Principles of Reflection in Higher-Order Logic

Hilary Term 2016

Sam Adam-Day
Final Honour School of Mathematics

Contents

1. Introduction

Set theory has been enormously successful. It provides a firm and precise foundation upon which the vast majority of mathematics can be built, and is now deeply embedded in the practice of modern mathematics. However, there are several natural questions for which classical set theory provides no answers. A famous example is the continuum hypothesis, which was shown to be independent of ZFC by Gödel and Cohen.

In fact Gödel proved that any consistent axiom system which is sufficiently strong to carry out arithmetic must be incomplete, in the sense that there must exist a sentence which can be neither proved nor refuted by the system. Since any reasonable system of set theory ought to be able to perform arithmetic, there will always exist questions unanswerable by a given axiomatisation. However, we need not let this deter us. Indeed as Gödel (1947, p. 520) writes:

> For first of all 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."

This motivates the introduction of new axioms for set theory, with the aim of affecting a "significant reduction in incompleteness" (Koellner, 2009). It would seem important to avoid arbitrariness in this introduction, so that these new axioms have some sort of intrinsic justification; they should in some sense express statements which are "true".[1] In the present dissertation we will consider one class of axioms which are good candidates for use as intrinsically justified additions to set theory: reflection principles. The general form of the reflection principles considered here is, for some property P:

> If P holds in the universe, then there is a transitive set in which P also holds. (RP)

The motivation for considering reflection principles is that the universe should not be "definable" by any property. As Maddy (1988, p. 503) writes:

> [The idea behind reflection principles is that] the universe of sets is so complex that it cannot be completely described; therefore, anything true of the entire universe must already be true of some initial segment of the universe. In other words, any attempt to uniquely describe V also applies to smaller V_α's that "reflect" the property ascribed to V.[2]

In order to formulate RP in a language of set theory, we need to decide which properties P are to be considered. It turns out that if one restricts P to first-order properties, RP

[1] Gödel's (1947) initiated this programme of finding and justifying new axioms.
[2] In the original passage, Maddy uses R_α instead of V_α.

already holds in first-order ZFC (see Theorem 2.1). The aim of this dissertation is to investigate RP when P can range over second-order and higher-order properties,[3] and determine the result of adding such principles to various axiomatisations of set theory.

Throughout this dissertation, we will be assuming the completeness theorem for first-order logic, and also Gödel's first and second incompleteness theorems for the systems of set theory considered here (all of which will be effective and strong enough to express arithmetic).

We will use the symbol † next to a result or proof which is believed to be new.

[3]Usually, "higher-order" will refer to orders greater than two.

2. Set Theory

In this chapter, we present an nth-order set theory for each $n \in \omega$.

2.1 First-Order Set Theory

First-order set theory knows about only one kind of object: sets. This is manifested by the fact that its formal language has only one sort of variable. We will use the lowercase $x, y, z, \ldots$ as the variables of our first-order language, and build up formulae in the usual way using the connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$ and $\leftrightarrow$, the quantifiers $\forall$ and $\exists$, the identity relation $=$, and the binary relation $\in$. Let $\mathcal{L}_1$ be this language.

Note that while first-order set theory formally only recognises sets, there is another kind of object which is indispensable for any informal understanding of set theory, namely classes. For the purposes of this dissertation, a class is thought of as any collection of sets. Classes are useful since there are some collections of sets about which one would like to reason (at least informally), but which cannot be coextensive with sets without engendering paradoxes (the collection of all sets is a simple example). The way in which these classes are handled in first-order set theory is by thinking of a formula $\phi(x)$ (possibly containing parameters) as specifying a class. We can encode this using the following (informal) principle:

> Let $\phi(x, s_1, \ldots, s_n)$ be a formula with all free variables shown, and let $t_1, \ldots, t_n$ be sets. There is a class X satisfying $\forall y(y \in X \Leftrightarrow \phi(y, t_1, \ldots, t_n))$. (CS)

The idea is then that $y \in X$ is shorthand for $\phi(y, t_1, \ldots, t_n)$, and that any theorem which involves "$y \in X$" can ultimately be translated to one involving only "$\phi(y, t_1, \ldots, t_n)$". This principle allows one to reason about classes without formally positing their existence. When we come to consider second-order set theory, classes will be bona fide citizens, and this principle will be formalised and enshrined as an axiom.

The first-order axioms we will be considering are those of standard ZFC, and these will be supported by the standard first-order deductive system, which we denote by D.

For later reference, we list the axioms of ZFC here:[1]

$$\forall z(z \in x \leftrightarrow z \in y) \to x = y \qquad \text{(Extensionality)}$$
$$\exists x.x = \varnothing \qquad \text{(Emptyset)}$$
$$\exists z.z = \{x, y\} \qquad \text{(Pairing)}$$
$$\exists y.y = \bigcup x \qquad \text{(Union)}$$
$$\exists y.y = \mathcal{P}(x) \qquad \text{(Powerset)}$$
$$x \neq \varnothing \to \exists y \in x.\forall z \in x.z \notin y \qquad \text{(Foundation)}$$
$$\forall x \exists z.z = \{y \in x \mid \phi(x, y)\} \qquad \text{(Separation)}$$
$$\forall w(\forall x \in w.\exists! y.\phi(x, y) \to \exists z.z = \{y \mid \exists x \in w.\phi(x, y)\}) \qquad \text{(Replacement)}$$
$$\exists x.(\exists y \in x.y = \varnothing \wedge \forall y \in x.y \cup \{y\} \in x) \qquad \text{(Infinity)}$$
$$\forall x(\forall y \in x.y \neq \varnothing \to \exists b \forall y \in x.\exists! z \in y.\, \langle y, z\rangle \in b) \qquad \text{(Choice)}$$

Here we are using standard abbreviations for brevity. Note that Separation and Replacement are axiom schemata, where the formula ϕ in each case may contain additional parameters, but cannot have z free. Also note that we can think of these schemata as quantifying over classes, given our informal identification of formulae with classes.[2] Since we have no direct access to classes, we must give one axiom instance for each formula.

Let us briefly look at reflection principles in ZFC. We consider the instances of RP in which the property P is a first-order statement with parameters. In order to formalise these we need two notions. First, for a formula ϕ and a set x, we let ϕ^x denote the relativisation of ϕ to x, i.e. the result of bounding each quantifier in ϕ by x. Second, we let $\mathsf{tran}(x)$ be the first-order formula which expresses that x is a transitive set. Now the first-order reflection schema for ZFC is

$$\phi(t_1, \ldots, t_n) \to \exists x(\mathsf{tran}(x) \wedge t_1 \in x \wedge \cdots \wedge t_n \in x \wedge \phi^x(t_1, \ldots, t_n)) \qquad (\Pi^1_0\text{-Reflection}_1)$$

We require that ϕ have only the free variables shown, and that x not be among them. Now, Lévy (1960) shows that Π^1_0-Reflection$_1$ already holds in ZFC.

Theorem 2.1 (Lévy Reflection Principle)**.** ZFC *proves every instance of* Π^1_0*-Reflection*$_1$.

[1] See Suabedissen (2015).

[2] Actually, the principle CS only gives one direction of this identification. We might consider the following stronger principle, which gives an explicit definition of classes:

X is a class if and only if there is a formula $\phi(x, s_1, \ldots, s_n)$ with all free variables shown, and sets $t_1, \ldots, t_n$ such that $\forall y(y \in X \Leftrightarrow \phi(y, t_1, \ldots, t_n))$. (CD)

It seems slightly artificial to restrict classes in this way, given our idea that a class is *any* collection of sets. In any case, this principle cannot be included in any set theory. This is because the forwards direction requires the theory to consider the existence of a formula and determine the truth of that formula, which violates Tarski's undefinability theorem (Tarski, 1933).

2.2 Second-Order Set Theory

We now embark on a description of second-order set theory.[3] In general, first-order logic has one kind of variable, and it is intended that these variables should range over the elements of the domain of whatever model is being considered. On the other hand, (monadic) second-order logic extends first-order logic by introducing a new kind of variable, which is intended to range over *collections* of elements of the domain. As a real-world motivation for considering second-order logic, note that while natural language propositions tend to be first-order in nature, there are some statements which have an essential second-order character. The Geach-Kaplan sentence is one such example.[4]

There are some critics who admire only each other. (GK)

Second-order set theory is set theory interpreted using second-order logic. Classes are now full-fledged citizens, and the theory can reason directly about them; second-order set theory thus knows about two kinds of objects: sets and classes. As above, we use the lowercase $x, y, z, \ldots$ to denote *first-order* variables, which we think of as ranging over sets; we use the uppercase $X, Y, Z, \ldots$ to denote *monadic second-order* variables, and we think of these as ranging over classes. For technical reasons, we also introduce the *dyadic* variable $\mathcal{B}$, which is thought of as ranging over binary relations; this can be done away with later.[5] For any first-order variable x and any second-order variable Y, we introduce the atomic formula Yx of *predication*, which we think of as expressing "x is a member of Y". Similarly, for any two first-order variables x and y, we introduce the atomic formula $x\mathcal{B}y$, which we think of as expressing "x is related to y by $\mathcal{B}$". We will allow any two first-order variables to participate in an $=$-expression, and in an $\in$-expression. We then build up formulae in the usual way, allowing for quantification over second-order variables in addition to first-order. Let $\mathcal{L}_2$ denote the language just described.

We now turn to the semantics of second-order logic.[6] To simplify the presentation somewhat, we will only describe the semantics for $\mathcal{L}_2$-formulae. An *$\mathcal{L}_2$-model* is a tuple

[3] The exposition given here is similar to that found in Shapiro (1991).

[4] Boolos (1984) gives an ingenious proof of the nonfirstorderisability of this sentence, which he attributes to Kaplan. For simplicity, we assume that the domain of discourse contains only critics. Using the language of second-order logic presented in the forthcoming paragraphs, and letting xAy denote the relation of admiration, the sentence can be formalised in second-order logic as

$$\exists X(\exists x.Xx \wedge \forall x \forall y((Xx \wedge xAy) \rightarrow (x \neq y \wedge Xy))) \tag{2.1}$$

If we now reinterpret xAy as the arithmetic predicate $(x = 0 \vee x = y + 1)$, (2.1) becomes

$$\exists X(\exists x.Xx \wedge \forall x \forall y((Xx \wedge (x = 0 \vee x = y + 1)) \rightarrow (x \neq y \wedge Xy))) \tag{2.2}$$

This sentence is true of all and only non-standard models of arithmetic (since it is equivalent to the existence of an infinite sequence of numbers (x_n) such that $x_{n+1} = x_n - 1$). If this were firstorderisable, then adding its negation to the first-order axioms of Peano Arithmetic would produce an axiom system with only one model (up to isomorphism): the standard model of arithmetic. By the completeness theorem for first-order logic, this system would be deductively complete, violating Gödel's first incompleteness theorem. Therefore GK has no first-order formalisation.

[5] This is perhaps somewhat unorthodox: usually languages enjoy an infinite supply of variables of each sort. However, we will only use $\mathcal{B}$ briefly, and in all such cases only one instance of a variable of its kind is needed. So for simplicity of exposition, only one is introduced.

[6] We will only consider so-called *standard semantics* here.

$M = \langle A, R\rangle$, where A is the domain,[7] and R is a binary relation on A. A *variable-assignment* for M is a tuple $I = \langle I_1, I_2, I_{\mathcal{B}}\rangle$ such that I_1 is a map from the first-order variables to A, I_2 is a map from the second-order variables to $\mathcal{P}(A)$, and $I_{\mathcal{B}}$ is an element of $\mathcal{P}(A \times A)$. We can now recursively define the *valuation* $\mathrm{val}_M^I\colon \mathrm{Formulae}(\mathcal{L}_2) \to \{\mathsf{false}, \mathsf{true}\}$, in an analogous way to the first-order case, using the following rules (in addition to the usual rules for connectives):

$$
\begin{aligned}
\mathrm{val}_M^I(x = y) = \mathsf{true} \quad &\Leftrightarrow \quad I_1(x) = I_1(y)\\
\mathrm{val}_M^I(x \in y) = \mathsf{true} \quad &\Leftrightarrow \quad R(I_1(x), I_1(y))\\
\mathrm{val}_M^I(Yx) = \mathsf{true} \quad &\Leftrightarrow \quad I_1(x) \in I_2(Y)\\
\mathrm{val}_M^I(x\mathcal{B}y) = \mathsf{true} \quad &\Leftrightarrow \quad \langle I_1(x), I_1(y)\rangle \in I_{\mathcal{B}}\\
\mathrm{val}_M^I(\forall x.\phi) = \mathsf{true} \quad &\Leftrightarrow \quad \forall h \in A\colon \mathrm{val}_M^{(I_1[x\mapsto h], I_2, I_{\mathcal{B}})}(\phi) = \mathsf{true}\\
\mathrm{val}_M^I(\exists x.\phi) = \mathsf{true} \quad &\Leftrightarrow \quad \exists h \in A\colon \mathrm{val}_M^{(I_1[x\mapsto h], I_2, I_{\mathcal{B}})}(\phi) = \mathsf{true}\\
\mathrm{val}_M^I(\forall X.\phi) = \mathsf{true} \quad &\Leftrightarrow \quad \forall H \in \mathcal{P}(A)\colon \mathrm{val}_M^{(I_1, I_2[X\mapsto H], I_{\mathcal{B}})}(\phi) = \mathsf{true}\\
\mathrm{val}_M^I(\exists X.\phi) = \mathsf{true} \quad &\Leftrightarrow \quad \exists H \in \mathcal{P}(A)\colon \mathrm{val}_M^{(I_1, I_2[X\mapsto H], I_{\mathcal{B}})}(\phi) = \mathsf{true}\\
\mathrm{val}_M^I(\forall \mathcal{B}.\phi) = \mathsf{true} \quad &\Leftrightarrow \quad \forall J \in \mathcal{P}(A \times A)\colon \mathrm{val}_M^{(I_1, I_2, J)}(\phi) = \mathsf{true}\\
\mathrm{val}_M^I(\exists \mathcal{B}.\phi) = \mathsf{true} \quad &\Leftrightarrow \quad \exists J \in \mathcal{P}(A \times A)\colon \mathrm{val}_M^{(I_1, I_2, J)}(\phi) = \mathsf{true}
\end{aligned}
$$

where, for a function f, we let:

$$f[p \mapsto w](q) := \begin{cases} f(q) & \text{if } q \neq p,\\ w & \text{if } q = p \end{cases}$$

For an $\mathcal{L}_2$-formula ϕ, we write $M \models \phi$ if $\mathrm{val}_M^I(\phi) = \mathsf{true}$ for every variable-assignment I. For a collection Γ of $\mathcal{L}_2$-sentences and an $\mathcal{L}_2$-formula ϕ, we write $\Gamma \vDash \phi$ if $M \models \phi$ for every $\mathcal{L}_2$-model M such that $\forall\gamma \in \Gamma\colon M \models \gamma$.

Notice that we haven't allowed second-order variables to participate in $=$-expressions.[8] However, it is possible to use the following "identity of indiscernibles" abbreviation for class equality, which we will do from now on:

$$X = Y := \forall x(Xx \leftrightarrow Yx)$$

So much for semantics, but what about syntax? Can we furnish second-order logic with an adequate deductive system? It turns out that no effective, sound deductive system can be complete with respect to standard semantics, even when we restrict to $\mathcal{L}_2$.

Theorem 2.2 (Inherent Incompleteness of Second-Order Logic with Standard Semantics)**.** *Let E be an effective deductive system in $\mathcal{L}_2$ which is sound with respect to standard semantics (i.e. if $\Gamma \vdash_E \phi$ then $\Gamma \vDash \phi$). Then E is incomplete in the sense that there is a set of $\mathcal{L}_2$-sentences Γ and some $\mathcal{L}_2$-sentence ϕ such that $\Gamma \vDash \phi$ while $\Gamma \nvdash_E \phi$.*

[7]Note that the consideration of metamathematical concepts discussed here must take place in a suitable metatheory. This metatheory is required to be able to talk about sets and various operations on them. The choice of metatheory is not too important here; we can take it to be first-order ZFC or even second-order ZFC_2 (the various philosophical issues related to this are beyond the scope of the present discussion). Once a suitable metatheory is selected, we can specify that the domain A must be a *meta-set* (i.e. something which the metatheory thinks of as a set).

[8]This is to simplify the process of induction on $\mathcal{L}_2$-formulae.

We postpone a proof of this until we have some more technology.

So, we cannot hope to capture the full power of second-order logic in a deductive system. Nevertheless, we do need some kind of system for the discussion that follows, hence we present one which is suitable for our needs. Again, we'll restrict our attention to $\mathcal{L}_2$-formulae. We'll call the deductive system D_2; it is an altered version of the one given in Shapiro (1991, pp. 66–67). The system D_2 extends the first-order predicate calculus; we add the Generalisation$_2$ rule of inference, the axiom schemata of Instantiation and Π^2_0-Comprehension$_2$, and the axiom of Strong-Choice.

The Generalisation$_2$ rule is a straightforward extension of the corresponding rule in the first-order D.

$$\frac{\phi}{\forall X \phi} \qquad \text{(Generalisation}_2\text{)}$$

Instantiation comes in two parts, as follows, where Y is free for X in ϕ:

$$\forall X \phi(X) \to \phi(Y) \qquad \text{(Instantiation}_2\text{)}$$

$$\forall \mathcal{B} \phi(\mathcal{B}) \to \phi(\in) \qquad (\mathcal{B}\text{-Instantiation}_2)$$

The comprehension axiom schema is the following, in which we require that X not occur free in ϕ (however ϕ may contain other first- and second-order parameters):

$$\exists X \forall x (Xx \leftrightarrow \phi(x)) \qquad (\Pi^2_0\text{-Comprehension}_2)$$

Note in particular that for every set there is a class with the same extension. We call this the *impredicative* comprehension schema, since ϕ may contain second-order variables. If we restrict ϕ to formulae in which all quantified variables are first-order, we produce the following weaker *predicative* comprehension schema:

$$\exists X \forall x (Xx \leftrightarrow \phi(x)) \qquad (\Pi^1_0\text{-Comprehension}_2)$$

Notice that this is precisely a formalisation of the principle CS.

Finally, we add a variant of the axiom of choice which is stronger than the usual axiom since it applies to classes in addition to sets.

$$\forall X (\forall x.(Xx \to \exists y. y \in x) \to \exists \mathcal{B}. \forall x (Xx \to \exists! y \in x. x \mathcal{B} y)) \qquad \text{(Strong-Choice)}$$

Note that each axiom of D_2 holds in every model (here we're using properties of the metatheory), and the inference rules are valid; therefore D_2 is sound with respect to our semantics.

With our suitable deductive system in place, we can now formulate our second-order set theory. This will be the system ZFC_2, a second-order version of ZFC (Shapiro, 1991, p. 85). The first axioms are the following first-order axioms inherited from ZFC:

Extensionality, Emptyset, Pairing, Union, Powerset, Foundation, Infinity

Note that these axioms are sufficient to show that $\langle x, y \rangle = \{\{x\}, \{x, y\}\}$ works as an ordered pair, hence by applying Π^2_0-Comprehension$_2$ to $\phi(x) := \exists y \exists z (x = \langle y, z \rangle \wedge y \mathcal{B} z)$, we can eliminate the need for $\mathcal{B}$. To complete the axiomatisation, we add a secondorderised version of the Replacement axiom schema. Using the power of second-order logic, this is now a single sentence.

$$\forall X \forall w (\forall x \in w. \exists! y. X \langle x, y \rangle \to \exists z. z = \{y \mid \exists x \in w. X \langle x, y \rangle\}) \qquad (\text{Replacement}_2)$$

ZFC_2 is finitely axiomatised by these sentences; let Z_2 be their conjunction.

Comparing ZFC_2 with ZFC, we notice that former does not directly include analogues of Separation and Choice. For reassurance, we show that ZFC_2 with our deductive system D_2 is indeed capable of proving Separation_2 and Choice, where the former is the following secondorderised version of Separation:

$$\forall X \forall x \exists y \forall z (z \in y \leftrightarrow (z \in x \wedge Xz)) \qquad (\text{Separation}_2)$$

Lemma 2.3. $\mathrm{ZFC}_2 \vdash$ *Separation_2.*

Proof (ZFC_2). Take a class X and a set x. If there is nothing in their common extension (i.e. if $X \cap x$ is empty) then $\varnothing$ is what we want. So we can assume that there is some t in their common extension. Apply Π^2_0-Comprehension_2 to the formula

$$\phi(u) := \exists p \exists q (u = \langle p, q \rangle \wedge p \in x \wedge ((Xp \wedge q = p) \vee (\neg Xp \wedge q = t)))$$

This gives a class F which is defined by $\phi(u)$. This is a class function on x, so by Replacement_2 its image is a set. But its image is precisely $\{z \mid z \in x \wedge Xz\}$. ☐

Lemma 2.4. $\mathrm{ZFC}_2 \vdash$ *Choice.*

Proof (ZFC_2). Let x be a set of non-empty sets. By applying Π^2_0-Comprehension_2 with $\phi(y) := y \in x$, we get a class X with the same extension as x. Then there is a class relation B on X as in the statement of Strong-Choice (we've replaced $\mathcal{B}$ by B using the elimination above). Now applying Separation_2 with B and $\mathcal{P}(x \times \bigcup x)$, we get a relation b with the desired properties. ☐

We also show that in ZFC_2, the following stronger form of Foundation holds:

$$\exists y . Xy \rightarrow \exists y (Xy \wedge \forall z (Xz \rightarrow z \notin y)) \qquad (\text{Foundation}_2)$$

Lemma 2.5. $\mathrm{ZFC}_2 \vdash$ *Foundation_2.*

Proof (ZFC_2). Assume X is a non-empty class, so there is some w in X. Letting w^* denote the transitive closure of w, by Foundation, $X \cap w^*$ (which is a set by Separation_2) has an $\in$-minimal element y. Now if there is z in X with $z \in y$, then by transitivity $z \in w^*$ so $z \in X \cap w^*$, contradicting the minimality of y. ☐

What now of reflection? Does a version of the Lévy reflection principle hold in ZFC_2? The second-order reflection schema is as follows (where again we require that ϕ have only the free variables shown):

$$\phi(T_1, \ldots, T_n) \rightarrow \exists x (\mathsf{tran}(x) \wedge \phi^x (T_1 \cap x, \ldots, T_n \cap x)) \qquad (\Pi^2_0\text{-Reflection}_2)$$

The process of relativising ϕ to x to produce ϕ^x is an extension of the first-order case. In addition to bounding first-order variables to x, we "bound" second-order variables to $\mathcal{P}(x)$. To do this, we introduce the following predicate:

$$\mathsf{InPowerset}(x, X) := \forall y (Xy \rightarrow y \in x)$$

Then for example $\forall X \psi$ becomes $\forall X (\mathsf{InPowerset}(x, X) \rightarrow \psi^x)$. Note also that we must now "relativise" each parameter T_i to x; it is natural to take this relativisation to be $T_i \cap x$.

While this principle officially allows only second-order parameters, we can simulate first-order parameters as follows. Let $\psi(s_1, \ldots, s_m, S_{m+1}, \ldots, S_n)$ be a formula, let $t_1, \ldots, t_m$ be sets and let $T_{m+1}, \ldots, T_n$ be classes. Then we can reflect $\psi(t_1, \ldots, t_m, T_{m+1}, \ldots, T_n)$ by applying Π^2_0-Reflection$_2$ to

$$\phi(S_1, \ldots, S_n) := \exists s_1 \in S_1 \cdots \exists s_m \in S_m . \psi(s_1, \ldots, s_m, S_{m+1}, \ldots, S_n)$$

with the parameters $\{t_1\}, \ldots, \{t_m\}, T_{m+1}, \ldots, T_n$.

Now, this is one place in which ZFC_2 differs from ZFC: ZFC_2 doesn't prove its version of the reflection principle (assuming that it's consistent). This follows from the finite-axiomatisability of ZFC_2. If ZFC_2 were to prove every instance of Π^2_0-Reflection$_2$, then in particular it would prove the instance in which ϕ is the sentence Z_2. Since Z_2 holds in ZFC_2, the latter would then believe in the existence of a model of Z_2, i.e. a model of ZFC_2. Now, by the soundness of the deductive system D_2, the existence of a model of a theory implies the consistency of said theory. If we internalise this argument into ZFC_2, we get that ZFC_2 proves its own consistency, violating Gödel's second incompleteness theorem. Later we will investigate the result of adding the schema Π^2_0-Reflection$_2$ to ZFC_2.[9]

2.3 Higher-Order Set Theory

Once we have made the step from first-order to second-order set theory, a reasonable continuation is to generalise to even higher orders of logic; so we have set theories which know about collections of classes (2-*classes*), collections of collections of classes (3-*classes*), and so on.

For $n \in \omega$, we specify the language $\mathcal{L}_n$. We have the usual first-order variables $x, y, z, \ldots$, and in addition we have kth-order monadic variables $X^{(k)}, Y^{(k)}, Z^{(k)}, \ldots$ for each $k \in \{2, \ldots, n\}$. We also initially need a kth-order binary relation variable $\mathcal{B}^{(k)}$, as before. We then build up formulae by analogy with the second-order case, introducing atomic formulae $X^{(k)}Y^{(k-1)}$ and allowing quantification over variables of any order.

With the language in place, the standard semantics and a deductive system for nth-order logic may be sketched. As before, we restrict our attention to $\mathcal{L}_n$-formulae. As in the second-order case, an $\mathcal{L}_n$-*model* of nth-order logic is a tuple $M = \langle A, R \rangle$ in which R is a binary relation on A. A *variable-assignment* is a tuple $I = \langle I_1, \ldots, I_n, I_{\mathcal{B}^{(2)}}, \ldots, I_{\mathcal{B}^{(n)}} \rangle$ in which I_k is a map from the kth-order variables to $\mathcal{P}^{k-1}(A)$,[10] and $I_{\mathcal{B}^{(k)}} \in \mathcal{P}(\mathcal{P}^{(k-2)}(A) \times \mathcal{P}^{(k-2)}(A))$. The definition of *valuation* is extended in the obvious way to nth-order logic, and the notions of $M \models \phi$ and $\Gamma \vDash \phi$ are defined in the same way.

The deductive system D_n presented here is constructed by recursion on n, so that D_n extends D_{n-1}. Starting with D_{n-1}, add the following axioms and axiom schemata (letting

[9] Note that while ZFC_2 doesn't prove every instance of Π^2_0-Reflection$_2$, it need not disprove any instance either (since it's syntactically incomplete). We will show later that $\mathrm{ZFC}_2 + \Pi^2_0$-Reflection$_2$ is consistent on the assumption of the existence of a certain large cardinal.

[10] This is the $(k-1)$th iterated powerset operation on A.

$X^{(1)}$ denote a first-order variable):

$$\forall X^{(n)}\phi(X^{(n)}) \to \phi(Y^{(n)}) \qquad \text{(Instantiation}_n\text{)}$$

$$\exists X^{(n)}\forall Y^{(n-1)}(X^{(n)}Y^{(n-1)} \leftrightarrow \phi(Y^{(n-1)})) \qquad (\Pi^n_0\text{-Comprehension}_n)$$

$$\begin{aligned}\forall X^{(n)}(\forall Y^{(n-1)}(X^{(n)}Y^{(n-1)} &\to \exists Z^{(n-2)}.Y^{(n-1)}Z^{(n-2)})\\ \to \exists \mathcal{B}^{(n)}\forall Y^{(n-1)}(X^{(n)}Y^{(n-1)} &\to \exists! Z^{(n-2)}(Y^{(n-1)}Z^{(n-2)}\\ &\wedge \exists W^{(n-1)}(\forall V^{(n-2)}(W^{(n-1)}V^{(n-2)} \leftrightarrow V^{(n-2)} = Z^{(n-2)})\\ &\wedge Y^{(n-1)}\mathcal{B}^{(n)}W^{(n-1)}))))\end{aligned} \qquad \text{(Strong-Choice}_n\text{)}$$

As in the second-order case, nth-order logic with standard semantics is inherently incomplete, so this deductive system is necessarily weaker than the semantics.

Finally, with the language, semantics, and deductive system in place, we can consider the axioms of nth-order set theory. However, looking at the axioms of ZFC_2, there is no real scope for enhancement in a natural way using the additional power of nth-order logic (the first seven axioms are first-order, and Replacement_2 involves quantification over class functions, which doesn't comfortably extend to utilise higher-order objects). Hence we simply let our nth-order set theory be the system $\text{ZFC}_n := \text{ZFC}_2$, and notice that any extra strength comes solely from the incorporation of Instantiation_n, Π^n_0-Comprehension_n, and Strong-Choice_n.

The last system of set theory which we consider here is ω-order set theory, which uses ω-order logic. Once we have logics of order n for every $n \in \omega$, it is reasonable to consider a logic which includes all finite orders of variables; when we do so we meet ω-order logic. Our language $\mathcal{L}_\omega$ is the union of the languages $\mathcal{L}_n$ for $n \in \omega$. The semantics and deductive system are extended in the natural way,[11] and we let $\text{ZFC}_\omega := \text{ZFC}_2$, as above.

We now introduce a hierarchy of $\mathcal{L}_\omega$-formulae, which will be needed shortly. The following definition is given in Tait (2005) and follows the ideas of the hierarchy introduced in Lévy (1965).

Definition 2.6.

- A formula whose quantified variables are of order n or lower is a Π^n_0-formula and a Σ^n_0-formula.
- If $\phi(Y^{(n+1)})$ is a Π^n_m-formula, then $\exists Y^{(n+1)}\phi(Y^{(n+1)})$ is a Σ^n_{m+1}-formula.
- If $\phi(Y^{(n+1)})$ is a Σ^n_m-formula, then $\forall Y^{(n+1)}\phi(Y^{(n+1)})$ is a Π^n_{m+1}-formula.

In the context of ZFC_ω, it is possible to prove that every Π^n_0-formula is equivalent to a Π^{n-1}_m-formula for some $m \in \omega$ (we define a pairing of nth-order objects and proceed in the usual way, following the method in Lévy (1965)).

[11] Note that the deductive system now has infinitely many schemata, namely we have Instantiation_n and Comprehension_n for each $n \in \omega$ (here ϕ ranges over $\mathcal{L}_\omega$-formulae). This is still effective however.

2.4 Notation and Shorthands

With our various set theories in place, we indicate the notation used for the remainder of this dissertation, and introduce some shorthands which will make our formulae more comprehensible.

Let On denote the class of all ordinals (this is only informal in the first-order case), and let Card denote the class of all cardinals. We will denote the standard von Neumann hierarchy by $\mathrm{V}_\alpha \mid \alpha \in \mathrm{On}$; since all set theories considered here include Foundation, we will use V to denote the class of all sets. We let L denote the constructible universe. We will use standard set-builder notation to construct sets, classes and so on, indicating the desired order of the constructed object as follows: $\{\cdots \mid \cdots\}^{(n)}$. In second- and higher-order set theory, given a set x and a class X, by Separation$_2$ and Π^2_0-Comprehension$_2$, $X \cap x$ may be regarded as a set and a class. We will let context determine what kind of object this should be. Sometimes, we will use the notation $X^{(1)}$ to denote a first-order variable.

A key ingredient in the proof of Gödel's incompleteness theorems is *Gödel numbering*. This is an effective enumeration of the formulae of a language. Since we will need this in our toolbox, we set up the notation here. Every language considered here admits an enumeration of this kind, but of course it will be different for each language. For a formula ϕ, we let $\ulcorner \phi \urcorner$ be the natural number which corresponds to ϕ.

The following shorthands will also be used.

- As above, we use $\mathsf{tran}(x)$ to express that x is a transitive set. We use $\mathsf{Tran}(X)$ to express that X is a transitive class (so $\mathsf{Tran}(X) \leftrightarrow \forall x(Xx \rightarrow \forall y \in x.Xy))$.

- Given $\alpha \in \mathrm{On}$, a subset $c \subseteq \alpha$ is *club in* α if it is unbounded in α and is closed in α (i.e. for every $G \subseteq c$, if $\sup(G) \in \alpha$ then $\sup(G) \in c$). This we express by $\mathsf{club}(\alpha, c)$. A class $C \subseteq \mathrm{On}$ is *club* if it is unbounded and closed in On. We express this by $\mathsf{Club}(C)$.

- For $\alpha \in \mathrm{On}$, a subset $s \subseteq \alpha$ is *stationary in* α if it intersects every set $c \subseteq \alpha$ which is club in α; this we express by $\mathsf{stationary}(\alpha, s)$. And similarly a class $C \subseteq \mathrm{On}$ is *stationary* if it intersects every club $C \subseteq \mathrm{On}$, which we express by $\mathsf{Stationary}(C)$.

- For a system S, we let $\mathsf{Con}(S)$ express the consistency of S.

3. Large Cardinals

Our next item of business is the description of a portion of the large cardinal hierarchy. For our purposes, a *large cardinal property* is a property of cardinal numbers such that the existence of infinite cardinals exhibiting said property cannot be shown in first-order ZFC. The fact that such cardinals are "beyond the reach" of ZFC indicates that they are indeed quite "large".

The reason for considering large cardinals here is that they provide a very convenient means of measuring the strength of systems of set theory. Indeed, we can define the following ordering on large cardinal properties. For properties P and Q, we define

$$P \preccurlyeq Q \quad \Leftrightarrow \quad \mathrm{ZFC} \vdash (\mathsf{Con}(\mathrm{ZFC} + \exists\kappa Q(\kappa)) \rightarrow \mathsf{Con}(\mathrm{ZFC} + \exists\kappa P(\kappa)))$$

Under the following equivalence, $\preccurlyeq$ is a weak partial order:

$$P \approx Q \quad \Leftrightarrow \quad \mathrm{ZFC} \vdash (\mathsf{Con}(\mathrm{ZFC} + \exists\kappa P(\kappa)) \leftrightarrow \mathsf{Con}(\mathrm{ZFC} + \exists\kappa Q(\kappa)))$$

Now, when restricted to the properties considered here, this ordering is total.[1] So we can imagine arranging these properties in a line to form a yardstick of consistency strength. We will use this yardstick to measure the strength of the reflection principles considered here, by determining which large cardinals are implied by the addition of these principles to a relevant version of ZFC.

We now embark on a whistle-stop tour of the large cardinals that will be used here.[2] To start us off, we have the inaccessible cardinals.

Definition 3.1. A cardinal κ is *inaccessible* if it is uncountable, regular, and for every $\lambda < \kappa$ we have $2^\lambda < \kappa$.

Shepherdson (1952) shows the following important theorem relating inaccessibles with ZFC_2.[3]

Theorem 3.2. *$\langle A, R\rangle$ is a model of ZFC_2 if and only if it is isomorphic to $\langle \mathrm{V}_\kappa, \in\rangle$ for some inaccessible κ.*

With this theorem, we are now in a position to readily prove that second-order logic is inherently incomplete.

Proof of Theorem 2.2. (We assume that ZFC_2 does actually have a model.) Our set of sentences Γ will be ZFC_2 plus the statement that there are no inaccessibles[4] (note

[1] A remarkable fact is that this order seems to be total for *all* large cardinal properties of interest (though for some pairs of properties the matter hasn't been determined). See Shelah's (1991).

[2] All the definitions given in this chapter, apart from that of ineffability, appear in Kanamori (1994).

[3] Zermelo (1930) gives a similar but more informal proof.

[4] This idea is from Karagila (2013).

that the property of being an inaccessible cardinal can be expressed by a first-order formula). By Theorem 3.2, every model of Γ is isomorphic to $\langle \mathrm{V}_\kappa, \in \rangle$ where κ is the first inaccessible (note that the property of being inaccessible is absolute between V and V_λ for λ inaccessible). Now suppose for a contradiction that E is a deductive system for second-order logic which is sound and complete. In particular, E includes first-order predicate calculus. Now, Γ with E is a set theory strong enough to perform arithmetic. Hence by Gödel's first incompleteness theorem, there is a sentence G such that $\Gamma \nvdash_E G$ and $\Gamma \nvdash_E \neg G$. But G either holds in $\langle \mathrm{V}_\kappa, \in \rangle$ or it doesn't; without loss of generality assume that G holds. But then G holds in all models of Γ, so that $\Gamma \vDash G$. This contradicts the completeness of E. □

The definition of inaccessible cardinals can be extended inductively to capture larger and larger cardinals.

Definition 3.3.

- κ is 0-*inaccessible* if it is inaccessible.
- κ is α-*inaccessible* for $\alpha \in \mathrm{On}$ if for every $\beta \in \alpha$ the set $\{\lambda < \kappa \mid \lambda \text{ is } \beta\text{-inaccessible}\}$ is unbounded in κ.
- κ is *hyper-inaccessible* if it is κ-inaccessible.

We can continue this process and define α-*hyper-inaccessible* cardinals and so on.

Definition 3.4. A cardinal κ is *Mahlo* if it is inaccessible and the set $\{\lambda < \kappa \mid \lambda \text{ is inaccessible}\}$ is stationary in κ.

Our next stop will be the weakly compact cardinals. These have many equivalent definitions; the definition considered here uses infinite binary trees.

Definition 3.5.

- A class of functions $T \subseteq \bigcup_{\alpha \in \mathrm{On}} 2^\alpha$ is a *binary tree* if for every $f \colon \alpha \to 2$ in T and $\beta \in \alpha$, we have $f|_\beta \in T$.
- A binary tree T is *path-bounded* if for every class function $F \colon \mathrm{On} \to 2$ there is $\alpha \in \mathrm{On}$ such that $F|_\alpha \notin T$.
- A binary tree T is *bounded* if there is $\alpha \in \mathrm{On}$ such that for every class function $F \colon \mathrm{On} \to 2$ we have $F|_\alpha \notin T$.
- The binary tree property is the following:

  For every binary tree T, if T is path-bounded then it is bounded. (BTP)

For a binary tree T, we can think of $f \colon \alpha \to 2$ as encoding (part of) a branch of T, such that at each step (i.e. ordinal in α) f makes a decision about whether to continue right or left.

Definition 3.6. A cardinal κ is *weakly compact* if it is inaccessible and V_κ satisfies BTP.

There is an equivalent formulation in terms of a generalisation of compactness for infinitary languages, which gives some motivation for considering these cardinals. For infinite cardinals κ, λ and a set S of constant, function and predicate symbols, the infinitary language $\mathcal{L}_{(\kappa\lambda)}[S]$ extends the first-order language with non-logical symbols S by including $\kappa \cdot \lambda$ many variable symbols, allowing quantification over any collection of fewer than λ many distinct variables, and allowing the conjunction and disjunction of fewer than κ many formulae. The notion of a model is the same, and the notion of satisfaction extends in the natural way.

Theorem 3.7. *A cardinal κ is weakly compact if and only if it is uncountable and whenever $|S| \leqslant \kappa$, if Γ is a set of $\mathcal{L}_{(\kappa\kappa)}[S]$-sentences for which every subset of size less than κ has a model, then Γ has a model.*[5]

We move on to a class of cardinals with clear ties to reflection principles.

Definition 3.8. A cardinal κ is Π^n_m-*indescribable* for $n, m \in \omega$ if for every Π^n_m-formula $\phi(X^{(2)})$ (with only $X^{(2)}$ free) and every $A \subseteq \mathrm{V}_\kappa$, if $\mathrm{V}_\kappa \models \phi(A)$ then there is $\lambda \in \kappa$ such that $\mathrm{V}_\lambda \models \phi(A \cap \mathrm{V}_\lambda)$.[6]

The following theorem (Kanamori, 1994) shows that indescribability already includes inaccessibility.

Theorem 3.9. *A cardinal κ is inaccessible if and only if it is Π^1_0-indescribable.*

To prove this, we first need a weakened version of the Tarski-Vaught criterion.[7]

Lemma 3.10 (Tarski-Vaught Criterion)**.** *Let $\mathcal{L}$ be a first-order language, and let M and N be $\mathcal{L}$-models with N a substructure of M. Assume that for every $\mathcal{L}$-formula $\phi(x, s_1, \ldots, s_n)$ (with free variables shown) and $t_1, \ldots, t_n \in M$, if $M \models \exists y.\phi(y, t_1, \ldots, t_n)$ then there is $x \in N$ such that $M \models \phi(x, t_1, \ldots, t_n)$. Then for every $\mathcal{L}$-sentence ψ we have $M \models \psi \Leftrightarrow N \models \psi$.*

We also prove a tiny lemma about inaccessibles.

Lemma 3.11. *Let κ be an inaccessible cardinal. If $\alpha \in \kappa$ then $|\mathrm{V}_\alpha| < \kappa$.*

Proof. This is by induction on α. The base case is obvious. For successor steps we use that $|\mathrm{V}_{\alpha+1}| = 2^{|\mathrm{V}_\alpha|} < \kappa$ since by induction hypothesis $|\mathrm{V}_\alpha| < \kappa$. For limit steps, we note that $|\mathrm{V}_\gamma| = \sup_{\alpha \in \gamma} |\mathrm{V}_\alpha| < \kappa$ by regularity. □

Proof of Theorem 3.9. Let κ be inaccessible. Let $\mathcal{L}_1[\mathrm{A}]$ be the first-order language $\mathcal{L}_1$ with the additional unary predicate symbol A. We note that any Π^1_0-formula $\phi(X)$ (with only X free) can be transformed to a $\mathcal{L}_1[\mathrm{A}]$-sentence ϕ^* (by replacing Xx by $\mathrm{A}(x)$) such that for any $A \subseteq \mathrm{V}_\kappa$ we have $\langle \mathrm{V}_\kappa, \in \rangle \models \phi(A) \Leftrightarrow \langle \mathrm{V}_\kappa, \in, A \rangle \models \phi^*$. Hence to prove that κ is Π^1_0-indescribable, it is sufficient to show that for every $\mathcal{L}_1[\mathrm{A}]$-sentence ψ and every $A \subseteq \mathrm{V}_\kappa$ such that $\langle \mathrm{V}_\kappa, \in, A \rangle \models \psi$ there is $\lambda \in \kappa$ such that $\langle \mathrm{V}_\lambda, \in, A \cap \mathrm{V}_\lambda \rangle \models \psi$. In fact we show something stronger.

Take $A \subseteq \mathrm{V}_\kappa$. Define (λ_n) a sequence of ordinals in V_κ by recursion on ω as follows. Let $\lambda_0 := 0$. Assume that we have λ_n. For every $\mathcal{L}_1[\mathrm{A}]$-formula $\psi(y, s_1, \ldots, s_n)$ and

[5] A proof is in Kanamori (1994, pp. 76–77).

[6] Only allowing a second-order parameter here might seem a bit arbitrary. We will see however that coming up with a version of this definition using a third-order parameter is not so straightforward.

[7] See Suabedissen (2015).

$t_1, \ldots, t_n \in \mathrm{V}_{\lambda_n}$ such that $\langle \mathrm{V}_\kappa, \in, A\rangle \models \exists y.\psi(y, t_1, \ldots, t_n)$, let $\beta_{\psi,t_1,\ldots,t_n}$ be the minimal δ such that there is $x \in \mathrm{V}_\delta$ for which $\langle \mathrm{V}_\kappa, \in, A\rangle \models \psi(x, t_1, \ldots, t_n)$. Let λ_{n+1} be the supremum of these $\beta_{\psi,t_1,\ldots,t_n}$. Now since $\lambda_n \in \kappa$, we have by Lemma 3.11 that $|\mathrm{V}_{\lambda_n}| < \kappa$, and hence (as κ is uncountable)

$$|\{(\psi, t_1, \ldots, t_n) \mid \langle \mathrm{V}_\kappa, \in, A\rangle \models \exists y.\psi(y, t_1, \ldots, t_n)\}| \leqslant \omega \times |\mathrm{V}_{\lambda_n}{}^{<\omega}| < \kappa$$

Then by regularity we must have $\lambda_{n+1} < \kappa$. Let $\lambda := \sup_{n\in\omega}\lambda_n$; this is in V_κ by regularity. Now $\langle \mathrm{V}_\lambda, \in, A \cap \mathrm{V}_\lambda\rangle$ and $\langle \mathrm{V}_\kappa, \in, A\rangle$ satisfy the Tarski-Vaught Criterion, so that for every $\mathcal{L}_1$[A]-sentence ψ we have $\langle \mathrm{V}_\kappa, \in, A\rangle \models \psi \Leftrightarrow \langle \mathrm{V}_\lambda, \in, A \cap \mathrm{V}_\lambda\rangle \models \psi$.

Conversely, let κ be Π^1_0-indescribable. Note first that κ must be uncountable; the finite case gives an obvious contradiction, and if we had $\kappa = \omega$ we could apply the indescribability property with a sentence which expresses 'every ordinal has a successor'. Next, κ must be regular. Otherwise, there is $F \subseteq \mathrm{V}_\kappa$ an unbounded function $\alpha \to \kappa$ for some $\alpha < \kappa$. Apply the Π^1_0-indescribability property to

$$\exists x\,(X\langle x, 0\rangle \wedge \forall y \in x.\exists z.X\langle\langle y, z\rangle, 1\rangle)$$

with parameter $X = \{\langle\alpha, 0\rangle\} \cup F \times \{1\}$ (this is a subset of V_κ since κ is a cardinal) to get some $\lambda \in \kappa$. Then we must have that $\alpha \in \mathrm{V}_\lambda$; but then $F \subseteq \mathrm{V}_\lambda$, a contradiction. Finally $\forall\epsilon < \kappa.2^\epsilon < \kappa$. For suppose that there is some $\epsilon < \kappa$ and a surjection $F\colon \mathcal{P}(\epsilon) \to \kappa$. Note that $\epsilon \in \mathrm{V}_\kappa$, so that $\mathcal{P}(\epsilon) \in \mathrm{V}_\kappa$. Apply the Π^1_0-indescribability property to

$$\exists x(X\langle x, 0\rangle \wedge \exists y(x = y \cup \{y\} \wedge \forall z \in \mathcal{P}(y).\exists w.X\langle\langle z, w\rangle, 1\rangle)$$

with parameter $X = \{\langle\epsilon + 1, 0\rangle\} \cup F \times \{1\}$ to get some $\lambda \in \kappa$. Then $\epsilon + 1 \in \mathrm{V}_\lambda$ so that $\mathcal{P}(\epsilon) \in \mathrm{V}_\lambda$. But then also $F \subseteq \mathrm{V}_\lambda$ (note $\mathcal{P}^{\mathrm{V}_\lambda}(\epsilon) = \mathcal{P}(\epsilon)$), which is a contradiction. □

Furthermore, another of the above definitions is captured by indescribability.

Theorem 3.12. *A cardinal κ is weakly compact if and only if it is Π^1_1-indescribable.*

Only the backwards direction will be shown here though, and not just yet. The other direction is in Kanamori (1994, pp. 59–60).

To define our next two classes of cardinals, we need a few prior definitions.

Definition 3.13.

- For any set x and $\lambda \in$ Card, we denote by $[x]^\lambda$ the set of λ-element subsets of x, and we denote by $[x]^{<\lambda}$ the set of subsets of x whose size is less than λ.
- A function J with domain $[\kappa]^n$ is a *thin n-sequence* if for every $s \in [\kappa]^n$ we have $J(s) \subseteq \min(s)$.
- If J is a thin n-sequence on κ, then $q \subseteq \kappa$ is *homogeneous* for J if there is $b \subseteq \kappa$ such that for every $s \in [q]^n$ we have $J(s) = b \cap \min(s)$.

Definition 3.14. A cardinal κ is *n-ineffable* for $n > 0$ if every thin n-sequence J has a stationary homogeneous set.

Definition 3.15. For $\alpha \in$ On the αth *Erdős cardinal*, denoted $\kappa(\alpha)$ is the least cardinal κ such that for every function $f\colon [\kappa]^{<\omega} \to 2$ there is a subset $s \subseteq \kappa$ of order type α such that for every $n \in \omega$, f restricted to $[s]^n$ is constant.

Definition 3.16. An infinite cardinal κ is *measurable* if there exists a non-trivial, κ-additive, $\{0,1\}$-valued measure on $\mathcal{P}(\kappa)$. In other words, there is a function $\mu\colon \mathcal{P}(\kappa) \to 2$ such that:

- $\mu(\varnothing) = 0$,
- $\mu(\kappa) = 1$,
- for every $\alpha \in \kappa$, we have $\mu(\{\alpha\}) = 0$,
- for every collection of pairwise-disjoint subsets $\mathcal{Q} \subseteq \mathcal{P}(\kappa)$ with $|\mathcal{Q}| < \kappa$, we have $\mu(\bigcup \mathcal{Q}) = \sum_{s \in \mathcal{Q}} \mu(s)$.

The existence of measurable cardinals is an important statement in set theory. For one thing, it implies that $\mathrm{V} \neq \mathrm{L}$. Furthermore, it has consequences regarding the Lebesgue measurability of certain sets of reals.[8]

There is an alternative way of capturing measurable cardinals, one which lends itself well to generalisation. This alternative characterisation requires two additional ancillary notions.

Definition 3.17. Let M and N be transitive classes, and $\pi\colon M \to N$. Then π is an *elementary embedding* if it is injective and for every first-order formula $\phi(s_1, \ldots, s_n)$

$$\forall t_1 \in M \cdots \forall t_n \in M\colon (M \models \phi(t_1, \ldots, t_n) \Leftrightarrow N \models \phi(\pi(t_1), \ldots, \pi(t_n)))$$

Note that π must take ordinals to ordinals, since the property of being an ordinal is absolute for transitive classes.

Definition 3.18. Let M and N be transitive classes and let $\pi\colon M \to N$ be an elementary embedding. The least δ such that $\pi(\delta) > \delta$, if this exists, is the *critical point* of π.

Definition 3.19. Let M be a class and κ be a cardinal. We write ${}^{\kappa}M \subseteq M$ if for every $S \subseteq M$ with $|S| \leqslant \kappa$ we have $S \in M$.

Now, with these additional definitions in place, we have the following theorem, due to Scott and Keisler.[9]

Theorem 3.20. *A cardinal κ is measurable if and only if it is the critical point of some elementary embedding from the universe* V *into a transitive class M. In this case also ${}^{\kappa}M \subseteq M$.*

With this characterisation in mind, we produce a class of large cardinal properties which generalise measurability.

Definition 3.21. Let κ be a cardinal, and let λ be a cardinal such that $\kappa \leqslant \lambda$. Then κ is λ-*supercompact* if it is the critical point of an elementary embedding $\pi\colon \mathrm{V} \to M$, for a transitive M, such that $\pi(\kappa) > \lambda$ and ${}^{\lambda}M \subseteq M$.

The final property considered here lies on the top rung of our hierarchy.

[8] See Kanamori (1994, pp. 49, 179).

[9] A proof is in Kanamori (1994, pp. 49–50).

Definition 3.22. Let κ be a cardinal, and let η be a non-zero ordinal. Then κ is η-*extendible* if there is some ordinal α and some elementary embedding $\pi\colon \mathrm{V}_{\kappa+\eta} \to \mathrm{V}_\alpha$ with critical point κ such that $\kappa + \eta < \pi(\kappa) < \alpha$.

To give a grounding for the previous definition, we note that the existence of a 1-extendible cardinal implies the axiom of projective determinacy, which is an important statement in descriptive set theory.[10]

As a visual anchor for the preceding definitions, Figure 3.1 presents the large cardinal properties in order of consistency strength (in which $n + m > 1$). Note that in most cases we actually have the stronger direct implication (i.e. ZFC $\vdash (\exists\kappa Q(\kappa) \to \exists\kappa P(\kappa))$) between two properties.

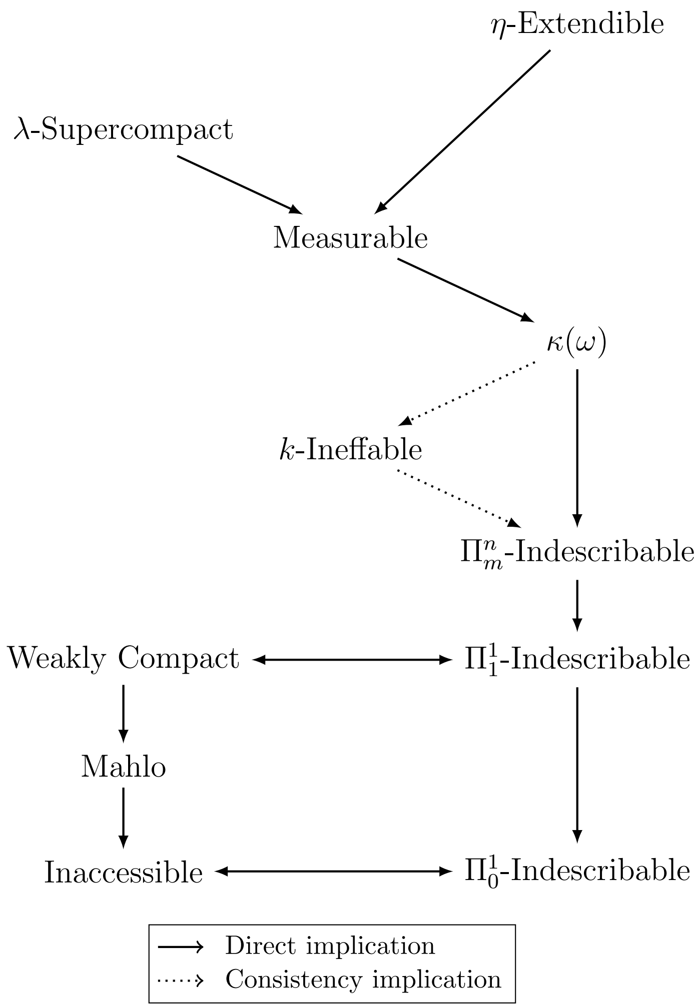

Figure 3.1: Our large cardinals in order.[11]

As far as supercompactness and extendibility go, we have the following result by Solovay et al. (1978).

Theorem 3.23. *If κ is η-extendible and $\delta + 2 \leqslant \eta$, then κ is $|\mathrm{V}_{\kappa+\delta}|$-supercompact.*

[10] In fact this follows from the weaker statement (Kanamori, 1994, pp. 360–361) that there are infinitely many so-called *Woodin cardinals* (Martin and Steel, 1989).

[11] All of the direct implications are shown in Kanamori (1994, pp. 41, 38, 83, 84, 316). The consistency implications are given by Friedman (2001).

4. Reflection

With our preliminary notions set up, we can now proceed with the primary investigation. So far, we have met the following two formalisations of the reflection principle RP:

$$\phi(t_1, \dots, t_n) \to \exists x(\mathsf{tran}(x) \wedge t_1 \in x \wedge \dots \wedge t_n \in x \wedge \phi^x(t_1, \dots, t_n)) \qquad (\Pi^1_0\text{-Reflection}_1)$$
$$\phi(T_1, \dots, T_n) \to \exists x(\mathsf{tran}(x) \wedge \phi^x(T_1 \cap x, \dots, T_n \cap x)) \qquad (\Pi^2_0\text{-Reflection}_2)$$

Note that in the former case ϕ ranges over Π^1_0-formulae, while in the latter ϕ ranges over Π^2_0-formulae. We saw that while ZFC proves every instance of Π^1_0-Reflection$_1$, the same cannot hold for ZFC_2 and Π^2_0-Reflection$_2$.

Π^2_0-Reflection$_2$ is perhaps the most natural generalisation of Π^1_0-Reflection$_1$ to second-order logic, since it allows for second-order parameters; as such this will be our main interest as far as second-order set theory goes. However, we will also briefly investigate the following, which only permits first-order parameters. Here ϕ ranges over Π^2_0-formulae.

$$\phi(t_1, \dots, t_n) \to \exists x(\mathsf{tran}(x) \wedge t_1 \in x \wedge \dots \wedge t_n \in x \wedge \phi^x(t_1, \dots, t_n)) \qquad (\Pi^2_0\text{-Reflection}_1)$$

Once we've considered second-order reflection, an obvious next step is to generalise to higher orders of set theory. The problem here is that it is not so clear how one should relativise higher-order variables to the witnessing set. As we shall see, several apparently natural methods of relativisation result in inconsistent theories. We will investigate various strategies of getting around these inconsistencies, eventually discovering a principle, due to Marshall R., which is very strong indeed.

4.1 Second-Order Reflection

We begin with the schema Π^2_0-Reflection$_2$. When this is added to ZFC_2, how powerful does the resulting system become? Let $\mathrm{ZFC}_2\mathrm{R}_2$ be the system which results from appending Π^2_0-Reflection$_2$ to ZFC_2. As indicated, we will be quantifying the strength of this axiom system by determining which large cardinals can be shown to exist under it. Note that since second-order logic (with $\mathcal{L}_2$) is incomplete (Theorem 2.2), there may be cardinals whose existence is a semantic — but not deductive — consequence of $\mathrm{ZFC}_2\mathrm{R}_2$.

We first observe that Π^2_0-Reflection$_2$ is self-strengthening. For example, we can refine it as follows.

$$\phi(T_1, \dots, T_n) \to \exists \alpha \in \mathrm{On}.\phi^{\mathrm{V}_\alpha}(T_1 \cap \mathrm{V}_\alpha, \dots, T_n \cap \mathrm{V}_\alpha) \qquad (\Pi^2_0\text{-V-Reflection}_2)$$

Lemma 4.1. $\mathrm{ZFC_2R_2}$ *proves every instance of* Π^2_0-V-Reflection$_2$.[1]

Proof ($\mathrm{ZFC_2R_2}$). Assume $\phi(T_1, \ldots, T_n)$. By applying Π^2_0-Reflection$_2$ to $\mathrm{Z_2} \wedge \phi(T_1, \ldots, T_n)$, we get a transitive set x such that ${\mathrm{Z_2}}^x \wedge \phi^x(T_1 \cap x, \ldots, T_n \cap x)$. By the first conjunct $\langle x, \in \rangle \models \mathrm{ZFC_2}$. Hence by Theorem 3.2, there is an inaccessible cardinal κ such that $\langle x, \in \rangle \cong \langle \mathrm{V}_\kappa, \in \rangle$; let υ be the isomorphism. Note that $\upsilon(\varnothing) = \varnothing$ and $p \in q \leftrightarrow \upsilon(p) \in \upsilon(q)$. Hence by induction on $\in$, we have $x = \mathrm{V}_\kappa$. □

Now, notice that the proof also shows that $\mathrm{ZFC_2R_2}$ proves the existence of an inaccessible cardinal. The following lemma (Tait, 2005) gives us Mahlo cardinals.

Lemma 4.2. *For any* Π^2_0*-formula* $\phi(S_1, \ldots, S_n)$ *with free variables shown,* $\mathrm{ZFC_2R_2}$ *proves the following:*[2]

$$\phi(T_1, \ldots, T_n) \to \mathsf{Stationary}(\{\alpha \mid \phi^{\mathrm{V}_\alpha}(T_1 \cap \mathrm{V}_\alpha, \ldots, T_n \cap \mathrm{V}_\alpha)\}^{(2)})$$

Proof ($\mathrm{ZFC_2R_2}$). Assume $\phi(T_1, \ldots, T_n)$. Take C a closed and unbounded class of ordinals. Apply Π^2_0-V-Reflection$_2$ to the formula $\phi(X) \wedge \mathsf{Club}(C)$, to get $\alpha \in \mathrm{On}$ such that

$$\phi^{\mathrm{V}_\alpha}(T_1 \cap \mathrm{V}_\alpha, \ldots, T_n \cap \mathrm{V}_\alpha) \wedge \mathsf{Club}^{\mathrm{V}_\alpha}(C \cap \mathrm{V}_\alpha)$$

Now, regarding $C \cap \mathrm{V}_\alpha$ as a set, $\mathsf{Club}^{\mathrm{V}_\alpha}(C \cap \mathrm{V}_\alpha)$ is equivalent to $\mathsf{club}(\mathrm{V}_\alpha, C \cap \mathrm{V}_\alpha)$, so in particular $C \cap \mathrm{V}_\alpha$ is unbounded in α. Since C is closed, $\alpha \in C$. □

Applying Lemma 4.2 to $\mathrm{Z_2}$, we get that the class of inaccessibles is stationary, then applying Π^2_0-V-Reflection$_2$ to the formula which expresses this gives us a κ (which we can assume is inaccessible by adding $\mathrm{Z_2}$) such that the class of inaccessibles in V_κ is stationary in $\mathrm{V}_\kappa \cap \mathrm{On} = \kappa$ (note "λ is an inaccessible" is absolute between V_κ and V as in the proof of Theorem 2.2); in other words, κ is Mahlo.

But as Tait explains, we can go further. Given a binary tree T, we can apply Π^2_0-V-Reflection$_2$ to get the following:

$$T \text{ is path-bounded} \to \exists \alpha \in \mathrm{On}.(T \cap \mathrm{V}_\alpha \text{ is path-bounded})^{\mathrm{V}_\alpha} \tag{4.1}$$

This means that BTP holds for V (the α in the consequent gives a bound on the tree), and hence it holds for some V_κ. Since (by including a $\mathrm{Z_2}$) we can assume that κ is also inaccessible, this κ is then weakly compact. Further, with this in mind, we can now deliver on our promise to provide a proof that Π^1_1-indescribable cardinals are weakly compact.

Half-proof of Theorem 3.12. Let κ be Π^1_1-indescribable, and let $T \subseteq \mathrm{V}_\kappa$ be a (path-bounded binary tree)$^{\mathrm{V}_\kappa}$. Note that the formula expressing that T is path-bounded is Π^1_1 (the only quantification over classes is "for every class function $F\colon \mathrm{On} \to 2 \ldots$"). Therefore V_κ believes in (4.1), and thus T is bounded. Also note that κ is in particular Π^1_0-indescribable so inaccessible. □

Tharp (1967) gives more information about the large cardinals whose existence is implied by $\mathrm{ZFC_2R_2}$, and gives an upper bound for such results.

[1] Chuaqui (1978, p. 52) attempts to prove a similar lemma by reflecting the supertransitivity of V (X is *supertransitive* if $\forall y \in x.\forall z \subseteq y.z \in X$). Unfortunately, not all supertransitive sets are equal to some V_α; for example $\{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}\}$ is supertransitive. The proof is easily fixed however, as done here.

[2] Recall that $\{\cdots \mid \cdots\}^{(2)}$ is the set-builder notation for a class.

Theorem 4.3. *For each $m \in \omega$, $\mathrm{ZFC_2R_2}$ proves that the collection of Π^1_m-indescribable cardinals is a proper class.*

Proof (Metatheory). Take $m \in \omega$. Tharp notes that it is possible to give a partial truth definition $\mathsf{Sat}_m(X, k)$, such that for every Π^1_m-formula $\phi(X)$ (whose only free variable is X), the following holds:[3]

$$\mathrm{ZFC_2} \vdash \forall X(\mathsf{Sat}_m(X, \ulcorner\phi\urcorner) \leftrightarrow \phi(X))$$
$$\mathrm{ZFC_2} \vdash \forall x(\mathsf{tran}(x) \to \forall X(\mathsf{Sat}^x_m(X, \ulcorner\phi\urcorner) \leftrightarrow \phi^x(X)))$$

By an application of Π^2_0-V-Reflection$_2$ we get

$$\mathrm{ZFC_2R_2} \vdash \forall X \forall k \in \omega.(\mathsf{Sat}_m(X, k) \to \exists \alpha \in \mathrm{On}.\mathsf{Sat}^{\mathrm{V}_\alpha}_m(X \cap \mathrm{V}_\alpha, k))$$

This expresses that V is Π^1_m-indescribable. Then applying Lemma 4.2 gives that $\mathrm{ZFC_2R_2}$ proves the existence of a proper class of Π^1_m-indescribables. □

Let $\mathsf{indes}(n, m, \kappa)$ express that κ is Π^n_m-indescribable.

Theorem 4.4. $\mathrm{ZFC_2R_2} \vdash \forall m \in \omega.\exists \kappa \in \mathrm{Card}.\mathsf{indes}(1, m, \kappa) \to \mathsf{Con}(\mathrm{ZFC_2R_2})$.

Proof ($\mathrm{ZFC_2R_2}$). Note that the collection of Π^2_0-formulae is equal to the union of the collections of Π^1_m-formulae for $m \in \omega$ (this follows as in Lévy (1960)). Furthermore, the definition of a Π^1_m-indescribable cardinal κ (with the use a class pairing operation) directly gives that V_κ satisfies Π^1_m-Reflection$_2$ (that is, Π^2_0-Reflection$_2$ restricted to Π^1_m-formulae). Therefore, for every instance χ of Π^2_0-Reflection$_2$, there is $m \in \omega$ such that for every $k \geqslant m$ and every Π^1_k-indescribable cardinal κ, we have $\mathrm{V}_\kappa \models \chi$ (note that if κ is Π^1_k-indescribable and $m \leqslant k$, then κ is Π^1_m-indescribable). Moreover, by Theorem 3.9, every such cardinal is inaccessible, so satisfies $\mathrm{ZFC_2}$.

Now suppose $\mathrm{ZFC_2R_2}$ is inconsistent. Then there is a proof of an inconsistency, and this uses only finitely many axioms of $\mathrm{ZFC_2R_2}$. Hence by the above there is $m \in \omega$ such that for every Π^1_m-indescribable cardinal κ we have that V_κ is a model of these axioms. Therefore (by the soundness of second-order logic) such a cardinal cannot exist. □

Therefore, by Gödel's second incompleteness theorem, if $\mathrm{ZFC_2R_2}$ is consistent, then $\mathrm{ZFC_2R_2} \nvdash \forall m \in \omega.\exists \kappa \in \mathrm{Card}.\mathsf{indes}(1, m, \kappa)$.

But this is not quite the end of the line, because if we now switch tracks so that we consider semantic consequences, we can go a little further (Shapiro, 1987). Let $\langle \mathrm{V}_\kappa, \in \rangle$ be a model of $\mathrm{ZFC_2}$ (see Theorem 3.2). Then by soundness:

$$\forall m \in \omega\colon\ \langle \mathrm{V}_\kappa, \in \rangle \models \exists \lambda \in \mathrm{Card}.\mathsf{indes}(1, m, \lambda)$$

Hence as $\omega^{\mathrm{V}_\kappa} = \omega$:

$$\langle \mathrm{V}_\kappa, \in \rangle \models \forall m \in \omega.\exists \lambda \in \mathrm{Card}.\mathsf{indes}(1, m, \lambda)$$

[3]This is an extension of a construction given by Lévy (1965, p. 21–26). The idea is as follows. We can define $\mathsf{Sat}_0(X, k)$ by working with a (second-order) object which codes the decomposition of the formula into its composite parts and attaches to a each part a truth value for each possible assignment for the free variables within it. Then $\mathsf{Sat}_m(X, k)$ is defined by recursion on m, (roughly) using $\neg\mathsf{Sat}_m(-, \ulcorner\neg\theta\urcorner)$ to determine the truth of the Σ^1_{m-1} formula θ after the first universal quantifier, and using a class pairing operation.

This means that we do in fact have $\text{ZFC}_2\text{R}_2 \vDash \forall m \in \omega.\exists\kappa \in \text{Card}.\mathsf{indes}(1, m, \kappa)$.[4]

The line of large cardinals that we get really does end here though. This is because, by definition, $\langle \text{V}_\kappa, \in\rangle$ is a model of ZFC_2R_2 if and only if κ is Π^2_0-indescribable (the inaccessibility deals with ZFC_2 by Theorem 3.9, and the rest of the indescribability deals with Π^2_0-Reflection$_2$). Therefore ZFC_2R_2 cannot entail the existence of a Π^2_0-indescribable cardinal (consider the smallest model of ZFC_2R_2, and use that indescribability is absolute for models of ZFC_2). Note that we have also shown that the existence of a Π^2_0-indescribable cardinal implies the consistency of ZFC_2R_2.

We have thus been able to identify quite precisely the consistency strength of ZFC_2R_2 with respect to our large cardinal yardstick. There is more to be said here however. Before we continue our march up the orders, we will shortly take a more lateral step and consider another system, BL, which will turn out to be equivalent (syntactically) to ZFC_2R_2.[5] This will give us a flavour of the utility of our reflection principle.

But first however, we have a promise to deliver upon. As mentioned above, we will briefly look into the following weak reflection principle:

$$\phi(t_1, \dots, t_n) \to \exists x(\mathsf{tran}(x) \wedge t_1 \in x \wedge \cdots \wedge t_n \in x \wedge \phi^x(t_1, \dots, t_n)) \qquad (\Pi^2_0\text{-Reflection}_1)$$

We let $\text{ZFC}_2\text{R}_1 := \text{ZFC}_2 + \Pi^2_0$-Reflection$_1$. First, we obtain the following version of the principle, using the the method from Lemma 4.1.

$$\phi(t_1, \dots, t_n) \to \exists\alpha \in \text{On}.(t_1 \in \text{V}_\alpha \wedge \cdots \wedge t_n \in \text{V}_\alpha \wedge \phi^{\text{V}_\alpha}(t_1, \dots, t_n)) \quad (\Pi^2_0\text{-V-Reflection}_1)$$

And as before, we also get inaccessible cardinals. We can go a little further as follows.

Proposition† 4.5. ZFC_1 *proves the existence of* (a) *α-inaccessible cardinals for all $\alpha \in \text{On}$,* (b) *hyper-inaccessible cardinals.*

Proof† ($\text{ZFC}_2 + \Pi^2_0$-Reflection$_1$).

(a) We prove by induction on α that On is α-inaccessible.[6] Note that On is 0-inaccessible (it's certainly uncountable, there is no unbounded set function on On, and $\forall\lambda \in \text{Card}.2^\lambda \in \text{Card}$). So take $\alpha > 0$. Let $\beta \in \alpha$ and take any ordinal γ. Then by induction hypothesis On is β-inaccessible, and also $\gamma \in \text{On}$. Applying Π^2_0-V-Reflection$_1$ to the conjunction of the previous two statements (with parameters β and γ), we get $\kappa \in \text{On}$ such that κ is β-inaccessible (this is absolute as $\beta \in \text{V}_\kappa$) and $\kappa > \gamma$. Therefore the class of β-inaccessibles in On is unbounded in On, so that On is α-inaccessible.

Finally, one last application of Π^2_0-Reflection$_1$ gives the existence of an α-inaccessible for any $\alpha \in \text{On}$.

(b) It follows from the above that for every $\alpha \in \text{On}$ the class of α-inaccessible cardinals in On is unbounded. If we apply Π^2_0-V-Reflection$_1$ to this statement, we find $\kappa \in \text{On}$ such that for every $\alpha < \kappa$ the class of α-inaccessibles below κ is unbounded—in other words, κ is κ-inaccessible; κ is then by definition hyper-inaccessible.

□

We can continue this process, and obtain α-hyper-inaccessible cardinals and so on.

[4]Note further that $\text{ZFC}_2\text{R}_2 \vDash \mathsf{Con}(\text{ZFC}_2\text{R}_2)$, a real contrast with the first-order case.

[5]Shapiro (1987) and Tait (2005) use a system very much like ZFC_2R_2. Tharp (1967), on the other hand, uses the system BL.

[6]This notion is the obvious continuation of the notion of α-inaccessibility for $\kappa \in \text{Card}$.

4.2 The System BL

The system BL was introduced by Bernays, making strong use of the reflection principles studied by Lévy (hence the name BL for Bernays-Lévy). The formulation given here (Chuaqui, 1978; Marshall R., 1989) is ostensibly first-order; however with its intended interpretation it can talk about both sets and classes. The idea is that instead of the variables ranging over sets, as in first-order ZFC, the variables range over classes. Then it is possible to separate out those objects which are to be thought of as sets by using the condition $\exists y. x \in y$. Notice that now, in contrast to ZFC_2, objects are determined by their extension: in ZFC_2 every set has a sister class with the same extension, which is necessarily distinct from it, whereas in BL all sets *are* classes.[7]

The axioms for BL are given below.[8] We use the language of first-order predicate logic with the binary relation symbol $\in$ (call it $\mathcal{L}_{\mathrm{BL}}$; note that it's the same as the language of ZFC). We denote by V the class of all sets, so that $x \in \mathrm{V}$ is shorthand for $\exists y. x \in y$.

$$\forall z (z \in x \leftrightarrow z \in y) \to x = y \qquad \text{(Extensionality)}$$
$$(x \in \mathrm{V} \wedge y \subseteq x) \to y \in \mathrm{V} \qquad (\text{Subset}_{\mathrm{BL}})$$
$$x \neq \varnothing \to \exists y \in x. \forall z \in x. z \notin y \qquad \text{(Foundation)}$$
$$\exists y \forall x \in \mathrm{V}. (x \in y \leftrightarrow \phi(x)) \qquad (\text{Comprehension}_{\mathrm{BL}})$$
$$\phi(t_1, \ldots, t_n) \to \exists x \in \mathrm{V}. \big(\mathsf{tran}(x) \wedge \phi^{\mathcal{P}(x)}(t_1 \cap x, \ldots, t_n \cap x)\big) \qquad (\text{Reflection}_{\mathrm{BL}})$$
$$\forall x \in \mathrm{V}. (\forall y \in x. y \neq \varnothing \to \exists b \in \mathrm{V}. \forall y \in x. \exists! z \in y. \langle y, z \rangle \in b) \qquad (\text{Choice}_{\mathrm{BL}})$$

Some remarks. (1) $\text{Comprehension}_{\mathrm{BL}}$ and $\text{Reflection}_{\mathrm{BL}}$ are axiom schemata. We let ϕ range over the formulae of $\mathcal{L}$ subject the following conditions: in $\text{Comprehension}_{\mathrm{BL}}$ ϕ may contain additional parameters, but may not contain y free; in $\text{Reflection}_{\mathrm{BL}}$, ϕ's free variables must be contained in $t_1, \ldots, t_n$, and x must not occur free. (2) Notice that Extensionality and Foundation are exactly as in ZFC. They have a stronger flavour here however, since they apply homogeneously to sets and classes. (3) In the $\text{Reflection}_{\mathrm{BL}}$ principle, the relativisation of ϕ to the set x is given by $\phi^{\mathcal{P}(x)}$ (i.e. the result of bounding every quantified variable by $\mathcal{P}(x)$). This elegantly gives us everything we want, since the collection of classes with respect to x is $\mathcal{P}(x)$, and $y \in x$ just if $y \in \mathcal{P}(x)$ and $(y \in \mathrm{V})^{\mathcal{P}(x)}$. Note that $\text{Comprehension}_{\mathrm{BL}}$ ensures that $\mathcal{P}(x)$ really does exist; we will see below that moreover $\mathcal{P}(x)$ is a set. (4) $\text{Choice}_{\mathrm{BL}}$ is the weaker form of choice, applying only to sets. Somewhat surprisingly, however, using reflection we can derive a strong form of choice.

Our goal is to show the (syntactic) equivalence of BL and $\mathrm{ZFC}_2\mathrm{R}_2$. To that end, we need to verify that appropriate versions of the other axioms of ZFC_2 hold in BL. We

[7] According to Shepherdson (1962), Bernays did not take this approach, favouring instead the two-sorted approach given above for ZFC_2.

[8] This particular axiomatisation is given at the beginning of Marshall R.'s (1989).

formulate these below.

$$\exists x \in \mathrm{V}. x = \varnothing \tag{Emptyset$_{\mathrm{BL}}$}$$

$$\forall x \in \mathrm{V}. \forall y \in \mathrm{V}. \exists z \in \mathrm{V}. z = \{x, y\} \tag{Pairing$_{\mathrm{BL}}$}$$

$$\forall x \in \mathrm{V}. \exists y \in \mathrm{V}. y = \bigcup x \tag{Union$_{\mathrm{BL}}$}$$

$$\forall x \in \mathrm{V}. \exists y \in \mathrm{V}. y = \mathcal{P}(x) \tag{Powerset$_{\mathrm{BL}}$}$$

$$\begin{aligned} &\forall f \forall w \in \mathrm{V}. (\forall x \in w. \exists! y \in \mathrm{V}. \langle x, y \rangle \in f \\ &\qquad \rightarrow \exists z \in \mathrm{V}. z = \{y \in \mathrm{V} \mid \exists x \in w. \langle x, y \rangle \in f\}) \end{aligned} \tag{Replacement$_{\mathrm{BL}}$}$$

$$\exists x \in \mathrm{V}. (\exists y \in x. y = \varnothing \wedge \forall y \in x. y \cup \{y\} \in x) \tag{Infinity$_{\mathrm{BL}}$}$$

Lemma 4.6. BL *proves each of the following:* (a) $\mathrm{Emptyset}_{\mathrm{BL}}$, (b) $\mathrm{Pairing}_{\mathrm{BL}}$, (c) $\mathrm{Union}_{\mathrm{BL}}$, (d) $\mathrm{Powerset}_{\mathrm{BL}}$, (e) $\mathrm{Replacement}_{\mathrm{BL}}$, (f) $\mathrm{Infinity}_{\mathrm{BL}}$.

Proof (BL).[9]

(a) Note that $\mathrm{Reflection}_{\mathrm{BL}}$ guarantees the existence of a transitive set x. If this is empty then we're done. Otherwise apply Foundation to get an $\in$-minimal element y. If there is $z \in y$ then by transitivity $z \in x$ contradicting the minimality of y.

(b) Take sets x and y. Apply $\mathrm{Reflection}_{\mathrm{BL}}$ to $\exists x \in a \wedge \exists y \in b$ with parameters $a = \{x\}$ and $b = \{y\}$ (these exist by $\mathrm{Comprehension}_{\mathrm{BL}}$). This gives us a set u such that $x, y \in u$. Now $\{x, y\}$ exists by $\mathrm{Comprehension}_{\mathrm{BL}}$, and $\{x, y\} \subseteq u$, whence by $\mathrm{Subset}_{\mathrm{BL}}$ we have that $\{x, y\}$ is a set.

(c) Take a set x. Apply $\mathrm{Reflection}_{\mathrm{BL}}$ to $\exists x \in a$ with parameter $a = \{x\}$, obtaining a transitive set u such that $x \in u$. Then for each $z \in y \in x$, by transitivity $z \in u$. Whence $\bigcup x \subseteq u$ so by $\mathrm{Subset}_{\mathrm{BL}}$ we have that $\bigcup x$ is a set.

(d) Let x be a set. Apply $\mathrm{Reflection}_{\mathrm{BL}}$ to $\exists x \in a \wedge \mathrm{Subset}_{\mathrm{BL}}$ with parameter $a = \{x\}$ to get a transitive u with $x \in u$ such that $\forall y \in u. \forall z \in \mathcal{P}(u). (z \subseteq y \rightarrow z \in u)$. Now, if $z \subseteq x$ then $z \subseteq u$ so $z \in \mathcal{P}(u)$; hence we have $z \in u$. Therefore $\mathcal{P}(x) \subseteq u$ so that $\mathcal{P}(x)$ is a set.

(e) Let f be a (set) function whose domain is the set w. By $\mathrm{Comprehension}_{\mathrm{BL}}$, there exists a class z such that $\forall y \in \mathrm{V}. (y \in z \leftrightarrow \exists x \in w. \langle x, y \rangle \in f)$. Apply $\mathrm{Reflection}_{\mathrm{BL}}$ to the formula $\exists f \in a$ with parameter $a = \{f\}$, to get a transitive u such that $f \in u$. Then by transitivity $z \subseteq u$, hence by $\mathrm{Subset}_{\mathrm{BL}}$ we have that z is a set as required.

(f) By $\mathrm{Comprehension}_{\mathrm{BL}}$, there is a class z whose elements are precisely the (set) ordinals. Now $\varnothing$ is an ordinal, and if y is an ordinal then so is $y \cup \{y\}$. Apply $\mathrm{Reflection}_{\mathrm{BL}}$ to the formula $\exists y \in z. y = \varnothing \wedge \forall y \in z. y \cup \{y\} \in z$ with parameter z, to get a transitive class u such that $\exists y \in (z \cap u). y = \varnothing \wedge \forall y \in (z \cap u). y \cup \{y\} \in (z \cap u)$. Let $x := (z \cap u)$. This has the required properties, and moreover $x \subseteq u$ so that x is a set by $\mathrm{Subset}_{\mathrm{BL}}$.

□

[9]Many of the ideas here are found in Gloede (1976) and Chuaqui (1978).

We will also need that BL satisfies the following strong form of the axiom of choice:

$$\forall x(\forall y \in x.y \neq \varnothing \to \exists b \forall y \in x.\exists! z \in y.\, \langle y, z\rangle \in b) \qquad (\text{Strong-Choice}_{\text{BL}})$$

The next lemma, whose proof is adapted from one in Gloede (1976), gives us what we want.

Lemma 4.7. $\text{BL} \vdash \text{Strong-Choice}_{\text{BL}}$.

Proof (BL). Take a class x such that $\forall y \in x.y \neq \varnothing$. Suppose for a contradiction that $\neg\exists b \forall y \in x.\exists! z \in y.\, \langle y, z\rangle \in b$. Then we can apply $\text{Reflection}_{\text{BL}}$ to the following formula:

$$\neg\exists b \forall y \in x.\exists! z \in y.\, \langle y, z\rangle \in b \wedge \text{Pairing}_{\text{BL}}$$

To get a set u such that $\mathcal{P}(u) \models \text{Pairing}_{\text{BL}}$ and

$$\neg\exists b \in \mathcal{P}(u).\forall y \in (x \cap u).\exists! z \in y.\, \langle y, z\rangle \in b \qquad (4.2)$$

We need $\mathcal{P}(u) \models \text{Pairing}_{\text{BL}}$ for $\langle y, z\rangle \in b$ to make sense.

Now, by $\text{Subset}_{\text{BL}}$ we have that $x \cap u \in \text{V}$, and further, by assumption on x, we have $\forall y \in (x \cup u).y \neq \varnothing$. Thence by $\text{Choice}_{\text{BL}}$

$$\exists b \in \text{V}.\forall y \in (x \cap u).\exists! z \in y.\, \langle y, z\rangle \in b$$

And so as u is transitive and $\mathcal{P}(u) \models \text{Pairing}_{\text{BL}}$

$$\exists b \in \text{V}.\forall y \in (x \cap u).\exists! z \in y.\, \langle y, z\rangle \in (b \cap u)$$

But for such a b we have $b \cap u \in \mathcal{P}(u)$, which contradicts (4.2). □

With these lemmas in place, all that remains to do in order to complete the equivalence between BL and ZFC_2R_2 is to describe a suitable translation of formulae (since BL and ZFC_2R_2 speak different languages). We will describe (effective) translations $\gimel\colon \mathcal{L}_{\text{BL}} \to \mathcal{L}_2$ and $\daleth\colon \mathcal{L}_2 \to \mathcal{L}_{\text{BL}}$ by recursion on the complexity of formulae.[10] The motivation behind $\gimel$ is that in BL all variables are, initially, thought of as classes. Since, up to extensionality, the classes of ZFC_2R_2 are all that exist in its domain of discourse (by Π^2_0-Comprehension_2), we can translate all variables of BL to second-order variables in ZFC_2R_2; the first-order variables of the latter become "auxiliary" to the second-order variables, as below. So, $\gimel$ sends every (lowercase) variable of $\mathcal{L}_{\text{BL}}$ to its uppercase partner in $\mathcal{L}_2$. The translation then proceeds as follows (for simplicity we only include the connectives $\neg$ and $\wedge$):

$$\begin{aligned}
\gimel(x = y) &:= X = Y \\
\gimel(x \in y) &:= \exists x(\forall z(z \in x \leftrightarrow Xz) \wedge Yx) \\
\gimel(\neg\phi) &:= \neg\gimel(\phi) \\
\gimel(\phi \wedge \psi) &:= \gimel(\phi) \wedge \gimel(\psi) \\
\gimel(\forall x \phi) &:= \forall X \gimel(\phi) \\
\gimel(\exists x \phi) &:= \exists X \gimel(\phi)
\end{aligned}$$

[10] $\gimel$ and $\daleth$ are the Hebrew letters 'gimel' and 'daleth', respectively.

For the other direction, $\daleth$ must "flatten" the formulae of $\mathcal{L}_2$ to produce formulae involving only one sort of variable. To do this, we need an (effective) injection $\mathfrak{i}\colon \{x, y, z, \ldots, X, Y, Z, \ldots\} \to \{x, y, z, \ldots\}$. The translation is then as follows:

$$
\begin{aligned}
\daleth(x = y) &:= \mathfrak{i}_x = \mathfrak{i}_y \\
\daleth(x \in y) &:= \mathfrak{i}_x \in \mathfrak{i}_y \\
\daleth(Yx) &:= \mathfrak{i}_x \in \mathfrak{i}_Y \\
\daleth(\neg\phi) &:= \neg\daleth(\phi) \\
\daleth(\phi \wedge \psi) &:= \daleth(\phi) \wedge \daleth(\psi) \\
\daleth(\forall x\phi) &:= \forall \mathfrak{i}_x \in \mathrm{V}.\daleth(\phi) \\
\daleth(\exists x\phi) &:= \exists \mathfrak{i}_x \in \mathrm{V}.\daleth(\phi) \\
\daleth(\forall X\phi) &:= \forall \mathfrak{i}_X \daleth(\phi) \\
\daleth(\exists X\phi) &:= \exists \mathfrak{i}_X \daleth(\phi)
\end{aligned}
$$

Lemma† 4.8.

(a) *For every sentence ϕ of $\mathcal{L}_{\mathrm{BL}}$*

$$\mathrm{BL} \vdash (\phi \leftrightarrow \daleth(\gimel(\phi)))$$

(b) *For every sentence ψ of $\mathcal{L}_2$*

$$\mathrm{ZFC_2R_2} \vdash (\gimel(\daleth(\psi)) \leftrightarrow \psi)$$

Proof† (Metatheory).

(a) Note that $\daleth(\gimel(x = y)) = \forall \mathfrak{i}_z(\mathfrak{i}_z \in \mathfrak{i}_X \leftrightarrow \mathfrak{i}_z \in \mathfrak{i}_Y)$ which is equivalent (by Extensionality) to $\mathfrak{i}_X = \mathfrak{i}_Y$, and $\daleth(\gimel(x \in y)) = \exists \mathfrak{i}_x \in \mathrm{V}.(\forall \mathfrak{i}_z \in \mathrm{V}.(\mathfrak{i}_z \in \mathfrak{i}_x \leftrightarrow \mathfrak{i}_z \in \mathfrak{i}_X) \wedge \mathfrak{i}_x \in \mathfrak{i}_Y)$ which is equivalent under BL to $\mathfrak{i}_X \in \mathfrak{i}_Y$. The other cases all go though with only a change of variable symbols. Hence for every $\mathcal{L}_{\mathrm{BL}}$-sentence ϕ, $\daleth(\gimel(\phi))$ is equivalent under BL to a sentence ϕ^* in which some of the variable symbols may have been systematically changed. This is equivalent to ϕ.

(b) Let $\mathfrak{I}\colon \{x, y, z, \ldots, X, Y, Z, \ldots\} \to \{X, Y, Z, \ldots\}$ be the result of composing $\mathfrak{i}$ with capitalisation. To ease the notation somewhat, we introduce the shorthand $x = Y^{\downarrow}$ for $\forall z(z \in x \leftrightarrow Yz)$. We will show that $\mathrm{ZFC_2R_2}$ proves the following statement, by induction on the complexity of $\zeta(x_1, \ldots, x_n, Y_1, \ldots, Y_m)$ (where all free variables are shown):

$$
\begin{aligned}
&\forall \mathfrak{I}_{x_1} \cdots \forall \mathfrak{I}_{x_n} \forall \mathfrak{I}_{Y_1} \cdots \forall \mathfrak{I}_{Y_m}((\exists x_1.x_1 = \mathfrak{I}^{\downarrow}_{x_1} \wedge \cdots \wedge \exists x_n.x_n = \mathfrak{I}^{\downarrow}_{x_n}) \\
&\qquad \to (\gimel(\daleth(\zeta))(\mathfrak{I}_{x_1}, \ldots, \mathfrak{I}_{x_n}, \mathfrak{I}_{Y_1}, \ldots, \mathfrak{I}_{Y_m}) \\
&\qquad\qquad \leftrightarrow \exists x_1 \cdots \exists x_n(x_1 = \mathfrak{I}^{\downarrow}_{x_1} \wedge \cdots \wedge x_n = \mathfrak{I}^{\downarrow}_{x_n} \wedge \zeta(x_1, \ldots, x_n, \mathfrak{I}_{Y_1}, \ldots, \mathfrak{I}_{Y_m}))))
\end{aligned}
\tag{4.3}
$$

In each case we assume that we have $\mathfrak{I}_{x_1}, \ldots, \mathfrak{I}_{x_n}, \mathfrak{I}_{Y_1}, \ldots, \mathfrak{I}_{Y_m}$ such that $\exists x_1.x_1 = \mathfrak{I}^{\downarrow}_{x_1} \wedge \cdots \wedge \exists x_n.x_n = \mathfrak{I}^{\downarrow}_{x_n}$. In a slight abuse of notation, we will be denoting by $\mathfrak{I}^{\downarrow}_{x_k}$ the (unique) x_k such that $x_k = \mathfrak{I}^{\downarrow}_{x_k}$.

- *Base case:* $x = y$. Here $\daleth(\urcorner(\zeta))$ is $\mathfrak{I}_x = \mathfrak{I}_y$, and by definition of $=$ on class variables and Extensionality we have

$$\mathfrak{I}_x = \mathfrak{I}_y \leftrightarrow \mathfrak{I}_x^{\downarrow} = \mathfrak{I}_y^{\downarrow}$$

- *Base case:* $x \in y$. Now we have that $\daleth(\urcorner(\zeta))$ is $\exists p(\forall z(z \in p \leftrightarrow \mathfrak{I}_x z) \wedge \mathfrak{I}_y p)$ (where p and q are distinct from $\mathfrak{I}_x$ and $\mathfrak{I}_y$). If this holds, then by Extensionality, $p = \mathfrak{I}_x^{\downarrow}$, and furthermore, by definition of $\mathfrak{I}_y^{\downarrow}$ we have $p \in \mathfrak{I}_y^{\downarrow}$; whence $\mathfrak{I}_x^{\downarrow} \in \mathfrak{I}_y^{\downarrow}$. Conversely, if $\mathfrak{I}_x^{\downarrow} \in \mathfrak{I}_y^{\downarrow}$ holds then by definition $\mathfrak{I}_y \mathfrak{I}_x^{\downarrow}$ and $\forall z(z \in \mathfrak{I}_x^{\downarrow} \leftrightarrow \mathfrak{I}_x z)$.
- *Base case:* Yx. Here $\daleth(\urcorner(\zeta))$ is $\exists p(\forall z(z \in p \leftrightarrow \mathfrak{I}_x z) \wedge \mathfrak{I}_Y p)$. Similarly to the previous case, this is equivalent to $\mathfrak{I}_Y \mathfrak{I}_x^{\downarrow}$.
- *Induction steps:* $\neg\eta$ *and* $\eta \wedge \theta$. These follow straightforwardly by induction.
- *Induction step:* $\forall x.\eta(x)$. This time $\daleth(\urcorner(\zeta))$ is

$$\forall \mathfrak{I}_x(\exists T \exists p(\forall z(z \in p \leftrightarrow \mathfrak{I}_x z) \wedge p \in T) \rightarrow \daleth(\urcorner(\eta))(\mathfrak{I}_x))$$

First-off, by Π^2_0-Comprehension$_2$, this is equivalent to

$$\forall \mathfrak{I}_x(\exists x.x = \mathfrak{I}_x^{\downarrow} \rightarrow \daleth(\urcorner(\eta))(\mathfrak{I}_x))$$

Then, by the induction hypothesis, this is equivalent to

$$\forall \mathfrak{I}_x(\exists x.x = \mathfrak{I}_x^{\downarrow} \rightarrow \eta(\mathfrak{I}_x^{\downarrow}))$$

Since for any set, by Π^2_0-Comprehension$_2$, there is a class with the same extension, this is equivalent to $\forall x.\eta(x)$.
- *Induction step:* $\forall X.\eta(X)$. This is clear since $\daleth(\urcorner(\zeta))$ is $\forall \mathfrak{I}_X.\eta(\mathfrak{I}_X)$.

Note that both $\exists$ cases are taken care of by using $\neg$ and $\forall$.

And thus, by (4.3), we have that $\mathrm{ZFC_2R_2} \vdash (\daleth(\urcorner(\psi)) \leftrightarrow \psi)$ for all $\mathcal{L}_2$-sentences ψ.

□

Theorem 4.9.

(a) *For every sentence* ϕ *of* $\mathcal{L}_{\mathrm{BL}}$

$$\mathrm{BL} \vdash \phi \Leftrightarrow \mathrm{ZFC_2R_2} \vdash \daleth(\phi)$$

(b) *For every sentence* ψ *of* $\mathcal{L}_2$

$$\mathrm{ZFC_2R_2} \vdash \psi \Leftrightarrow \mathrm{BL} \vdash \urcorner(\psi)$$

Proof† (Metatheory).[11] First, we will show that $\mathrm{BL} \vdash \phi \Rightarrow \mathrm{ZFC_2R_2} \vdash \daleth(\phi)$. We note that by definition of $\mathrm{D_2}$, the translation of every axiom and rule of inference of first-order predicate calculus by $\daleth$ is deducible in $\mathrm{D_2}$. So it is sufficient to show that the translation

[11]This equivalence is known (see e.g. Shapiro, 1987).

of every axiom of BL is deducible in $\mathrm{ZFC_2R_2}$. The translated axioms are as follows (after some simplification using the axioms of $\mathrm{ZFC_2R_2}$):

$$\forall z(Xz \leftrightarrow Yz) \rightarrow X = Y \qquad (\gimel(\text{Extensionality}))$$

$$\exists x(x = X^{\downarrow} \wedge \forall z.(Yz \rightarrow z \in x)) \rightarrow \exists y.y = Y^{\downarrow} \qquad (\gimel(\text{Subset}_{\mathrm{BL}}))$$

$$\exists y.Xy \rightarrow \exists y.(Xy \wedge \forall z(Xz \rightarrow z \notin y)) \qquad (\gimel(\text{Foundation}))$$

$$\exists Y \forall x(Yx \leftrightarrow \gimel(\phi)(x)) \qquad (\gimel(\text{Comprehension}_{\mathrm{BL}}))$$

$$\phi(T_1, \ldots, T_n) \rightarrow \exists x.(\mathsf{tran}(x) \wedge \gimel(\phi)^x(T_1 \cap x, \ldots, T_n \cap x)) \qquad (\gimel(\text{Reflection}_{\mathrm{BL}}))$$

$$\forall x(\forall y \in x.y \neq \varnothing \rightarrow \exists b \forall y \in x.\exists! z \in y. \langle y, z\rangle \in b) \qquad (\gimel(\text{Choice}_{\mathrm{BL}}))$$

$\gimel$(Extensionality) follows by definition of = on second-order variables. $\gimel(\text{Subset}_{\mathrm{BL}})$ follows from Separation_2, which is deducible in $\mathrm{ZFC_2R_2}$ by Lemma 2.3. $\gimel$(Foundation) and $\gimel(\text{Choice}_{\mathrm{BL}})$ are Foundation_2 and Choice, which are deducible in $\mathrm{ZFC_2R_2}$ by Lemma 2.5 and Lemma 2.4, respectively. Each instance of $\gimel(\text{Comprehension}_{\mathrm{BL}})$ is an instance of Π^2_0-Comprehension_2. Finally each instance of $\gimel(\text{Reflection}_{\mathrm{BL}})$ is an instance of Π^2_0-Reflection_2 (note that the relativisation of ϕ translates properly since in $\gimel(\phi)^x$, the second-order variables get bounded by $\mathcal{P}(x)$, and in BL $(p \in \mathrm{V})^{\mathcal{P}(q)}$ precisely when $p \in q$).

Next, we show that $\mathrm{ZFC_2R_2} \vdash \psi \Rightarrow \mathrm{BL} \vdash \daleth(\psi)$ (we can assume that we've eliminated $\mathcal{B}$). The axioms and rules of inference other than Π^2_0-Comprehension_2 and Strong-Choice translate into statements easily derivable in BL. The remaining axioms of $\mathrm{ZFC_2R_2}$ with $\mathrm{D_2}$ are translated as follows (again after some simplification in BL):

$$\exists y \forall x \in \mathrm{V}.(x \in y \leftrightarrow \daleth(\phi)(x)) \qquad (\daleth(\Pi^2_0\text{-Comprehension}_2))$$

$$\forall x(\forall y \in x.y \neq \varnothing \rightarrow \exists b \forall y \in x.\exists! z \in y. \langle y, z\rangle \in b) \qquad (\daleth(\text{Strong-Choice}))$$

$$\forall x \in \mathrm{V}.\forall y \in \mathrm{V}.(\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y) \qquad (\daleth(\text{Extensionality}))$$

$$\exists x \in \mathrm{V}.x = \varnothing \qquad (\daleth(\text{Emptyset}))$$

$$\forall x \in \mathrm{V}.\forall y \in \mathrm{V}.\exists z \in \mathrm{V}.z = \{x, y\} \qquad (\daleth(\text{Pairing}))$$

$$\forall x \in \mathrm{V}.\exists y \in \mathrm{V}.y = \bigcup x \qquad (\daleth(\text{Union}))$$

$$\forall x \in \mathrm{V}.\exists y \in \mathrm{V}.y = \mathcal{P}(x) \qquad (\daleth(\text{Powerset}))$$

$$\forall x \in \mathrm{V}.(x \neq \varnothing \rightarrow \exists y \in x.\forall z \in x.z \notin y) \qquad (\daleth(\text{Foundation}))$$

$$\exists x \in \mathrm{V}.(\exists y \in x.y = \varnothing \wedge \forall y \in x.y \cup \{y\} \in x) \qquad (\daleth(\text{Infinity}))$$

$$\begin{aligned} &\forall f \forall w \in \mathrm{V}.(\forall x \in w.\exists! y \in \mathrm{V}. \langle x, y\rangle \in f \\ &\qquad \rightarrow \exists z \in \mathrm{V}.z = \{y \in \mathrm{V} \mid \exists x \in w. \langle x, y\rangle \in f\}) \end{aligned} \qquad (\daleth(\text{Replacement}_2))$$

Each of these either follows directly from an axiom of BL, from Lemma 4.6, or from Lemma 4.7.

Finally, the other implications follow by Lemma 4.8. □

The deductive equivalence between BL and $\mathrm{ZFC_2R_2}$ is now complete. We have demonstrated the usefulness of the second-order reflection, since it allowed us to derive all the normal axioms of set theory with minimal additional ingredients. Also, the sets-as-classes perspective — and its translatability (as far a provability goes) to the more traditional approach of ZFC — will be important shortly when we attempt to formulate third-order reflection.

4.3 Higher-Order Reflection

Now that we have explored reflection in ZFC_2, we would like to consider higher-order principles. We will present a number of attempts at breaching the third-order, several of which turn out to be inconsistent, and towards the end we will state without proof a few results concerning the strength and limitations of the more successful principles.

Koellner (2009) gives a reasonable way in which we might formulate the principle RP for higher-order logic. The idea is that to relativise a 2-class,[12] we relativise each class within it. More generally, the relativisation $X^{(n);\,x}$ of $X^{(n)}$ to x is given as follows:

$$X^{(2);\,x} := X^{(2)} \cap x$$
$$X^{(n);\,x} := \left\{Y^{(n-1);\,x} \colon X^{(n)}Y^{(n-1)}\right\}^{(n)} \qquad \text{for } n > 2$$

We also need a way to relativise higher-order formulae, but this is done in a way which is a straightforward generalisation of the second-order case (i.e. quantified third-order variables get bound by $\mathcal{PP}(x)$ and so on).

The nth-order reflection principle is then given as follows (in which ϕ is a Π^n_0-formula):

$$\phi\Big(T_1^{(n)}, \ldots, T_k^{(n)}\Big) \to \exists x\Big(\mathsf{tran}(x) \wedge \phi^x\Big(T_1^{(n);\,x}, \ldots, T_k^{(n);\,x}\Big)\Big) \qquad (\Pi^n_0\text{-Reflection}^\#_n)$$

However, Koellner notes that this principle, when added to ZFC_n for $n > 2$ results in an inconsistent theory. First, we can derive the following Π^n_0-V-Reflection$^\#_n$ principle from Π^n_0-Reflection$^\#_n$ as in Lemma 4.1:

$$\phi\Big(T_1^{(n)}, \ldots, T_k^{(n)}\Big) \to \exists\alpha \in \text{On}.\phi^{\text{V}_\alpha}\Big(T_1^{(n);\,\text{V}_\alpha}, \ldots, T_k^{(n);\,\text{V}_\alpha}\Big) \qquad (\Pi^n_0\text{-V-Reflection}^\#_n)$$

Now let $T = \{\{\beta \in \alpha\}^{(2)} \mid \alpha \in \text{On}\}^{(3)}$ and apply Π^3_0-V-Reflection$^\#_3$ to:

$$\forall Y^{(2)}\big(T^{(3)}Y^{(2)} \to \exists\gamma \in \text{On}.Y^{(2)} \subseteq \gamma\big)$$

This gives an $\alpha \in \text{On}$ such that:

$$\forall Y^{(2)}\big(T^{(3);\text{V}_\alpha}Y^{(2)} \to \exists\gamma \in \alpha.Y^{(2)} \subseteq \gamma\big)$$

But this gives us our contradiction, since $\{\beta \in \alpha\}^{(2)}$ is in $T^{(3);\text{V}_\alpha}$.

An alternative formulation is given by Marshall R. (1989). In this paper, she generalises the system BL to the third-order to produce the system BL_2. The only axiom whose third-order version is not obvious is $\text{Reflection}_{\text{BL}}$. We will eventually come to consider the formulation which becomes part of BL_2, but first we mention a more natural formulation which Marshall R. notes to be inconsistent. The system BL_2 extends the notion of sets-as-classes, so that the only objects which BL_2 knows about are 2-classes. From this perspective, a good relativisation of a 2-class t to a set x is $t \cap \mathcal{P}(x)$, since quantified variables which (effectively) range over classes get bound by $\mathcal{P}(x)$. Now, since the translation between the languages of BL and ZFC_2R_2 given above extends to the third-order, we can present this reflection principle from the BL_2 perspective, knowing that this can be converted back to the language of ZFC_3.

$$\phi(t_1, \ldots, t_k) \to \exists x \in \text{V}.\big(\mathsf{tran}(x) \wedge \phi^{\mathcal{PP}(x)}(t_1 \cap \mathcal{P}(x), \ldots, t_k \cap \mathcal{P}(x))\big) \qquad (\text{Reflection}^\#_{\text{BL}_2})$$

To see that this principle is inconsistent, we reflect the formula $\exists y.y \in t$ with the parameter $t = \{\text{V}\}$. For any set x, we have $\text{V} \notin \mathcal{P}(x)$, to that $t \cap \mathcal{P}(x) = \varnothing$. The problem is that a 2-class can contain objects which are not sets,[13] so intersecting it with a set (which

[12]Recall that a 2-class is a collection of classes.

[13]Or objects not coextensive with sets, depending on one's perspective.

contains only sets) need not behave as we would like.

So these fairly natural continuations of Π^2_0-Reflection$_2$ turned out to be inconsistent, but perhaps we were too hasty. It may be that third- and higher-order reflection is just a whole different beast, one which should be approached more cautiously by first considering weaker principles. We examine this route briefly by considering Π^n_0-Reflection$_2$ and $\Gamma^{(2)}_k$-Reflection.

The first weakened principle we deal with is a partially generalised version of Π^2_0-Reflection$_2$. We ran into trouble when we allowed both Π^n_0-formulae and nth-order parameters, so how about only allowing Π^n_0-formulae? This results in the following schema:

$$\phi(T_1, \ldots, T_k) \to \exists x(\mathsf{tran}(x) \wedge \phi^x(T_1 \cap x, \ldots, T_k \cap x)) \qquad (\Pi^n_0\text{-Reflection}_2)$$

Is this consistent, and if so, how far can it reach on our yardstick? By considering the definition, $\mathrm{V}_\kappa \models \Pi^n_0$-Reflection$_2$ if and only if κ is Π^n_0-indescribable. Thus the existence of a Π^n_0-indescribable cardinal implies the consistency of $\mathrm{ZFC}_2 + \Pi^n_0$-Reflection$_2$, and Π^n_0-Reflection$_2$ can't give us these cardinals (by Gödel's second incompleteness theorem). However, in a similar manner to Theorem 4.3, we can show that Π^n_0-Reflection$_2$ does give us Π^{n-1}_m-indescribable cardinals for every $m \in \omega$. We will omit the argument here however.

The next principle we consider is given by Tait (2005). A formula of $\mathcal{L}_\omega$ is *positive* if it is built up from the atoms $x = y$, $x \neq y$, $x \in y$, $x \notin y$, $Y^{(2)}x$, $\neg Y^{(2)}x$, $X^{(k)} = Y^{(k)}$ and $Y^{(k+1)}X^{(k)}$ where $k > 1$, using the binary connectives $\wedge$ and $\vee$, and the quantifiers $\forall$ and $\exists$. For $n > 0$ and $k \geqslant 2$, we let $\Gamma^{(n)}_k$ be the class of formulae with the following form:

$$\forall X^{(n)}_1 \exists Y^{(a_1)}_1 \cdots \forall X^{(n)}_k \exists Y^{(a_k)}_k \chi\Big(X^{(n)}_1, Y^{(a_1)}_1, \ldots, X^{(n)}_k, Y^{(a_k)}_k\Big)$$

in which $a_1, \ldots, a_k \in \{2, 3, 4, \ldots\}$ and χ is a positive Π^1_0-formula. We will consider the following reflection principle, in which ϕ ranges over $\Gamma^{(2)}_k$-formulae and $m > 1$.

$$\phi\Big(T^{(m)}_1, \ldots, T^{(m)}_l\Big) \to \exists x\Big(\mathsf{tran}(x) \wedge \phi^x\Big(T^{(m);\,x}_1, \ldots, T^{(m);\,x}_l\Big)\Big) \qquad (\Gamma^{(2)}_k\text{-Reflection})$$

This has quite a high (maximum) consistency strength when measured with our yardstick, as the following theorem (Tait, 2005) demonstrates.

Theorem 4.10. *If* $\mathrm{V}_\kappa \models \Gamma^{(2)}_k$-Reflection *then κ is k-ineffable.*

However, $\Gamma^{(2)}_k$-Reflection doesn't crawl that far up the large cardinal hierarchy, as the next theorem (Koellner, 2009) shows.

Theorem 4.11. *If the ωth Erdős cardinal $\kappa(\omega)$ exists, then there is $\delta < \kappa(\omega)$ such that for every $k > 0$ we have* $\mathrm{V}_\delta \models \Gamma^{(2)}_k$-Reflection.

$\Gamma^{(2)}_k$-Reflection has allowed us to inch further along the large cardinal yardstick to produce the strongest theory yet in terms of consistency strength. The obvious next step is to generalise it to produce the principle below, where ϕ can be any $\Gamma^{(n)}_k$-formula.

$$\phi\Big(T^{(m)}_1, \ldots, T^{(m)}_l\Big) \to \exists x\Big(\mathsf{tran}(x) \wedge \phi^x\Big(T^{(m);\,x}_1, \ldots, T^{(m);\,x}_l\Big)\Big) \qquad (\Gamma^{(n)}_k\text{-Reflection})$$

As Koellner shows though, this again causes us to fall into inconsistency.

Theorem 4.12. $\mathrm{ZFC}_\omega + \Gamma_1^{(3)}$-Reflection *is inconsistent.*

And so finally, we reach the last and strongest reflection principle considered here.[14] It is this principle which Marshall R. uses in the generalisation of BL. In fact, BL gets generalised to the nth-order (to produce the system BL_{n-1}), and so we present the general reflection principle here. We need a few preliminary notions.

The process of variable relativisation is slightly different; we no longer stipulate that the object to which a variable is relativised need be a set. We relativise a class a to u as $a^u := a \cap u$. Then for a general object a, the relativisation to u is defined by recursion on the $\in$-relation $a^u := \{b^u \mid b \in a \cap u\}$. Next, we can define the ordered pair of two classes as

$$[p, q] := p \times \{0\} \cup q \times \{1\}$$

Then for m-classes p and q, we define the ordered pair by recursion as follows:

$$[p, q] := \{[d, 0] \mid d \in p\} \cup \{[e, 1] \mid e \in q\}$$

Now, the condition of transitivity used in previous principles is replaced by the "niceness" predicate:

$$\mathsf{nice}(u) := u \cap \mathrm{V} \in \mathrm{V} \wedge \mathsf{tran}(u \cap \mathrm{V}) \wedge \forall p \forall q((p \in u \wedge q \in u) \leftrightarrow [p, q] \in u)$$

Finally, the BL_n reflection principle is as follows:

$$\phi(t_1, \ldots, t_k) \to \exists u.\big(\mathsf{nice}(u) \wedge \phi^{\mathcal{P}^n(u \cap \mathrm{V})}({t_1}^u, \ldots, {t_k}^u)\big) \qquad (\text{Reflection}_{\mathrm{BL}_n})$$

The idea is that our transitive set is now $u \cap \mathrm{V}$, but we augment it by adding extra "appendages" (the remaining elements of u), which produces a more complex variable relativisation. This relativisation allows $\text{Reflection}_{\mathrm{BL}_n}$ to dodge our previously successful attacks. For example the strategy against Π_0^n-$\text{Reflection}_n^{\#}$ can be carried out for $\text{Reflection}_{\mathrm{BL}_2}$ (we can assume that $u \cap \mathrm{V} = \mathrm{V}_\alpha$, and we can add a class to our parameter to make sure it gets relativised as a 2-class). But this will not be successful, since we can manipulate u so that $\{s \cap u \mid s \in t \cap u\}$ doesn't contain the troublesome unbounded class.

So, how strong is this new principle? Let us now form BL_n by adding $\text{Reflection}_{\mathrm{BL}_n}$ to the generalised versions of Extensionality, $\text{Subset}_{\mathrm{BL}}$, Foundation, $\text{Comprehension}_{\mathrm{BL}}$ and $\text{Choice}_{\mathrm{BL}}$. We conclude by stating the two main results about BL_n (Marshall R., 1989), which locate its consistency strength quite precisely on our hierarchy.

Theorem 4.13. $\mathrm{BL}_n \vdash \mathsf{Stationary}(\{\alpha \in \mathrm{On} \mid \alpha$ *is* $(n-1)$*-extendible*$\})$.

So $\text{Reflection}_{\mathrm{BL}_n}$ is very strong indeed; it reaches right to the top rung of the large cardinal hierarchy given here. We can also give quite a tight upper bound on the large cardinals produced, and establish a consistency result.

Theorem 4.14. *If* κ *is* n*-extendible or* $|\mathrm{V}_{\kappa+n-1}|$*-supercompact, then* $\mathrm{V}_{\kappa+n} \models \mathrm{BL}_n$.

This shows that BL_n is consistent relative to the existence of an n-extendible cardinal. Moreover by Gödel's second incompleteness theorem, BL_2 cannot prove the existence of a n-extendible cardinal, if it is consistent.

[14]Bar the inconsistent ones of course.

5. Conclusion

We began our journey with a challenge: to strengthen standard set theory by adding new axioms to affect a "significant reduction in incompleteness". We focused on reflection principles as good candidates for axioms that are intrinsically justified. In investigating these principles, we partitioned them based on the order of the property being reflected. First-order reflection turned out to be already included in standard ZFC. Second-order principles, on the other hand, gave us something new, and their addition to ZFC_2 was fruitful and interesting. However, the strength of the resulting theory was limited as far as our large cardinal hierarchy was concerned, and in our thirst for stronger set theories we turned to third- and higher-order versions of the principles thus far considered. This is where the going got tougher; our initial attempts at formulation resulted in inconsistency, and so we were forced to proceed more carefully, formulating axioms diverging somewhat from the original first-order principle. This eventually led us to the very strong principle $\mathrm{Reflection}_{\mathrm{BL}_n}$.

Now, the power of $\mathrm{Reflection}_{\mathrm{BL}_n}$ is very seductive, but in our eagerness to obtain stronger and stronger theories, we might have lost sight of the motivation for looking at reflection in the first place: is $\mathrm{Reflection}_{\mathrm{BL}_n}$ intrinsically justified? It doesn't have the elegance of Π^1_0-$\mathrm{Reflection}_1$ or $\mathrm{Reflection}_{\mathrm{BL}}$; the transitive set is now augmented by strange appendages, and there is a more complex condition which the n-class u must satisfy. It seems that we would have to work quite hard to convince ourselves of the "truth" of this principle.

The conclusion we might draw here is that reflection does not generalise well to the third-order. Indeed, there is a sense in which third-order set theory has a different nature from that of first- and second-order. While the concept of a class can be considered as an extension of the concept of a set — classes are collections of sets just like sets themselves — 2-classes have a different character as they contain objects other than sets. But possibly we just don't have enough experience of third-order set theory to determine what the "true" statements are; after all, it's relatively rare to see a third-order concept in the wild.[1] So perhaps the jury is still out with regard to a third-order version of RP. It is plausible that we might eventually come to see a third-order reflection principle as an intrinsically justified addition to set theory.

Regardless of the status of third-order reflection, our exploration in this dissertation has provided us with a greater understanding of the nature of set theory, and it is one step along the infinite path of strengthening this essential tool of mathematics.

[1]I would like to thank Rolf Suabedissen for suggesting this idea to me.

Bibliography

Boolos, G. (1984). To be is to be a value of a variable (or to be some values of some variables). *The Journal of Philosophy*, 81(8):430–449.

Chuaqui, R. (1978). Bernays' class theory. In *Mathematical logic: proceedings of the first Brazilian conference*, volume 39 of *Lecture Notes in Pure and Applied Mathematics*, pp. 31–55. Marcel Dekker, New York.

Friedman, H. (2001). Subtle cardinals and linear orderings. *Annals of Pure and Applied Logic*, 107:1–34.

Gloede, K. (1976). Reflection principles and indescribability. In *Sets and Classes: on The Work by Paul Bernays*, volume 84 of *Studies in Logic and the Foundations of Mathematics*, pp. 277–323. North-Holland Publishing Company, Amsterdam.

Gödel, K. (1947). What is Cantor's continuum problem? *The American Mathematical Monthly*, 54(9):515–525.

Kanamori, A. (1994). *The higher infinite : large cardinals in set theory from their beginnings*. Springer–Verlag, Berlin; London.

Karagila, A. (2013). What axioms need to be added to second-order ZFC before it has a unique model (up to isomorphism)? Mathematics Stack Exchange.

Koellner, P. (2009). On reflection principles. *Annals of Pure and Applied Logic*, 157(2):206–219.

Lévy, A. (1960). Axiom schemata of strong infinity in axiomatic set theory. *Pacific Journal of Mathematics*, 10:223–238.

Lévy, A. (1965). A hierarchy of formulas in set theory. *Memoirs of the American Mathematical Society*, (57).

Maddy, P. (1988). Believing the axioms. I. *Journal of Symbolic Logic*, 53(2):481–511.

Marshall R., M. V. (1989). Higher order reflection principles. *Journal of Symbolic Logic*, 52(2):474–489.

Martin, D. A. and Steel, J. R. (1989). A proof of projective determinacy. *Journal of the American Mathematical Society*, 2(1):71–125.

Shapiro, S. (1987). Principles of reflection and second-order logic. *Journal of Philosophical Logic*, 16(3):309–333.

Shapiro, S. (1991). *Foundations without foundationalism: A case for second-order logic*. Oxford: Clarendon Press.

Shelah, S. (1991). The future of set theory. In Judah, H. (editor), *Israel Mathematical Conference Proceedings*, volume 6, pp. 1–12.

Shepherdson, J. C. (1952). Inner models for set theory–part II. *Journal of Symbolic Logic*, 17(4):225–237.

Shepherdson, J. C. (1962). Review of *Zur Frage der Unendlichkeitsschemata in der Axiomatischen Mengenlehre*. *Journal of Symbolic Logic*, 27(3):353–354.

Solovay, R. M., Reinhardt, W. N., and Kanamori, A. (1978). Strong axioms of infinity and elementary embeddings. *Annals of Mathematical Logic*, 13(1):73–116.

Suabedissen, R. (2015). C1.4 axiomatic set theory. Lecture Notes.

Tait, W. W. (2005). Constructing cardinals from below. In *The Provenance of Pure Reason: essays in the philosophy of mathematics and its history*, pp. 133–154. Oxford University Press.

Tarski, A. (1933). *Pojęcie prawdy w językach nauk dedukcyjnych [The concept of truth in formalized languages]*. Towarszystwo Naukowe Warszawskie, Warsaw. Translation in Tarski (1956, pp. 152–278).

Tarski, A. (1956). *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Woodger, J. H. (editor). Oxford: Clarendon Press.

Tharp, L. H. (1967). On a set theory of Bernays. *Journal of Symbolic Logic*, 32(3):319–321.

Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre [On boundary numbers and domains of sets: New investigations in the foundations of set theory]. *Fundamenta Mathematicæ*, 16:29–47. Translation in Zermelo (2010, pp. 399–430).

Zermelo, E. (2010). *Enrst Zermelo, Collected Works: Volume I*. Ebbinghaus, H.-D., Kanamori, A., and Fraser, C. G. (editors). Springer, Berlin.