Independence and Large Cardinals

Peter Koellner

April 16, 2010

The independence results in arithmetic and set theory led to a proliferation of mathematical systems. One very general way to investigate the space of possible mathematical systems is under the relation of interpretability. Under this relation the space of possible mathematical systems forms an intricate hierarchy of increasingly strong systems. Large cardinal axioms provide a canonical means of climbing this hierarchy and they play a central role in comparing systems from conceptually distinct domains.

This article is an introduction to independence, interpretability, large cardinals and their interrelations. Section 1 surveys the classic independence results in arithmetic and set theory. Section 2 introduces the interpretability hierarchy and describes some of its basic features. Section 3 introduces the notion of a large cardinal axiom and discusses some of the central examples. Section 4 brings together the previous themes by discussing the manner in which large cardinal axioms provide a canonical means for climbing the hierarchy of interpretability and serve as an intermediary in the comparison of systems from conceptually distinct domains. Section 5 briefly touches on some philosophical considerations.

1 Independence

Let us begin with the notion of an *axiom system*. To motivate this notion consider the manner in which justification traditionally proceeds in mathematics. In reasoning about a given domain of mathematics (or, in fact, any domain) the question of justification is successively pushed back further and further until ultimately one reaches principles that do not admit more fundamental justification. The statements at this terminal stage are elected as *axioms* and the subject is then organized in terms of derivability from the base of axioms. In the case of arithmetic this led to the axiom system PA (Peano arithmetic) and in the case of set theory it led to the axiom system ZFC (Zermelo-Frankel set theory with the Axiom of Choice).

Two natural questions arise: (1) If the axioms do not admit of more fundamental justification then how does one justify them? (2) Is the base of axioms sufficiently rich that one can settle every sentence on this basis?

There are two traditional views concerning the epistemological status of axioms. On the first view the axioms do not admit further justification since they are *self-evident*. On the second view the axioms do not admit further justification since they are *definitive of the subject matter*. Each of these views concerning our first question leads to an associated optimistic view concerning our second question—-according to the first optimistic view, all mathematical truths are derivable (in first-order logic) from self-evident truths, while according to the second optimistic view, all mathematical truths are derivable (in first-order logic) from statements that are definitive of the subject matter. Should either of these optimistic views turn out to be correct, then the question of justification in mathematics would take on a particularly simple form: Either a statement would be an axiom (in which case it would be self-evident or definitive of the subject matter (depending on the view under consideration)) or it would be derivable in first-order logic from some such statements.

Unfortunately, these optimistic views came to be challenged in 1931 by Gödel's incompleteness theorems. Here is one version of the second incompleteness theorem:

Theorem 1.1 (Gödel, 1931)**.** *Assume that* PA *is consistent. Then* PA *does not prove* Con(PA)*.*

Here Con(PA) is a statement of arithmetic that expresses the informal statement that PA is consistent.[1] Under slightly stronger assumptions (for example, that PA is Σ^0_1-sound[2]) one can strengthen the conclusion by adding that PA does not prove $\neg$Con(PA); in other words, under this stronger assumption, Con(PA) is *independent* of PA. Thus, we have here a case of a statement of arithmetic (and, in fact, a very simple one) that cannot be settled on the basis of the standard axioms. Moreover, the theorem is completely general—it holds not just for PA but for any sufficiently strong formal system T.

This raises a challenge for the two aforementioned optimistic views concerning the nature of mathematical truth. To begin with it shows that we cannot work with a *fixed* axiom system T. We will always need to introduce new axioms. More importantly, it raises the question of how one is to justify these new axioms, for as one continues to add stronger and stronger axioms the claim that they are either self-evident or definitive of the subject matter will grow increasingly more difficult to defend.

Already in 1931 Gödel pointed out a natural way to justify new axioms. He pointed out that if one moves beyond the natural numbers and climbs the hierarchy of types (the sets of natural numbers, the sets of sets of natural numbers, etc.) one arrives at axioms (the axioms of second-order arithmetic PA_2, the axioms of third-order arithmetic PA_3, etc.) that settle the undecided statements that he discovered. The axiom system for the second level, PA_2, settles the statement left undecided at the first level, namely $\mathrm{Con}(\mathrm{PA})$; in fact, PA_2 proves $\mathrm{Con}(\mathrm{PA})$, which is the desired result. But now we have a problem at the second level. For the second incompleteness theorem shows that (under similar background assumptions to those above) PA_2 does not settle $\mathrm{Con}(\mathrm{PA}_2)$. Fortunately, the axiom system for the third level, PA_3, settles the statement left undecided at the second level, namely $\mathrm{Con}(\mathrm{PA}_2)$. This pattern continues. For every problem there is a solution and for every solution there is a new problem. In this way, by climbing the hierarchy of types one arrives at systems that successively settle the consistency statements that arise along the way.

The above hierarchy of types can be recast in the uniform setting of set theory. The set-theoretic hierarchy is defined inductively by starting with the emptyset, taking the powerset at successor stages $\alpha + 1$, and taking the union at limit levels λ:

$$\begin{aligned} V_0 &= \varnothing \\ V_{\alpha+1} &= P(V_\alpha) \\ V_\lambda &= \textstyle\bigcup_{\alpha<\lambda} V_\alpha \end{aligned}$$

The *universe of sets* V is the union of all such stages: $V = \bigcup_{\alpha\in\mathrm{On}} V_\alpha$, where On is the class of ordinals. The first infinite level V_ω consists of all of the hereditarily finite sets[3] and this level satisfies $\mathrm{ZFC} - \mathrm{Infinity}$. The sets at this level can be coded by natural numbers and in this way one can show that PA and $\mathrm{ZFC} - \mathrm{Infinity}$ are mutually interpretable.[4] The second infinite level $V_{\omega+1}$ is essentially $P(\mathbb{N})$ (or, equivalently, $\mathbb{R}$) and this level satisfies (a theory that is mutually interpretable with) PA_2. The third infinite level $V_{\omega+2}$ is essentially $P(P(\mathbb{N}))$ (or, equivalently, as the set of functions of real numbers) and this level satisfies (a theory that is mutually interpretable with) PA_3. The first three infinite levels thus encompass arithmetic, analysis and functional analysis and therewith most of standard mathematics. In this fashion, the hierarchy of sets and associated set-theoretic systems encompasses the objects and systems of standard mathematics.

Now, should it turn out to be the case that the consistency sentences (and the other, related sentences discovered by Gödel in 1931) were the *only* instances of undecidable statements, then the sequence of systems in the above hierarchy would catch every problem that arises. And although we would never have a *single* system that gave us a complete axiomatization of mathematical truth, we would have a *series* of systems that collectively covered the totality of mathematical truths.

Unfortunately, matters were not to be so simple. The trouble is that when one climbs the hierarchy of sets in this fashion the greater expressive resources that become available lead to more intractable instances of undecidable sentences and this is true already of the second and third infinite levels. For example, at the second infinite level one can formulate the statement PM (that all projective sets are Lebesgue measurable) and at the third infinite level one can formulate CH (Cantor's continuum hypothesis).[5] These statements were intensively investigated during the early era of set theory but little progress was made. The explanation was ultimately provided by the subsequent independence techniques of Gödel and Cohen.

Gödel invented (in 1938) the method of *inner models* by defining the minimal inner model L. This model is defined just as V is defined except that at successor stages instead of taking the *full powerset* of the previous stage one takes the *definable powerset* of the previous stage, where for a given set X the definable powerset $\mathrm{Def}(X)$ of X is the set of all subsets of X that are definable over X with parameters from X:

$$\begin{aligned} L_0 &= \varnothing \\ L_{\alpha+1} &= \mathrm{Def}(L_\alpha) \\ L_\lambda &= \textstyle\bigcup_{\alpha<\lambda} L_\alpha \end{aligned}$$

The inner model L is the union of all such stages: $L = \bigcup_{\alpha\in\mathrm{On}} L_\alpha$. Gödel showed that L satisfies (arbitrarily large fragments of) ZFC along with CH. It follows that ZFC cannot refute CH. Cohen complemented this result by invented (in 1963) the method of *forcing* (or *outer models*). Given a complete Boolean algebra he defined a model $V^{\mathbb{B}}$ and showed that $\neg$CH holds in $V^{\mathbb{B}}$.[6] This had the consequence that ZFC could not prove CH. Thus, these results together showed that CH is independent of ZFC. Similar results hold for PM and a host of other questions in set theory.

These instances of independence are more intractable in that no simple iteration of the hierarchy of types leads to their resolution. They led to a more profound search for new axioms.

Once again Gödel provided the first steps in the search for new axioms. In 1946 he proposed as new axioms *large cardinal axioms*—axioms of infinity that assert that there are very large levels of the hierarchy of types—as new axioms and he went so far as to entertain a generalized completeness theorem for such axioms, according to which all statements of set theory could be settled by such axioms. See Gödel (1946), p. 151.

The purpose of the remainder of this entry is to describe the nature of independence (along with the hierarchy of interpretability) and the connection between independence and large cardinal axioms. The more involved issues of the search for new axioms and the nature of justification in mathematics are treated in the entries "Determinacy and Large Cardinals" (which focuses on PM) and "The Continuum Hypothesis" (which focuses on CH and discusses the current state of the art).

Further Reading: For more on the incompleteness theorems see Smoryński (1977), Buss (1998a), and Lindström (2003). For more on the independence techniques in set theory see Jech (2003) and Kunen (1980).

2 The Interpretability Hierarchy

Our aim is to investigate the space of mathematical theories (construed as recursively enumerable axiom systems). The ordering on the space of such theories that we will consider is that of *interpretability*. The informal notion of interpretability is ubiquitous in mathematics; for example, Poincaré provided an interpretation of two dimensional hyperbolic geometry in the Euclidean geometry of the unit circle; Dedekind provided an interpretation of analysis in set theory; and Gödel provided an interpretation of the theory of formal syntax in arithmetic.

We shall use a precise formal regimentation of this informal notion. Let

T_1 and T_2 be recursively enumerable axiom systems. We say that T_1 is *interpretable* in T_2 ($T_1 \leqslant T_2$) when, roughly speaking, there is a translation τ from the language of T_1 to the language of T_2 such that, for each sentence φ of the language of T_1, if $T_1 \vdash \varphi$ then $T_2 \vdash \tau(\varphi)$.[7] We shall write $T_1 < T_2$ when $T_1 \leqslant T_2$ and $T_2 \nleqslant T_1$ and we shall write $T_1 \equiv T_2$ when both $T_1 \leqslant T_2$ and $T_2 \leqslant T_1$. In the latter case, T_1 and T_2 are said to be *mutually interpretable.* The equivalence class of all theories mutually interpretable with T is called the *interpretability degree of* T.

For ease of exposition we shall make three simplifying assumptions concerning the theories under consideration. First, we shall assume that all of our theories are couched in the language of set theory. There is no loss of generality in this assumption since every theory is mutually interpretable with a theory in this language. For example, as noted earlier, PA and ZFC − Infinity are mutually interpretable. Second, we shall assume that all of our theories contain ZFC − Infinity. Third, we shall assume that all of our theories are Σ^0_1-sound.

The *interpretability hierarchy* is the collection of all theories (satisfying our three simplifying assumptions) ordered under the relation $\leqslant$. We now turn to a discussion of the structure of this hierarchy.

To begin with, there is a useful characterization of the relation $\leqslant$. Let us write $T_1 \subseteq_{\Pi^0_1} T_2$ to indicate that every Π^0_1-statement provable in T_1 is also provable in T_2. A central result in the theory of interpretability is that (granting our simplifying assumptions) $T_1 \leqslant T_2$ iff $T_1 \subseteq_{\Pi^0_1} T_2$. It follows from this characterization and the second incompleteness theorem that for any theory T the theory $T + \mathrm{Con}(T)$ is strictly stronger than T, that is, $T < T + \mathrm{Con}(T)$. Moreover, it follows from the arithmetized completeness theorem that the theory $T + \neg\mathrm{Con}(T)$ is interpretable in T, hence, $T \equiv T + \neg\mathrm{Con}(T)$.

In terms of interpretability there are three possible ways in which a statement φ can be independent of a theory T.

(1) (Single Jump) Only one of φ or $\neg\varphi$ leads to a jump in strength, that is,

$$T + \varphi > T \quad \text{and} \quad T + \neg\varphi \equiv T$$

(or likewise with φ and $\neg\varphi$ interchanged).

(2) (No Jump) Neither φ nor $\neg\varphi$ lead to a jump in strength, that is,

$$T + \varphi \equiv T \quad \text{and} \quad T + \neg\varphi \equiv T.$$

(3) (Double Jump) Both φ and $\neg\varphi$ lead to a jump in strength, that is,

$$T+\varphi > T \text{ and } T+\neg\varphi > T.$$

It turns out that each of these possibilities is realized. For the first it suffices to take the Π^0_1-sentence $\mathrm{Con}(T)$. For the second it is easy to see that there is no example that is Π^0_1; the simplest possible complexity of such a sentence is Δ^0_2 and it turns out that there are such examples; examples of this type of independence are called *Orey sentences*. For the third kind of independence there are Π^0_1 instances. (This is a corollary of Lemma 14 on pages 128–129 of Lindström (2003).)

These are all metamathematical examples, the kind of example that only a logician would construct. It is natural to ask whether there are "natural" examples, roughly the sort of example occurring in the normal course of mathematics. In the set theoretic case, such examples are abundant for the first two cases. For example, PM is an example of the first kind of independence and CH is an example of the second kind of independence. There are no known "natural" examples of the third kind of independence. In the arithmetical case, such examples are rare. There are examples of the first kind of independence (the most famous of which is a classic example due to Paris and Harrington) but none of the second or third kind of independence.

Notice that in the case of the third example the two theories above T are incomparable in the interpretability order. To construct a pair of such Π^0_1-statements one uses a reciprocal form of the diagonal lemma to construct two Π^0_1-statements that refer to one another. Using such techniques can show that the interpretability order is quite complex. For example, for any two theories T_1 and T_2 such that $T_1 < T_2$ there is a third theory T such that $T_1 < T < T_2$. Thus, the order on the degrees of interpretability is neither linearly ordered nor well-founded. (See Feferman (1960).)

Remarkably, it turns out that when one restricts to those theories that "arise in nature" the interpretability ordering is quite simple: There are no descending chains and there are no incomparable elements—the interpretability ordering on theories that "arise in nature" is a wellordering. In particular, although there are natural examples of the first and second kind of independence (e.g. PM and CH, respectively, something to which we will return to below), there are no known natural examples of the third kind of independence.

So, for theories that "arise in nature", we have a wellordered hierarchy under the interpretability ordering. At the base of the ordering one has the degree that is represented by our minimal theory ZFC − Infinity and there is only one way to proceed, namely, upward in terms of strength.

We have already seen one way of climbing the hierarchy of the degrees of interpretability, namely, by adding consistency statements. There are two drawbacks to this approach. First, if one starts with a theory that "arises in nature" and adds the consistency statement one lands in a degree that has no known representative that "arises in nature". Second, the consistency statement does not take one very far up the hierarchy. Both of these drawbacks are remedied by a very natural class of axioms—the large cardinal axioms.

Further Reading: For more on the structure of the interpretability hierarchy see chapters 6–8 of Lindström (2003).

3 Large Cardinal Axioms

Let Z_0 be the theory ZFC − Infinity − Replacement. (This theory is logically equivalent to our base theory ZFC − Infinity.) We shall successively strengthen Z_0 by reflectively adding axioms that assert certain levels of the universe of sets exist.

The standard model of Z_0 is V_ω. The Axiom of Infinity (in one formulation) simply asserts that this set exists. So, when we add the Axiom of Infinity, the resulting theory Z_1 (known as Zermelo set theory with Choice) not only proves the consistency of Z_1; it proves that there is a standard model of Z_1. Now the standard model of Z_1 is $V_{\omega+\omega}$. The Axiom of Replacement implies that this set exists. So, when we add the Axiom of Replacement, the resulting theory Z_2 (known as ZFC), not only proves the consistency of Z_1; it proves that there is a standard model of Z_1.

A standard model of Z_2 has the form V_κ where κ is a regular cardinal such that for all $\alpha < \kappa$, $2^\alpha < \kappa$. Such a cardinal is called a (*strongly*) *inaccessible* cardinal. The next axiom in the hierarchy under consideration is the statement asserting that such a cardinal exists. The resulting theory ZFC + "There is a strongly inaccessible cardinal" proves that there is a level of the universe that satisfies ZFC. Continuing in this fashion one arrives at stronger and stronger axioms that assert the existence of larger and larger levels of the universe of sets. Before continuing with an outline of such axioms let us first draw the connection with the hierarchy of interpretability.

Recall our classification of the three types of independence. We noted that there are no known natural examples of the third kind of independence but that there are natural examples of the first and second kind of independence.

Natural examples of the second kind of independence are provided by the dual method of inner and outer models. For example, these methods show that the theories $\mathrm{ZFC}+\mathrm{CH}$ and $\mathrm{ZFC}+\neg\mathrm{CH}$ are mutually interpretable with ZFC, that is, all three theories lie in the same degree. In other words, CH is an Orey sentence with respect to ZFC. What about that other sentence we introduced: PM?

Using the method of inner models Gödel showed that $\neg$PM holds in L. It follows that $\mathrm{ZFC}+\neg\mathrm{PM}$ is mutually interpretable with ZFC. But what about PM? To show that $\mathrm{ZFC}+\mathrm{PM}$ is mutually interpretable with ZFC a natural approach would be to follow the approach used for CH and build an outer model of ZFC that satisfies PM. However, it is known that this cannot be done starting with ZFC alone. For it turns out (by a result of Shelah (1984)) that ZFC + PM implies the consistency of ZFC and this implies, by the second incompleteness theorem, that $\mathrm{ZFC}+\mathrm{PM}$ is not interpretable in ZFC. In a sense we have here a case of the independence of independence. More precisely, even if we assume that ZFC is consistent we cannot (in contrast to the case of CH) prove that PM is independent of ZFC. To establish the independence of PM from ZFC we need to assume the consistency of a *stronger* theory, namely, that of ZFC + "There is a strongly inaccessible cardinal". For it turns out that ZFC + PM lies not in the interpretability degree of ZFC but rather in that of ZFC + "There is a strongly inaccessible cardinal". To summarize: While CH is a case of the first type independence, PM is a case of the second type independence; it is similar to Con(ZFC) in that it is a sentence φ such that only one of φ or $\neg\varphi$ leads to a jump in strength, only now there are two differences; the jump lands in a degree that is much stronger and it is represented by a natural theory.

In general, the (known) sentences of set theory are either like CH or PM. Some are like CH in that both $\mathrm{ZFC}+\varphi$ and $\mathrm{ZFC}+\neg\varphi$ lie in the degree of ZFC. Others are like PM in that one of $\mathrm{ZFC}+\varphi$ and $\mathrm{ZFC}+\neg\varphi$ lies in the degree of ZFC while the other lies in the degree of an extension of ZFC via a large cardinal axiom.

Let us now return to our overview of large cardinal axioms. After strongly inaccessible cardinals there are Mahlo cardinals, indescribable cardinals, and ineffable cardinals. All of these large cardinal axioms can be derived in a uniform way using the traditional variety of reflection principles (see Tait (2005a)) but there are limitations on how far this variety of reflection principles can take one. For under a very general characterization of such principles it is known that they cannot yield the Erdös cardinal $\kappa(\omega)$. See Koellner (2009).

The large cardinals considered thus far (including $\kappa(\omega)$) are known as *small* large cardinals. A large cardinal is *small* if the associated large cardinal axiom can hold in Gödel's constructible universe L, that is, if "$V \models \kappa$ is a φ-cardinal" is consistent, then "$L \models \kappa$ is a φ-cardinal" is consistent. Otherwise the large cardinal is *large*.

There is a simple template for formulating (large) large cardinal axioms is in terms of elementary embeddings. In general such an axiom asserts that there is a transitive class M and a non-trivial elementary embedding

$$j : V \to M.$$

To say that the embedding is non-trivial is just to say that it is not the identity, in which case there must be a least ordinal that is moved. This ordinal is called the *critical point of* j and denoted $\mathrm{crit}(j)$. The critical point is (typically) the large cardinal associated with the embedding. A cardinal κ is said to be *measurable* iff it is the critical point of some such embedding.[8]

It is easy to see that for any such embedding $V_{\kappa+1} \subseteq M$ where $\kappa = \mathrm{crit}(j)$. This amount of agreement enables one to show that κ is strongly inaccessible, Mahlo, indescribable, ineffable, etc. To illustrate this let us assume that we have shown that κ is strongly inaccessible and let us show that κ has much stronger large cardinal properties. Since κ is strongly inaccessible in V and since $(V_{\kappa+1})^M = V_{\kappa+1}$, M also thinks that κ is strongly inaccessible. In particular, M thinks that there is a strongly inaccessible cardinal (namely, κ) below $j(\kappa)$. But then by the elementarity of j, V must think the same think of the preimage of $j(\kappa)$, namely, κ, that is, V must think that there is a strongly inaccessible below κ. So κ cannot be the least strongly inaccessible cardinal. Continuing in this manner one can show that there are many strongly inaccessibles below κ and, in fact, that κ is Mahlo, indescribable, ineffable, etc. So measurable cardinals subsume the small large cardinals.

In fact, Scott showed that (in contrast to the small large cardinals) measurable cardinals cannot exist in Gödel's constructible universe. Let us be precise about this. Let $V{=}L$ be the statement that asserts that all sets are constructible. Then for each small large cardinal axiom φ (to be precise, those listed above) if the theory ZFC $+\,\varphi$ is consistent then so is the theory ZFC $+\,\varphi + V{=}L$. In contrast, the theory ZFC + "There is a measurable cardinal” proves $\neg V = L$. This may seem somewhat counterintuitive since L contains all of the ordinals and so if κ is a measurable cardinal then κ is an ordinal in L. The point is that L cannot “recognize” that κ is a measurable cardinal since it is too “thin” to contain the ultrafilter that witnesses the measurability of κ.

One way to strengthen a large cardinal axiom based on the above template is to demand greater agreement between M and V. For example, if one demands that $V_{\kappa+2} \subseteq M$ then the fact that κ is measurable (something witnessed by a subset of $P(\kappa)$) can be recognized by M. And so, by exactly the same argument that we used above, there must be a measurable cardinal below κ.

This leads to a progression of increasingly strong large cardinal axioms. It will be useful to discuss some of the major stepping stones in this hierarchy.

If κ is a cardinal and $\eta > \kappa$ is an ordinal then κ is *η-strong* if there is a transitive class M and a non-trivial elementary embedding $j : V \to M$ such that $\text{crit}(j) = \kappa$, $j(\kappa) > \eta$ and $V_\eta \subseteq M$. A cardinal κ is *strong* iff it is η-strong for all $\eta > \kappa$. One can also demand that the embedding preserve certain classes: If A is a class, κ is a cardinal, and $\eta > \kappa$ is an ordinal then κ is *η-A-strong* if there exists a $j : V \to M$ which witnesses that κ is η-strong and which has the additional feature that $j(A \cap V_\kappa) \cap V_\eta = A \cap V_\eta$. The following large cardinal notion plays a central role in the search for new axioms.

Definition 3.1. A cardinal κ is a *Woodin cardinal* if κ is strongly inaccessible and for all $A \subseteq V_\kappa$ there is a cardinal $\kappa_A < \kappa$ such that

$$\kappa_A \text{ is } \eta\text{-}A\text{-strong},$$

for each η such that $\kappa_A < \eta < \kappa$.[9]

One can obtain stronger large cardinal axioms by forging a link between the embedding j and the amount of resemblance between M and V. For example, a cardinal κ is *superstrong* if there is a transitive class M and a non-trivial elementary embedding $j : V \to M$ such that $\text{crit}(j) = \kappa$ and $V_{j(\kappa)} \subseteq M$. If κ is superstrong then κ is a Woodin cardinal and there are arbitrarily large Woodin cardinals below κ.

One can also obtain strong large cardinal axioms by placing closure conditions on the target model M. For example, letting $\gamma \geqslant \kappa$ a cardinal κ is *γ-supercompact* if there is a transitive class M and a non-trivial elementary embedding $j : V \to M$ such that $\text{crit}(j) = \kappa$ and ${}^{\gamma}M \subseteq M$, that is, M is closed under γ-sequences. (It is straightforward to see that if M is closed under γ-sequences then $V_{\gamma+1} \subseteq M$; so this approach subsumes the previous approach.) A cardinal κ is *supercompact* if it is γ-supercompact for all $\gamma \geqslant \kappa$. Now, just as in the previous approach, one can strengthen these axioms by forging a link between the embedding j and the closure conditions on the target model. A cardinal κ is *n-huge* if there is a transitive class M and a non-trivial elementary embedding $j : V \to M$ such that ${}^{j^n(\kappa)}M \subseteq M$, where $\kappa = \text{crit}(j)$ and $j^{i+1}(\kappa)$ is defined to be $j(j^i(\kappa))$.

One can continue in this vein, demanding greater agreement between M and V. The ultimate axiom in this direction would, of course, demand that $M = V$. This axiom was proposed by Reinhardt and shortly thereafter shown to be inconsistent (in ZFC) by Kunen. In fact, Kunen showed that, assuming ZFC, there can be a transitive class M and a non-trivial elementary embedding $j : V \to M$ such that $j``\lambda \in M$, where $\lambda = \sup_{n<\omega} j^n(\kappa)$ and $\kappa = \text{crit}(j)$. In particular, there cannot exists such an M and j such that $V_{\lambda+1} \subseteq M$. This placed a limit on the amount of closure of the target model (in relation to the embedding).[10]

Nevertheless, there is a lot of room below the above upper bound. For example, a very strong axiom is the statement that there is a non-trivial elementary embedding $j : V_{\lambda+1} \to V_{\lambda+1}$. The strongest large cardinal axiom in the current literature is the axiom asserting that there is a non-trivial elementary embedding $j : L(V_{\lambda+1}) \to L(V_{\lambda+1})$ such that $\text{crit}(j) < \lambda$. In recent work, Woodin has discovered axioms much stronger than this.

Further Reading: For more on large cardinal axioms see Kanamori (2003).

4 Large Cardinal Axioms and the Interpretability Hierarchy

The large cardinal axioms discussed above are naturally well-ordered in terms of strength.[11] This provides a natural way of climbing the hierarchy of interpretability. At the base we start with the theory ZFC $-$ Infinity and then we climb to ZFC and up through ZFC $+ \Phi$ for various large cardinal axioms Φ. Notice that for two large cardinal axioms Φ and Ψ, if Ψ is stronger than Φ then Ψ implies that there is a standard model of Φ and so we have a natural interpretation of ZFC $+ \Phi$ in ZFC $+ \Psi$.

We have already noted that $\mathrm{ZFC} + \neg\mathrm{PM}$ is mutually interpretable with $\mathrm{ZFC} + \mathrm{LC}$ where LC is the large cardinal axiom "There is a strongly inaccessible cardinal" and that this is shown using the dual techniques of inner and outer model theory. It is a remarkable empirical fact that for *any* "natural" statement in the language of set theory φ one can generally find a large cardinal axiom Φ such that $\mathrm{ZFC} + \varphi$ and $\mathrm{ZFC} + \Phi$ are mutually interpretable. Again, this is established using the dual techniques of inner and outer model theory only now large cardinals enter the mix. To establish that $\mathrm{ZFC} + \Phi$ interprets $\mathrm{ZFC} + \varphi$ one generally starts with a model of $\mathrm{ZFC} + \Phi$ and uses forcing to construct a model of $\mathrm{ZFC} + \varphi$. In many cases the forcing construction involves "collapsing" the large cardinal associated with Φ and arranging the collapse in such a way that φ holds in the "rubble". In the other direction, one generally starts with a model of $\mathrm{ZFC} + \varphi$ and then constructs an inner model (a model resembling L but able to accommodate large cardinal axioms) that contains the large cardinal asserted to exist by Φ. The branch of set theory known as *inner model theory* is devoted to the construction of such "L-like" models for stronger and stronger large cardinal axioms.

In this way the theories of the form $\mathrm{ZFC} + \mathrm{LC}$, where LC is a large cardinal axiom, provide a yardstick for measuring the strength of theories. They also act as intermediaries for comparing theories from conceptually distinct domains: Given $\mathrm{ZFC} + \varphi$ and $\mathrm{ZFC} + \psi$ one finds large cardinal axioms Φ and Ψ such that (using the methods of inner and outer models) $\mathrm{ZFC} + \varphi$ and $\mathrm{ZFC} + \Phi$ are mutually interpretable and $\mathrm{ZFC} + \psi$ and $\mathrm{ZFC} + \Psi$ are mutually interpretable. One then compares $\mathrm{ZFC} + \varphi$ and $\mathrm{ZFC} + \psi$ (in terms of interpretability) by mediating through the natural interpretability relationship between $\mathrm{ZFC} + \Phi$ and $\mathrm{ZFC} + \Psi$. So large cardinal axioms (in conjunction with the dual method of inner and outer models) lie at the heart of the remarkable empirical fact that natural theories from completely distinct domains can be compared in terms of interpretability.

Further Reading: For an introduction to inner model theory see Section 4 of the entry "Large Cardinals and Determinacy" and for further details see the references therein.

5 Some Philosophical Considerations

The main question that arises in light of the independence results is whether one can justify new axioms that settle the statements left undecided by the standard axioms. There are two views. On the first view, the answer is taken to be negative and one embraces a radical form of pluralism in which one has a plethora of equally legitimate extensions of the standard axioms. On the second view, the answer it taken (at least in part) to be affirmative, and the results simply indicate that ZFC is too weak to capture the mathematical truths. This topic is quite involved and lies outside the scope of the present article. For more on the subject see the entries on "Large Cardinals and Determinacy" and "The Continuum Hypothesis".

But there are other philosophical questions more directly related to the themes of this article. First, what is the significance of the empirical fact that the large cardinal axioms appear to be wellordered under interpretability? Second, what is the significance of the empirical fact that large cardinal axioms play a central role in comparing many theories from conceptually distinct domains. Let us consider these two questions in turn.

One might try to argue that the fact that the large cardinal axioms are wellordered under interpretability is a consideration in their favour. However, this would be a weak argument. For, as we have noted above, *all* "natural" theories appear to be wellordered under interpretability and this includes theories that are incompatible with one another. For example, it is straightforward to select "natural" theories from higher and higher degrees of theories in the wellordered sequence that are incompatible with one another. It follows that the feature of being wellordered under interpretability, while remarkable, can not be a point in favour of truth.

But large cardinal axioms have additional features that singles them out from the class of natural theories in the wellordered sequence of degrees. To begin with they provide the most natural way to climb the hierarchy of interpretability—they are the simplest and most natural manifestation of pure mathematical strength. But more important is the second component mentioned above, namely, the large cardinal axioms act as intermediaries in comparing theories from conceptually distinct domains. For recall how this works: Given ZFC + φ and ZFC + ψ one finds large cardinal axioms Φ and Ψ such that (using the methods of inner and outer models) ZFC + φ and ZFC + Φ are mutually interpretable and ZFC + ψ and ZFC + Ψ are mutually interpretable. One then compares ZFC + φ and ZFC + ψ (in terms of interpretability) by mediating through the natural interpretability relationship between $\mathrm{ZFC}+\Phi$ and $\mathrm{ZFC}+\Psi$.

It turns out that in many cases this is the *only known way* to compare $\mathrm{ZFC}+\varphi$ and $\mathrm{ZFC}+\psi$, that is, in many cases there is no direct interpretation in either direction, instead one *must pass through the large cardinal axioms.* Can this additional feature be used to make a case for large cardinal axioms? The answer is unclear. However, what is clear is the absolute centrality of large cardinal axioms in set theory.

Notes

[1]It will be useful at this point to introduce some basic syntactic notions. The *language of first-order arithmetic* is the language consisting of the usual *logical symbols* (propositional connectives, quantifiers, and equality $(=)$) and the following *non-logical symbols*: The constant symbol 0, the unary successor function S, the binary operators $+$ and $\cdot$, and the binary relation $\leqslant$. The *terms* are generated by starting with 0 and the variables and iteratively applying S, $+$ and $\cdot$. A quantifier is *bounded* if it is of the form $\exists x \leqslant t$ or $\forall x \leqslant t$ where t is a term not involving x. A formula is a *bounded formula* (denoted Δ^0_0) if all of its quantifiers are bounded. For $n \geqslant 0$ the classes of formulas Σ^0_n and Π^0_n are defined as follows: $\Sigma^0_0 = \Pi^0_0 = \Delta^0_0$. Σ^0_{n+1} is the set of formulas of the form $\exists\vec{x}\varphi$ where φ is Π^0_n and $\vec{x}$ is a (possibly empty) list of variables. Π^0_{n+1} is the set of formulas of the form $\exists\vec{x}\varphi$ where φ is Σ^0_n and $\vec{x}$ is a (possibly empty) list of variables. The classes Σ^0_n, Π^0_n constitute the *arithmetical hierarchy of formulas.*

For a formal system T of arithmetic, a formula φ is $(\Pi^0_n)^T$ if it is provably equivalent in T to a Π^0_n formula. Likewise, for $(\Sigma^0_n)^T$. A formula φ is $(\Delta^0_n)^T$ if it is provably equivalent in T to both a Σ^0_n formula and a Π^0_n formula. In many cases, when the context is clear, we shall drop reference to T.

A formal system T of arithmetic is Σ^0_1-*complete* if it proves every true Σ^0_1-statement. The system PA is Σ^0_1-complete. The raises the question of whether PA can capture the truths at the next level of the arithmetical hierarchy, that is, the Π^0_1 truths. The statement Con(PA) is a Π^0_1-statement (informally it asserts "for all natural numbers n, n does not code a proof of a contradiction from the axioms of PA"). Thus, granting the consistency of PA, the incompleteness theorem shows that there is a Π^0_1-truth that cannot be proved in PA.

[2]A system T is Σ^0_1 *sound* iff for every Σ^0_1-statement φ, if T proves φ then φ is true.

[3]In general, for a cardinal κ, a set x is of hereditarily cardinality of size less than κ, if x has size less than κ, all of the members of x have size less than κ, etc., in other words, the transitive closure of x has size less than κ. The set of all sets of hereditary cardinality less than κ is denoted $H(\kappa)$. In particular, $H(\omega) = V_\omega$.

[4]For a more detailed treatment of this notion see the next section. The interpretation of PA in ZFC$-$Infinity was established during the early days of set theory. The interpretation of ZFC $-$ Infinity in PA was established by Ackermann in his (1937).

[5]In descriptive set theory it is convenient to work with the "logician's reals", that is, Baire space, ω^ω. This is the set of all infinite sequences of natural numbers (with the product topology (taking ω to be discrete)).

The *projective* sets of reals (or sets of k-tuples of reals) are obtained by starting with the closed subsets of $(\omega^\omega)^k$ and iterating the operations of complement and projection. For $A \subseteq (\omega^\omega)^k$, the *complement* of A is simply $(\omega^\omega)^k - A$. For $A \subseteq (\omega^\omega)^{k+1}$, the *projection* of A is

$$p[A] = \{\langle x_1, \ldots, x_k\rangle \in (\omega^\omega)^k \mid \exists y\, \langle x_1, \ldots, x_k, y\rangle \in A\}.$$

(Think of the case where $k+1 = 3$. Here A is a subset of three-dimensional space and $p[A]$ is the result of "projecting" A along the third axis onto the plane spanned by the first two axes.) We can now define the hierarchy of projective sets as follows: At the base level let $\underset{\sim}{\Sigma}^1_0$ consist of the open subsets of $(\omega^\omega)^k$ and let $\underset{\sim}{\Pi}^1_0$ consist of the closed subsets of $(\omega^\omega)^k$. For each n such that $0 < n < \omega$, recursively define $\underset{\sim}{\Pi}^1_n$ to consist of the complements of sets in $\underset{\sim}{\Sigma}^1_n$ and $\underset{\sim}{\Sigma}^1_{n+1}$ to consist of the projections of sets in $\underset{\sim}{\Pi}^1_n$. The projective sets of (k-tuples of) reals are the sets appearing in this hierarchy. This hierarchy is also a proper hierarchy, as can be seen using universal sets and diagonalization.

It is a classical result of analysis (due to Luzin and Suslin in 1917) that the $\underset{\sim}{\Sigma}^1_1$ sets are Lebesgue measurable.

[6]There are a number of different ways of formalizing forcing. One can forgo talk of models and treat the independence results proof-theoretically. One can also give a model-theoretic treatment and here there are two approaches. The first approach starts with a countable transitive model M of a sufficiently large fragment of ZFC and then (working in V) one actually builds an outer model $M[G]$ that satisfies a large fragment of set theory and the statement one wishes to show is independent. The second approach involves working with class-size Boolean-valued models. Cohen took the first approach but we have found it simpler to take the second in our treatment. (We caution the reader that set theorists often speak of $V[G]$ as if there were a model larger than the universe of sets. When they do so they are either thinking of V as a countable model (like M above) or thinking of $V[G]$ as a Boolean valued model (like $V^{\mathbb{B}}$). See Kunen (1980), pp. 232–235 for further discussion.

[7]For a precise definition see chapter 6 of Lindström (2003).

[8]The above formulation of large cardinal axioms invokes classes but there are equivalent formulations in terms of sets. For example, in the case of a measurable cardinal κ there is a subset of $P(\kappa)$ (namely, a κ-complete ultrafilter over κ) witnessing that κ is measurable.

[9]It should be noted that in contrast to measurable and strong cardinals, Woodin cardinals are not characterized as the critical point of an embedding or collection of embeddings. In fact, a Woodin cardinal need not be measurable. However, if κ is a Woodin cardinal, then V_κ is a model of ZFC and V_κ satisfies that there is a proper class of strong cardinals.

[10]It is still open whether large cardinal axioms at the level of Reinhardt and beyond are inconsistent in ZF alone (that is, without invoking AC). Should they turn out to be consistent then one would have a hierarchy of "choiceless" large cardinal axioms that would climb the hierarchy of interpretability beyond the large cardinal axioms formulated in the context of ZFC.

[11]We should note that there are many large cardinal axioms that we have not discussed. The large cardinal axioms that have been investigated to date are *for the most part* known

to be naturally well-ordered. However, there are some comparisons that remain open; for example, although it is known that a strongly compact cardinal is weaker than a supercompact cardinal it is not know how it compares with Woodin cardinals.

References

Ackermann, W. (1937). Die Widerspruchsfreiheit der allgemeinen Mengenlehre, *Mathematische Annalen* (114): 305–315.

Barwise, J. K. (ed.) (1977). *Handbook of Mathematical Logic*, Vol. 90 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam.

Buss, S. R. (1998a). First-order proof theory of arithmetic, Vol. 137 of *Sudies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam, chapter in Buss (1998b), pp. 79–147.

Buss, S. R. (ed.) (1998b). *Handbook of Proof Theory*, Vol. 137 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam.

Feferman, S. (1960). Arithmetization of metamathematics in a general setting, *Fundamenta Mathematicae* **49**: 35–92.

Foreman, M. & Kanamori, A. (2009). *Handbook of Set Theory*, Springer-Verlag.

Gödel, K. (1946). Remarks before the Princeton bicentennial conference on problems in mathematics, *in (Gödel 1990)*, Oxford University Press, pp. 150–153.

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

Gödel, K. (1990). *Collected Works, Volume II: Publications 1938–1974*, Oxford University Press, New York and Oxford.

Jech, T. J. (2003). *Set Theory: Third Millennium Edition, Revised and Expanded*, Springer-Verlag, Berlin.

Kanamori, A. (2003). *The Higher Infinite: Large Cardinals in Set Theory from their Beginnings*, Springer Monographs in Mathematics, second edn, Springer, Berlin.

Koellner, P. (2009). On reflection principles, *Annals of Pure and Applied Logic* (157): 206–219.

Kunen, K. (1980). *Set theory: An Introduction to Independence Proofs*, Vol. 102 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam.

Lindström, P. (2003). *Aspects of Incompleteness*, Vol. 10 of *Lecture Notes in Logic*, second edn, Association of Symbolic Logic.

Shelah, S. (1984). Can you take Solovay's inaccessible away?, *Israel Journal of Mathematics* **48**(1): 1–47.

Smoryński, C. A. (1977). The incompleteness theorems, *in Barwise (1977)*, pp. 821–865.

Tait, W. W. (2005a). Constructing cardinals from below, *in (Tait 2005b)*, Oxford University Press, pp. 133–154.

Tait, W. W. (2005b). *The Provenance of Pure Reason: Essays in the Philosophy of Mathematics and Its History*, Logic and Computation in Philosophy, Oxford University Press.