation to the above intuitive principle by restricting the properties we talk about and the the forcing extensions we use. ([Magidor, U], pp. 15–16)

Magidor is clear that the idea is rough, and sees forcing axioms as a way of making this precise. However, it seems that his motivations apply equally well to the idea of width absoluteness, which is the focus of this paper.

Definition 28. [Bagaria, 2000] (ZFC) *Absolute*-BMM. We say that V satisfies *Absolute*-BMM iff whenever ϕ is a Σ_1 sentence with parameters from $H(\omega_2)$, if ϕ holds a forcing extension $V[G]$ obtained by a forcing $\mathbb{P}$ that preserves stationary subsets of ω_1, then ϕ holds in V.

Each of these axioms shows how one can encapsulate bounded forcing axioms using absoluteness principles. One might think that this provides evidence for their truth, or at least their *naturalness*:

> In the case of MA and some weaker forms of PFA and MM, some justification for their being taken as true axioms is based on the fact that they are equivalent to principles of generic absoluteness. That is, they assert, under certain restrictions that are necessary to avoid inconsistency, that everything that might exist, does exist. More precisely, if some set having certain properties could be forced to exist over V, then a set having the same properties already exists (in V). ([Bagaria, 2008], pp. 319–320)

These formulations and remarks make it particularly perspicuous the sense in which some bounded forcing axioms can be thought of as maximising the universe under 'possible' sets; if we could force there to be a set of kind ϕ (for a particular kind of ϕ and $\mathbb{P}$), one already exists in V.[44] There is a clear sense in which such an intuition corresponds to a natural idea about mathematics: If it is possible to have an object such that ϕ, then there actually is such an object—mathematics should not be constrained by the limits of what is actual rather than possible.

Importantly for us, the CIMH is clearly a kind of width absoluteness principle, asserting that anything true in an inner model of an outer model is already true in an inner model of V. Moreover, it conforms to criteria laid out by Bagaria (in [Bagaria, 2005]) on what it is to be a *natural* axiom of set theory. His criteria (which he calls *meta*-axioms of set theory) he terms *Consistency*, *Maximality*, and *Fairness*.[45] We look at each of them in turn. First:

[44] For some discussion of the coding of Absolute-MA (and similar principles) for the philosopher inclined towards a "universist" picture of set-theoretic ontology see [Barton, 2020] and [Antos et al., 2021]. .

[45] He also mentions the criterion of *Success* for evaluating axioms determined to be natural on the basis of Consistency, Maximality, and Fairness. I won't discuss this here since (a) I have reservations about how we assess the 'success' of an axiom, and (b) at this stage, we're just assessing whether or not the CIMH is a 'dud' for the purposes of the restrictiveness argument, and arguing that is natural is presumably sufficient for showing non-dud-ness. In any case, the CIMH has several interesting consequences, and provides a cohesive (if controversial) perspective on the nature of V, and so is successful in some sense.

(Consistency) The new axiom should be consistent with ZFC.

By the results of [Friedman et al., 2008], if the existence of a Woodin cardinal with an inaccessible above is consistent, then the CIMH is consistent. Thus, the CIMH passes this test if certain large cardinal axioms are consistent.

Bagaria's second meta-axiom we have already seen in §1. We have:

(Maximality) The more sets the axiom asserts to exist, the better.

Bagaria acknowledges that this criteria is somewhat vague, and makes it precise as follows:

> To attain a more concrete and useful form of the Maximality criterion it will be convenient to think about maximality in terms of models. Namely, suppose V is the universe of all sets as given by ZFC, and think of V as being properly contained in an ideal larger universe W which also satisfies ZFC and contains, of course, some sets that do not belong to V—and it may even contain V itself as a set—and whose existence, therefore, cannot be proved in ZFC alone. Now the new axiom should imply that some of those sets existing in W already exist in V, i.e., that some existential statements that hold in W hold also in V. [Bagaria, 2005]

Bagaria thus holds that we can cash out maximality in terms of absoluteness, asserting that existential sentences true in extensions are already true in V. Indeed, Bagaria argues that in addition to bounded forcing axioms, many *large cardinals* should be counted as natural on this basis, since they can be thought of as capturing absoluteness through the use of reflection principles or embeddings to inner models. Bagaria's condition does not then immediately tell in favour of the CIMH *over* large cardinals. However, if we are trying to determine whether the CIMH is a natural axiom (rather than a dud), its status as a width absoluteness principle attempting to capture set-existence maximality counts in its favour.

As Bagaria notes, one cannot have width absoluteness principles without some restrictions, since both CH and $\neg$CH can both be formulated as existential sentences; the former by postulating the existence of sets of reals and the latter by asserting that functions exist between $\mathcal{P}(\omega)$ and subsets thereof. In order to maintain consistency, Bagaria recommends a restriction to Σ_1-sentences.

Whilst the CIMH is not exactly of this form, it is close. Instead of restricting to Σ_1-sentences, we obtain (probable) consistency by asserting that sentences that are realisable in tame class forcing extension are already realised in *inner models* of V. Moreover, if we allow predicative second-order quantification there are existential formulations of the CIMH that do have a formulation in terms of Σ^1_1-sentences. For example we can characterise the CIMH using the following definition:

Definition 29. (NBG) A formula is *persistent-*Σ^1_1 iff it is of the following form:

$$(\exists M)(\text{“}M \text{ is a transitive class”} \wedge M \models \psi)$$

where ψ is first-order.

Definition 30. (NBG) *Tame parameter-free persistent* Σ^1_1*-absoluteness* is the claim that if ϕ is persistent-Σ^1_1 and true in a tame class-generic extension of V, then ϕ is true in V.

Theorem 31. [Friedman, 2006] (NBG) The CIMH is equivalent to tame parameter-free persistent Σ^1_1-absoluteness.[46]

In this way, we can view the CIMH as a generalisation of the following theorem of ZFC (as [Friedman, 2006] notes):

Theorem 32. (ZFC) *Parameter-Free Lévy-Shoenfield Absoluteness*. Let ϕ be a parameter-free Σ_1-sentence. If ϕ is true in an outer model of V, then ϕ is true in V.

Thus the CIMH can be thought of as a principle along the lines that Bagaria suggests—asserting that anything (of a particular kind) that 'could' have existed already has a witness. Moreover, it does so by generalising an idea already present in ZFC. In this respect, it resembles a reflection principle for height: A standard principle of absoluteness true in ZFC is generalised to a language of higher-order.

Bagaria's third condition concerns how maximality through absoluteness is applied. Given that there are no a priori reasons for accepting one existential statement true in some extension over another, we should accept all statements of the same complexity. This motivates the following criterion:

[46] Friedman in fact proves (Theorem 14 of [Friedman, 2006]) that the IMH is equivalent to parameter-free persistent Σ^1_1-absoluteness (i.e. there is no restriction to tame forcings in the statement of his theorem). However, one can simply run his argument replacing 'outer model' with 'tame class forcing extension' everywhere and get the same result for the CIMH and tame parameter-free persistent Σ^1_1-absoluteness, and so we attribute the result to him.

(Fairness) One should not discriminate against sentences of the same logical complexity and (where parameters are concerned) one should not discriminate against sets of the same complexity.

I contend that the CIMH also satisfies the Fairness condition, or at least comes very close. Since the CIMH does not allow the use of parameters, the constraint to not discriminate against different parameters is vacuously satisfied.[47] Concerning the discrimination against sentences, the usual version of the CIMH concerns sentences of arbitrary complexity, and so does not discriminate on these grounds.

Of course one might then object that the usual version of the CIMH, whilst it does not discriminate in terms of Fairness, does do so in terms of Maximality, as formulated by Bagaria, since it reflects truth in outer models to inner model of V, not V itself. On the other hand, the formulation of the CIMH in terms of tame parameter-free persistent Σ^1_1-absoluteness, whilst it does not discriminate on the basis of Maximality (since it reflects directly to truth in V), does discriminate on the basis of only reflecting the *persistent* Σ^1_1-sentences, rather than all of them.

I think there are a couple of responses here. The first point to bear in mind is that that we are currently trying to determine whether or not the CIMH is a dud, in order to run the argument concerning Maddy-restrictiveness. In this context, we might think that even if the CIMH does not *exactly* satisfy Bagaria's requirements, it does come desperately close, and this is perhaps sufficient for it to clear the bar of non-dud-ness.

Second, we might point out that there are other respects in which the CIMH is *less* discriminatory than bounded forcing axioms. In particular, all the bounded forcing axioms that Bagaria considers discriminate against tame class forcing extensions, and we might think that fairness in the kind of extension considered is a requirement overlooked by Bagaria. In this way the CIMH incorporates a more liberal and less discriminatory account of possibility than its bounded (distant) cousins.[48]

[47]The use of parameters in the CIMH is prohibited because one could quickly collapse ω_1 in an inner model and hence in V, contradicting ZFC. Nonetheless, there are variants of the IMH that consider the careful introduction of parameters, such as the *Strong Inner Model Hypothesis*, see [Friedman, 2006].

[48]Indeed, we might think that set-forcing is a relatively mild kind of extension. Bukovský's Theorem (in [Bukovský, 1973]) states that if $\mathfrak{M}$ an inner model of $\mathfrak{N}$ definable in $\mathfrak{N}$, and κ a regular uncountable cardinal in $\mathfrak{M}$, then $\mathfrak{M}$ κ-globally covers $\mathfrak{N}$ if and only if $\mathfrak{N}$ is a κ-c.c. set-generic extension of $\mathfrak{M}$. This theorem suggests that set-forcing is relatively mild, since if one model is a set forcing extension of

I thus think that the CIMH is minimally in the running as a contender for an axiom, albeit a controversial one, at least insofar as one accepts NBG class theory. Of course one could reject the use of NBG, but this strikes me as an overly harsh restriction (though [Bagaria, 2005] is keen to make sure all axioms are first-order). For the sake of argument, let us assume from this point on that the CIMH is a natural enough axiom of set theory that can be motivated along *similar* lines to bounded forcing axioms.

Immediately though, we run into an apparent problem. Given that both the CIMH and bounded forcing axioms represent natural set-existence maximising axioms, we may wish to use them in tandem. However, this is not possible:

Proposition 33. (NBG) The CIMH is inconsistent with the BPFA (and hence BMM).

Proof. Say that a cardinal κ is *reflecting* iff κ is regular and $V_\kappa \preceq_{\Sigma_2} V$. [Goldstern and Shelah, 1995] showed that over ZFC, BPFA implies that ω_2^V is reflecting in L, and their arguments hold for any model of the form $L[x]$ where x is a real. Thus, under BPFA, any inner model of the form $L[x]$ contains a reflecting cardinal (and hence an inaccessible) namely ω_2^V. This straightforwardly contradicts the claim that the CIMH implies that there is a model of the form $L[x]$ that contains no inaccessibles. □

What is going on here? The key issue is that any width absoluteness principle depends on a careful calibration between the following factors:[49]

another (by some κ-c.c. forcing), then every function in the extension is already κ-covered by some function in the ground model, which stands in contrast to class forcing (though, whether there could be an analogue for class forcing is still open). See [Friedman et al., F] for further discussion of the Bukovský Theorem.

[49] I am grateful to Matteo Viale for discussion of the nature of the calibration of width absoluteness principles. Whilst a full examination of the whole space of possibilities for absoluteness principles is outside the scope of this article, some alternatives deserve mentioning. [Bagaria, 2006] provides a detailed survey of how the various variables of principles of generic absoluteness can be tweaked to yield different results, and how the consequences of the absoluteness principles depend on the ambient properties of the model. In this vein, Viale has provided a fine-grained analysis of various forcing axioms including (a) how Martin's Maximum can be strengthened and how this relates to category forcings, (b) how many forcing axioms can be characterised as principles of resurrection (c) how many principles of set theory such as AC, Łoś' Theorem, and some large cardinals, can be characterised as forcing axioms, and (c) how these results are able to yield the kind of absoluteness suggested as desirable by Bagaria. See [Viale, 2016], [Viale, 2016a], [Viale, 2016b] for these results. A different approach is suggested by [Venturi, 2020] and taken up by

(1.) What extensions you consider for width absoluteness. For example we might allow set forcing extensions (possibly with some additional constraints—e.g. obtained by forcings with the countable chain condition or proper posets), tame class forcing extensions, arbitrary extensions etc.

(2.) What complexity of sentences you reflect, and in what language.

(3.) What parameters you allow in the reflected sentences.

(4.) Where we reflect the sentences (e.g. to V, to an inner model $M \subseteq V$, to a structure of the form H_κ, etc.).

The problem is that over-generalisation across different areas will result in inconsistency. It is obvious, for instance, that allowing arbitrary parameters and arbitrary set forcing extensions is immediately inconsistent by collapsing ω_1. Or that having Σ_2-sentences reflected to V is inconsistent (since both CH and $\neg$CH are Σ_2). More generally, we know that no two transitive models $\mathfrak{M}$ and $\mathfrak{N}$ with the same ordinals can be fully Σ_1-elementary (with parameters) in one another, since $\underset{\sim}{\Sigma}_1$-elementarity entails that $V_\alpha^{\mathfrak{M}} = V_\alpha^{\mathfrak{N}}$ for every α.[50] Combining the BPFA with the CIMH is, unfortunately, asking for too much; the absoluteness given by BPFA produces large cardinals in all models of the form $L[x]$, but the absoluteness given by the CIMH kills large cardinals in at least one such model.

5 Maddy-restrictiveness redux: Width absoluteness principles

Is there a way to break the deadlock? Since both the CIMH and bounded forcing axioms count as axioms that maximise in set-existence, one attractive idea is to examine how they compare with respect to maximality in interpretive power, and in particular Maddy-restrictiveness. Here we can prove:

Proposition 34. (NBG) The CIMH strongly maximizes over the BPFA, in the sense that ZFC$^{\text{CIMH}}$ strongly maximizes over ZFC + $BPFA$.

[Venturi and Viale, 2019] and [Viale, 2020]; to use Robinson infinite forcing in combination with an analysis of model completion and model companionship in characterising absoluteness properties.

[50] These points, as well as some other easy impossibility results, are made by [Bagaria, 2006], §3. For a proof of the folklore result that $\underset{\sim}{\Sigma}_1$-elementarity entails identity for transitive models with the same ordinals, see Observation 2.4 of [Barton et al., 2020].

Proof. We first need to show that $\mathsf{ZFC}^{\mathsf{CIMH}}$ can provide a fair interpretation of ZFC + BPFA (this will immediately give us the maximzation of $\mathsf{ZFC}^{\mathsf{CIMH}}$ over ZFC + BPFA since they are mutually inconsistent). As $\mathsf{ZFC}^{\mathsf{CIMH}}$ implies the existence of $0^\sharp$, within L we have a reflecting cardinal κ (i.e. a regular cardinal such that $V_\kappa \preceq_{\Sigma_2} V$, and in this case an L-regular cardinal such that $V_\kappa^L \preccurlyeq_{\Sigma_2} L$). [Goldstern and Shelah, 1995] (Theorem 2.11) showed that if κ is reflecting, then there is a proper forcing iteration $\mathbb{P} \subseteq V_\kappa$ of length κ forcing BPFA. Standard facts about L under the existence of $0\sharp$ imply that if L thinks that κ is the least reflecting cardinal, then κ is countable.[51] Thus, there is an L-generic G for said forcing in $L[0\sharp]$. $L[G]$ is then an inner model of ZFC + BPFA. To eliminate the parameter G (and hence get a fair interpretation within $\mathsf{ZFC}^{\mathsf{CIMH}}$) note that since (i) G is in $L[0\sharp]$, (ii) $0\sharp$ is definable if it exists, and (iii) $L[0\sharp]$ has a definable well-order, we can eliminate G via quantification and putting together definable notions into the following formula (informally expressed):

$x \in L[G]$ iff "There is a G such that G is the $L[0\sharp]$-least generic for the Goldern-Shelah forcing over the least L-reflecting cardinal and there is an α such that $x \in L_\alpha[G]$"

Since BPFA is inconsistent with $\mathsf{ZFC}^{\mathsf{CIMH}}$, we get maximization immediately, and inconsistent maximization if we can show that ZFC + BPFA does not maximize over $\mathsf{ZFC}^{\mathsf{CIMH}}$.

We prove this by showing that ZFC + BPFA cannot provide a fair interpretation of $\mathsf{ZFC}^{\mathsf{CIMH}}$. By [Goldstern and Shelah, 1995], we know that over ZFC, BPFA implies that ω_2^V is reflecting in L, and their arguments hold for any model of the form $L[x]$ where x is a real. Thus, under BPFA, any inner model of the form $L[x]$ contains an inaccessible (and hence a reflecting cardinal) namely ω_2^V. This straightforwardly contradicts the claim that the CIMH implies that there is a model of the form $L[x]$ that contains no inaccessibles, and the $L[x]$ of any inner model (possibly satisfying the CIMH) is also the $L[x]$ of V (by the absoluteness of the construction of $L[x]$). Clearly, truncation at an inaccessible leaves the argument unaffected.

For exactly this reason, no consistent extension of ZFC + BPFA can maximise over $\mathsf{ZFC}^{\mathsf{CIMH}}$, since any extension of ZFC + BPFA proves that every model of the form $L[x]$ contains a reflecting cardinal, by the BPFA alone. We therefore get the strong maximization of $\mathsf{ZFC}^{\mathsf{CIMH}}$ over ZFC + BPFA for free. □

Thus, despite the inconsistency between the CIMH and BPFA, the

[51]See Corollary 18.2 on p. 312 of [Jech, 2002].

CIMH appears appears to Maddy-maximize over the BPFA. Does the CIMH strongly maximize over BMM? The following theorem indicates that this may be difficult:

Theorem 35. [Schindler, 2006] (ZFC) BMM implies that for every set X there is an inner model with a strong cardinal containing X.

BMM thus implies that there are inner models containing cardinals outside the reach of the CIMH using current technology. It is thus unclear if the CIMH alone can maximize over BMM. However, the fact that the CIMH implies the existence of a model of the form $L[x]$ with no worldly cardinals has ramifications for interactions with BMM and other generic absoluteness principles (such as those mentioned above). We can immediately identify:

Proposition 36. No theory T that implies that there is an inaccessible (or even worldly) cardinal in $L[x]$ for every real x can ever maximise over $\mathsf{ZFC}^{\mathsf{CIMH}}$. Hence, if we extend $\mathsf{ZFC}^{\mathsf{CIMH}}$ to a consistent extension $\mathsf{ZFC}^{\mathsf{CIMH}*}$ that has a parameter-free definable inner model for T, $\mathsf{ZFC}^{\mathsf{CIMH}*}$ will strongly maximize over T.

Proof. By assumption, any such $\mathsf{ZFC}^{\mathsf{CIMH}*}$ inconsistently maximizes over T. But also by assumption, no consistent extension of T can find a fair interpretation of $\mathsf{ZFC}^{\mathsf{CIMH}*}$ (exactly as in Proposition 34), and so we have strong maximization. □

Since BMM is exactly one such T, if we find a reasonable extension of $\mathsf{ZFC}^{\mathsf{CIMH}}$ then such an extension will strongly maximize over ZFC + BMM.

Thus, whilst the CIMH is inconsistent with other width absoluteness principles, it is the other putative axioms that seem restrictive. Moreover, it is the fact that the CIMH has such *strong* anti-large cardinal properties (prohibiting even principles that imply that all inner models of the form $L[x]$ for x a real have large cardinals) that gives it these maximisation properties.

6 Foundational roles of large cardinal axioms under the CIMH

We are now in a position where we have seen that:

(a) The CIMH can be motivated along lines similar to other principles of absoluteness such as bounded forcing axioms, and hence should count as a natural set-existence maximising principle (along the lines of Bagaria's account of natural axioms).

(b) If true, the CIMH implies that some large cardinal axioms (up to the level of many measurable cardinals) are Maddy-restrictive. Moreover, if the CIMH is extended to include definable inner models for stronger large cardinals, then we get strong maximisation over these cardinals too.

(c) Whilst the CIMH is inconsistent with many bounded forcing axioms, it Maddy-maximises over some of them (e.g. BPFA), and if extended to include inner models of others (e.g. BMM) maximises over them too.

We thus seem to have a legitimate perspective on set theory on which large cardinal axioms are false and restrictive, but consistent. Earlier however (§1) we identified the following features of large cardinals:

(1.) Large cardinals appear to be linearly ordered by consistency strength and are the standard indicies of consistency strength.

(2.) They are used in various technical model-building constructions in inner model theory.

(3.) They can be used to prove axioms of definable determinacy.

(4.) They are often regarded as species of 'maximality' axiom, in particular with respect to *maximality in interpretive power* and *maximality in set-existence*.

Our previous arguments put pressure on elements of (4.): There are set-theoretic frameworks on which large cardinal axioms, whilst counting as set-existence maximality axioms, are in fact *restrictive* with respect to interpretive power. In this section, we'll argue that nonetheless large cardinal axioms can still fulfil roles (1.)–(2.), and (3.) remains open.

Point (1.) can be dealt with very quickly. In order to study the consistency strengths of mathematical theories, we only require that the theories be true in *some* model or other, not necessarily in V. More generally, there are the following 'levels' for where an axiom Φ can be true:

(i) Φ could be true in V.

(ii) Φ could be true in an inner model.

(iii) Φ could be true in a transitive model.

(iv) Φ could be true in a countable transitive model.

(v) Φ could be true in some model (whatever it may be).

For consistency statements, any model will do, and so any of (i)–(v) are acceptable places for considering Φ. There is no obstacle to having any of (ii)–(v) for the friend of the CIMH (or any other anti-large cardinal principle). Indeed they may well want to accept the consistency (in some model or other) of ZFC + "There is a Woodin cardinal with an inaccessible above", since this allows them to prove their theory consistent. There is no incoherence here; it is just that for the friend of the CIMH, large cardinal 'axioms' form a body of false but useful principles.

This, as we saw, played out in the role of large cardinals concerning interpretive power. In order to maximise interpretive power we just need some 'appropriate' (e.g. well-founded, containing all ordinals) place where the relevant mathematics can be developed. But our earlier observations concerning the maximising properties of the CIMH show that we can perfectly well have large cardinals in inner models, and indeed this is positively *implied* for many large cardinals.[52] Thus far from their being a *loss* of interpretive power there may in fact be a *gain*. Any interpretability work that could be done using a large cardinal axiom can be done in an inner model, without requiring that the axiom be true.

Since they are interrelated, let's examine (2.) (model-building) and (3..) (the case for axioms of definable determinacy) in tandem. For (2.) we should begin by noting that there are a wide variety of model building enterprises that set theorists engage in. In many cases, we try to build models that are L-like in that we can determine a rich variety of their properties (e.g. satisfying the GCH), but also satisfy some large cardinal axiom. Often such models are of the form $L[E]$ where E is a set or a class, and in this vein we can consider $L[\emptyset] = L$ (the vanilla constructible hierarchy), $L[\mathcal{M}]$ (where $\mathcal{M}$ is the class of all mice—this is the Dodd-Jensen core model), $L[U]$ (where U is an ultrafilter on the least measurable cardinal), $L[\mathcal{U}]$ (where $\mathcal{U}$ is a proper class of ultrafilters; one for each measurable cardinal), and so on.[53] For many of these models we can simply build them in V, exactly as from the perspective of the large cardinal theorist. For example we can construct L as normal, and since the CIMH implies the existence of $0^\sharp$ we can build $L[0^\sharp]$ too.

[52][Arrigoni and Friedman, 2013] also make this point.

[53]See [Mitchell, 2010] for an outline of inner model theory.

There are, however, some limitations here. The CIMH implies that the reals are not closed under $\sharp$, and so there is some real x for which $x^\sharp$ (and hence $L[x^\sharp]$) does not exist. However, in these cases we can (if we so desire) interpret the construction as conducted within an inner model on which the reals *are* closed under sharp. This possibility shows how one can interpret a construction as building a *smaller* inner model *within* a proper inner model of V with the required properties.[54] Within this perspective all the usual technical work can be carried out (such as comparing different ultrapower iterations and so on). In this respect, there is still a place to interpret inner model theory in a *natural* way.

However, there is a sense then in which the CIMH provides a different picture of the kinds of L-like model that can be built from reals compared to the large cardinal theorist. For the large cardinal theorist there is often a *unique* real corresponding to the mouse/mice from which we want to build the model. For the friend of the CIMH the real we choose will only have its properties relative to a *perspective* provided by a *proper inner model*, and so there is not in general a unique real corresponding to the building of some L-like model.[55] I remain agnostic about whether or not this is merely a matter of taste or represents an objection.

For some philosophical applications of inner model theory, however, this difference in how models relate to reals is immaterial. For example, one philosophical application of the existence of L-like models is given by John Steel who writes:

> Canonical inner models admit a systematic, detailed, "fine structure theory" much like Jensen's theory of L. Such a thorough and detailed description of what a universe satisfying H might look like provides evidence that H [a large cardinal axiom] is indeed consistent, for a voluble witness with an inconsistent story is more likely to contradict himself than a reticent one. ([Steel, 2014], p. 156)

Steel's point is the following. Given a large cardinal axiom H, we might (rightly) be concerned about its consistency. However, if we can construct an L-like inner model $\mathfrak{M}_H$ with the requisite structure theory (often this is founded on some form of condensation) then we have a huge amount of information about $\mathfrak{M}_H$, for example such models usually satisfy the GCH and versions of principles like $\diamondsuit$ and $\square$. This

[54] This point is also made by [Arrigoni and Friedman, 2013].

[55] Though see below for some possible modifications to the CIMH that might avoid this feature.

should give us confidence that H is consistent, since we know that *if* H is consistent *then* it is also consistent with the GCH etc. (and indeed anything that can be forced over $\mathfrak{M}_H$ whilst preserving H). We might, therefore, expect any inconsistency encoded by H to turn up in one of these models (which would imply the inconsistency of H by modus tollens), and since we have not discovered any inconsistency in the inner model we can be more confident that H is indeed consistent. This contrasts with those large cardinals for which we do not yet have an inner model theory, since we do not have such information-rich contexts in which to examine them. We can now simply note that Steel's argument does not depend on there being a *unique* real in any way, it is enough for his argument to work that there is *some* information-rich context(s) in which H is satisfied, and these models can perfectly well all be constructed within different proper inner models.

The discussion of inner models immediately brings us on to (3.) the case for axioms of definable determinacy. Whilst it is *not* the case that a principle having anti-large cardinal features *immediately* disqualifies the justificatory case for PD found in the literature, we will see that there are again limitations when it comes to the CIMH.

Anti-large cardinal frameworks can incorporate axioms of definable determinacy because they do not require the *literal truth* of large cardinal axioms, but rather only the truth of the large cardinals axioms in inner models. Generally speaking this is where there are equivalences (rather than strict implications from the large cardinals to axioms of definable determinacy). For example[56]:

Theorem 37. (Woodin) The following are equivalent:

(a) Projective Determinacy (schematically rendered).

(b) For every $n < \omega$, there is a fine-structural, countably iterable inner model $\mathfrak{M}$ such that $\mathfrak{M} \models$ "There are n Woodin cardinals".

Thus it may very well be the case that PD holds, there are plenty of Woodin cardinals in inner models, but no actual Woodin cardinals in V. More must be done to argue why the existence of such models must be *explained* by truth of the large cardinals, rather than the apparent consistency of the practice.[57]

[56] For a list see [Koellner, 2011].

[57] This is perhaps what lies behind the following idea of Woodin:

> "**A Set Theorist's Cosmological Principle:** The large cardinal axioms for which there is an inner model theory are consistent; the corresponding predictions of unsolvability are true because the axioms are true."

Nonetheless, for the specific case of the CIMH (rather than anti-large cardinal principles in general), we have some limitations. This is due to the fact that the CIMH implies that PD is false outright, because (as noted above) the CIMH implies that it is not the case that for every real x, $x^\sharp$ exists and boldface (i.e. with parameters) $\underset{\sim}{\Pi}^1_1$-determinacy fails. In spite of this, we do have some definable determinacy; the CIMH is consistent with (and in fact implies) *lightface* (i.e. parameter-free) Π^1_1-determinacy.[58] Moreover, it is open whether there could be CIMH-like principles with *some* anti-large cardinal features that are nonetheless consistent with strong axioms of definable determinacy like PD.

For example, suppose that one is moved by justifications for Woodin cardinals and adopts ZFC + "There is a proper class of Woodin cardinals" as one's canonical theory of sets and one is not prepared to give up on this theory in the face of our earlier observations about restrictiveness. Suppose further that one holds that some CIMH-like principle should hold on the basis of absoluteness considerations. We might then formulate the following principle:

Definition 38. (NBG) Let $(V, \in, \mathcal{C})$ be a NBG structure containing a proper class of Woodin cardinals. The Class-Generic Inner Model Hypothesis for Woodins CIMH^W states that if a (first-order, parameter-free) sentence ϕ holds in an inner model of a tame class forcing extension $(V[G], \in, \mathcal{C}[G])$ containing a proper class of Woodin cardinals, then ϕ holds in an inner model of V.

Assuming this axiom is consistent, we would have a version of the CIMH that is consistent with PD (since a proper class of Woodin cardinals implies PD), and the CIMH^W trivially implies that there is such a class. The CIMH^W might still have *some* anti-large cardinal features though. The usual ways of killing large cardinals under the CIMH involve moving to an outer model of the form $L[x]$ such that $L[x] \models \mathsf{ZFC}$, but every level $L_\alpha[x]$ violates ZFC, and pulling this back to V using the CIMH. Assuming then that the existence of a proper class of Woodin cardinals can be given an inner model theory (i.e. there is a model of the form $L[E]$ with sufficient fine structure such that $L[E] \models$ "There

([Woodin, 2011], p. 458)

Woodin's idea is that on the basis of consistency statements, we can make predictions. For example, "There will be no discovery of an inconsistency in the theory ZFC+"There is a Woodin cardinal" in the next 10'000 years" is a prediction ratified by the truth of the theory ZFC + "There is a proper class of Woodin cardinals". I see no reason why this prediction should be explained by the truth of the large cardinal axiom rather than its consistency (possibly witnessed by an inner model).

[58]See [Friedman, 2018], p. 91.

is a proper class of Woodin cardinals"), the results of [Friedman, 2006] (in particular Theorem 15) might well then be generalised to show that over the base theory ZFC + "There is a proper class of Woodin cardinals", the Inner Model Hypothesis for Woodins implies that there is no inaccessible limit of Woodin cardinals in V in the presence of PD. The details appear difficult, especially since the construction of these inner models is complex, and so we leave the question open in the conclusion. However, if consistent, we might thereby obtain an axiom with *some* anti-large cardinal properties, but nonetheless consistent with stronger axioms of definable determinacy.

7 Open questions and concluding remarks

In this paper I have argued that:

(1.) There is a set-theoretic principle of absoluteness (the CIMH) which should count as both *natural* and a *set-existence maximising*.

(2.) Under this principle, some large cardinal axioms (and some bounded forcing axioms) are Maddy-restrictive.

(3.) Large cardinals can still play many of their usual foundational roles on this framework, despite their falsity. Nonetheless, there are some specific questions about how much definable determinacy is desirable.

I'll close with a few philosophical upshots and directions for future research.

Tension with 'height' absoluteness. Throughout this paper we've been considering principles of absoluteness (e.g. BPFA, CIMH) that are 'width-like' in the sense that they consider what is absolute between the universe and some extension of the universe with the same ordinals but different subsets. In this way, the rough motivation is to make V as 'wide' as possible by ensuring that witnesses for certain claims true in extensions exist. These contrast with 'height' absoluteness principles (often called 'reflection principles') that assert that sentences satisfied by the universe are satisfied (suitably relativised) by substructures thereof (usually some V_α). But even second-order height absoluteness (i.e. the claim that if $\phi(A)$ holds then there is a V_α such that $(V_\alpha, \in, A \cap V_\alpha) \models \phi^{V_\alpha}(A)$) implies the existence of inaccessible cardinals. Both count as principles of set-existence maximality. This shows that there is a tension between height and width absoluteness (and thus perhaps set-existence maximality in general); one cannot have both in full generality.[59]

At this point, we might wonder if there are natural weakenings of the CIMH that yield a greater degree of consistency with large cardinals, thereby incorporating the best of both worlds. Some have already been considered, for example [Friedman, 2016] considers the IMH♯, a principle combining the IMH with a certain amount of reflection from height extensions. For stronger large cardinals, however, the question is still open. In this direction we recall questions raised by earlier discussion:

Questions. Is the CIMH^W consistent? If so, does it have substantial anti-large cardinal consequences?

Answering these questions positively would not only provide us with a width absoluteness principle consistent with many large cardinals but destroying others, but would also provide a version of the CIMH consistent with axioms of definable determinacy.

Connection to the iterative conception of set. One might instead push back on the claim that we should be trying to incorporate height absoluteness at all at the expense of width absoluteness. Whilst we have mostly concerned ourselves with the idea of *restrictiveness*, an argument in favour of width absoluteness as privileged as compared to height absoluteness can be obtained by considering the *iterative conception of set*. This tells us to:

1. Take all *possible* sets at successor stages.
2. Continue this process for as long as *possible*.

If one thinks then that the CIMH is a *good* measure of taking all subsets at successor stages, we might simply say that it is *not possible* to take all subsets at successor stages and iterate the stages far enough to satisfy strong height reflection principles or large cardinal axioms. In

[59]This is especially interesting since certain other kinds of width-absoluteness principles—such as Woodin's results concerning the absoluteness of $Th(L(\mathbb{R}))$ or the Inner Model Reflection Principles of [Barton et al., 2020]—are positively *implied* by large cardinal axioms. The situation is complicated by the fact that this tension seems to generalise further: [Barton and Friedman, MS] shows that there are principles related to the CIMH that (i) imply that there are *no* uncountable sets, and (ii) under a certain Maddy-inspired definition of maximisation, show existence of uncountable cardinals to be restrictive (see [Barton, F]).

this sense, width absoluteness is *privileged* in relation to height absoluteness and/or large cardinal existence.[60]

Should we repudiate large cardinals on this basis as definitively false? I want to emphasise that this is *not* my intention. All we have seen is that there are certain perspectives on which large cardinal axioms appear restrictive, and that this calls into question the idea that the *truth* (as opposed to *consistency in inner models*) of large cardinals is needed for maximality in interpretive power. Moreover, there is a question as to whether they *really* postulate 'large' sets in a straightforward sense. Rather, large cardinal axioms postulate a careful calibration between the largeness of ordinals and the kinds of subset that exist within the universe. Even the existence of inaccessibles—large cardinal axioms right at the bottom of the hierarchy—can be viewed as asserting that certain functions (i.e. subsets) witnessing accessibility do not exist.

Nonetheless, there are plenty of places where one can object to the arguments given in this paper, and I am mindful of the phrase "One person's modus ponens is another's modus tollens." One might take my observations to show that instead width absoluteness is not a good measure of subset maximisation. Another option is to take this as a further false positive for Maddy's theory of restrictiveness. Either way, I think that (i) the sense in which large cardinal axioms are taken to be clear examples of maximisation principles, and (ii) the idea that the truth (rather than consistency in inner models) of large cardinals is an essential ingredient of any successful foundational programme are both deserving of serious philosophical scrutiny and require further foundational support.

References

[Antos et al., 2021] Antos, C., Barton, N., and Friedman, S.-D. (2021). Universism and extensions of V. *The Review of Symbolic Logic*, 14(1):112–154.

[60] An analogous (but somewhat speculative, and hence relegated to a footnote) argument runs as follows: What would happen if we gained good evidence of the consistency of Reinhardt cardinals (and other choiceless cardinals) with ZF? Should we accept that AC is limitative and 'prevents' Reinhardt cardinals from being formed? I argue no: The Axiom of Choice is well-motivated on the basis of the iterative conception and the idea of taking *all* subsets at previous stages, and the conclusion should be that the hypothetical consistency of a choiceless cardinal is witnessed by leaving out choice sets somehow (either in a proper inner model, or if the consistency is witnessed by a forcing extension then in a countable transitive model).

[Antos et al., 2015] Antos, C., Friedman, S.-D., Honzik, R., and Ternullo, C. (2015). Multiverse conceptions in set theory. *Synthese*, 192(8):2463–2488.

[Arrigoni and Friedman, 2013] Arrigoni, T. and Friedman, S.-D. (2013). The Hyperuniverse Program. *Bulletin of Symbolic Logic*, 19:77–96.

[Bagaria, 1997] Bagaria, J. (1997). A characterization of Martin's Axiom in terms of absoluteness. *The Journal of Symbolic Logic*, 62(2):366–372.

[Bagaria, 2000] Bagaria, J. (2000). Bounded forcing axioms as principles of generic absoluteness. *Archive for Mathematical Logic*, 39(6):393–401.

[Bagaria, 2005] Bagaria, J. (2005). Natural axioms of set theory and the continuum problem. In *Proceedings of the 12th International Congress of Logic, Methodology, and Philosophy of Science*, pages 43–64. King's College London Publications.

[Bagaria, 2006] Bagaria, J. (2006). *Axioms of generic absoluteness*, page 28–47. Lecture Notes in Logic. Cambridge University Press.

[Bagaria, 2008] Bagaria, J. (2008). Set theory. In *The Princeton Companion to Mathematics*, pages 302–321. Princeton University Press.

[Bagaria and Brooke-Taylor, 2013] Bagaria, J. and Brooke-Taylor, A. (2013). On colimits and elementary embeddings. *The Journal of Symbolic Logic*, 78(2):562–578.

[Barton, 2020] Barton, N. (2020). Forcing and the universe of sets: Must we lose insight? *Journal of Philosophical Logic*, 49(4):575–612.

[Barton, F] Barton, N. (F). Is (un)countabilism restrictive? To appear in the *Journal of Philosophical Logic*. Preprint: `https://philpapers.org/rec/BARIUR`.

[Barton et al., 2020] Barton, N., Caicedo, A. E., Fuchs, G., Hamkins, J. D., Reitz, J., and Schindler, R. (2020). Inner-model reflection principles. *Studia Logica*, 108(3):573–595.

[Barton and Friedman, MS] Barton, N. and Friedman, S. (MS). Countabilism and maximality principles. Manuscript under review. Preprint: `https://philpapers.org/rec/BARCAM-5`.

[Barton and Friedman, 2017] Barton, N. and Friedman, S.-D. (2017). Maximality and ontology: How axiom content varies across philosophical frameworks. *Synthese*, 197(2):623–649.

[Bukovský, 1973] Bukovský, L. (1973). Characterization of generic extensions of models of set theory. *Fundamenta Mathematicae*, 83(1):35–46.

[Drake, 1974] Drake, F. R. (1974). *Set Theory: An Introduction to Large Cardinals*. North Holland Publishing Co.

[Friedman, 2000] Friedman, S.-D. (2000). *Fine Structure and Class Forcing*. de Gruyter. de Gruyter Series in Logic and its Applications, Vol. 3.

[Friedman, 2006] Friedman, S.-D. (2006). Internal consistency and the inner model hypothesis. *Bulletin of Symbolic Logic*, 12(4):591–600.

[Friedman, 2016] Friedman, S.-D. (2016). Evidence for set-theoretic truth and the Hyperuniverse Programme. *IfCoLog J. Log. Appl.*, 4("Proof, Truth, Computation" 3):517–555.

[Friedman, 2018] Friedman, S.-D. (2018). *Evidence for Set-Theoretic Truth and the Hyperuniverse Programme*, pages 75–107. Springer International Publishing, Cham.

[Friedman et al., F] Friedman, S.-D., Fuchino, S., and Sakai, H. (F). On the set-generic multiverse. *Proceedings of the Singapore Programme on Sets and Computations*. Forthcoming.

[Friedman et al., 2008] Friedman, S.-D., Welch, P., and Woodin, W. H. (2008). On the consistency strength of the Inner Model Hypothesis. *The Journal of Symbolic Logic*, 73(2):391–400.

[Gödel, 1931] Gödel, K. (1931). On formally undecidable propositions of *Principia Mathematica* and related systems I. In *[Gödel, 1986]*, pages 144–195. Oxford University Press.

[Gödel, 1947] Gödel, K. (1947). What is Cantor's continuum problem? In *[Gödel, 1990]*, pages 176–187. Oxford University Press.

[Gödel, 1986] Gödel, K. (1986). *Collected Works, Volume I: Publications 1929-1936*. Oxford University Press. Edited by: Solomon Feferman (Editor-in-chief), John W. Dawson, Jr., Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, Jean van Heijenoort.

[Gödel, 1990] Gödel, K. (1990). *Collected Works, Volume II: Publications 1938-1974*. Oxford University Press. Edited by: Solomon Feferman (Editor-in-chief), John W. Dawson, Jr., Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, Jean van Heijenoort.

[Goldstern and Shelah, 1995] Goldstern, M. and Shelah, S. (1995). The bounded proper forcing axiom. *J. Symbolic Logic*, 60(1):58–73.

[Hamkins, 2014] Hamkins, J. D. (2014). A multiverse perspective on the axiom of constructibility. In *Infinity and Truth*, volume 25, pages 25–45. World Sci. Publ., Hackensack, NJ.

[Hauser, 2001] Hauser, K. (2001). Objectivity over Objects: a Case Study in Theory Formation. *Synthèse*, 128(3).

[Holy et al., 2019] Holy, P., Lücke, P., and Njegomir, A. (2019). Small embedding characterizations for large cardinals. *Annals of Pure and Applied Logic*, 170(2):251–271.

[Incurvati, 2017] Incurvati, L. (2017). Maximality principles in set theory. *Philosophia Mathematica*, 25(2):159–193.

[Incurvati and Löwe, 2016] Incurvati, L. and Löwe, B. (2016). Restrictiveness relative to notions of interpretation. *The Review of Symbolic Logic*, 9(2):238–250.

[Jech, 2002] Jech, T. (2002). *Set Theory*. Springer.

[Kanamori, 2009] Kanamori, A. (2009). *The Higher Infinite: Large Cardinals in Set Theory from Their Beginnings*. Springer, 2nd edition.

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

[Koellner, 2011] Koellner, P. (2011). Independence and large cardinals. In *The Stanford Encyclopedia of Philosophy*. `http://plato.stanford.edu/archives/sum2011/entries/independence-large-cardinals/`, Summer 2011 edition.

[Koellner, 2014] Koellner, P. (2014). Large cardinals and determinacy. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2014 edition.

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

[Löwe, 2001] Löwe, B. (2001). A first glance at non-restrictiveness. *Philosophia Mathematica*, 9(3):347–354.

[Löwe, 2003] Löwe, B. (2003). A second glance at non-restrictiveness. *Philosophia Mathematica*, 11(3):323–331.

[Mac Lane, 1971] Mac Lane, S. (1971). *Categories for the Working Mathematician*. Springer.

[Maddy, 1997] Maddy, P. (1997). *Naturalism in Mathematics*. Oxford University Press.

[Maddy, 1998] Maddy, P. (1998). $V = L$ and maximize. In Makowsky, J. A. and Ravve, E. V., editors, *Proceedings of the Annual European Summer Meeting of the Association of Symbolic Logic*, pages 134–152. Springer.

[Maddy, 2011] Maddy, P. (2011). *Defending the Axioms*. Oxford University Press.

[Magidor, U] Magidor, M. (U). Some set theories are more equal. Unpublished manuscript available from `http://logic.harvard.edu/EFI_Magidor.pdf`.

[McLarty, 1992] McLarty, C. (1992). *Elementary Categories, Elementary Toposes*. Clarendon Press.

[McLarty, 2010] McLarty, C. (2010). What does it take to prove Fermat's Last Theorem? Grothendieck and the logic of number theory. *The Bulletin of Symbolic Logic*, 16(3):359–377.

[Mitchell, 2010] Mitchell, W. J. (2010). Beginning inner model theory. In *Handbook of Set Theory*, pages 1449–1495. Springer Netherlands.

[Sargsyan, 2013] Sargsyan, G. (2013). Descriptive inner model theory. *Bulletin of Symbolic Logic*, 19(1):1–55.

[Schindler, 2006] Schindler, R. (2006). *Bounded Martin's Maximum and Strong Cardinals*, pages 401–406. Birkhäuser Basel, Basel.

[Schindler, 2014] Schindler, R. (2014). *Set Theory: Exploring Independence and Truth*. Springer.

[Steel, 2014] Steel, J. (2014). Gödel's program. In Kennedy, J., editor, *Interpreting Gödel*. Cambridge University Press.

[Steel, 2010] Steel, J. R. (2010). *An Outline of Inner Model Theory*, pages 1595–1684. Springer Netherlands, Dordrecht.

[Venturi, 2020] Venturi, G. (2020). Infinite forcing and the generic multiverse. *Studia Logica*, 108(2):277–290.

[Venturi and Viale, 2019] Venturi, G. and Viale, M. (2019). The model companions of set theory.

[Viale, 2016a] Viale, M. (2016a). Category forcings, MM^{+++}, and generic absoluteness for the theory of strong forcing axioms. *J. Amer. Math. Soc*, 29:675–728.

[Viale, 2016b] Viale, M. (2016b). Martin's maximum revisited. *Archive for Mathematical Logic*, 55(1-2):295–317.

[Viale, 2016] Viale, M. (2016). Useful axioms. *ArXiv e-prints*.

[Viale, 2020] Viale, M. (2020). Tameness for set theory i.

[Welch, 2017] Welch, P. (2017). *Obtaining Woodin's cardinals*, pages 161–170. Contemporary Mathematics. American Mathematical Society, United States.

[Woodin, 2001] Woodin, H. (2001). The continuum hypothesis, Part I. *Notices of the American Mathematical Society*, 48(6):569–576.

[Woodin, 2011] Woodin, W. H. (2011). The transfinite universe. In Baaz, M., editor, *Kurt Gödel and the Foundations of Mathematics: Horizons of Truth*, page 449. Cambridge University Press.

[Woodin, 2017] Woodin, W. H. (2017). In search of Ultimate-L: The 19th Midrasha Mathematicae lectures. *The Bulletin of Symbolic Logic*, 23(1):1–109.