The independence of the Axiom of Choice in Set Theory

Submitted by

C. Elliott Belbin, B.Sc. (Honours)

A thesis submitted to

The Faculty of Graduate Studies and Research

in partial fulfillment of

the requirements for the degree of

Master of Science

School of Mathematics and Statistics

Ottawa-Carleton Institute of Mathematics and Statistics

Carleton University

Ottawa, Ontario, Canada K1S 5B6

January 2006

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

We present a demonstration that the Axiom of Choice is consistent with and independent of the usual Zermelo-Fraenkel (ZF) axioms for set theory. That is, it cannot be proven either true or false on the basis of those axioms alone. We start by discussing the unique role that the Axiom of Choice has played in set theory due to its non-constructive nature. We present Kurt Gödel's 1938 proof that the Axiom of Choice is consistent with the ZF axioms. The concept of constructibility is central to this argument. We then present a more recent take on Paul Cohen's 1963 proof that the Axiom of Choice is independent of the ZF axioms, due to Raymond Smullyan and Melvin Fitting. Cohen's forcing technique is reformulated in the context of modal logic in this approach.

iii

Acknowledgements

I would like to thank my supervisor, Dr. John Poland, for all of the guidance he has given me over the years. Also, I would like to thank Estelle Vincent-Fleurs and Danielle Langlois for their patience and kindness.

Contents

1   Introduction

The Axiom of Choice is a statement about the set-theoretical universe. Like most of the other statements presented as axioms of the various formulations of set theory, it asserts the existence of a certain kind of set. Let us state one form of it here for the record.

Axiom of Choice. *For any set A there is a function F (a "choice function" for A) such that the domain of F is the set of nonempty subsets of A, and such that $F(B) \in B$ for every nonempty $B \subseteq A$. [End77, p.151]*

The word "axiom" is used in two different senses in English: in its formal mathematical sense it means a statement in a formal system that is adopted as true without proof so that its consequences can be used or investigated. In its informal everyday sense, it simply means a self-evident truth. Modern mathematics employs only formal axioms. Usually, the formulation of the axioms in mathematics is guided by the informal notion of self-evident truth, especially when the area in question is thought to reflect the "real world" in some way, such as with geometry or set theory. The standard Zermelo-Fraenkel axioms for set theory, presented here in section 2.4.1, are commonly thought to satisfy both definitions, at least in the statements they make about finite sets. Their extension of our intuition about sets from the finite to the infinite is extremely useful in mathematics and has never been shown to entail any contradictions.

The Axiom of Choice, however, is a different kind of statement. The Zermelo-Fraenkel axioms make straightforward assertions such as "if a and b are sets, then there is a set containing a and b". The Axiom of Choice is different; its status as an axiom is tainted by the fact that it is not universally considered a self-evident truth—it is an axiom only in the formal sense. First, it is only useful when relationships between infinite collections are being considered, and infinity has always been difficult to handle with intuition alone. Second, it is a more complicated statement than the Zermelo-Fraenkel axioms. Like Euclid's parallel postulate, it has the air of a theorem rather than an axiom. More explanation of the claim it makes will be given in section 2.2.

This paper will focus on the relationship between the Axiom of Choice (sometimes referred to here as "AC" or "the Axiom") and the Zermelo-Fraenkel axioms for set theory (or "ZF"). The Axiom of Choice is not part of this latter collection. By "ZFC" we mean ZF augmented with the Axiom of Choice.

The Axiom of Choice became important in the early 20th century after it was explicitly formulated and proposed for the first time as an axiom in both senses by Ernst Zermelo in a 1904 article

in which he assumed that it held and proved that every set could be well-ordered. He included an informal justification of it at the end of the article in which he said that "...it is applied without hesitation everywhere in mathematical deduction." ([Zer04]) Reaction from those who objected to the Axiom was so strong that Zermelo was forced to publish a follow-up paper addressing the criticism. More will be said about this controversy in section 2.3. It is worth remarking that this initial controversy took place before there was a formal axiomatisation of set theory; the argument was not about formal consequences of AC but whether AC held true in the "real world"—the informal sense of the word "axiom". With the benefit of 100 years' hindsight and the development of mathematical logic in the early 20th century, we can paint a clearer picture today of the Axiom's role in mathematics.

It is our objective to paint a part of this picture. Our general approach will be as follows. We settle on ZF as our system of set theory, and assume that it has at least one model (i.e. that it is consistent). Now the question is: is it desirable to adopt the Axiom of Choice as an axiom of set theory or not? There are two broad questions that need to be answered:

1. (Consistency) Is it possible to adopt AC? We need to know that it doesn't conflict with ZF.

2. (Independence) Is it necessary to adopt AC as a separate axiom to guarantee that it always holds? We need to know that it isn't already implied by some combination of the axioms of ZF.

We will show that the answer to both questions is yes; that is, AC is both consistent with ZF and independent of it.

Our primary source for the independence proofs is "Set Theory and the Continuum Problem" by Raymond Smullyan and Melvin Fitting [SF96]. An updated errata list for this text is available at http://comet.lehman.cuny.edu/fitting/errata/errata.html.

The author has corrected [SF96] on two minor points in this text and provided proofs in several places where [SF96] omits them (see lemma 4.44 and theorems 4.17, 4.23 and 4.47). The author has also attempted to provide, in section 4, an explanation of some properties of S4 models (especially equality and identity) that complements the original source text by Smullyan and Fitting and provides a new glimpse into the modal universe. It is hoped that this would be of help to a student using the Smullyan and Fitting text to learn forcing.

2 Set Theory, Logic, and the Axiom of Choice

2.1 Equivalent forms of AC

A wide range of mathematical statements have been proven equivalent to the Axiom of Choice. Some of them are stated in the theorem below. Many more can be found in [RR63].

Theorem 2.1. *The following statements are equivalent:*

1. *For any set A there is a function $F : \mathcal{P}(A) \setminus \emptyset \to A$ such that $F(B) \in B$ for every nonempty $B \subseteq A$. We refer to F as a choice function for A.*

2. *For any set A there is a function $G : A \setminus \emptyset \to \cup A$ such that for all $x \in A$, $G(x) \in x$. Some authors use the term "choice function" to refer to this function rather than F as defined in item 1. We use it only for F, never G.*

3. *The Cartesian product of nonempty sets is always nonempty. That is, if H is a function with domain I and if $(\forall i \in I)H(i) \neq \emptyset$, then there is a function f with domain I such that $(\forall i \in I)f(i) \in H(i)$.*

4. *Let $\mathcal{A}$ be a set such that (a) each member of $\mathcal{A}$ is a nonempty set, and (b) any two distinct members of $\mathcal{A}$ are disjoint. Then there exists a set C containing exactly one element from each member of $\mathcal{A}$ (i.e. for each $B \in \mathcal{A}$ the set $C \cap B$ is a singleton $\{x\}$ for some x).*

5. *For any relation R, there is a function $F \subseteq R$ with dom F = dom R.*

6. *(Cardinal comparability) For any sets C and D, either $C \preceq D$ or $D \preceq C$. For any two cardinal numbers κ and λ either $\kappa \leq \lambda$ or $\lambda \leq \kappa$.*

7. *(Zorn's lemma) Let $\mathcal{A}$ be a set such that for every chain $\mathcal{B} \subseteq \mathcal{A}$, we have $\bigcup \mathcal{B} \in \mathcal{A}$. ($\mathcal{B}$ is called a chain iff for any C and D in $\mathcal{B}$, either $C \subseteq D$ or $D \subseteq C$.) Then $\mathcal{A}$ contains an element M (a "maximal" element) such that M is not a subset of any other set in $\mathcal{A}$.*

8. *(Well-ordering theorem) For any set A, there is a well ordering on A.*

The first five of these forms of AC will be referred to by number as they are needed (e.g. "Form 2 of the Axiom of Choice"). Proof of the equivalences can be found in many set theory books. Enderton proves all except item 2 in [End77].

2.2 What the Axiom of Choice means

2.2.1 Looking at AC more closely

In some circumstances, ZF is adequate to provide a choice function. For example, if $A = \omega$, the set of natural numbers, then a choice function on A can be found without assuming the Axiom since it is provable in ZF that ω is well-ordered by the natural order using the properties of ω. Thus each subset of ω has a least element, and we are justified by the replacement axiom schema in assuming the existence of a function $F : \mathcal{P}(\omega) \to \omega$ that maps each set of natural numbers to its least element. F meets the definition of choice function in form 1 of the Axiom.

The Axiom of Choice becomes necessary when (a) a proof requires a set containing a simultaneous choice from infinitely many sets; (b) this infinite collection of sets is not a proper class as defined in section 2.4.1, i.e. there exists a set containing all of them; and (c) the existence of the choice function cannot be proven from ZF together with what is already known about the infinitely many sets. In form 1 of the Axiom, the infinite collection of sets is the powerset of a given (infinite) set. In form 3, it is the set of images of each point in the domain of a relation. In form 4, it is any infinite set of sets. And in form 5, it is an infinite set of mutually disjoint sets.

As a concrete example, we demonstrate the role of the Axiom in Zermelo's proof of the Well-Ordering Theorem. Given a non-empty set M, we wish to prove the existence of a well-ordering on M. Invoking the Axiom of Choice, we let F be a choice function on M. Also, let $s(b, x)$ denote the set of elements that precede b in the well-ordered set x. Zermelo introduced the idea of a γ-set, a well-ordered subset of M such that b is the distinguished element of $M \setminus s(b, M_\gamma)$ for every b in M_γ. That is, $b = F(M \setminus s(b, M_\gamma))$ for all b in M_γ. In this way, F can be thought of as repeatedly (and transfinitely) choosing the next element of the well-ordering under construction. Zermelo went on to show that γ-sets existed; of two γ-sets one was always an initial segment of the other; the set L_γ of all elements of γ-sets is itself a γ-set and includes all the elements of M; and L_γ is a well-ordering of M. The role of the Axiom of Choice is clear here: we need a function that picks a distinguished element out of each subset of M in order to create L_γ, and we need it to exist regardless of the size or other properties of M.

2.2.2 Uses of AC in other areas of mathematics

The everyday objects of mathematics—integers, real numbers, functions, relations, sequences, and so on—can be built up as certain types of sets, and statements about them can be interpreted as

statements about these sets. In this sense, set theory is said to provide a foundation for mathematics. For the working mathematician, these objects are not normally thought of as sets, but when they are defined precisely, the precision is achieved by applying the language of set theory, and when properties of these objects are in dispute, the solution is an appeal to the underlying assumptions of set theory.

With this heavy burden on its shoulders, set theory has a responsibility to adopt an adequate set of axioms to play this foundational role without allowing anything that would lead to a situation like the discovery of Russell's paradox in the early 20th century. The Zermelo-Fraenkel collection of axioms was explicitly formulated to achieve this balance. The addition of the Axiom of Choice caused many complaints that the balance had been lost.

When the debate over the Axiom of Choice began in the early 20th century, the consequences of the Axiom were used to argue for or against its adoption. This way of solving the debate was unsuccessful; there are several results that require AC without which mathematics would be much weaker, and there are several results that follow from AC that have undesirable (or at least counter-intuitive) consequences.

Before giving a brief account of the early debate over AC, we demonstrate some of these results. We begin with two simple, plausible results that require the Axiom.

Theorem 2.2. *Every vector space has a basis.*

We don't prove this in detail here; rather, we focus on the part of the proof that requires AC. First, it is clearly only required in the infinite-dimensional case. For finite-dimensional spaces, a basis can be produced without much work. For infinite-dimensional spaces, Zorn's Lemma is used with $\mathcal{A}$ taken to be the collection of all linearly independent sets of vectors. $\mathcal{A}$ is partially ordered by inclusion, and for any chain $\mathcal{B}$, $\cup\mathcal{B}$ is a linearly independent set and thus a member of $\mathcal{A}$. So by Zorn's Lemma, $\mathcal{A}$ has a maximal element, which can be shown to be a basis. Gregory Moore notes in [Moo82, p.112] that the Axiom has been shown to be essential to this proof; that is, the result cannot be proven without some form of AC.

Theorem 2.3. *The union of a countable family of countable sets is countable.*

Proof. Let $\mathcal{A} = \{A_0, A_1, A_2, \dots\}$ where each A_i is a countable, non-empty set. For each i, there may be many functions from ω onto A_i, so we require the Axiom of Choice to simultaneously choose one such function for each $i \in \omega$. When the choice is made, we can easily use it to create a function

from $\omega \times \omega$ onto $\mathcal{A}$ similar to the one constructed in the usual proof that the rational numbers are countable. $\qquad\square$

Next, we show that not all consequences of AC are as easy to accept.

In analysis, the concept of *measure* extends the ideas of length, area, and volume to more complicated sets. In simple cases, the idea of measure coincides with our intuition—the measure of the real interval $[a, b]$ is $b - a$, for example. It would be nice if a suitable measure could exist for every set of real numbers, but in the presence of the Axiom of Choice, it cannot.

Theorem 2.4. *There is a subset of $\mathbb{R}$ that is not Lebesgue measurable.*

Proof. Partition the interval $[0, 1]$ into equivalence classes by taking $x \equiv y$ if $x - y$ is rational. The Axiom of Choice lets us choose one representative from each equivalence class to create a new set $M \subseteq [0, 1]$. M can be shown to be non-measurable as follows. Assume the contrary. The real line is then partitioned into countably many disjoint sets M_r, one for each rational r, and each of the form $M_r = \{m + r \mid m \in M\}$. Each of these M_r has the same measure as M since measure is translation-invariant. Let $\mu(x)$ denote the measure of set x. It is not possible that $\mu(M) = 0$ since then, by countable additivity of Lebesgue measure we would have $\mu(\mathbb{R}) = 0$. So then $\mu(M) > 0$. But then consider the set S defined by

$$S = \bigcup \{M_r \mid r \text{ rational and } -1 \leq r \leq 1\}.$$

Since $S \subseteq [-1, 2]$, its measure must be finite. But the measure of S is equal to the sum of infinitely many positive quantities (since the measures of the M_r are all equal to the measure of M), which is impossible. $\qquad\square$

This theorem leads to even more alarming consequences: the famous Banach-Tarski paradox says that it is possible in three-dimensional space to decompose a sphere into finitely many non-measurable pieces, move these pieces around using only rigid rotations and translations, and re-assemble them into two spheres, each as big as the original one. This result is in open revolt against our intuition about physical space, but importantly, it doesn't cause any mathematical contradiction.

2.3 Historical development of the Axiom of Choice

Before the 19th century, there was very little need for something like the Axiom of Choice because there had been no generally accepted method of employing infinite collections in a rigorous way.

There had been much debate about the nature of infinity, going at least as far back as Zeno's paradoxes. Aristotle distinguished between potential infinity—the property of being finite but able to grow to any larger finite size—and actual, or completed, infinity, as we now think of sets like ω, the set of all natural numbers. He declared that the number system could be extended as far as necessary for the purposes of mathematics, but needn't ever be considered a completed infinity.

Nevertheless, situations arose where the infinitely large and the infinitely small were essential to some new mathematical technique or proof, and in each of these situations a method of avoiding the actual infinite was sought. The infinitesimal quantities used by Newton and Leibniz in the invention of calculus, for example, were eliminated with the rigorous definition of the limit in the second half of the 19th century.

By that time, infinite sequences and series were so important to analysis and their applications so successful that the underlying assumptions about infinity were starting to be questioned. Georg Cantor, the inventor of set theory, was originally working on the problem of unique representations of functions as infinite trigonometric series when he began to examine the role of infinity in his work. Over the course of the last 30 years of the 19th century, Cantor published a series of papers in which he introduced many of the concepts still central to set theory, such as cardinal numbers, ordinal numbers, the continuum hypothesis, and the proof that there is no bijection between the natural numbers and the real numbers. With Cantor's work, it was finally possible to assume the existence of a completed infinity. However, there was no formal axiomatisation of set theory until 1908, so before that year, the question of what infinite sets were admissible depended on intuition and precedent.

In 1900, part of the first item on David Hilbert's famous list of unsolved problems, delivered to the Second International Congress of Mathematicians in Paris, was the problem of finding a well-ordering for $\mathbb{R}$. Zermelo showed in [Zer04] that the possibility of such a well-ordering follows from the Axiom of Choice. He intended it as a proof that the Well-Ordering Theorem followed from self-evident principles; he argued that AC had been used implicitly by mathematicians for years. In [Moo82], Gregory Moore traces this back as far as 1871 when Eduard Heine published a result about continuity of real-valued functions and assumed an infinite number of simultaneous choices for which no rule was possible.

The assumption of AC in Zermelo's proof was not universally accepted and was roundly criticised for years. This came at a time when the new discipline of set theory was still recovering from the paradoxes of Russell and Burali-Forti and even the underlying logic was not well understood.

Between 1904 and 1908, a debate over the Axiom raged in the academic journals of Europe. Émile Borel rejected the proof because he rejected AC in the case of non-denumerable sets. Henri Lebesgue asked whether it was possible to prove the existence of a mathematical object without explicitly defining it. Moore provides a detailed account of the debate. Finally, in 1908, Zermelo responded to his critics in a new paper, [Zer08], in which he presented a new, more detailed proof of the Well-Ordering Theorem, including his explicit axiomatization of set theory, and responded to the criticisms that had been levelled at his previous proof in the preceding four years. It was this axiomatization which would eventually be extended and improved by A.A. Fraenkel and Thoralf Skolem and become the Zermelo-Fraenkel axioms in common use today, presented formally in the next section.

As time went on and other areas of the foundations of mathematics were stablized, the controversy over the Axiom of Choice subsided. By the 1930's Kurt Gödel and others were proving metamathematical results that made it possible to separate the investigation of the consequences of AC from the question of whether the Axiom was "really true". Mathematicians had the option of working in models of ZF or ZFC; if they chose the latter they were required by convention to explicity state it. After 100 years we still see the Axiom of Choice highlighted in many textbooks where it is used to prove theorems in other areas of mathematics.

From [Moo82]:

> As the controversy gradually drew new mathematicians into it, it became more technical. Little by little, the number of mathematicians who exploited the Axiom as a legitimate mathematical tool increased. Such supporters tended to adopt the cautious but useful position of stating when a proof depended on the Axiom – an attitude which Zermelo himself was to propose in his axiomatization.

2.4 Zermelo-Fraenkel set theory

The modern form of Zermelo's axiomatization is presented here, along with some notes on our approach to the logical language in which it is expressed. The intent is to prevent later confusion that could arise due to the many variations on ZF used by different authors.

2.4.1 Syntax and semantics

Throughout this paper, we assume the standard results of classical first-order logic (such as, for example, Gödel's Completeness Theorem), without proof. We do not present a list of logical con-

nectives or logical axioms here; any standard treatment will do. We make no attempt to restrict ourselves to a minimal set of logical connectives or axioms unless it serves to make a proof or definition shorter. We freely use '∀' and '∃' in their normal senses according to our needs. We pay close attention to this matter only when it becomes necessary. In section 4, modal logic is also used. The relevant differences between classical and modal logic are highlighted in that section.

By the "language of set theory" we mean the formal first-order language consisting of a single two-place predicate symbol (denoted '∈'). Whenever reference is made to a formula or sentence in this paper, it is intended to be a formula or sentence of the language of set theory. When sentences in this language are interpreted in a model, we intend the universe of discourse to be a collection of sets, and the symbol "∈" to represent set membership.

Some of the assertions we make are intended as formal theorems, which can be proven from the axioms of ZF, and some are made in the *metatheory*, that is, they are about the ZF system rather than within it and their proof techniques are different. Where necessary, these are labelled.

We note explicitly that equality is not part of the language; we take it to be a defined notion, as follows.

Definition 2.5. $x = y$ is an abbreviation for $\forall z (x \in z \leftrightarrow y \in z)$.

Theorem 2.6. *In any model of ZF, equality as defined above is reflexive, symmetric, transitive, and substitutive. That is, it behaves like we want equality to behave.*

Proof. Reflexivity, symmetry, and transitivity are obvious from the definition. Substitutivity requires the Power set axiom. Assume that $x = y$. Then if $z = P(y)$, we have $x \in z \leftrightarrow y \in z$ so $x \in P(y) \leftrightarrow y \in P(y)$. But y is always a subset of itself so we must have $x \in P(y)$, i.e. $x \subseteq y$. Similarly for $y \subseteq x$. So x and y are members of the same sets and have the same extension. It follows that they can be substituted in any atomic statement of set theory, and an induction on formula length shows they can be substituted in any other formula. Fraenkel provides a detailed discussion of equality and extensionality in [Fra58]. □

We will use $(\forall x \in \mathbf{M})\phi(x)$ as an abbreviation of $\forall x (M(x) \to \phi(x))$ and $(\exists x \in \mathbf{M})\phi(x)$ as an abbreviation of $\exists x (M(x) \land \phi(X))$. Similarly, we will use $(\forall x \in y)\phi(x)$ as an abbreviation of $\forall x (x \in y \to \phi(x))$ and $(\exists x \in y)\phi(x)$ as an abbreviation of $\exists x (x \in y \land \phi(x))$. We say these quantifiers are *bounded*.

Formulas in the language of set theory can be arranged into a hierarchy according to their complexity, introduced by Azriel Levy in [Lev65]. It will save us some work in both the sections

on consistency and independence to be familiar with three of the simplest types of formulas. A Δ_0 formula is one in which all the quantifiers are bounded. Σ_1 formulas allow an unbounded existential quantifier and Π_1 formulas an unbounded universal quantifier.

Definition 2.7. A Δ_0 *formula* is a formula in the language of set theory defined inductively as follows:

- $x \in y$ is Δ_0.

- If ϕ, ψ are Δ_0, so are $\neg\phi$ and $\phi \wedge \psi$ (and thus also $\phi \vee \psi$, $\phi \rightarrow \psi$, and $\phi \leftrightarrow \psi$).

- If ϕ is Δ_0, so is $\exists x(x \in y \wedge \phi)$ (and thus also $\forall x(x \in y \rightarrow \phi)$).

Definition 2.8. A Σ_1 *formula* is a formula in the language of set theory defined inductively as follows:

- If a formula ϕ is Δ_0, then ϕ is also Σ_1.

- If ϕ, ψ are Σ_1, so are $\phi \wedge \psi$ and $\phi \vee \psi$.

- If ϕ is Σ_1, so is $\exists x\phi$.

- If ϕ is Σ_1, so are $\exists x(x \in y \wedge \phi)$ and $\forall x(x \in y \rightarrow \phi)$.

Definition 2.9. A Π_1 *formula* is a formula in the language of set theory defined inductively as follows:

- If a formula ϕ is Δ_0, then ϕ is also Π_1.

- If ϕ, ψ are Π_1, so are $\phi \wedge \psi$ and $\phi \vee \psi$.

- If ϕ is Π_1, so is $\forall x\phi$.

- If ϕ is Π_1, so are $\exists x(x \in y \wedge \phi)$ and $\forall x(x \in y \rightarrow \phi)$.

A model of Zermelo-Fraenkel set theory is a non-empty structure consisting of a domain and a two-place relation on the domain, (intended to give meaning to the set membership symbol $\in$) in which the axioms listed below are true. They are stated here formally because there are several different but equivalent versions of them in use. At this time we commit to use the versions below for all formal proofs, or at least to start with them and derive equivalent versions if it proves more convenient. The non-logical symbols other than '$\in$' used below are equality ($=$), the empty set ($\emptyset$), the successor of a set ($S(x) = x \cup \{x\}$), and the subset relation ($\subseteq$). Equality has already been defined. Each of the others is a straightforward abbreviation for a formula that only uses $\in$.

1. Extensionality

$$\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y)$$

2. Foundation

$$\forall x [\exists y (y \in x) \rightarrow \exists y (y \in x \wedge \neg \exists z (z \in x \wedge z \in y))]$$

3. Pairing

$$\forall x \forall y \exists z \forall w (w \in z \leftrightarrow (w = x \vee w = y))$$

4. Union

$$\forall x \exists y \forall z (z \in y \leftrightarrow (\exists w (z \in w \wedge w \in x)))$$

5. Replacement (schema). For each formula ϕ with free variables among $x, y, A, w_1, \ldots, w_n$,

$$\forall w (\forall x \forall y \forall z (\phi(x,y) \wedge \phi(x,z) \rightarrow y = z) \rightarrow \exists v \forall y (y \in v \leftrightarrow \exists x (x \in w \wedge \phi(x,y))))$$

6. Infinity

$$\exists x (\emptyset \in x \wedge \forall y \in x (S(y) \in x))$$

7. Power set

$$\forall x \exists y \forall z (z \subseteq x \leftrightarrow z \in y)$$

We assume without justification that there exists at least one model of these axioms (that they are consistent) and that this model is non-empty. We use the bold letter $\mathbf{V}$ to denote an arbitrary non-empty model of ZF. The elements of the domain of $\mathbf{V}$ are formal sets. We know $\mathbf{V}$ is a model of ZF and all the consequences of ZF, but we don't know whether AC holds in $\mathbf{V}$ or not. When the consistency and independence proofs are complete, we will see that we can never know whether AC holds in an arbitrary model of ZF.

We use the word *class* in the metatheory to refer to any collection of sets in $\mathbf{V}$. In particular, every set is a class. We are mostly interested in classes $\mathbf{M}$ that can be defined by a formula ϕ in the language of set theory with one free variable. For example, if ϕ is "$x = x$" then $\mathbf{M} = \mathbf{V}$. If ϕ is "$x \neq x$", then $\mathbf{M} = \emptyset$. And if ϕ is "$\exists y (y \in x)$" then $\mathbf{M}$ is all of $\mathbf{V}$ except the empty set. When it makes sense we allow ourselves to speak about classes in the metalanguage, even though they have no formal existence in $\mathbf{V}$. Also in the metalanguage we speak of operations on classes by analogy with those on sets. For example, we may say that one class is a superclass or subclass of another, or

that a set is a subset of a class. The meaning of each of these statements is obvious and informal; they will never be used in a formal context without explanation or justification.

In what follows, we assume familiarity with the concept of an ordinal number. For the record, we define an ordinal as a set that is transitive and well-ordered by $\in$. In particular, all natural numbers are ordinals, but there also exist infinite ordinal numbers, such as ω. Each ordinal is the set of all preceding ordinals.

Some classes (such as the empty class or the class of all natural numbers) may contain exactly the same elements as some formal set in the universe; classes without this property are called *proper classes*. Proper classes must be carefully distinguished from formal sets. A formula with a free variable may define either one of the two. For example, the formula that says "x is an ordinal" defines the proper class **ON** of all the ordinals. This can be seen to be a proper class since if it were a set it would meet the definition of ordinal number, and would thus be a member of itself, violating the axiom of Foundation. Clearly, any superclass of **ON** is also a proper class. So the formula "x is an ordinal" defines a proper class, but "x is a finite ordinal" does not (it defines ω, the set of natural numbers).

Because we are assuming the Axiom of Foundation, it can be shown that each set in the universe of **V** appears at some level of the following transfinite recursion:

$$
\begin{aligned}
V(0) &= \emptyset \\
V(\alpha + 1) &= \mathcal{P}(V(\alpha)) \\
V(\alpha) &= \bigcup_{\lambda < \alpha} (V(\lambda)) \quad \text{when } \alpha \text{ is a limit ordinal}
\end{aligned}
$$

where $\mathcal{P}(x)$ is the power set of x in **V**. This is referred to as the *cumulative hierarchy* of sets. It is a convenient way to think of the structure of **V**. For example, we can assign a rank to every set based on where it appears in the construction above.

Definition 2.10. The *rank* of a set $x \in \mathbf{V}$ is the least α such that $x \in V(\alpha + 1)$ but $x \notin V(\alpha)$.

The substitutivity of equality immediately gives us:

Theorem 2.11. *The converse of Extensionality, i.e.:*

$$
\forall x \forall y (x = y \rightarrow \forall z (z \in x \leftrightarrow z \in y))
$$

Many texts on set theory include an axiom schema of *separation*; this was the original attempt by Zermelo to exclude Russell's paradox from set theory. In the axiomatization given above, it is not necessary to assume separation since it can be derived from the axiom schema of replacement.

Theorem schema 2.12. *(Separation) For each formula ϕ with one free variable, the following is a theorem of ZF:*

$$\forall x \exists y \forall z (z \in y \leftrightarrow (z \in x \wedge \phi(z)))$$

Proof. See, for example, [TZ70]. The idea is that for any definable subset y of x, a formula can be found that acts as a function from x into y. $\square$

We will need the following important property of sets:

Definition 2.13. A set x is said to be *transitive* if the following statement holds of x:

$$(\forall y \in x)(\forall z \in y)(z \in x)$$

Equivalently, x is transitive if every element of x is also a subset of x. For example, ω is transitive since each natural number contains only other natural numbers.

2.4.2 Relativization and Absolute formulas

Two concepts that will be important in our arguments are *relativization* and *absoluteness*. Often in mathematics, when structures are defined by axioms (groups, vector spaces, etc.), it is useful to know whether or not some proper substructure also satisfies the axioms (subgroups, subspaces, etc.) Something similar can be done with models of ZF. Working within $\mathbf{V}$, we can sometimes identify a subclass $\mathbf{M}$ that is also a model of ZF, i.e. $\mathbf{M} \vDash \phi$ for each axiom ϕ of ZF. We will explore such a class in section 3. But in order to achieve relative consistency results, we need to be able to express this as a truth of $\mathbf{V}$. We know that each axiom ϕ of ZF is true in $\mathbf{V}$, and contains no free variables or constants. Intuitively, if we restrict each of the quantifiers of ϕ to the subclass $\mathbf{M}$ (that is, if we *relativize* the formula to $\mathbf{M}$), the result is a formula that is true in $\mathbf{V}$ but can only make assertions about the members of $\mathbf{M}$. Of course this will have the same truth value in $\mathbf{V}$ as the unrelativized formula has in $\mathbf{M}$. Any formula in the language of set theory with constants in $\mathbf{M}$ can be relativized to $\mathbf{M}$. A more formal definition follows.

Definition 2.14. Let $M(x)$ be a formula in the language of set theory with one free variable, and $\mathbf{M}$ be the class defined by $M(x)$ in $\mathbf{V}$. For any formula ϕ in the language of set theory, we define the *relativization of ϕ to $\mathbf{M}$*, denoted $\phi^{\mathbf{M}}$, as the formula ϕ with all quantifiers ranging over $\mathbf{M}$ instead of $\mathbf{V}$. The definition is by induction on the length of the formula ϕ. Using $\{\exists, \wedge, \neg\}$ as primitive connectives, we define for $x, y \in \mathbf{M}$ (after [Kun80]):

- $(x \in y)^{\mathbf{M}}$ is $(x \in y)$.

- $(\exists x \phi(x))^{\mathbf{M}}$ is $\exists x (M(x) \wedge \phi(x)^{\mathbf{M}})$. As noted in section 2.4.1, we also write $(\exists x \in \mathbf{M}) \phi(x)^{\mathbf{M}}$.

- $(\phi \wedge \psi)^{\mathbf{M}}$ is $(\phi^{\mathbf{M}} \wedge \psi^{\mathbf{M}})$.

- $(\neg \phi)^{\mathbf{M}}$ is $\neg(\phi^{\mathbf{M}})$.

It is easy to verify that the following hold for the other sentential connectives:

- $(\forall x \phi(x))^{\mathbf{M}}$ is $\forall x (M(x) \to \phi(x)^{\mathbf{M}})$. We also write $(\forall x \in \mathbf{M}) \phi(x)^{\mathbf{M}}$.

- $(\phi \vee \psi)^{\mathbf{M}}$ is $(\phi^{\mathbf{M}} \vee \psi^{\mathbf{M}})$.

- $(\phi \to \psi)^{\mathbf{M}}$ is $(\phi^{\mathbf{M}} \to \psi^{\mathbf{M}})$.

Note that if ϕ is a well-formed formula in the language of set theory, then so is $\phi^{\mathbf{M}}$.

Relativization allows a formal sentence to make statements that are restricted to subclasses of any model in which it is interpreted. If A is a set of axioms (sentences) then the completeness theorem says that a statement such as

$$A \vdash \phi^{\mathbf{M}}$$

means that every model, say $\mathbf{A}$, of A is a model of $\phi^{\mathbf{M}}$, which in turn means that the subclass of $\mathbf{A}$ described by $\mathbf{M}$ is a model of the original formula ϕ. That is, if

$$\mathbf{A} \models A$$

then

$$\mathbf{A} \models \phi^{\mathbf{M}}$$

and

$$\mathbf{M} \models \phi,$$

where $\mathbf{M}$ is the collection of elements of $\mathbf{A}$ for which $\mathbf{M}(\mathbf{x})$ holds. We sometimes make statements like "$\mathbf{A}$ thinks $\mathbf{M}$ is a model of ϕ" for $\mathbf{A} \vdash \phi^{\mathbf{M}}$ when they help make a point clear.

It is not generally true that ϕ and $\phi^{\mathbf{M}}$ will have the same truth value in $\mathbf{V}$. For example, if ϕ is of the form $(\exists x)\psi$ then it may be true that there is some $x \in \mathbf{V}$ such that ψ is true in $\mathbf{V}$, but there

is no $x \in \mathbf{M}$ such that ψ is true in $\mathbf{M}$. In other cases, relativizing a formula to a class $\mathbf{M}$ will have no effect on whether or not it holds in $\mathbf{V}$, no matter what values (of $\mathbf{M}$) are substitued for its free variables. In this case we say the formula is *absolute for* $\mathbf{M}$ *in* $\mathbf{V}$. A more formal definition:

Definition 2.15. In the metatheory, let $\mathbf{N}$ be a class and $\mathbf{M}$ be a subclass of $\mathbf{N}$, and let these classes be defined by the formulas $N(x)$ and $M(x)$, respectively. Let ϕ be a formula in the language of set theory whose free variables are among $x_1, \ldots, x_n$. We say that ϕ is *absolute for* $\mathbf{M}$ *in* $\mathbf{N}$ if for any elements $x_1, \ldots, x_n \in \mathbf{M}$, we have $\mathbf{V} \vDash \phi^{\mathbf{M}}(x_1, \ldots, x_n) \leftrightarrow \phi^{\mathbf{N}}(x_1, \ldots, x_n)$. If $\mathbf{N} = \mathbf{V}$, we just say that ϕ *is absolute for* $\mathbf{M}$. If ϕ is absolute for every transitive class, we just say ϕ is *absolute*.

Theorem 2.16. *The formula $x \subseteq y$ is absolute.*

Proof. To see this, expand the formula to eliminate the defined "$\subseteq$" symbol:

$$\forall z(z \in x \to z \in y) \tag{2.1}$$

Then for any transitive class $\mathbf{M}$, and any $x, y \in \mathbf{M}$, the relativized statement is:

$$\forall z \in \mathbf{M}((z \in x)^{\mathbf{M}} \to (z \in y)^{\mathbf{M}})$$

By the definition of relativization for atomic statements, this is equivalent to

$$\forall z \in \mathbf{M}(z \in x \to z \in y)$$

Clearly if formula 2.1 holds for all $z \in \mathbf{V}$, it must be true for any particular collection of z. In the other direction, since $x \in \mathbf{M}$, any $z \in x$ is also in $\mathbf{M}$, so the statement $\forall z(z \in x \to z \in y)$ is vacuously true for $z \notin \mathbf{M}$ and thus true over all of $\mathbf{V}$. $\square$

Theorem 2.17. *Equality is absolute, i.e., if $\mathbf{M}$ is any transitive class then for any $x, y \in \mathbf{M}$,*

$$\mathbf{V} \vDash (\mathbf{x} = \mathbf{y})^{\mathbf{M}} \leftrightarrow (\mathbf{x} = \mathbf{y})$$

Proof. In the presence of the Axiom of Extensionality, this is equivalent to

$$(x \subseteq y \land y \subseteq x)^{\mathbf{M}} \leftrightarrow (x \subseteq y \land y \subseteq x),$$

which holds by theorem 2.16 and the definition of relativization. $\square$

In some cases, we know that a formula is absolute based on its form alone, so we can avoid reasoning through every formula individually.

Theorem 2.18. *All Δ_0 formulas are absolute.*

Proof. We know by definition that $x \in y$ is absolute. Clearly if ϕ and ψ are absolute then so are $\neg\phi$ and $\phi \wedge \psi$. So quantifier-free formulas are absolute. This just leaves the case where we have a bounded existential quantifier in front of a Δ_0 formula. So let ϕ be a Δ_0 formula that may have free variables x, y, and $z_1, \ldots, z_n$. Then for any $y, z_1, \ldots, z_n \in \mathbf{M}$,

$$[\exists x(x \in y \wedge \phi(y, z_1, \ldots, z_n))]^{\mathbf{M}}$$

$$\leftrightarrow \quad (\exists x \in \mathbf{M})[(x \in y)^{\mathbf{M}} \wedge \phi(y, z_1, \ldots, z_n)^{\mathbf{M}}] \quad \text{by definition of relativization}$$

$$\leftrightarrow \quad (\exists x \in \mathbf{M})[x \in y \wedge \phi(y, z_1, \ldots, z_n)] \quad \text{by absoluteness of } x \in y \text{ and absoluteness of } \phi$$

$$\leftrightarrow \quad \exists x(x \in y \wedge \phi(y, z_1, \ldots, z_n)) \quad \text{since } y \in \mathbf{M} \text{ and } \mathbf{M} \text{ is transitive}$$

$\square$

Theorem 2.19. *The formulas representing the following statements are Δ_0 and thus absolute:*

1. $x \in (y \cup z)$	*14. x is a successor ordinal*
2. $x = (y \cup z)$	*15. x is a limit ordinal*
3. $x = (y \cap z)$	*16. x is a natural number*
4. x is empty	*17. x is an unordered pair*
5. $z = \{x, y\}$	*18. x is an ordered pair*
6. $y = \{x\}$	*19. x is a binary relation*
7. $y = S(x)$	*20. x is a function*
8. $\emptyset \in x$	*21. y is the domain of x*
9. $S(y) \in x$	*22. y is an element of the domain of x*
10. $y = \cup x$	*23. y is the range of x*
11. $z = x \times y$	*24. f is a function from x into y*
12. x is transitive	*25. f is a function from x onto y*
13. x is an ordinal	*26. f is a 1-1 function from x onto y*

Proof. The Δ_0 formulas that define these can all be found in [SF96, pp. 145-148]. As an example, we show item 7, "y is the successor of x". This can be characterized by the Δ_0 formula:

$$x \in y \wedge x \subseteq y \wedge (\forall z \in y)(z \in x \vee z = x).^1$$

$\square$

Theorem 2.20. *All Σ_1 formulas ϕ are absolute upwards, i.e. $\mathbf{V} \vDash \phi^{\mathbf{M}} \rightarrow \phi$ for any transitive class* $\mathbf{M}$.

Proof. Of the four criteria for a formula to be Σ_1 given in definition 2.8, this theorem is obvious for the first three, and for $(\exists x \in y)\phi$.

To show $(\forall x \in y)\phi$ is absolute upwards if ϕ is, we write out ϕ with its free variables as $\phi(x, y, z_1, \ldots, z_n)$. Now let $\mathbf{M}$ be a transitive class and let $a, b_1, \ldots, b_n \in \mathbf{M}$ such that $[(\forall x \in a)\phi(x, a, b_1, \ldots, b_n)$ holds. This gives us $(c \in a \rightarrow \phi(c, a, b_1, \ldots, b_n))^{\mathbf{M}}$ for any $c \in M$. But ϕ is absolute upwards so we have $c \in a \rightarrow \phi(c, a, b_1, \ldots, b_n)$. Thus, we have $(\forall x \in a)\phi(x, a, b_1, \ldots, b_n)$. $\square$

Theorem 2.21. *All Π_1 formulas ϕ are absolute downwards, i.e. $\mathbf{V} \vDash \phi \rightarrow \phi^{\mathbf{M}}$ for any transitive class* $\mathbf{M}$.

Proof. Similar to the proof of theorem 2.20. $\square$

Definition 2.22. Call a formula with two free variables $\phi(x, y)$ *function-like* over a transitive class $\mathbf{M}$ if $\mathbf{M} \vDash (\forall x)(\exists y)[\phi(x, y) \wedge (\forall z)(\phi(x, z) \rightarrow (z = y))]$. Thus ϕ acts informally like a total function on $\mathbf{M}$.

Theorem 2.23. *If $\phi(x, y)$ is Σ_1 and function-like over both the universe $\mathbf{V}$ and a transitive subclass $\mathbf{M}$, then $\phi(x, y)$ is absolute for $\mathbf{M}$.*

Proof. We already know that ϕ is absolute upwards since it is Σ_1. So suppose we have $\phi(a, b)$ for some $a, b \in \mathbf{M}$. We need to show that $\phi^{\mathbf{M}}(a, b)$. Since $a \in \mathbf{M}$ and ϕ is function-like over $\mathbf{M}$, there must be some $c \in \mathbf{M}$ such that $\phi^{\mathbf{M}}(a, c)$ and thus $\phi(a, c)$. But then $b = c$. Since equality is absolute, b and c are the same set in $\mathbf{M}$ so we conclude $\phi^{\mathbf{M}}(a, b)$. $\square$

Observe that the axiom of Extensionality holds in any transitive class.

Theorem 2.24. *If $\mathbf{M}$ is a transitive class, then*

$$\mathbf{V} \vDash \forall x \in \mathbf{M} \forall y \in \mathbf{M}(\forall z \in \mathbf{M}(z \in x \leftrightarrow z \in y) \rightarrow (x = y)^{\mathbf{M}})$$

---

[1]The proof of this item given in [SF96, p. 146, item 11] is incorrect.

Proof. Since x and y are both in **M**, theorem 2.17 applies and we can replace $(x = y)^{\mathbf{M}}$ with $(x = y)$.

Fix x and y as sets in **M**, not equal to each other. Since we have extensionality in the universe, there is an element of x not in y or vice versa. Call it a. Then $a \in x$ or $a \in y$. But $x \in \mathbf{M}$ and $y \in \mathbf{M}$ and **M** is transitive, so $a \in \mathbf{M}$ and therefore

$$\exists z \in \mathbf{M}(z \in x \leftrightarrow z \in y).$$

□

Since equality is absolute, we may also conclude that the converse implication holds in **M**.

For classes **Z** defined in a certain way, and for a formula ϕ, the following theorem guarantees the existence of a subclass for which ϕ is absolute in **Z**. It will be used to prove theorem 3.6. The proof can be found in [Kun80, p.137].

Theorem 2.25. *Suppose* **Z** *is a class and, for each ordinal α in* **V** *$Z(\alpha)$ is a set, and assume*

1. $\alpha < \beta \rightarrow Z(\alpha) \subseteq Z(\beta)$.

2. If γ is a limit ordinal, $Z(\gamma) = \bigcup_{\alpha < \gamma} Z(\alpha)$.

3. $\mathbf{Z} = \bigcup_{\alpha} Z(\alpha)$, where the union is over all ordinals α.

Then for any formula ϕ,

$$(\forall \alpha)(\exists \beta > \alpha)(\phi \text{ is absolute for } Z(\beta) \text{ in } \mathbf{Z}).$$

3 Consistency

Here we prove Gödel's 1938 result (see [Göd40]) that the Axiom of Choice is consistent with the axioms of ZF set theory, using a model of ZF called the Constructible Universe. It is formally provable from the axioms of ZF that all the ZFC axioms hold when relativized to the Constructible Universe. The demonstration of a model which ZF 'thinks' is a model of ZFC shows that if ZF is consistent, then so is ZFC. Our main references are [Kun80] and [SF96].

3.1 Relative consistency

If a consistent set of sentences A asserts that $\mathbf{M}$ is a model of ϕ it cannot assert that $\mathbf{M}$ is also a model of anything that contradicts ϕ. Put another way, if A is going to make statements about $\mathbf{M}$, it must accept the consequences of those statements. To see this, assume that

$$A \vdash \phi^{\mathbf{M}}$$

and

$$A \vdash (\neg\phi)^{\mathbf{M}}.$$

Then clearly

$$A \vdash \phi^{\mathbf{M}} \wedge (\neg\phi)^{\mathbf{M}}.$$

By the definition of relativization above, this is equivalent to

$$A \vdash \phi^{\mathbf{M}} \wedge \neg(\phi^{\mathbf{M}}).$$

Thus A is inconsistent, contradicting our original assumption.

Now, consider two finite collections of sentences in the language of set theory, A and B. Let

$$A = \{\phi_1, \ldots, \phi_n\}$$

and

$$B = \{\psi_1, \ldots, \psi_m\}.$$

Then if

$$A \vdash \psi_1^{\mathbf{M}} \wedge \cdots \wedge \psi_m^{\mathbf{M}}$$

for some $\mathbf{M}$, then we have the metamathematical result

$$Con(A) \to Con(B),$$

where "$Con(A)$" is the statement "A is a consistent collection of formulas".

For by the above reasoning, if a contradiction can be derived in B, it can be derived from some subset of the ψ_i, and that list, relativized to $\mathbf{M}$, is a contradiction provable from A.

We now formalize this reasoning. The only piece of the puzzle missing before we can apply this to ZF is to ensure that we can handle an infinite set of axioms. But this is not a problem: since all proofs are finite, any result provable from ZF can be proven from a finite subcollection of the sentences in ZF. Thus, in the reasoning above, if B is an infinite collection of axioms, we can still ensure our result by establishing metamathematically that A is capable of asserting any finite collection of ψ_i, relativized to B. First we prove the following lemma, stated without proof in [Kun80, p.141].

Lemma 3.1. *If* $\phi_1, \phi_2, ..., \phi_n, \psi$ *are sentences in the language of set theory,* $\phi_1, \phi_2, ..., \phi_n \vdash \psi$, *and* $\mathbf{M}$ *is a class, then*

$$\vdash [(\phi_1^M \wedge \phi_2^M \wedge ... \wedge \phi_n^M) \to \psi^M]$$

Proof. The conclusion follows from the Completeness Theorem if we show that every structure in the language is a model of $(\phi_1^M \wedge \phi_2^M \wedge ... \wedge \phi_n^M) \to \psi^M$, i.e. that every model of $\phi_1^M \wedge \phi_2^M \wedge \cdots \wedge \phi_n^M$ is a model of ψ^M. Consider a model $\mathbf{V}$ of $\phi_1^M \wedge \phi_2^M \wedge ... \wedge \phi_n^M$. In this model we have a subclass of the universe, $\mathbf{M}$, which is a model of $\phi_1 \wedge \phi_2 \wedge \cdots \wedge \phi_n$. In $\mathbf{M}$ we may apply the assumption of the theorem and conclude that ψ holds. Thus $\mathbf{V}$ must be a model of ψ^M. $\square$

Theorem 3.2. *Let A and B be possibly infinite collections of sentences in the language of set theory. If we know that there is some nonempty class $\mathbf{M}$ such that for every finite collection $\phi_1, \phi_2, ..., \phi_n$ of sentences in B, we have*

$$A \vdash (\phi_1^{\mathbf{M}} \wedge \phi_2^{\mathbf{M}} \wedge \cdots \wedge \phi_n^{\mathbf{M}}),$$

then B is consistent if A is.

Proof. We prove the contrapositive. If B is inconsistent, then some contradiction, say $\chi \wedge \neg\chi$, can be proven from it. But formal deductions are finite so there exists some finite set of axioms of B, $\phi_1, \phi_2, ..., \phi_n$, from which the contradiction can be proven. That is, we have

$$\phi_1, \phi_2, ..., \phi_n \vdash \chi \wedge \neg\chi.$$

So by Lemma 3.1, we must have

$$\vdash [(\phi_1^M \wedge \phi_2^M \wedge ... \wedge \phi_n^M) \to (\chi \wedge \neg\chi)^M],$$

which is the same as

$$\vdash [(\phi_1^M \wedge \phi_2^M \wedge ... \wedge \phi_n^M) \to \chi^M \wedge \neg(\chi^M)].$$

Clearly then $\phi_1^M, \phi_2^M, \ldots, \phi_n^M$ is an inconsistent collection of sentences. But by assumption we have

$$A \vdash (\phi_1^M \wedge \phi_2^M \wedge ... \wedge \phi_n^M).$$

Thus A is inconsistent. $\qquad\square$

3.2 Constructible sets

We now turn to the task of defining the Constructible Universe, which we will denote as **L**. This will be a slice of **V** that will model ZFC. The construction of **L** is similar to that of the usual class of well-ordered sets, but with a restriction to subsets that are definable from a previously defined set by a formula relativized to that set. Intuitively, **L** is the result of removing as much of **V** as possible without violating the ZF axioms. The elements of **L** are referred to as *constructible sets*.

Let $\mathcal{D}(x)$ be the set of all subsets of x definable by a formula relativized to x. Clearly $\mathcal{D}(x)$ is a set if x is. To see what **L** looks like, we first define the set $L(\alpha)$ for each ordinal α in **V**:

$$L(0) = \emptyset$$
$$L(\alpha + 1) = \mathcal{D}(L(\alpha))$$
$$L(\alpha) = \bigcup_{\lambda < \alpha} (L(\lambda)) \quad \text{when } \alpha \text{ is a limit ordinal}$$

Definition 3.3. **L**$= \bigcup L(\alpha)$, where the union is taken over all ordinals α in **V**.

The reader may wonder why we created the definable powerset operation for this definition rather than using the unrestricted powerset $\mathcal{P}(x)$. Let $V(0) = \emptyset$, and for each successor and limit ordinal α in $\mathbf{V}$, define $V(\alpha)$ as we defined $L(\alpha)$ except using $\mathcal{P}(x)$ instead of $\mathcal{D}(x)$. Then the union of all the $V(\alpha)$ is in fact the entire universe $\mathbf{V}$ (this statement is equivalent to the axiom of Foundation—see [Kun80, p.101]). The important difference between $L(\alpha+1)$ and $V(\alpha+1)$ is that $V(\alpha+1)$ contains all subsets of $V(\alpha)$ that happen to exist in the model $\mathbf{V}$ (though some of these may not exist in other models) whereas $L(\alpha+1)$ contains only those subsets of $L(\alpha)$ that can be defined from $L(\alpha)$ by a formula relativized to $L(\alpha)$. By starting this process at $\alpha = 0$ (which exists in all models of ZF) and relying on formulas to define new sets, we are not doing anything model-specific—that is, the construction of $\mathbf{L}$ given above results in the same class in any model $\mathbf{V}$ of ZF since the $\mathcal{D}(x)$ operation isn't affected by any extraneous objects in $\mathbf{V}$. It is easy to see that at each α we have $L(\alpha) \subseteq V(\alpha)$.

When we use $\mathbf{L}$ to prove the consistency of ZF and AC, we will need a formula to define it. In the next section, we will construct a formula that holds true of a set x if and only if x is in $\mathbf{L}$ as defined above.

Some of the simpler properties of the L-hierarchy can be established by analogy with the usual transfinite construction of the well-founded universe, using powersets rather than $\mathcal{D}$. For example, there is a straightforward notion of rank in $\mathbf{L}$: if $x \in \mathbf{L}$, the $\mathbf{L}$-rank of x is the least ordinal α such that $x \in L(\alpha + 1)$.

In order to show that $\mathbf{L}$ is a model of ZFC, we show separately that it is a model of ZF and that AC holds in $\mathbf{L}$. We first establish three important properties of $\mathbf{L}$ and then use those to prove that $\mathbf{L}$ is a model of ZF. In the next section, we show that $\mathbf{L}$ is also a model of AC.

Lemma 3.4. $\mathbf{L}$ *is transitive.*

Proof. We show that each $L(\alpha)$ is transitive. This establishes that $\mathbf{L}$ is transitive since every set in $\mathbf{L}$ appears in some $L(\alpha)$. The proof is by transfinite induction. Assume that $L(\beta)$ is transitive for all $\beta < \alpha$. If $\alpha = \emptyset$ the result is trivial. If α is a limit ordinal then the result follows because the union of transitive sets is transitive. If α is a successor ordinal, then let β be such that $\alpha = \beta + 1$. Then $L(\beta)$ is transitive and $L(\alpha) = \mathcal{D}(L(\beta))$. Since each element a of a set A is definable from A by the formula $x \in a$ if A is transitive, we have $L(\beta) \subseteq L(\alpha) \subseteq \mathcal{P}(L(\beta))$. So if $x \in y$ and $y \in L(\alpha)$ then $y \subseteq L(\beta)$ and $x \in L(\beta)$. Thus $x \in L(\alpha)$ and $L(\alpha)$ is transitive. $\square$

It follows from transitivity of $\mathbf{L}$ that if $x \in L(\alpha)$ then $x \in L(\beta)$ for all ordinals $\beta > \alpha$. That is,

$L(\alpha) \subseteq L(\beta)$. Also, $L(\alpha) \in L(\beta)$ since a set is definable from itself by the formula $x = x$.

Lemma 3.5. *For any set $x \in \mathbf{V}$ of constructible sets, there is a set $y \in \mathbf{L}$ such that $x \subseteq y$ is true in $\mathbf{V}$. We say $\mathbf{L}$ is "almost universal."*

Proof. Let x be any set in $\mathbf{V}$, all of whose members are in $\mathbf{L}$. Then for some $L(\alpha)$, $x \subseteq L(\alpha)$. And $L(\alpha) \in L(\alpha + 1)$ so $L(\alpha) \in \mathbf{L}$. Thus any subset of $\mathbf{L}$ is included in an element of $\mathbf{L}$. $\square$

Lemma 3.6. *The separation theorems (theorem schema 2.12) hold when relativized to $\mathbf{L}$.*

Proof. This should be easy to believe, since $\mathbf{L}$ was clearly constructed to satisfy separation. For each formula $\phi(x, z, w_1, \ldots, w_n)$ with its free variables among those displayed, we need to show that the relativized separation theorem holds as follows:

$$\forall z, w_1, \ldots, w_n \in \mathbf{L} \exists y \in \mathbf{L} \forall x \in \mathbf{L}(x \in y \leftrightarrow x \in z \wedge \phi^{\mathbf{L}}(x, z, w_1, \ldots, w_n)).$$

So given $z, w_1, \ldots, w_n \in \mathbf{L}$, we need to show that the set

$$A = \{x \in z : \phi^{\mathbf{L}}(x, z, w_1, \ldots, w_n)\}.$$

is in $\mathbf{L}$. We don't need to explicitly say that $x \in \mathbf{L}$ since $x \in z$ and $z \in \mathbf{L}$ and $\mathbf{L}$ is transitive.

So fix $z, w_1, \ldots, w_n \in \mathbf{L}$. Then fix α such that $z, w_1, \ldots, w_n \in L(\alpha)$. Any α greater than the highest of the $\mathbf{L}$-ranks of the $z, w_1, \ldots, w_n$ will do. We know that for all ordinals $\gamma > \alpha$, we have $z, w_1, \ldots, w_n \in L(\gamma)$.

If we can find an ordinal $\beta \geq \alpha$ such that ϕ is absolute for $L(\beta)$ in $\mathbf{L}$, then A is equal to the set

$$B = \{x \in z : \phi^{L(\beta)}(x, z, w_1, \ldots, w_n)\}.$$

Since $\mathbf{L}$ meets the criteria for theorem 2.25, this ordinal β must exist.

Since $z \in L(\beta)$ and $L(\beta)$ is transitive (see the proof of theorem 3.4), B is equal to

$$C = \{x \in L(\beta) : x \in z \wedge \phi^{L(\beta)}(x, z, w_1, \ldots, w_n)\}.$$

This modified formula is still absolute for $L(\beta)$ since x and z are in $L(\beta)$ and the conjunction of two absolute formulas is absolute. By the definition of relativization C is equal to

$$D = \{x \in L(\beta) : (x \in z \wedge \phi(x, z, w_1, \ldots, w_n))^{L(\beta)}\}.$$

This set is defined from $L(\beta)$ by a formula relativized to $L(\beta)$ and is thus a member of $\mathcal{D}(L(\beta)) = L(\beta + 1)$ and thus of **L**. Since $A = B = C = D$ and $D \in \mathbf{L}$, we have $A \in \mathbf{L}$ as required.

$\square$

These three lemmas are enough for us to prove that:

Theorem 3.7. *All of the axioms of ZF hold when relativized to* **L**. *Equivalently, ZF says that* **L** *is a model of ZF. Equivalently, every model of ZF, assuming there are some, includes a class* **L** *which is a submodel of ZF, though we don't know whether it is properly included or not.*

(After this proof, the only missing piece is to show that ZF proves that AC holds in **L**, for then we will be able to invoke theorem 3.2 with $A = $ ZF and $B = $ ZFC to prove the consistency of AC with ZF.)

Proof. We work in ZF and show that each axiom of ZF holds in **L**.

Extensionality

L is transitive so Extensionality holds by theorem 2.24.

Foundation

The Axiom of Foundation can be written in the following form, which is Π_1 and thus absolute downwards:

$$\forall x[\exists y \in x(x = x) \to (\exists y \in x(\neg \exists z \in x(z \in y)))]$$

Pairing

Fix x and y in **L**. Then $\{x, y\}$ is a set in the universe and $\{x, y\} \subseteq \mathbf{L}$. So $\exists z \in \mathbf{L}\{x, y\} \subseteq z$. So $\exists z \in \mathbf{L}(x \in z \wedge y \in z)$. The result follows by separation.

Union

Since **L** is transitive, all elements of elements of **L** are in **L**. Thus if $x \in \mathbf{L}$ then $\cup x \subseteq \mathbf{L}$ so $\exists y \in \mathbf{L}(\cup x \subseteq y)$. The result follows by separation.

Replacement

Let w be any set in $\mathbf{L}$, ϕ be any formula with two free variables, and assume that $\phi^{\mathbf{L}}$ acts as a function in $\mathbf{L}$, i.e.

$$\forall x \in \mathbf{L} \forall y \in \mathbf{L} \forall z \in \mathbf{L}(\phi^{\mathbf{L}}(x,y) \wedge \phi^{\mathbf{L}}(x,z) \rightarrow y = z)$$

Then we need to show that those members of $\mathbf{L}$ that occur in the range of $\phi^{\mathbf{L}}$ form a set in $\mathbf{L}$, i.e.

$$\exists v \in \mathbf{L} \forall y \in \mathbf{L}(y \in v \leftrightarrow \exists x \in \mathbf{L}(x \in w \wedge \phi^{\mathbf{L}}(x,y)))$$

Applying replacement in the universe to the same w and the functional formula

$$y \in \mathbf{L} \wedge \phi^{\mathbf{L}}(x,y),$$

we are assured

$$\exists v \forall y(y \in v \leftrightarrow \exists x(x \in w \wedge y \in \mathbf{L} \wedge \phi^{\mathbf{L}}(x,y))).$$

This implies

$$\exists v \forall y \in \mathbf{L}(y \in v \leftrightarrow \exists x \in \mathbf{L}(x \in w \wedge \phi^{\mathbf{L}}(x,y))),$$

which is almost what we want. It only remains to show that v is in $\mathbf{L}$. But this is easy: clearly $v \subseteq \mathbf{L}$. Then by lemmas 3.5 and 3.6, $v \in \mathbf{L}$.

Infinity

We know $\emptyset \in \mathbf{L}$ (it is in $L(1)$). Now, for some $k \in \mathbf{L}$, since we have already proven that pairing follows from the premises of the theorem, we can take $\{k,k\} = \{k\} \in \mathbf{L}$ and $\{k,\{k\}\} \in \mathbf{L}$. Then from the Union axiom we get $k \cup \{k\} \in \mathbf{L}$. So $\mathbf{L}$ contains the empty set and is closed under taking successors. Thus $\omega \subseteq \mathbf{L}$ and so ω is included in some element of $\mathbf{L}$.

Powerset

For an element x of $\mathbf{L}$, we need to show that there is an element y of $\mathbf{L}$ containing all subsets of x that are in $\mathbf{L}$. In the universe we have $\mathcal{P}(x)$. We want to show that the set of elements of $\mathcal{P}(x)$ that are also elements of $\mathbf{L}$ is a set in $\mathbf{L}$, i.e. we want to show that $\mathcal{P}(x) \cap \mathbf{L} \in \mathbf{L}$. Clearly $\mathcal{P}(x) \cap \mathbf{L} \subseteq \mathbf{L}$. So since $\mathbf{L}$ is almost universal, we get $\exists Y \in \mathbf{L}(\mathcal{P}(x) \cap \mathbf{L} \subseteq Y)$. The result follows by lemma 3.6. $\square$

A class such as **L** which is constructed as a subclass of a larger model and is also a model itself, is referred to as an *inner model*.

Thus we have proven that **V** has a submodel **L** that is also a model of ZF. For any finite collection F of the axioms of ZF, we can conclude formally that the conjunction of the axioms in F, relativized to **L**, is provable from ZF.

3.3 A model in which AC holds

In this section we complete the proof of the consistency of AC with ZF by showing that **L** is a model of AC. We do this in a slightly indirect way. The direct way would be to express AC as a formal first-order sentence and show that its relativization to **L** is provable in ZF. Since any formal first-order formula that represents the Axiom of Choice would be too long and complex to work with directly, we find another simpler statement X and prove $X \to AC$ and $X^{\mathbf{L}}$. The result is the same but the work is easier. The statement that implies AC is, informally, "every set is constructible", denoted $\mathbf{V} = \mathbf{L}$ and referred to as the Axiom of Constructibility. Our main task will be to show that the following two sentences are theorems of ZF:

- $(\mathbf{V} = \mathbf{L})^{\mathbf{L}}$; and

- $\mathbf{V} = \mathbf{L} \to AC$

From this we conclude $AC^{\mathbf{L}}$ by the definition of relativization and the fact that **L** is a model of ZF. This will complete the consistency proof.

3.3.1 The Axiom of Constructibility is true in **L**

In order to formally prove $(\mathbf{V} = \mathbf{L})^{\mathbf{L}}$, we need to replace references to the class **L** with a formula that defines it. Let us refer to the formula as "Const(x)". Then the Axiom of Constructibility will have the form $(\forall x)\text{Const}(x)$. **L** is a proper class (it contains all the ordinals) so the formula for "x is constructible" will have to be more clever than just $x \in \mathbf{L}$.

The process we use to create the formula Const(x) is as follows: we will map each logical symbol $(\to, \wedge, \forall, \text{etc})$ to a set in **L** and then use set operations to define formulas that say things like "x is the code of a formula with constants in a", "y is the code of a formula that defines the set x over a", and "x is a definable subset of y". The key result is the last one. From a formula that defines the relation $x \in \mathcal{D}(y)$, we can create a formula that says $x \in L(\alpha)$ and one that says $(\exists \alpha)(x \in L(\alpha))$. This last formula says "x is constructible".

Most of the rest of this section is dedicated to defining the formula $\text{Const}(x)$. It is essential that this formula be absolute for **L** because then **V** and **L** agree on which sets are constructible—exactly those sets that are in **L**. In this case **L** is a model of $(\forall x)\text{Const}(x)$ so $[(\forall x)\text{Const}(x)]^{\mathbf{L}}$ is a theorem of ZF, as required.

There are several ways to build this formula. Kenneth Kunen in [Kun80] defines for each $n \in \omega$ the set of all definable n-tuples and uses it to provide a formula that says "x is definable from A". Thomas Jech in [Jec73] and [Jec03] extracts definable subsets using closure under "Gödel operations" (so called because this was the technique Gödel used in [Göd40]). The particular scheme given below is due to Raymond Smullyan and Melvin Fitting with only minor modifications (and many proof omissions) from [SF96, pp. 166-171].

First, in order for formulas to talk about other formulas, we must have a way of representing them as formal objects, that is, sets. So we define the following encoding of formulas in the language of set theory.

Definition 3.8. We define an *individual* as a variable v_i or a set a. The *code* of v_i, denoted $\ulcorner v_i \urcorner$, will be the ordered pair $\langle 0, i \rangle$. The code of a set a, denoted $\ulcorner a \urcorner$, will be $\langle 1, a \rangle$.

Using these codes as building blocks, we now define the code of a formula ϕ inductively.

1. If ϕ is atomic then it is of the form $m_1 \in m_2$ for two individuals m_1 and m_2, and its code is $\langle \ulcorner m_1 \urcorner, \ulcorner m_2 \urcorner, 0 \rangle$.

2. $\ulcorner \neg \phi \urcorner = \langle \ulcorner \phi \urcorner, 1 \rangle$. If x is the code for a formula ϕ, we refer to the code $\langle x, 1 \rangle$ as $Neg(x)$.

3. $\ulcorner \phi \wedge \psi \urcorner = \langle \ulcorner \phi \urcorner, \ulcorner \psi \urcorner, 2 \rangle$. If x and y are the codes for formulas ϕ and ψ respectively, we refer to the code $\langle x, y, 2 \rangle$ as $Con(x, y)$.

4. $\ulcorner (\exists v_i)\phi \urcorner = \langle i, \ulcorner \phi \urcorner, 3 \rangle$. If x is the code of a formula of the form $\exists v_i \phi$, we refer to the code $\langle i, x, 3 \rangle$ as $E_i(x)$.

Definition 3.9. By $\Pi(a)$ we mean the set of all elements of a; all elements $Neg(x)$ such that $x \in a$; all elements $Con(x, y)$ such that $x, y \in a$; and all elements $E_i(x)$ such that $x \in a$ and $i \in \omega$.

Theorem 3.10. *The relation $y = \Pi(x)$ can be represented by a Σ_1 formula.*

Proof. $\Pi(x) = x \cup (x \times \{1\}) \cup ((x \times x) \times \{2\}) \cup ((\omega \times x) \times \{3\})$. The relation $z = x \times y$ is Δ_0 (theorem 2.19) so the relation $w = (x \times y) \times z$ is Σ_1 since it is equivalent to $\exists v(x \times y = v \wedge v \times z = w)$. $\qquad\Box$

Definition 3.11. For any set a let $\mathcal{E}_0^a$ be the set of all codes of atomic formulas with constants in a. Let $\mathcal{E}_{n+1}^a = \Pi(\mathcal{E}_n^a)$. Finally, let $\mathcal{E}^a$ be the union of all the $\mathcal{E}_i^a$, for $i \in \omega$. $\mathcal{E}^a$ is the set of codes of all formulas with constants in a. We refer to these formulas as "a-formulas".

Theorem 3.12. *The relations $y = \mathcal{E}^a$ and $y \in \mathcal{E}^a$ can be represented by Σ_1 formulas.*

Next we need a formal way to distinguish between codes of formulas that have different numbers of free variables. Formulas with no free variables are sentences and can be assigned a truth value. Formulas with one free variable can be used to define classes.

Definition 3.13. The type $t(x)$ of the code $x = \ulcorner \phi \urcorner$ of a formula ϕ is the set of indices of its free variables. If a and b are constants,

- $t(\ulcorner x_i \in x_j \urcorner) = \{i, j\}$

- $t(\ulcorner x_i \in a \urcorner) = \{i\}$

- $t(\ulcorner a \in x_i \urcorner) = \{i\}$

- $t(\ulcorner a \in b \urcorner) = \emptyset$

- $t(\ulcorner \neg \phi \urcorner) = t(\ulcorner \phi \urcorner)$

- $t(\ulcorner \phi \wedge \psi \urcorner) = t(\ulcorner \phi \urcorner) \cup t(\ulcorner \psi \urcorner)$

- $t(\ulcorner (\exists v_i)\phi \urcorner) = t(\ulcorner \phi \urcorner) \setminus \{i\}$

Clearly $t(x)$ is always a finite set of natural numbers.

Theorem 3.14. *The relation $z \in \mathcal{E}^a \wedge y = t(z)$, i.e. "$z$ is the code of a formula of type y with constants in a", can be represented by a Σ_1 formula.*

Definition 3.15. For a set a and a finite set of natural numbers b, we write $\mathcal{E}_b(a)$ for the set of all codes of a-formulas of type b. We write $\overline{\mathcal{E}_b}(a)$ for the set of all codes of a-formulas not of type b.

Theorem 3.16. *The following relations can all be represented by Σ_1 formulas:*

- $y \in \mathcal{E}_b(a)$

- $y \in \overline{\mathcal{E}_b}(a)$

- $y = \mathcal{E}_b(a)$

Now we know that an a-formula ϕ is a sentence if $\ulcorner\phi\urcorner \in \mathcal{E}_\emptyset(a)$, and ϕ has one free variable if $\ulcorner\phi\urcorner \in \mathcal{E}_{\{i\}}(a)$ for some $i \in \omega$. These formulas are both Σ_1.

Theorem 3.17. *There is a Σ_1 formula that represents the 5-place relation $Sub_a(x, i, y) = z$. This means "z is the code of the a-formula that results from replacing all free occurences of v_i in x with y."*

Definition 3.18. $Sub(x, i, y) = z$ is defined as $(\exists a)(x \in \mathcal{E}^a \wedge Sub_a(x, i, y) = z)$. Clearly this relation is Σ_1.

Definition 3.19. We write $\phi(a) = b$ to mean that ϕ is a formula with one free variable and b is the result of substituting a for the free variable. We write $x(y) = z$ to mean that x is the code of a formula ϕ with one free variable and $\phi(y) = z$, i.e. $z = Sub(\phi, v_i, y)$ where v_i is the free variable of ϕ.

Theorem 3.20. *The relation $x(y) = z$ can be represented by a Σ_1 formula.*

Proof. The formula can be written

$$(\exists a)(\exists i \in \omega)(x \in \mathcal{E}_{\{i\}}(a) \wedge Sub(x, i, y) = z).$$

$\square$

We have sets that represent formulas of set theory. We can identify the ones whose constants are in a given set a. We can identify their free variables. We can substitute for their free variables. Next we need to introduce the concept of a truth valuation for these (codes of) formulas. To do this, we encode the concept of truth of a sentence ϕ over a set a. For $a_1, a_2 \in a$ and ϕ, ψ closed a-formulas, the definition of truth over a is:

1. $(a_1 \in a_2$ is true over $a) \leftrightarrow (a_1$ is a member of $a_2)$.

2. $\neg\phi$ is true over $a \leftrightarrow \phi$ is not true over a.

3. $\phi \wedge \psi$ is true over $a \leftrightarrow \phi$ and ψ are both true over a.

4. $(\exists v_i)\phi(v_i)$ is true over $a \leftrightarrow$ for some $a_1 \in a$, $\phi(a_1)$ is true over a.

Given the code of an a-sentence ϕ, we can define a formula that represents the relation "ϕ is true over a". Informally, we can think of defining a function $V_a(x)$ for each a where x is the code of an

a-sentence. The range of V_a is $\{0,1\}$. V_a is defined using the definition of truth above. For example, V_a satisfies the condition:

$$(\forall a_1 \in a)(\forall a_2 \in a)[(a_1 \in a_2 \wedge V_a(\langle \ulcorner a_1 \urcorner, \ulcorner a_2 \urcorner, 0 \rangle) = 1) \vee (a_1 \notin a_2 \wedge V_a(\langle \ulcorner a_1 \urcorner, \ulcorner a_2 \urcorner, 0 \rangle) = 0)].$$

This ensures that the value of $V_a(x)$ is correct when x is the code of an atomic formula. By encoding the other three truth criteria given above, we define a formula that represents the relation $V_a(x) = y$ where x is the code of an a-sentence and $y \in \{0,1\}$. As usual, the formula is Σ_1.

Definition 3.21. Let $Def(x, y, a)$ mean that x is the code of an a-formula ϕ with one free variable such that ϕ defines y over a, i.e. y is the set of all $b \in a$ such that $\phi(b)$ is true over a.

Based on previous results, it can be shown that the formula representing the relation $Def(x, y, a)$ is Σ_1. But our definable powerset operation, used in the construction of **L**, is not much different than $Def(x, y, a)$. In fact, we have:

Theorem 3.22. *The relation $y = \mathcal{D}(x)$ can be represented by a Σ_1 formula.*

Proof. 1. The relation $y \in \mathcal{D}(x)$ is represented by the Σ_1 formula

$$(\exists z)(z \text{ is a formula with one free variable whose constants are in } x \wedge Def(z, y, x))$$

2. The relation $y \subseteq \mathcal{D}(x)$ is represented by the Σ_1 formula $(\forall z \in y)(z \in \mathcal{D}(x))$.

3. The relation $\mathcal{D}(x) \subseteq y$ is represented by the Σ_1 formula

$$(\forall z)(z \text{ is an } x\text{-formula with one free variable} \rightarrow (\exists w \in y)(Def(z, w, x)))$$

Note that this is in the form $(\forall x \in y)(\exists w \ldots)$ since the $\forall z$ is bounded by the Σ_1 condition that z be an x-formula with one free variable.

4. The relation $y = \mathcal{D}(x)$ is represented by the Σ_1 formula $y \subseteq \mathcal{D}(x) \wedge \mathcal{D}(x) \subseteq y$.

$\square$

Theorem 3.23. *The relation $x = L(\alpha)$ can be represented by a Σ_1 formula. We write this formula as $\mathcal{M}(\alpha, x)$. This formula is absolute.*

Proof. The formula can be written in the following Σ_1 form:

$$\alpha \text{ is an ordinal} \wedge (\exists f)[f \text{ is a function} \wedge (\alpha + 1 \text{ is the domain of } f) \wedge f(\alpha) = x$$

$$\wedge f(0) = \emptyset \wedge (\forall \beta \in \alpha + 1)(f(\beta + 1) = \mathcal{D}(f \restriction \beta))$$

$$\wedge (\forall \beta \in \alpha + 1)(\beta \text{ is a limit ordinal} \rightarrow f(\beta) = \cup f''(\beta))]$$

We will show that $\mathcal{M}(\alpha, x)$ is absolute for **L** by extending it to a formula that is function-like on **L** and appealing to theorem 2.23. Let $\mathcal{N}(x, y)$ be the formula

$$[\text{Ordinal}(x) \wedge \mathcal{M}(x, y)] \vee [\neg\text{Ordinal}(x) \wedge y = 0]$$

$\mathcal{N}$ can be thought of informally as a function that maps x to $L(x)$ if x is an ordinal, or to 0 if not. Clearly $\mathcal{N}$ is function-like on the universe **V**. Since **L** is a model of ZF and contains 0, all the ordinals, and all the $L(\alpha)$, $\mathcal{N}$ is also function-like on **L**. Thus, $\mathcal{N}$ is absolute for **L** by theorem 2.23. And since $\mathcal{M}$ is equivalent to

$$\text{Ordinal}(x) \wedge \mathcal{N}(x, y)$$

it is also absolute for **L**. $\qquad\square$

Let $\mathcal{L}(x, \alpha)$ be the Σ_1 formula $(\exists z)(\mathcal{M}(\alpha, z) \wedge x \in z)$, i.e. $x \in L(\alpha)$. Then $\mathcal{L}(x, \alpha)$ is absolute upwards.

Theorem 3.24. $\mathcal{L}(x, \alpha)$ *is absolute for* **L**.

Proof. We only need to show that it is absolute downwards. So suppose $\mathcal{L}(x, \alpha)$ is true in **V** for some $x, \alpha \in$ **L**. We want to show $\mathcal{L}^{\mathbf{L}}(x, \alpha)$ is also true in **V**. We know α is an ordinal (in both **V** and **L**). Let $c = L(\alpha)$. So we have $\mathbf{V} \vDash \mathcal{M}(\alpha, c)$. Then $x \in c$ and $x, c \in$ **L**. So $\mathbf{L} \vDash x \in c$. $\mathcal{M}$ is absolute for **L** so we have $\mathbf{L} \vDash \mathcal{M}(\alpha, c)$ and thus $\mathbf{L} \vDash [\mathcal{M}(\alpha, c) \wedge x \in c]$. Since $c \in$ **L**, we have $\mathbf{L} \vDash [(\exists z)(\mathcal{M}(\alpha, z) \wedge x \in z)]$. This is the same as $\mathbf{L} \vDash \mathcal{L}(x, \alpha)$. So we have $\mathbf{V} \vDash \mathcal{L}^{\mathbf{L}}(x, \alpha)$

So $\mathcal{L}(x, \alpha)$ is absolute for **L** and we have $(x \in L(\alpha)) \leftrightarrow (x \in L(\alpha))^{\mathbf{L}}$ in **V**. $\qquad\square$

Definition 3.25. We denote by $\text{Const}(x)$ the formula $(\exists \alpha)\mathcal{L}(x, \alpha)$. Clearly $\text{Const}(x)$ is Σ_1. In **V**, it defines the property of being a constructible set.

Theorem 3.26. *The formula* $\text{Const}(x)$ *is absolute for* **L**.

Proof. Since it is a Σ_1 formula, it is already absolute upwards. So assume that for some a we have $\mathbf{V} \vDash \mathrm{Const}(a)$, i.e. $\mathbf{V}$ thinks that $a \in \mathbf{L}$. We must show $\mathbf{V} \vDash \mathrm{Const}^{\mathbf{L}}(a)$. We know $\mathbf{V} \vDash a \in L(\alpha)$ for some ordinal α (which is an ordinal in both $\mathbf{V}$ and $\mathbf{L}$), so we have $\mathbf{V} \vDash \mathcal{L}(a, \alpha)$ and thus $\mathbf{V} \vDash \mathcal{L}^{\mathbf{L}}(a, \alpha)$ by the previous theorem. Since $\alpha \in \mathbf{L}$, this means $\mathbf{V} \vDash (\exists \beta)\mathcal{L}^{\mathbf{L}}(a, \beta)$, which is the same as $\mathbf{V} \vDash \mathrm{Const}^{\mathbf{L}}(a)$. $\qquad \square$

Theorem 3.27. *The Axiom of Constructibility is true in $\mathbf{L}$. That is, $ZF \vdash (\forall x \in \mathbf{L}\, Const^{\mathbf{L}}(x))$.*

Proof. For any model $\mathbf{V}$ of ZF, we define $\mathbf{L}$ as above, and let x be in $\mathbf{L}$. Since $x \in \mathbf{L}$, we have $\mathbf{V} \vDash \mathrm{Const}(x)$ and thus $\mathbf{V} \vDash \mathrm{Const}^{\mathbf{L}}(x)$ by theorem 3.26. $\qquad \square$

3.3.2 The Axiom of Constructibility implies the Axiom of Choice

Here it is shown that the Axiom of Choice holds in $\mathbf{L}$ in the form of the Well-Ordering Theorem. This can be achieved by well-ordering $\mathbf{L}$. The logic is: $\mathbf{L}$ can be well-ordered, so all the elements of $\mathbf{L}$ can be well-ordered (since $\mathbf{L}$ is transitive) so if every set is in $\mathbf{L}$ then every set can be well-ordered. Thus the Axiom of Constructibility implies the Axiom of Choice if we can exhibit a well-ordering of $\mathbf{L}$. We start by showing that if we have a well-ordering of $L(\alpha)$ then we can well-order $L(\alpha + 1)$. Of course this means defining a well-ordering of $\mathcal{D}(A)$ from a well-ordering of A. We follow the treatment given in [SF96, pp.173-175].

We create the well-ordering of $\mathcal{D}(A)$ from the well-ordering of A by constructing a well-ordering $W_{\mathcal{E}^A}$ of the set of codes of formulas with constants in A (the set $\mathcal{E}^A$ as in definition 3.11) and setting $x < y$ for $x, y \in \mathcal{D}(A)$ if the formula that defines x over A comes before the formula that defines y over A according to $W_{\mathcal{E}^A}$. In order to well-order codes of formulas with constants in A, we need to be able to extend a well-ordering of A to a well ordering of a set containing all the elements of A and all ordered pairs of elements of A. This is because formula codes are ordered triples $\langle a, b, c \rangle$, which are defines as ordered pairs $\langle \langle a, b \rangle, c \rangle$.

Definition 3.28. For a set A, let A^* be defined from A as the countable union $A_0, A_1, \ldots, A_n, \ldots$ where

1. $A_0 = A$

2. $A_{n+1} = A_n \cup (A_n \times A_n)$

So for any $a, b \in A^*$, we have $\langle a, b \rangle \in A^*$.

Theorem 3.29. *Given a well-ordering W of A, there is a well-ordering W^* of A^*.*

Proof. First we show how to extend a well-ordering of A to a well-ordering of $A \cup (A \times A)$. So, given W, let W' be a well-ordering of $A \cup (A \times A)$ as follows:

- The elements of A are ordered by W.

- All elements of A come before all elements of $(A \times A)$ that are not in A.

- For any two distinct elements of $A \times A$ that are not in A, since they are ordered pairs of elements of A, we order them lexicographically using W, that is, if $x, y \in (A \times A) \setminus A$ and $x = \langle x_1, x_2 \rangle$ and $y = \langle y_1, y_2 \rangle$ then set $x < y$ in W' if $x_1 < y_1$ in W or $x_1 = y_1$ and $x_2 < y_2$ in W.

Clearly $W \subseteq W'$. Now, since $A^* = \bigcup(A_0, A_1, \ldots, A_n, \ldots)$ and $A_0 = A$, we can set $W_0 = W$ and use the W' construction to construct a sequence of well-orderings $\{W_0, W_1, W_2, \ldots, W_n, \ldots\}$ where for each $n \in \omega$, W_n is a well-ordering of A_n and $W_n \subseteq W_{n+1}$.

Let $W^* = \bigcup(W_0, W_1, \ldots, W_n, \ldots)$. W^* is a well-ordering of A^* since any two elements of A^* appear in some A_n which is well-ordered by W_n. $\qquad \square$

Theorem 3.30. *For any set A, let $B = A \cup \omega$. Then the set $\mathcal{E}^A \subseteq B^*$.*

Proof. If $x \in \mathcal{E}^A$ then x is the code of formula with constants in A. The code of a formula is constructed by a finite number of applications of definition 3.8. By setting $B_0 = B = A \cup \omega$ and applying definition 3.28, we get all codes of atomic statements with constants in A appearing in level B_2. If the code for ϕ is in level B_n, the code for $\neg \phi$ is in B_{n+1} and the code for $(\exists v_i)\phi$ is in B_{n+2}. If the code for ψ is also in B_n, the code for $\phi \wedge \psi$ is in B_{n+2}. $\qquad \square$

Now, let us fix A for the rest of the discussion, well-ordered by W, and build a well-ordering of $\mathcal{D}(A)$. Let $B = A \cup \omega$ and let the asterisk * operate as in definition 3.28. First we build a well-ordering of B^*.

Theorem 3.31. *There exists a well-ordering of B^*.*

Proof. We well-order B and apply theorem 3.29. Since $B = A \cup \omega$ and A and ω are already well-ordered, we well order B by putting all elements of A before any elements of $\omega \setminus A$. Elements of A are ordered by W. Elements of ω not in A are ordered by the usual order on the natural numbers. $\qquad \square$

By theorem 3.30, $\mathcal{E}^A \subseteq B^*$ so we can well-order $\mathcal{E}^A$ by the well-ordering of B^*. Thus the set of codes of formulas with constants in A can be well-ordered. Fix a well-ordering of $\mathcal{E}^A$ for the rest of the discussion, and call it $W_{\mathcal{E}^A}$. Now, finally, we can well-order $\mathcal{D}(A)$.

Theorem 3.32. *There exists a well-ordering of $\mathcal{D}(A)$.*

Proof. Every element $x \in \mathcal{D}(A)$ is defined by at least one formula ϕ such that $\ulcorner \phi \urcorner \in \mathcal{E}^A$. Let x_0 denote the least element of the well-ordered set $\mathcal{E}^A$ (using the well-ordering $W_{\mathcal{E}^A}$) such that x_0 is the code of a formula that defines x over A. The well-ordering $W_{\mathcal{D}}$ of $\mathcal{D}(A)$ is:

- All elements of $\mathcal{D}(A) \cap A$ are put before all elements of $\mathcal{D}(A) \setminus A$.

- For any two elements of A, we order them using W.

- For any two elements $x, y \in \mathcal{D}(A) \setminus A$, we put $x < y$ in $W_{\mathcal{D}}$ if $x_0 < y_0$ in $W_{\mathcal{E}^A}$.

$\square$

Clearly $W_{\mathcal{D}}$ is an extension of W. Now, consider the following sequence of well-orderings:

1. $W_0 = \emptyset$

2. $W_{\alpha+1} = (W_\alpha)_{\mathcal{D}}$

3. $W_\lambda = \cup_{\alpha < \lambda}$ (λ a limit ordinal)

Definition 3.33. Let $W_{\mathbf{L}} = \bigcup(W_\alpha)$ for all ordinals α.

Theorem 3.34. *Each W_α is a well-ordering of $L(\alpha)$. $W_{\mathbf{L}}$ is a well-ordering of $\mathbf{L}$.*

Proof. The second statement follows immediately from the first and the fact that for two ordinals α and β such that $\alpha < \beta$ we have $W_\alpha \subseteq W_\beta$. The first is a simple transfinite induction. Let $x, y \in L(\alpha)$ for some ordinal α. We know $\alpha \neq 0$ since $L(0)$ is empty. Then α is a successor or a limit. If it is a limit then $x, y \in L(\beta)$ for some successor $\beta < \alpha$. If it is a successor then let $\alpha = \beta + 1$. We know $L(\beta)$ is well-ordered by W_β. Then by definition, $W_\alpha = (W_\beta)_{\mathcal{D}}$, the well-ordering of $\mathcal{D}(L(\beta))$. But $\mathcal{D}(L(\beta)) = L(\alpha)$. $\square$

Theorem 3.35. *The Axiom of Constructibility implies the Axiom of Choice.*

Proof. Assume the Axiom of Constructibility. Then every set x is constructible (an absolute notion). Since $\mathbf{L}$ is transitive, all the elements of x are also constructible. By theorem 3.34, every set of constructible sets can be well-ordered. So the arbitrary set x can be well-ordered, a statement which implies the Axiom of Choice by theorem 2.1. $\square$

The following is immediate from theorems 3.27 and 3.35.

Theorem 3.36. *The Axiom of Choice is true in* **L**.

Thus we have proven:

Theorem 3.37. $Con(ZF) \rightarrow Con(ZFC)$.

Proof. Theorem 3.2 with A=ZF and B=ZFC.   □

Thus AC does not introduce any contradictions into ZF set theory.

4 Independence

In this section it is shown that

$$Con(ZF) \rightarrow Con(ZF + \neg AC).$$

Note that this is a statement of relative consistency, just as in section 3. The statement "*AC* is independent of ZF" means that the relative consistency of both *AC* and $\neg AC$ with ZF has been proven—the truth value of *AC* is left undetermined by the axioms.

Unfortunately, the method of inner models cannot be used for this relative consistency result. It can never supply a model of ZF in which AC fails. This follows from the fact that any model of ZF contains all the constructible sets (i.e. $\mathbf{L} \subseteq \mathbf{V}$ for any model $\mathbf{V}$ of ZF). To see why this rules out inner models as a method of establishing independence, suppose we had an inner model $\mathbf{M}$ of $\mathbf{V}$ in which AC fails. Then $AC^{\mathbf{L}}$ and $(\neg AC)^{\mathbf{M}}$ would both be provable in ZF. Since we are assuming that ZF is consistent, it follows that we could then prove $\mathbf{L} \neq \mathbf{M}$, and since $\mathbf{L} \subseteq \mathbf{M}$, this would give us a proof in ZF that there is a non-constructible set, but this is impossible because by theorem 3.27, ZF is consistent with the statement that every set is constructible.

A model of $ZF + \neg AC$ will not come that easily. In fact, contrary to our proof of consistency, and to the standard proof of independence due to Paul Cohen, we don't explicitly construct a model of $ZF + \neg AC$ to establish the desired result (in the classical sense, at least). Rather, we use a modified version of the standard approach that establishes the result by the argument outlined below. The essence of the argument is due to Cohen. This particular approach is due to Raymond Smullyan and Melvin Fitting, published in their 1996 book [SF96, Chapters 16-21]. This will be the plan for the rest of section 4.

1. Assume ZF is consistent. Then $ZF + \neg AC$ is inconsistent if and only if AC is provable from ZF if and only if there exists a finite collection of ZF axioms $ZF_1, \ldots, ZF_n$ such that $ZF_1 \wedge ZF_2 \wedge \cdots \wedge ZF_n \rightarrow AC$ is a valid sentence (i.e. not falsifiable under any interpretation);

2. We introduce a collection of models using modal rather than classical logic, and observe that a classical sentence ϕ in the language of set theory can be translated into a modal sentence $[\![\phi]\!]$ such that ϕ is classically valid if and only if $[\![\phi]\!]$ is valid in all such modal models (modal validity will be defined);

3. Finally, we construct a modal model in which $[\![\phi]\!]$ is valid for every axiom ϕ of ZF, and $[\![\neg AC]\!]$ is

also valid. We will show that in such a model it is impossible for $[\![ZF_1 \wedge ZF_2 \wedge \cdots \wedge ZF_n \to AC]\!]$ to be valid for any collection $ZF_1 \ldots ZF_n$ of ZF axioms.

From items 2 and 3, we conclude that $ZF_1 \wedge ZF_2 \wedge \cdots \wedge ZF_n \to AC$ is not classically valid for any finite collection $ZF_1 \ldots ZF_n$ of ZF axioms. Then, by item 1, $ZF + \neg AC$ is consistent if ZF is. This is the desired result.

Item 1 is clear. Items 2 and 3 will require more work. In order to achieve them, the technique called *forcing*, invented by Cohen and first published in [Coh63] in 1963, is used. Forcing can be approached in several ways, and many mathematicians have made important contributions to the theory since Cohen's original publication, such as the method of Boolean-valued models, due to Dana Scott and Robert Solovay (See [Bel85]), or the modal approach used here, due mostly to Fitting.

The modal approach may seem different than the standard treatment at first, but is really the same idea in disguise. Many of the constructions of standard forcing are recognizable in the argument below. The major difference is that Cohen used forcing to construct a classical model of ZF in which AC fails, but the modal approach is less constructive—we only advance our argument as far as necessary to establish the independence result. When our work is done, we will not be able to produce a classical set with no choice function without doing a little more work. But it establishes the independence result nevertheless, by the reasoning in items 1 to 3 above.

4.1 Classical and Modal Logic

The "models" used in section 4 are not actually models of ZF in the traditional sense. They are classes of sets that we construct in our usual set-theoretic universe **V**, but we construct them in such a way that they can be treated as models when the underlying logic is changed from classical to modal, as described below. This requires formal (classical) definitions of the domain of the modal model and how to interpret formulas of set theory in it.

In the form we will be using, modal logic extends the syntax of classical logic by introducing symbols for the modalities of necessity and possibility. In addition to a domain, a modal model also consists of a collection of possible worlds. Just what these are will become clear later. To keep things simple, in our modal models, the domain is shared by all possible worlds, so we have a clear idea of how to interpret quantifiers, but each possible world has its own idea of what formal statements are true or false of the domain. For example, if x and y are two elements of the domain, the statement "$x \in y$" may be true at one possible world but not at another. It will require some work to define

the domain and to interpret the $\in$ symbol so that these models look "set-theoretic", but when this is done, it will provide the types of models promised in item 2 of the plan for this section.

The statement "ϕ is true at world p" is denoted:

$$p \vDash \phi.$$

The possible worlds are related to each other by a two-place *accessibility relation R*. World q is said to be *accessible from world p* if pRq holds. In the language of modal logic, there is an extra logical symbol: the box, denoting necessity ($\Box$). For a formula ϕ (classical or modal), if the modal formula $\Box\phi$ is true at world p, this is interpreted to mean that ϕ is true at every world accessible from p. There is a dual notion of possibility, $\Diamond$, defined as $\neg\Box\neg\phi$. Thus if $\Diamond\phi$ is true at world p, then ϕ is true at some world accessible from p.

4.1.1 The Modal Logic S4

The modal models needed to establish the independence of AC have accessibility relations that are reflexive and transitive (the reasons for this will be explored in the next section). The modal logic with this property is called "S4". Since S4 is the only modal logic used in this paper, we use "S4 model" and "modal model" interchangeably. Most modal logic texts provide standard results about S4, some of which we will use here without proof.

It is important to note that when we refer to an S4 model, we are actually referring to a class of sets that exist in some classical model of ZF, such as **V** or **L**. The classes we will use for our S4 models in the next section can be represented by formulas, but we can't expect the axioms of ZF to hold and AC to fail when relativized to these classes. This would contradict the fact that inner models can't be used to prove the independence of AC. Rather, these classes are useful if we view them as modal models and operate on them with modal logic. But this is really just smoke and mirrors. The modal operations we use are abbreviations for longer classical formulas that do the hard work of checking, for example, that a boxed formula really is true at all accessible worlds. For this reason, the collection of possible worlds and the accessibility relation must be sets in the classical model from which we start. Let us begin to formalize these notions.

As noted above, the modal language of set theory is the same as the classical language of set theory, but with the extra symbol $\Box$, read "necessarily". We often write $\Diamond\phi$ as an abbreviation for $\neg\Box\neg\phi$ and we read "possibly phi".

In S4 we have the following modal axioms. The first three describe our use of modal logic; the last two describe the particular modal logic S4.

1. **Axiom K** $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$.

2. **Barcan formula** $(\forall x)\Box\phi(x) \to \Box(\forall x)\phi(x)$

3. **Necessitation rule** Conclude $\Box\phi$ from ϕ (since a proof of ϕ means ϕ holds in all worlds).

4. **Reflexivity axiom** $\Box\phi \to \phi$.

5. **Transitivity axiom** $\Box\phi \to \Box\Box\phi$.

A more detailed treatment of syntax and proofs in S4 can be found in [BdRV01]. We will be interested mostly in S4 models.

Definition 4.1. Given a model $\mathbf{M}$ of ZF, an *S4 model* $\mathfrak{M}$ on $\mathbf{M}$ is a structure $\langle \mathcal{G}, \mathcal{R}, \mathcal{D}, \mathcal{V} \rangle$ where $\mathcal{G}$ is a non-empty set of possible worlds, $\mathcal{R}$ is a reflexive, transitive binary relation on $\mathcal{G}$, $\mathcal{D}$ is a non-empty class (which may or may not be a set), and $\mathcal{V}$ is a mapping from $\mathcal{G}$ to closed atomic formulas in the modal language of set theory with constants from $\mathcal{D}$. $\mathcal{V}$ assigns formulas to worlds in which they are true. Note that $\mathcal{D}$ and $\mathcal{V}$ may be proper classes, in which case the model $\mathfrak{M}$ will also be a proper class inside $\mathbf{M}$. We only require $\mathcal{G}$ and $\mathcal{R}$ to be sets in $\mathbf{M}$.

Definition 4.2. Let $\mathfrak{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{D}, \mathcal{V} \rangle$ be an S4 model. Let $p \in \mathcal{G}$ and let ϕ and ψ be formulas in the modal language of set theory. We write $p \models_{\mathfrak{M}} \phi$ or simply $p \models \phi$ to mean that formula ϕ is true at world p, as follows:

1. if ϕ is closed atomic and $\phi \in \mathcal{V}(p)$ then $p \models \phi$;

2. if ϕ is closed then $p \models \neg\phi$ if and only if $p \not\models \phi$.

3. if ϕ and ψ are closed then $p \models (\phi \wedge \psi)$ if and only if $p \models \phi$ and $p \models \psi$.

4. if $\phi(x)$ has one free variable, then $p \models (\exists x)\phi(x)$ if and only if there is some $d \in \mathcal{D}$ such that $p \models \phi(d)$.

5. if ϕ is closed then $p \models \Box\phi$ if and only if for every world $q \in \mathcal{G}$ with $p\mathcal{R}q$, $q \models \phi$.

Definition 4.3. A modal formula ϕ is *valid in an S4 model* $\mathfrak{M}$ if it is true at every world of $\mathfrak{M}$, i.e. $p \models \phi$ for all possible worlds p in $\mathfrak{M}$. A modal formula is *S4-valid* if it is valid in all S4 models.

Later we will need the following properties of $\Box$ and $\Diamond$ in the presence of the S4 axioms. The proofs of them, which follow from the axioms above, can be found in [SF96, pp. 192, 197], and are standard results about S4.

1. If ϕ is valid in a modal model so is $\Box\phi$.

2. If $\phi \rightarrow \psi$ is valid in a modal model so is $\Box\phi \rightarrow \Box\psi$.

3. If $\phi \rightarrow \psi$ is valid in a modal model so is $\Diamond\phi \rightarrow \Diamond\psi$.

4. $\Box\Diamond\Box\Diamond\phi \leftrightarrow \Box\Diamond\phi$ is S4-valid.

5. $\Box\Diamond(\Box\phi \rightarrow \Box\Diamond\psi) \rightarrow (\Box\phi \rightarrow \Box\Diamond\psi)$ is S4-valid.

6. $(\Box\Diamond\phi \wedge \Box\Diamond\psi) \leftrightarrow \Box\Diamond(\Box\Diamond\phi \wedge \Box\Diamond\psi)$ is S4-valid.

4.1.2 Why S4 is Useful

As we will see in the next section, there is a large collection of S4 models in which the axioms of ZF can be embedded in a meaningful way. In these models, using the notation of definition 4.1, once we specify $\mathbf{M}$, $\mathcal{G}$, and $\mathcal{R}$, the construction of $\mathcal{D}$ and $\mathcal{V}$ is determined (this construction will be given in section 4.2). The key to using S4 for relative consistency results in set theory is to choose these items in a way that results in a modal model with the desired properties. Then, using an important property of the embedding of classical formulas into modal logic (see theorem 4.5 below), the relative consistency result is established. We now turn our attention to the embedding of classical logic into S4.

Intuitively, we will think of a possible world p in an S4 model $\mathfrak{M}$ as an approximation to a classical model, and we will be interested in finding a better approximation by investigating other worlds accessible from p that provide more information, in a certain sense. If the classical model we have in mind is a model of the statement ϕ then certainly we are interested in possible worlds where ϕ holds. It would be even better if we could identify those worlds where ϕ will continue to hold as we move to other accessible worlds. This suggests that we should look at worlds of the modal model in which $\Box\phi$ holds.

In S4 models, there are two nice consequences of $p \vDash \Box\phi$. First, by reflexivity, $p \vDash \phi$. And second, by transitivity, for any possible world p' such that pRp', we have not only $p' \vDash \phi$ but also $p' \vDash \Box\phi$. If we think of the worlds accessible from p as providing more information, then as we move from p to another accessible world, we can be sure we have guaranteed the truth of $\Box\phi$ permanently.

Unfortunately, this comes at too high a price. The problem is that in order to establish our results later we will need a similar kind of permanency when $p \nvDash \Box\phi$, that is, we want to be able to guarantee that negative information will also continue to be true as we move through possible worlds. But if we focus on formulas of the form $\Box\phi$, then the falsity of such a formula at world p only gives us $p \vDash \neg\Box\phi$, or equivalently $p \vDash \Diamond\neg\phi$. To improve this, we further restrict our attention to formulas of the form $\Box\Diamond\phi$. This gives us accumulation of both positive and negative information, if we pay close attention. If $p \vDash \Box\Diamond\phi$, then ϕ may not be true at every world accessible from p, but in the event that it becomes false at a world, it will be recoverable (that is, if $p \vDash \Box\Diamond\phi$, and pRp' and $p' \nvDash \phi$, we still have $p' \vDash \Diamond\phi$. In fact, because the accessibility relation is transitive, we also have $p' \vDash \Box\Diamond\phi$). So ϕ can never become false and remain false by following accessible worlds if $\Box\Diamond\phi$ has already been encountered. And as for negative information, if we have $p \nvDash \Box\Diamond\phi$ then this means $p \vDash \neg\Box\Diamond\phi$, which is equivalent to $p \vDash \Diamond\Box\neg\phi$. In English: if $p \nvDash \Box\Diamond\phi$, there is a world accessible from p where ϕ is false and will remain false at all worlds accessible from it. We can do even better by adding $\Box\Diamond$ in front of ϕ and all subformulas of ϕ. This is the embedding of classical formulas into modal logic that was promised in item 2 in the plan for this section. The embedding is defined inductively on the length of the formula ϕ.

Definition 4.4. ([SF96, p. 197]) For a first-order formula ϕ, we define $[\![\phi]\!]$ recursively as follows:

1. If ϕ is atomic, $[\![\phi]\!] = \Box\Diamond\phi$.

2. If ϕ is of the form $\neg\psi$, $[\![\phi]\!] = \Box\Diamond\neg[\![\psi]\!]$.

3. If ϕ is of the form $\psi \land \sigma$, $[\![\phi]\!] = \Box\Diamond([\![\phi]\!] \land [\![\sigma]\!])$.

4. If ϕ is of the form $(\exists x)\psi$, $[\![\phi]\!] = \Box\Diamond(\exists x)[\![\psi]\!]$.

We will refer to $[\![\phi]\!]$ as the *S4 translate* of ϕ. The effect is to put the symbols "$\Box\Diamond$" in front of ϕ and every subformula of ϕ. Note that this translate can operate on either modal or classical formulas.

Clearly every classical formula ϕ has an S4 translate $[\![\phi]\!]$, but not every modal formula is of the form $[\![\phi]\!]$ for some classical ϕ. Also, every modal formula of the form $[\![\phi]\!]$ begins with $\Box\Diamond$, so the discussion above applies. Let us say that p *forces* ϕ if $p \vDash [\![\phi]\!]$. Then the discussion in section 4.1 could be summed up as saying (1) if p forces ϕ then any world accessible from p also forces ϕ; and (2) if p doesn't force ϕ then there is a world q accessible from p such that q forces $\neg\phi$.

Now, we show that the S4-translate mapping provides the embedding promised.

Transfer theorem 4.5. *([SF96, p. 200]) If ϕ is a sentence in a language of classical logic (possibly the language of set theory), then ϕ is valid if and only if $[\![\phi]\!]$ is S4 valid.*

Proof. If $[\![\phi]\!]$ is true in all S4 models, ϕ is classically valid. Assume that ϕ fails in some classical model. Then we can build a one-world S4 model in which $[\![\phi]\!]$ fails. Call the world p. Take the domain of the modal model to be the same as the domain of the classical model in which ϕ fails. And set $p \models \psi$ if ψ is true in the classical model. Then clearly the formulas ψ, $\Diamond\psi$, and $\Box\psi$ are all equivalent at p. It follows easily that for any ψ, ψ is true in the classical model if and only if $p \models [\![\psi]\!]$. Thus since ψ fails classically, it fails in this S4 model and is not S4-valid.

Conversely, if ϕ is classically valid, then ϕ is provable from some axiomatisation of classical logic. So we pick such an axiomatisation, prove that the S4 translates of the axioms are all true in every S4 model, and show that the rules of inference hold in S4. From there, it is easy (but tedious) to show that $[\![\phi]\!]$ must be S4-valid. $\qquad\square$

Lemma 4.6. *The following are S4-valid:*

1. $[\![\phi]\!] \leftrightarrow \Box[\![\phi]\!]$

2. $[\![\phi]\!] \leftrightarrow \Box\Diamond[\![\phi]\!]$

3. $[\![\neg\phi]\!] \rightarrow \neg[\![\phi]\!]$

4. $[\![\phi \lor \psi]\!] \rightarrow ([\![\phi]\!] \lor [\![\psi]\!])$

5. $[\![\phi \land \psi]\!] \leftrightarrow ([\![\phi]\!] \land [\![\psi]\!])$

Item 1 of lemma 4.6 will be very useful. It says that if p forces ϕ and p' is accessible from p then p' forces ϕ.

Theorem 4.7.   *1. In an S4 model, $p \models [\![\phi_1 \land \phi_2 \land \cdots \land \phi_n]\!]$ if and only if*

       $p \models ([\![\phi_1]\!] \land [\![\phi_2]\!] \land \cdots \land [\![\phi_n]\!])$

   2. In an S4 model, if $p \models [\![\phi \rightarrow \psi]\!]$ then $p \models ([\![\phi]\!] \rightarrow [\![\psi]\!])$

Proof. The proofs follow from Lemma 4.6 above. $\qquad\square$

So our goal is to create an S4 model in which (1) for each ZF axiom ϕ, the S4 translate $[\![\phi]\!]$ is valid; and (2) the translate of the axiom of choice fails in at least one world. In this case the formula in item 3 of the plan for this section can't be valid for any finite set of ϕ_i, by the following

argument: if there were a finite set $\phi_1, \phi_2, \ldots, \phi_n$ of ZF axioms such that $[\![\phi_1 \wedge \phi_2 \wedge \cdots \wedge \phi_n \rightarrow AC]\!]$ were valid, then by theorem 4.7, we would have the validity of $([\![\phi_1]\!] \wedge [\![\phi_2]\!] \wedge \cdots \wedge [\![\phi_n]\!]) \rightarrow [\![AC]\!]$, which is impossible since we have a model with a world where each $[\![\phi_i]\!]$ holds but $[\![AC]\!]$ fails.

One more property of $[\![\ ]\!]$ will be proven here. Several other properties used in our proofs are given in the appendix.

Lemma 4.8. *If ϕ and ψ are classical formulas in the language of set theory and $\phi \leftrightarrow \psi$ is classically valid, then $[\![\phi]\!] \leftrightarrow [\![\psi]\!]$ is S4-valid.*

Proof. If $\phi \leftrightarrow \psi$ is classically valid, then by theorem 4.5, $[\![\phi \leftrightarrow \psi]\!]$ is S4-valid. Using theorem 4.7,

$$
\begin{aligned}
[\![\phi \leftrightarrow \psi]\!] &= [\![(\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)]\!] \\
&\leftrightarrow [\![\phi \rightarrow \psi]\!] \wedge [\![\psi \rightarrow \phi]\!] \\
&\rightarrow ([\![\phi]\!] \rightarrow [\![\psi]\!]) \wedge ([\![\psi]\!] \rightarrow [\![\phi]\!]) \\
&= [\![\phi]\!] \leftrightarrow [\![\psi]\!]
\end{aligned}
$$

$\square$

4.2 S4-ZF models

We now move on to item 3 in the plan for this section, starting with the goal of building S4 models in which the translate of each axiom of ZF holds at every world. We will describe a class of S4 models called S4-ZF models and show that they meet this goal. The main challenges will be the careful definition of the domain of the models, and the interpretation of the $\in$ symbol.

In order to take advantage of the multiple worlds of modal models, we want to define the domain and the $\in$ relation in such a way that different worlds see the same set as containing different elements. On one hand, this is simple enough – membership is the only atomic statement in set theory so the only way for different worlds to make different statements true is to see sets as having different elements; if they didn't we really wouldn't be gaining anything by using modal logic. But on the other hand, this is a radical departure from classical set theory, where a set is simply determined by its extension—the elements that it contains—and the extension of a set is completely determined by the $\in$ relation. In an S4 model with worlds p and q and elements x and y in the domain, we want to allow the possibility that $p \vDash x \in y$ and $q \vDash x \notin y$ simultaneously. To achieve this, it is (thankfully) not necessary to create a separate domain for each possible world. We just need a clever way of constructing the common domain and defining the $\in$ relation.

Here is the technique: in our modal models, a set x is determined by its members (so that these models still feel intuitively like set theory), but each member of x is associated with a possible world. So, for example, a classical set may be represented as $\{a, b, c\}$ where a, b, and c are sets of rank lower than x. But in an S4-ZF model with worlds p and q, a set x might look more like $\{\langle p, a \rangle, \langle p, b \rangle, \langle q, b \rangle, \langle q, c \rangle\}$, where a, b, and c are sets in the S4-ZF model that also associate possible worlds with previously-constructed elements. In this way, a set in an S4-ZF model specifies how it looks to each possible world. In this example, we will have $p \models a \in x$, $p \models b \in x$, $q \models b \in x$, and $q \models c \in x$.

We start with a transitive inner model $\mathbf{M}$ of ZF (such as $\mathbf{L}$, for example) and create the domain for our S4 model so that every element of the domain is an element of $\mathbf{M}$. The domain will thus be a subclass of $\mathbf{M}$.

The formal construction of this domain (from $\mathbf{M}$) is by transfinite recursion:

Definition 4.9. Let $\mathbf{M}$ be a transitive inner model of ZF and $G \in \mathbf{M}$ be a set of possible worlds. For each ordinal α in $\mathbf{M}$ we define a set R_α as follows:

1. $R_0 = \emptyset$.

2. $R_{\alpha+1}$ is the set of all subsets (in $\mathbf{M}$) of $G \times R_\alpha$.

3. For a limit ordinal λ, $R_\lambda = \cup_{\alpha < \lambda} R_\alpha$.

We define the domain D as $\cup_\alpha R_\alpha$, where α ranges over all ordinals in $\mathbf{M}$.

An element of D is a set of ordered pairs, where in each pair the first item is a possible world (an element of G) and the second item is a previously constructed element of the domain. Note that there is a notion of rank in this construction: the rank of $f \in D$ is the least α such that $f \in R_{\alpha+1}$ but $f \notin R_\alpha$.

In the presence of only a single possible world, the construction above is isomorphic to the standard construction of the well-ordered universe of classical set theory. With multiple possible worlds, we still have some sets that look the same at all possible worlds, but most appear different to different possible worlds. This will be the key to the flexibility we require to force a given statement to be true or false at certain worlds of a modal model.

We need to know how to interpret the membership relation $\in$ in an S4-ZF model. Clearly the classical interpretation won't do. Our goal is to define it in such a way that our models still feel like models of set theory; we don't want to let go of the intuition of set membership and inclusion. For

one thing, we want the $\in$ relation to be *extensional*. For the classical case, Kunen defines extensional relations as follows in [Kun80, p.105]:

Definition 4.10. A relation R is extensional on a (classical) class **N** if

$$\forall x \forall y \forall z (zRx \leftrightarrow zRy) \rightarrow x = y) \text{ is true in } \mathbf{N}.$$

Contrapositively, if x and y are not equal ($\mathbf{N} \vDash \neg(x = y)$), we can use R to distinguish between them. By theorem 2.24, $\in$ is extensional on any transitive class. The axiom of extensionality says that $\in$ is extensional on the entire universe. In our modal models we will want $\in$ to be extensional at each possible world.

Before continuing with our search for $\in$, we pause to consider some of the consequences of working in a modal model where the same set appears different in different worlds. Specifically, we must ask what makes it the "same" set. The concepts of identity and equality need to be carefully examined in modal models. In classical set theory, this is easy: they coincide. If we have a notion of equality "=" (see section 2.4.1 for a discussion) that is reflexive, transitive, symmetric, and substitutive in a classical model of ZF, then "$a = b$" means we are using a and b as two different names for the same element of the domain. They are *identical*. However, in our modal models, we haven't excluded the possibility that we could have two non-identical elements of the domain that just happen to "look equal" at a possible world p, in the sense that p sees exactly the same elements in both sets, i.e. $p \vDash \forall z (z \in x \leftrightarrow z \in y)$ but some other world q may see a difference between the sets, i.e. $q \vDash \exists z (z \in x \wedge z \notin y)$. In this case, x and y are not identical elements of the universe; if they were, they would look equal at every possible world. Let us use the symbol "$\approx$" for equality in S4 models. The most we can say about sets x and y if $p \vDash x \approx y$ is that p cannot distinguish between x and y. We will define $\approx$ more precisely below. The main points here are (1) equality and identity no longer necessarily coincide; and (2) it is possible for two sets to be distinguishable at one world but not at another.

Now, back to the modal membership relation. There is one natural interpretation of $\in$ that suggests itself based on the form of the elements of S4-ZF models. We could interpret it as ε , where ε is defined as follows.

Definition 4.11. For a possible world p and $f, g \in D$, define $p \vDash f \varepsilon g$ (read f is *immediately* in g at p) when $\langle p, f \rangle \in g$.

This definition will be a useful building block but doesn't meet all of our requirements for an

interpretation of $\in$. Let us momentarily assume that we have defined $\in$ as ε and explore why it doesn't work. Smullyan and Fitting's example is the following set-up:

$$f \;=\; \{\langle q,a\rangle\}$$
$$g \;=\; \{\langle p,a\rangle, \langle q,a\rangle\}$$
$$h \;=\; \{\langle q,f\rangle\}$$

World q cannot distinguish between f and g using this definition of $\in$—it thinks they both consist of the single element a—so in order for $\in$ to be extensional we must have $q \vDash f \approx g$. But we have $q \vDash f \in h$ and $q \nvDash g \in h$, so something is wrong, since we want to think of $\approx$ as a (substitutive) notion of equality. The solution will be simply to allow $q \vDash g \in h$ by virtue of the fact that at q, g is indistinguishable from something that is immediately in h, namely f. Extending membership from "immediate members" to "things indistinguishable from immediate members" will give us everything we want in an S4-ZF model, though proving this will take some work.

So ε is rejected as a definition of $\in$ and we turn to an example that will help us build a better one. However, ε will frequently be useful in the coming development.

Example 4.12. Consider a model with worlds p and q and elements f, g, c, and d as defined below. The model also contains elements a, b, x, and y, but their exact definitions are not important to the example.

$$f \;=\; \{\langle p,a\rangle, \langle p,b\rangle, \langle p,c\rangle, \langle q,a\rangle, \langle q,b\rangle, \langle q,c\rangle\}$$
$$g \;=\; \{\langle p,a\rangle, \langle p,b\rangle, \langle p,d\rangle, \langle q,a\rangle, \langle q,b\rangle, \langle q,d\rangle\}$$
$$c \;=\; \{\langle p,x\rangle, \langle q,y\rangle\}$$
$$d \;=\; \{\langle p,x\rangle\}$$

World p should think $c \approx d$ since it sees exactly the same elements in both sets. For this simple comparison, we don't need much more than ε and the usual notion of extensionality to see that the two sets are equal. Also, world q should not think that $c \approx d$ since it thinks d is empty and c is not. Now, p and q both agree that f consists of a, b, and c, and g consists of a, b, and d. But since p cannot distinguish between c and d, it also cannot distinguish between f and g—when we formally define $\approx$ and examine its properties we will see that there is no formal sentence in the classical language of set theory at world p that holds of f but not of g. On the other hand, q can distinguish

between f and g since it thinks, for example, that there is an element of g that is empty, but there is no such element of f. Note that when we define $\approx$ formally and show that it is substitutive, we will also be able to conclude that $p \vDash c \in g$ and $p \vDash d \in f$.

It should be clear (or at least plausible) by now that in a world of an S4-ZF model we should define $a \in b$ if $a \; \varepsilon \; b$ or if $a \approx c$ for some $c \; \varepsilon \; b$. This depends on the definition of $\approx$, which will be specially constructed so that $p \vDash a \approx b$ if p can't make any formal distinction between a and b. Since this is set theory, we want the equality of a and b to depend on their extensions, so we adopt a definition of $\approx$ that looks into a and b to see if it can match the elements of a to those of b. This usually means looking into the elements of elements of a and the elements of elements of b, and into their elements, and so on (in the same way that checking equality in example 4.12 requires examining whether c equals d in order to see if f equals g). The logic that makes this work is a modified version of the axiom of extensionality. Rather than "everything in a is in b and everything in b is in a", we will say that a equals b at world p if "everything in a is equal to something in b and everything in b is equal to something in a", where all these statements are interpreted at world p. This sounds circular, but an inductive definition will make it precise. Then, armed with this notion of equality, we can define $x \in y$ at p to mean "x is equal at p to something that is immediately in y at p".

Now, let us make these notions precise, and prove that they fulfill the original goal of creating S4-ZF models. We define $\approx$ first, and we approach it by an inductive definition of $\approx_\alpha$ for all ordinals α. The idea will be that if $f \approx_\alpha g$ for some ordinal, then $f \approx_\beta g$ for all ordinals $\beta > \alpha$ and in this case we say $f \approx g$.

This definition of $\approx$ below introduces one new idea to the previous discussion: it only allows cases where the equality is *forced*. It is possible that f and g could satisfy all our criteria for equality at a world p but fail to meet them at any world accessible from p; however, for our purposes, this will never be useful. We are only interested in cases where the equality is stable, which is the reason for the $[\![\ ]\!]$ brackets in the definition.

Definition 4.13. For a possible world p and $f, g \in D$,

1. $p \nvDash (f \approx_0 g)$.

2. $p \vDash (f \approx_{\alpha+1} g)$ if

$$p \vDash [\![(\forall x)[(x \; \varepsilon \; f) \to (\exists y)(y \; \varepsilon \; g \land y \approx_\alpha x)] \land (\forall x)[(x \; \varepsilon \; g) \to (\exists y)(y \; \varepsilon \; f \land y \approx_\alpha x)]]\!].$$

3. For a limit ordinal λ, $p \vDash (f \approx_\lambda g)$ if $p \vDash (f \approx_\alpha g)$ for some $\alpha < \lambda$.

We pause for some remarks on this definition.

First, clearly the important case here is the successor ordinal case—the least ordinal α at which $p \vDash (f \approx_\alpha g)$ will always be a successor.

Second, note the use of ε in the successor case, as given in definition 4.11. The effort we put into investigating that idea was not wasted.

Third, note that the successor case of the definition above is a formula, not in the language of set theory, but enclosed in $[\![\,]\!]$ brackets. Let us be very careful about what this means. As already discussed, for a classical or modal formula ϕ, all statements of the form $p \vDash \phi$ are classical formulas in disguise. By definition 4.4, the S4 translate adds the symbols $\square\lozenge$ in front of the formula and all its subformulas. This causes no problems; the result is just a different classical formula in disguise. Here we are using the S4 translate not to embed a classical formula into modal logic but rather to take advantage of the stability it brings. In this case, we write $p \vDash [\![f \approx_\alpha g]\!]$ instead of $p \vDash f \approx_\alpha g$. These two statements are equivalent since $\square\lozenge[\![\phi]\!] \leftrightarrow [\![\phi]\!]$ is S4-valid for any ϕ. This is easily shown from basic properties of $\square$ and $\lozenge$ in S4.

Lemma 4.14. *If for some world p of an S4 model with elements a and b we have $p \vDash a \approx_\alpha b$ then for all ordinals $\beta > \alpha$, $p \vDash a \approx_\beta b$.*

Proof. Let α be the least ordinal such that $p \vDash a \approx_\alpha b$. We prove by transfinite induction on α that each ordinal has the desired property. Since $p \nvDash a \approx_0 b$, α is a successor or a limit ordinal. If it is a limit ordinal, then by definition of $\approx_\alpha$, we must have $p \vDash a \approx_\gamma b$ for some $\gamma < \alpha$. But then $\gamma < \alpha < \beta$. By the induction hypothesis, all ordinals less than α have the desired property so we know that $p \vDash a \approx_\beta b$. So we turn to the case where α is a successor. Let α have the desired property; we show $\alpha + 1$ also has the property, i.e. that if $\gamma > \alpha + 1$ and $p \vDash a \approx_{\alpha+1} b$ then $p \vDash a \approx_\gamma b$. It is enough to prove the case for γ a successor since if it is a limit, the result holds for some successor less than γ.

Here we make use of the fact that $a \approx_\alpha b$ is true at a world if and only if $[\![a \approx_\alpha b]\!]$ is. We let $\alpha + 1 < \beta + 1$, assume that $p \vDash [\![a \approx_{\alpha+1} b]\!]$ but $p \nvDash [\![a \approx_{\beta+1} b]\!]$, and derive a contradiction.

Since $p \nvDash [\![a \approx_{\beta+1} b]\!]$, then by definition of $\approx_{\beta+1}$, we must have either

$$p \nvDash [\![(\forall x)[(x \; \varepsilon \; a) \rightarrow (\exists y)(y \; \varepsilon \; b \wedge y \approx_\beta x)]]\!]$$

or

$$p \nvDash \llbracket (\forall x)[(x \; \varepsilon \; b) \to (\exists y)(y \; \varepsilon \; a \land y \approx_\beta x)] \rrbracket.$$

or both. It doesn't matter which; say the first is true. Then there is a world p' accessible from p where $p' \nvDash \llbracket \neg(\forall x)[(x \; \varepsilon \; f) \to (\exists y)(y \; \varepsilon \; g \land y \approx_\beta x)] \rrbracket$, which implies that for some c we have

$$p' \vDash \llbracket (c \; \varepsilon \; a) \rrbracket \quad \text{and} \quad p' \vDash (\forall y)\llbracket y \; \varepsilon \; b \to \neg(y \approx_\beta c) \rrbracket,$$

By item 1 of lemma 4.6, since $p \vDash \llbracket a \approx_{\alpha+1} b \rrbracket$, we have $p' \vDash \llbracket a \approx_{\alpha+1} b \rrbracket$. This implies

$$p' \vDash \llbracket (\forall x)[(x \; \varepsilon \; a) \to (\exists y)(y \; \varepsilon \; b \land y \approx_\alpha x)] \rrbracket.$$

Since $p' \vDash \llbracket (c \; \varepsilon \; a) \rrbracket$ we have $p' \vDash \llbracket (\exists y)(y \; \varepsilon \; b \land y \approx_\alpha c) \rrbracket$. For some d, and some p'' accessible from p', we have

$$p'' \vDash \llbracket (d \; \varepsilon \; b) \rrbracket \quad \text{and} \quad p'' \vDash \llbracket (d \approx_\alpha c) \rrbracket.$$

Since $\alpha < \beta$ and we are assuming that the lemma holds for α, we must have $p'' \vDash \llbracket d \approx_\beta c \rrbracket$.

Since $p' \vDash (\forall y)\llbracket y \; \varepsilon \; b \to \neg(y \approx_\beta c) \rrbracket$, we must have $p' \vDash \llbracket \neg(d \approx_\beta c) \rrbracket$ and thus $p'' \vDash \llbracket \neg(d \approx_\beta c) \rrbracket$ and $p'' \nvDash \llbracket d \approx_\beta c \rrbracket$, a contradiction. $\square$

Definition 4.15. We say that $p \vDash a \approx b$ if there exists an ordinal α such that $p \vDash a \approx_\alpha b$.

So if $p \vDash a \approx b$, then there are infinitely many ordinals α at which $p \vDash a \approx_\alpha b$, but if $p \nvDash a \approx b$, there are none. Also, we have $p \vDash a \approx b$ if and only if $p \vDash \llbracket a \approx b \rrbracket$.

Now that $\approx$ has been defined, we can finally give a formal definition of $\in$ in an S4-ZF model.

Definition 4.16. For a possible world p and $f, g \in D$, we define $p \vDash f \in g$ to be true if for some h in D, $p \Vdash \llbracket h \approx f \rrbracket$ and $p \Vdash \llbracket h \; \varepsilon \; g \rrbracket$.

Since $\in$ depends on two formulas of the form $\llbracket X \rrbracket$, then at any world p we have $a \in b$ if and only if $\llbracket a \in b \rrbracket$.

In order to justify the use of the word "equality" when referring to the relation $\approx$, we must prove that it is reflexive, transitive, symmetric, and substitutive. We show reflexivity and substitutivity in detail. Symmetry is obvious from the definition of $\approx_\alpha$. Transitivity is proven in [SF96, p.209].

Theorem 4.17. *The relation $\approx$ is reflexive. ([SF96], p.212, exercise 2.1)*

Proof. We show for $a \in R_\alpha$, $p \vDash \llbracket a \approx_\alpha a \rrbracket$.

Assume that for all $\beta < \alpha$, if $b \in R_\beta$, then $p \vDash \llbracket b \approx_\beta b \rrbracket$. We proceed by transfinite induction.

- $\alpha \neq 0$, since R_0 is empty.

- If α is a limit ordinal, the result is immediate from the definition of $\approx_\alpha$ and the induction hypothesis.

- If α is a successor ordinal, let $\alpha = \beta + 1$. In this case the definition of $\approx_\alpha$ reduces to:

$$p \vDash [\![\forall x [x \; \varepsilon \; a \to \exists y (y \; \varepsilon \; a \wedge y \approx_\beta x)]]\!].$$

Now, if a is empty at p, the result is vacuously true. If not, for $x \in a$, take y to be the same set as x. Then $y \in R_\beta$ so we have $p \vDash [\![y \approx_\beta y]\!]$ and the result follows.

$\square$

In order to prove the substitutivity of $\approx$, we need the following two lemmas for the atomic cases, after which the result follows.

Lemma 4.18. *Let p be a possible world in an S4 model and a, b, and c be elements of the domain. If $p \vDash [\![a \approx b]\!]$ and $p \vDash [\![a \in c]\!]$ then $p \vDash [\![b \in c]\!]$.*

Proof. Assume that $p \vDash [\![a \approx b]\!]$ and $p \vDash [\![a \in c]\!]$ but $p \nvDash [\![b \in c]\!]$. We derive a contradiction.

Since $p \nvDash [\![b \in c]\!]$, we can expand the definition of $\in$ to get $p \vDash \neg [\![(\exists x)[x \approx b \wedge x \; \varepsilon \; c]]\!]$. So there is a world p' accessible from p such that $p' \vDash [\![\neg (\exists x)[x \approx b \wedge x \; \varepsilon \; c]]\!]$, i.e. $p' \vDash [\![(\forall x)[x \approx b \to \neg (x \; \varepsilon \; c)]]\!]$. We can move the universal quantifier outside the S4 transform to get $p' \vDash (\forall x)[\![x \approx b \to \neg (x \; \varepsilon \; c)]\!]$.

We know that $p' \vDash [\![a \in c]\!]$ since $p \vDash [\![a \in c]\!]$. Expanding the definition of $\in$ and using simple properties of S4, we see that for some f in the domain, $p' \vDash [\![f \approx a]\!]$ and $p' \vDash [\![f \; \varepsilon \; c]\!]$. Now, since $p' \vDash [\![f \approx a]\!]$ and $p' \vDash [\![a \approx b]\!]$, by transitivity we have $p' \vDash [\![f \approx b]\!]$. By the previous paragraph, this means that $p' \vDash [\![\neg (f \; \varepsilon \; c)]\!]$. $\square$

Similarly we can prove

Lemma 4.19. *If $p \vDash [\![a \approx b]\!]$ and $p \vDash [\![c \in a]\!]$ then $p \vDash [\![c \in b]\!]$.*

Theorem 4.20. *The relation $\approx$ is substitutive. In particular, if $p \vDash [\![a \approx b]\!]$ and ϕ and ϕ' are formulas in the classical language of set theory differing in that ϕ' has an occurence of b where ϕ has an occurence of a, then $p \vDash [\![\phi]\!] \leftrightarrow [\![\phi']\!]$.*

Proof. If ϕ and ϕ' are atomic then the result follows from the previous two lemmas. If not, we argue by induction on the complexity of ϕ. By lemma 4.8 and the fact that every classical formula is equivalent to one with a minimal set of logical connectives, we can restrict our induction to $\wedge, \forall$, and $\neg$.

If ϕ is $\phi_1 \wedge \phi_2$ then let ϕ' be $\phi'_1 \wedge \phi'_2$ and assume the result holds for the subformulas, i.e. $p \vDash [\![\phi_1]\!] \leftrightarrow [\![\phi'_1]\!]$ and similarly for ϕ_2. We have $p \vDash [\![\phi_1 \wedge \phi_2]\!]$ if and only if $p \vDash [\![\phi_1]\!] \wedge [\![\phi_2]\!]$ if and only if $p \vDash [\![\phi_1]\!]$ and $p \vDash [\![\phi_2]\!]$, so by our assumption this is true if and only if $p \vDash [\![\phi'_1]\!]$ and $p \vDash [\![\phi'_2]\!]$, i.e. $p \vDash [\![\phi'_1 \wedge \phi'_2]\!]$. So if the result holds for the subformulas, it holds for the conjunction of them. The argument for $\forall$ is similar. The argument for $\neg$ is by contradiction.

Assume that $[\![\phi]\!] \leftrightarrow [\![\phi']\!]$ is true at every world. Suppose for some p that $p \vDash [\![a \approx b]\!]$ and $p \nvDash [\![\neg\phi]\!] \leftrightarrow [\![\neg\phi']\!]$. Both directions are similar. Assume that $p \nvDash [\![\neg\phi']\!] \rightarrow [\![\neg\phi]\!]$. Then $p \vDash [\![\neg\phi']\!]$ and $p \vDash \neg[\![\neg\phi]\!]$. This means that for some p' accessible from p, $p' \vDash [\![\phi]\!]$. But $p' \vDash [\![\neg\phi']\!]$, which violates our assumption that $[\![\phi]\!] \leftrightarrow [\![\phi']\!]$ at all worlds. $\qquad\square$

This completes the definition of S4-ZF models. To create one, it is enough to choose a transitive inner model **M** of ZF (such as **L** for example), a collection of possible worlds $G \in \mathbf{M}$ and a reflexive, transitive two-place accessibility relation R on G ($R \in \mathbf{M}$). The domain of the model will be built from **M** as in definition 4.9 and the truth assignment of any formula of set theory at any world will be given by the definition of $\in$ above.

Our goal for the rest of section 4.2 will be to prove the following.

Theorem 4.21. *For each axiom ϕ of ZF, $[\![\phi]\!]$ is valid in every S4-ZF model.*

Then, in section 4.3, we will create an S4-ZF model in which $[\![AC]\!]$ is not valid.

Before proving theorem 4.21, we pause for one last discussion about the properties of equality in the S4-ZF models. Specifically, we must be clear on the relationship between this definition of equality and the one in section 2.4.1, i.e. that two sets are equal if and only if they are members of the same sets.

At first, we appear to have a mess on our hands; the definitions don't resemble each other in the slightest, and in classical models, we are defining equality in terms of membership whereas in modal models we are defining membership in terms of equality. But there is really no problem. The independence result we're seeking involves formulas in the classical language of set theory, which has $\in$ as the only relational symbol, no $=$ and certainly no $\approx$. In investigating classical models of set theory, $\in$ was understood unambiguously and we used $=$ as a simple syntactic substitution

for a longer statement involving only the $\in$ symbol, purely for convenience. We can take classical statements of the form $a = b$ and create the S4 translate $[\![a = b]\!]$ or $[\![\forall z(a \in z \leftrightarrow b \in z)]\!]$. We also used the symbol $\approx$, in modal models only, as a way of describing a more complex notion of equality in those structures that couldn't be described in classical set theory. Both $\approx$ and ε were created to facilitate the definition of $\in$ in modal models; they have no formal status in the language of set theory. So what is the relationship between $=$ and $\approx$?

To answer this question, we need the concept of the modal representation of a classical set in an S4 model. For a classical set x in $\mathbf{M}$, this is denoted $\hat{x}$.

Definition 4.22. For x in $\mathbf{M}$, and G the set of all possible worlds, we associate a member $\hat{x}$ in $\mathcal{D}$ defined by:

$$\hat{x} = G \times \{\hat{y} \mid y \in x\}$$

This inductive definition says that the modal representation of x in an S4 model is the element of the domain of that model in which every possible world can see every element of the set. If $a \in x$ then $p \vDash \hat{a} \; \varepsilon \; \hat{x}$ at all worlds p. Modal representations of classical sets have no secrets.

It should be clear that every classical x in $\mathbf{M}$ has a representative $\hat{x}$ in every S4-ZF model created from $\mathbf{M}$, but not every element of the domain of an S4-ZF model is the modal representative of a classical set—most are not. The proof of the following is easy:

Theorem 4.23. If $a, b \in \mathbf{M}$, then for an S4-ZF model $\mathfrak{M}$ constructed from $\mathbf{M}$, at any world p we have $p \vDash [\![\hat{a} \approx \hat{b}]\!]$ if and only if $a = b$ classically. ([SF96], p.214, exercise 3.1).

Proof. If $a = b$ in $\mathbf{M}$, then the result follows from the reflexivity of $\approx$. If $p \vDash [\![\hat{a} \approx \hat{b}]\!]$, then we must have $p \vDash [\![\hat{a} \approx_\alpha \hat{b}]\!]$ for some ordinal α. We proceed by transfinite induction on the ordinals. We assume that for all ordinals $\beta < \alpha$, if $p \vDash [\![\hat{a} \approx_\beta \hat{b}]\!]$ then $a = b$ in $\mathbf{M}$. The only case we need to consider is the one where α is a successor ordinal, so let $\alpha = \beta + 1$. By definition of $\approx_\alpha$, we have

$$p \vDash [\![(\forall x)[(x \; \varepsilon \; \hat{a}) \to (\exists y)(y \; \varepsilon \; \hat{b} \wedge y \approx_\beta x)] \wedge (\forall x)[(x \; \varepsilon \; \hat{b}) \to (\exists y)(y \; \varepsilon \; \hat{a} \wedge y \approx_\beta x)]]\!]$$

Consider the left side of the conjunction. Given x such that $p \vDash [\![x \; \varepsilon \; \hat{a}]\!]$, we know that x is $\hat{c}$ for some $c \in a$. We also know that there is some y and some world p' accessible from p such that $p' \vDash [\![y \; \varepsilon \; \hat{b}]\!]$ and $p' \vDash [\![y \approx_\beta \hat{c}]\!]$. Since $p' \vDash [\![y \; \varepsilon \; \hat{b}]\!]$, y must be $\hat{d}$ for some $d \in b$. Also, $p' \vDash [\![\hat{c} \; \varepsilon \; \hat{a}]\!]$. So $p' \vDash [\![\hat{c} \approx_\beta \hat{d}]\!]$. But by our inductive hypothesis, this means that $c = d$. Now, since the elements of $\hat{a}$ at any world are exactly the sets $\hat{c}$ for $c \in a$, we conclude than an arbitrary element of a is also an element of b, so $a \subseteq b$. By repeating the argument on the right side of the conjunction, we get $b \subseteq a$ and thus (since $\mathbf{M}$ is a model of ZF), $a = b$. $\qquad\square$

More than equality is preserved when we work with modal representations:

Lemma 4.24. *If $\phi(x_1, \ldots, x_n)$ is a formula in the classical language of set theory with all its free variables among those shown and $s_1, \ldots, s_n \in \mathbf{M}$ then for any world p of an S4-ZF model $\mathfrak{M}$ constructed from $\mathbf{M}$:*

1. If ϕ is Δ_0, then $\mathbf{M} \vDash \phi(s_1, \ldots, s_n)$ if and only if $p \vDash_{\mathfrak{M}} [\![\phi(\widehat{s_1}, \ldots, \widehat{s_n})]\!]$.

2. If ϕ is Σ_1, then $\mathbf{M} \vDash \phi(s_1, \ldots, s_n)$ implies $p \vDash_{\mathfrak{M}} [\![\phi(\widehat{s_1}, \ldots, \widehat{s_n})]\!]$.

Proof. See [SF96, pp. 212-213] □

Now, returning to the proof of theorem 4.21, we need to verify for each of the axioms ϕ given in section 2.4.1, that $[\![\phi]\!]$ is valid in every S4-ZF model. We provide details for Extensionality and Infinity.

For Extensionality, we need the following formula to be valid in all S4-ZF models:

$$[\![\forall x \forall y [\forall z(z \in x \leftrightarrow z \in y) \rightarrow \forall z(x \in z \leftrightarrow y \in z)]]\!]$$

We start by showing that if a and b are members of the domain of an S4-ZF model, and $p \vDash [\![(\forall x)(x \in a \leftrightarrow x \in b)]\!]$, then $p \vDash [\![a \approx b]\!]$. Let α be such that $a, b \in R_\alpha$. We show by transfinite induction on α that $p \vDash [\![a \approx_\alpha b]\!]$. The 0 and limit cases are trivial, so suppose we know the result holds for R_α and we have $f, g \in R_{\alpha+1}$. Also suppose that $p \vDash [\![(\forall x)(x \in a \leftrightarrow x \in b)]\!]$ but $p \nvDash [\![a \approx_{\alpha+1} b]\!]$. We derive a contradiction.

Since $p \nvDash [\![a \approx_{\alpha+1} b]\!]$, then at least one of the conjuncts in the definition of $\approx_{\alpha+1}$ must be false. The problem is symmetric, so assume the false conjunct is the first one and so for some c we have

$$p \nvDash [\![(c \,\varepsilon\, a) \rightarrow (\exists y)(y \,\varepsilon\, b \wedge c \approx_\alpha y)]\!].$$

for some c. This means that for some world p' accessible from p we have

$$p' \vDash [\![(c \,\varepsilon\, a)]\!] \quad \text{and} \quad p' \vDash [\![(\forall y)(y \,\varepsilon\, b \rightarrow \neg(c \approx_\alpha y))]\!]. \tag{4.1}$$

Now, since $p' \vDash [\![c \,\varepsilon\, a]\!]$, c must have lower rank than a, so $c \in R_\alpha$. Also, since equality is reflexive, then by definition of $\in$, we have $p' \vDash [\![c \in a]\!]$.

Since we are assuming $p \vDash [\![(\forall x)(x \in a \leftrightarrow x \in b)]\!]$, we have $p' \vDash [\![c \in a \leftrightarrow c \in b]\!]$. Thus $p' \vDash [\![c \in b]\!]$ so by definition of $\in$ there exists some d such that $p' \vDash [\![d \,\varepsilon\, b]\!]$ and $p' \vDash [\![c \approx d]\!]$. So d has lower rank than b and is in R_α.

Since the formula $\forall x(x \in c \leftrightarrow x \in c)$ is classically valid for any c, we may assume that the S4 translate of it is S4-valid. Using the substitutivity of $\approx$ and the fact that $p' \vDash [\![c \approx d]\!]$, we have $p' \vDash [\![\forall x(x \in c \leftrightarrow x \in d)]\!]$. Since $c, d \in R_\alpha$, the induction hypothesis applies and we conclude that $p' \vDash [\![c \approx_\alpha d]\!]$. Since $p' \vDash [\![d \varepsilon b]\!]$, this contradicts 4.1.

By the definition of $\approx$ it follows immediately that if $p \vDash [\![(\forall x)(x \in a \leftrightarrow x \in b)]\!]$, then $p \vDash [\![a \approx b]\!]$.

The translate of the axiom of extensionality can now be shown by contradiction. Assume that it fails at p. Then for some a, b and some p' accessible from p we have

$$p' \vDash [\![\forall z(z \in a \leftrightarrow z \in b)]\!]$$

and

$$p' \vDash [\![\neg \forall z(a \in z \leftrightarrow b \in z)]\!].$$

The first formula implies that $p' \vDash [\![a \approx b]\!]$. But since $\approx$ is substitutive, we now have $p' \vDash [\![\neg \forall z(a \in z \leftrightarrow a \in z)]\!]$, which is impossible.

Thus the translate of extensionality holds at all worlds of all S4-ZF models.

For Infinity, in the classical inner model $\mathbf{M}$ from which we constructed the S4-ZF model, we have a Σ_1 formula that says "The set of all natural numbers exists". Since this statement implies the axiom of infinity, we establish the result by showing that its translate is valid in S4-ZF models.

The formula is

$$\exists x(x \text{ is a limit ordinal} \wedge (\forall y \in x)\neg(y \text{ is a limit ordinal})).$$

It is Σ_1 because the formula for "x is a limit ordinal" is Δ_0 by theorem 2.19. The only set that satisfies this formula is ω, the set of natural numbers. Since the formula holds in $\mathbf{M}$, it is valid in any S4-ZF model constructed from $\mathbf{M}$ by lemma 4.24.

Proofs for the remaining axioms can be found in [SF96, pp.211-222], as can a proof (pp.222-225) of the following important fact:

Lemma 4.25. *If the Axiom of Choice holds in the classical class* $\mathbf{M}$, *the S4 translate* $[\![AC]\!]$ *holds in any S4-ZF model constructed from* $\mathbf{M}$.

This may seem like a major setback since it means that the translate of the Axiom of Choice can never fail in any S4-ZF model constructed by the method above. Instead of working with these

models, we search for submodels of them, still models of the translate of each ZF axiom, in which the translate of AC fails.

4.3 Submodels of S4-ZF models

The first step in creating an S4-ZF model is to choose the classical inner model $\mathbf{M}$ from which the model will be constructed. Let us call the S4-ZF model $\mathbb{M}$. We want $[\![AC]\!]$ to fail at some world or worlds of $\mathbb{M}$, but lemma 4.25 says that if AC holds in $\mathbf{M}$, the S4-translate of it will hold at every world of $\mathbb{M}$. We don't have ready access to a classical model in which AC fails—if we did, the independence of AC would follow immediately. Our strategy will be to identify a submodel of $\mathbb{M}$ that is still an S4-ZF model but in which there is some world where $[\![AC]\!]$ fails. This will complete the proof of the independence of the Axiom of Choice.

For the rest of this section, let D be the domain of $\mathbb{M}$, G the set of possible worlds, and R the accessibility relation. D is a subclass of $\mathbb{M}$, and G and R are sets in $\mathbb{M}$.

We use automorphisms (and groups of automorphisms, and sequences of groups of automorphisms) to identify the elements of D that will be part of the submodel. The idea is to impose so much symmetry that certain sets will fail to have choice functions. It turns out that if we focus only on those elements of an S4-ZF model that are left unmoved by certain automorphisms, these elements are also models of the translates of the ZF axioms. In these submodels it is possible to find elements of the domain that fail to have choice functions.

Definition 4.26. An automorphism of $\langle G, R \rangle$ is a bijection $\theta : G \to G$ such that pRq if and only if $\theta(p)R\theta(q)$. An automorphism group of $\langle G, R \rangle$ is a group of automorphisms under composition (i.e. a set of automorphisms that contains the identity map, is closed under composition, and contains with each member its inverse.)

Definition 4.27. The action of an automorphism can be extended to the domain of the S4-ZF model as follows. For any element f of the domain of $\mathbb{M}$, and for any automorphism θ of $\langle G, R \rangle$,

$$\theta(f) = \{\langle \theta(p), \theta(a) \rangle \mid \langle p, a \rangle \in f\}.$$

Definition 4.28. The action of an automorphism can also be extended to the formulas with constants. If ϕ is a formula with constants in D and θ is an automorphism of $\langle G, R \rangle$, then $\theta(\phi)$ is the formula ϕ with each constant f replaced by its image $\theta(f)$ as in definition 4.27.

The following is proven in [SF96, p.246]:

Theorem 4.29. *Let ϕ be a closed formula in the modal language of set theory with constants in $\mathcal{D}$. Let θ be an automorphism. Then $p \vDash \phi$ if and only if $\theta(p) \vDash \theta(\phi)$.*

Definition 4.30. Let $\mathfrak{G}$ be an automorphism group of $\langle G, R \rangle$. We say a member $f \in D$ is $\mathfrak{G}$-invariant if $[\![f \approx \theta(f)]\!]$ is valid in $\mathbb{M}$ for each $\theta \in \mathfrak{G}$.

Note that for classical sets x (in $\mathbf{M}$), $\hat{x}$ is $\mathfrak{G}$-invariant for every automorphism group $\mathfrak{G}$.

Definition 4.31. If θ_1 and θ_2 are automorphisms, and H is a set of automorphisms, we define $\theta_1 H \theta_2$ to be the set $\{ \theta_1 h \theta_2 \mid h \in H \}$.

Definition 4.32. Let θ be an automorphism and H a set of automorphisms. We say θ normalizes H if $\theta^{-1} H \theta = H$.

Definition 4.33. An automorphism sequence of $\langle G, R \rangle$ is a sequence $\mathfrak{F} = \mathfrak{G}_0, \mathfrak{G}_1, \ldots$ of automorphism groups of $\langle G, R \rangle$ such that:

1. $\mathfrak{G}_{n+1} \subseteq \mathfrak{G}_n$ for each n;

2. for every $\theta \in \mathfrak{G}_0$, there is some n such that θ normalizes each of $\mathfrak{G}_n, \mathfrak{G}_{n+1}, \mathfrak{G}_{n+2}, \ldots$.

Definition 4.34. Let $\mathfrak{F}$ be an automorphism sequence of $\langle G, R \rangle$. We say a member $f \in D$ is $\mathfrak{F}$-invariant if:

1. for some $\mathfrak{G}$ in $\mathfrak{F}$, f is $\mathfrak{G}$-invariant; and

2. for each $\langle p, g \rangle \in f$, g is $\mathfrak{F}$-invariant.

If $f \in D$ is $\mathfrak{F}$-invariant, we think of this as saying that $\mathbb{M}$ eventually can't distinguish between f and images of f under $\mathfrak{F}$.

For an automorphism sequence $\mathfrak{F}$, we can define a subclass $D_{\mathfrak{F}}$ of the domain D of $\mathbb{M}$ consisting of those elements of D that are invariant under $\mathfrak{F}$. The intended submodel of $\mathbb{M}$ in which $[\![AC]\!]$ will fail at some world will have a domain of the form $D_{\mathfrak{F}}$ for some $\mathfrak{F}$.

Definition 4.35. Let $\mathfrak{F}$ be an automorphism sequence. Set $D_{\mathfrak{F}}$ to be the subclass of D consisting of those members that are $\mathfrak{F}$-invariant. Set $p \vDash_{\mathfrak{F}} \phi$ if and only if

- ϕ is a formula given in the modal counterpart of the classical language of set theory, i.e. the language consisting only of the two-place relation "$\in$" as well as the modal operators.

- The constants in the formula ϕ (if any) are in $D_{\mathfrak{F}}$.

Let $\mathbb{M}_{\mathfrak{F}}$ be the S4 model with the same set of possible worlds G and accessibility relation R as $\mathbb{M}$, with domain $D_{\mathfrak{F}}$, and with truth assignment $\vDash_{\mathfrak{F}}$ as defined above.

We pause for a note on the language requirement in the definition. We have been allowing the symbols "$\approx$", "$\approx_{\alpha}$", "ε", etc in our modal formulas until now, in order to prove properties of S4-ZF models, and in order to define the symbol "$\in$". In the submodels defined above, we no longer explicitly allow formulas with these extra symbols. We are restricting ourselves to "$\in$" as the only relation in our language. The only symbol we allow outside the classical language of set theory is "$\Box$". The symbol "$\in$" will still have the same meaning and will still be defined the same way, but we will never break it down into statements involving "$\approx$" and "ε" in the submodel. The reason for this is that the formula we are trying to prove independent (the Axiom of Choice) and the formulas from which we want to prove the independence (the axioms of ZF) are all expressed in the classical language of set theory, so the S4 translates of them are all expressed in that language plus the symbol "$\Box$". This way we ensure that our proof doesn't rely on properties of the modal model that can't be expressed in the classical language of set theory.

Our objective for the rest of this section will be to show that whenever $\mathfrak{F}$ is a set in $\mathbf{M}$, the resulting $\mathbb{M}_{\mathfrak{F}}$ is an S4-ZF model, so for all axioms ϕ of ZF, $[\![\phi]\!]$ is valid in $\mathbb{M}_{\mathfrak{F}}$. It is necessary for $\mathfrak{F}$ to be a set in $\mathbf{M}$ in order to ensure that the axiom schema of Replacement holds in the submodel $\mathbb{M}_{\mathfrak{F}}$. Some relationships between truth in $\mathbb{M}$ and truth in $\mathbb{M}_{\mathfrak{F}}$ will help us demonstrate that the axioms continue to hold.

First, it is clear that $\mathbb{M}$ and $\mathbb{M}_{\mathfrak{F}}$ agree on atomic statments about the common domain $D_{\mathfrak{F}}$. This follows easily from the fact that all the possible worlds of $\mathbb{M}$ are present in $\mathbb{M}_{\mathfrak{F}}$. In fact, the two models agree on all Δ_0 statements with constants in the common domain. The proof of this will be omitted here. Also, the expected results about Σ_1 and Π_1 formulas hold:

Lemma 4.36. *Let $\phi(x_1, x_2, \ldots, x_n)$ be a formula in the classical language of set theory, with all free variables among those shown. Let $s_1, s_2, \ldots, s_n \in D_{\mathfrak{F}}$. Let p be a possible world in G.*

- *If ϕ is Δ_0, we have*

$$p \vDash [\![\phi(s_1, s_2, \ldots, s_n)]\!] \quad \text{if and only if} \quad p \vDash_{\mathfrak{F}} [\![\phi(s_1, s_2, \ldots, s_n)]\!].$$

- *If ϕ is Σ_1, we have*

$$\text{if } p \vDash_{\mathfrak{F}} [\![\phi(s_1, s_2, \ldots, s_n)]\!] \text{ then } p \vDash [\![\phi(s_1, s_2, \ldots, s_n)]\!].$$

- *If ϕ is Π_1, we have*

$$\textit{if } p \vDash [\![\phi(s_1, s_2, \ldots, s_n)]\!] \textit{ then } p \vDash_{\mathfrak{F}} [\![\phi(s_1, s_2, \ldots, s_n)]\!].$$

Theorem 4.37. *For each axiom ϕ of ZF, and for each world p of $\mathbb{M}_{\mathfrak{F}}$, we have $p \vDash_{\mathfrak{F}} [\![\phi]\!]$.*

Proof. We illustrate with three of the axioms.

Extensionality and Foundation can be expressed in Π_1 form, and since their translates hold in $\mathbb{M}$, by lemma 4.36, they also hold in $\mathbb{M}_{\mathfrak{F}}$.

For Infinity, we only need to show the existence of an infinite set in the model. Now, $\omega \in \mathbf{M}$ so $\hat{\omega} \in D$. Also, $\hat{\omega}$ is invariant for every automorphism sequence $\mathfrak{F}$. To see this, note that by definition $\hat{x}$ has the same extension at every world, so $\theta(\hat{x}) = \hat{x}$ for all automorphisms θ. $\qquad\square$

The discussion about language after definition 4.35 has left us without a way of discussing equality in our submodels. Since in these models we are using the classical language of set theory (plus $\square$), we are forced to express equality in the classical way, as in definition 2.5, or what is equivalent in any model of ZF, as extensional equivalence. When we use the symbol "=" in a modal formula interpreted at a world of a submodel of an S4-ZF model, it will be an abbreviation for either of these classical statements. The link between $\approx$ in $\mathbb{M}$ and equality in the submodel is provided by the following theorem.

Theorem 4.38. *Let $a, b \in \mathcal{D}_{\mathfrak{F}}$, let p be a possible world, and let "=" be defined as above. Then*

$$p \vDash [\![a \approx b]\!] \textit{ if and only if } p \vDash_{\mathfrak{F}} [\![a = b]\!].$$

4.4 A model in which AC fails

We now turn our attention to the construction of a particular S4-ZF model in which $[\![AC]\!]$ is not valid. The independence of AC from ZF will follow immediately. This section presents a concise version of Smullyan and Fitting's modifications of Dana Scott and Robert Solovay's modifications of Paul Cohen's modifications of Abraham Fraenkel's 1922 proof of the independence of the Axiom of Choice from a variation of set theory in which objects called "urelements" were allowed to exist in the domain but were not sets. Unfortunately, these objects were essential to Fraenkel's proof so the independence result was not established for the more general case where there are no urelements. Cohen showed in 1963 that forcing allowed the idea of Fraenkel's proof to be used in ZF itself.

Here we will follow Smullyan and Fitting ([SF96, pp.254-258]) in creating an S4-ZF model in which Fraenkel's technique is used to show that $[\![AC]\!]$ fails to be valid.

The first thing we need is a partition of ω into infinitely many disjoint sets $I_0, I_1, I_2, \ldots$. In order to construct a set in an S4-ZF model that has no choice function, we need the partition to have the following three properties: (1) Each natural number appears in exactly one of the I_n; (2) If $k \in I_n$ then $k \geq n$; and (3) Each I_n is infinite. The following partition will do. In it each of the I_n is equal to the set $\{(n+k)(n+k+1)/2 + k \mid k = 0, 1, 2, \ldots\}$.

$$I_0 = 0, 2, 5, 9, 14, 20, 27, \ldots$$

$$I_1 = 1, 4, 8, 13, 19, 26, 34, \ldots$$

$$I_2 = 3, 7, 12, 18, 25, 33, 42, \ldots$$

$$I_3 = 6, 11, 17, 24, 32, 41, 51, \ldots$$

$$I_4 = 10, 16, 23, 31, 40, 50, 61, \ldots$$

$$I_5 = 15, 22, 30, 39, 49, 60, 72, \ldots$$

$$I_6 = 21, 29, 38, 48, 59, 71, 84, \ldots$$

With this detail out of the way, we now define our S4-ZF model and identify some sets that are present in it. Then we create a $\mathfrak{F}$-invariant submodel in which one of these sets has no choice function.

To create an S4-ZF model, we need a classical transitive inner model of ZF, a set of possible worlds, and an accessibility relation. For the rest of this section, let $\mathbf{M}$ be any transitive inner model of ZF in which the axiom of choice is true, such as $\mathbf{L}$. We fix $\mathbb{M}$ as the S4-ZF model constructed from $\mathbf{M}$ with the following worlds and accessibility relation:

- A possible world in this model is a function p with domain ω such that $p(n) = \langle P_n, N_n \rangle$, where P_n and N_n are finite, disjoint sets of integers, and such that $p(n) = \langle \emptyset, \emptyset \rangle$ for all but a finite number of n.

- For two worlds p and q, pRq holds (q is accessible from p) if for all n we have: if $p(n) = \langle P_n, N_n \rangle$ and $q(n) = \langle P'_n, N'_n \rangle$, then pRq iff $P_n \subseteq P'_n$ and $N_n \subseteq N'_n$.

Let D be the domain of $\mathbb{M}$. By lemma 4.25, $[\![AC]\!]$ is valid in $\mathbb{M}$, so our work is not done yet. The plan is this: we show that certain sets are present in the domain D of $\mathbb{M}$, including a set we will

call T. Then we exhibit a submodel $\mathbb{M}_{\mathfrak{F}}$ of $\mathbb{M}$ whose domain $D_{\mathfrak{F}}$ also includes T. This requires a careful choice of the automorphism sequence $\mathfrak{F}$. By theorem 4.37, the S4 translate of each axiom of ZF is true in $\mathbb{M}_{\mathfrak{F}}$. We will show that it is impossible for any element of $D_{\mathfrak{F}}$ to be a choice function for T at any world of $\mathbb{M}_{\mathfrak{F}}$, and so $[\![AC]\!]$ fails to be valid in $\mathbb{M}_{\mathfrak{F}}$. Then $\mathbb{M}_{\mathfrak{F}}$ proves the independence of AC from ZF by the reasoning at the beginning of section 4.

Definition 4.39. For each $n \in \omega$, let

$$s_n = \{\langle p, \hat{k}\rangle \mid p \in G, p(n) = \langle P_n, N_n\rangle \text{ and } k \in P_n\}.$$

$$t_n = \{\langle p, s_k\rangle \mid p \in G \text{ and } k \in I_n\}$$

$$T = \{\langle p, t_n\rangle \mid p \in G \text{ and } n \in \omega\}$$

For each n, we have $s_n \in D$ and $t_n \in D$. Also, $T \in D$. We also have

Lemma 4.40. *For $n, k \in \omega$, if $n \neq k$ then $[\![\neg(s_n \approx s_k)]\!]$ is valid in $\mathbb{M}$.*

We now define the automorphism sequence $\mathfrak{F}$ that will give us the submodel $\mathbb{M}_{\mathfrak{F}}$. We do this in such a way that s_n, t_n (for each $n \in \omega$) and T are all $\mathfrak{F}$-invariant, and thus are members of the domain of $\mathbb{M}_{\mathfrak{F}}$. They will be used to demonstrate the failure of $[\![AC]\!]$ everywhere in $\mathbb{M}_{\mathfrak{F}}$.

We start by introducing permutations of the natural numbers ω and use these to permute the possible worlds. As usual in mathematics, by a permutation of ω we mean a bijection $\theta : \omega \to \omega$.

Definition 4.41. For each $n \in \omega$, let $\mathfrak{G}_n$ be the set of all permutations θ of ω that have the following properties:

1. θ is the identity on all integers $< n$;

2. θ is the identity on all but a finite set; and

3. θ respects the I partition of ω, i.e. for each $m \in \omega$, $\theta : I_m \to I_m$.

If $\theta \in \mathfrak{G}_n$ then θ may permute some of the naturals that are greater than or equal to n, but only finitely many of them, and only within their set in the I-sequence.

Definition 4.42. If θ is a permutation of ω, its action can be extended to G as follows. For each possible world $p \in G$,

$$\theta(p)(n) = p(\theta^{-1}(n)).$$

So θ shuffles around the range of p.

At this point we observe that for any permutation θ of ω, extended to act on possible worlds as in the definition above, θ is an automorphism of $\langle \mathcal{G}, \mathcal{R} \rangle$. We fix $\mathfrak{F}$ as the automorphism sequence consisting of $\mathfrak{G}_0, \mathfrak{G}_1, \mathfrak{G}_2, \ldots$ as defined above. We need to prove that it meets the requirements for an automorphism sequence given in definition 4.33.

Theorem 4.43. *$\mathfrak{F}$ is an automorphism sequence. So*

1. *$\mathfrak{G}_n$ is an automorphism group for each n;*

2. *$\mathfrak{G}_{n+1} \subseteq \mathfrak{G}_n$; and*

3. *for every $\theta \in \mathfrak{G}_0$, there is some n so that θ normalizes each of $\mathfrak{G}_n, \mathfrak{G}_{n+1}, \mathfrak{G}_{n+1}, \ldots$*

Proof.   1. Each permutation θ is an automorphism on G. It is clear from definition 4.41 that the identity permutation on ω is in each $\mathfrak{G}_n$, and that if a permutation θ meets the criteria for membership in $\mathfrak{G}_n$, so does θ^{-1}.

2. If θ is the identity on all natural numbers $< n + 1$, it is clearly also the identity on all the natural numbers $< n$. The other properties are trivial.

3. We know that θ only permutes finitely many natural numbers. Let n be the largest number moved by θ. By definition of θ, we know that the automorphisms θ and θ^{-1} both map $\{0, 1, 2, \ldots, n\}$ to itself and for all $m > n$, $\theta(m) = \theta^{-1}(m) = m$. Now, let $\pi \in \mathfrak{G}_{n+1}$, so that π and π^{-1} both map $\{n+1, n+2, \ldots\}$ to itself. For $m \leq n$, $\pi(m) = \pi^{-1}(m) = m$. We show that θ normalizes each of $\mathfrak{G}_{n+1}, \mathfrak{G}_{n+2}, \ldots$. First we show that it normalizes $\mathfrak{G}_{n+1}$. Since $\pi \in \mathfrak{G}_{n+1}$, we need to show that for a natural number m, $(\theta^{-1}\pi\theta)(m) = \pi(m)$. There are two cases:

   - If $m > n$ then $m = \theta(m)$, and $\pi(m) > n$ so $\theta^{-1}(\pi(m)) = \pi(m)$. Substituting $\theta(m)$ for m we get $(\theta^{-1}\pi\theta)(m) = \pi(m)$.

   - If $m \leq n$ then $\theta(m) \leq n$ [2]. So $\pi(\theta(m)) = \theta(m)$. Applying θ^{-1} to both sides, we get $(\theta^{-1}\pi\theta)(m) = (\theta^{-1}\theta)(m) = m = \pi(m)$.

Now, since θ normalizes $\mathfrak{G}_{n+1}$, it must normalize $\mathfrak{G}_m$ for $m > n+1$ since any automorphism in $\mathfrak{G}_m$ must be the identity on all natural numbers less than m, and thus on all natural numbers less than $n + 1$, and the above argument applies.

$\square$

---

[2] Here we correct a typo in Smullyan and Fitting by using $\leq$ instead of $<$.

Now, we have $\mathbb{M}$, our S4-ZF model constructed from $\mathbf{M}$. We also have an automorphism sequence $\mathfrak{F}$, so we can form the submodel $\mathbb{M}_{\mathfrak{F}}$ of $\mathfrak{F}$-invariant elements of the domain of $\mathbb{M}$. This submodel is also an S4-ZF model by theorem 4.37. The end is in sight. We now show that $s_n \in D_{\mathfrak{F}}$ and $t_n \in D_{\mathfrak{F}}$ for each $n \in \omega$, and also that $T \in D_{\mathfrak{F}}$. Finally, we use these facts to show that T has no choice function in $D_{\mathfrak{F}}$. First we need the following two lemmas.

Lemma 4.44. *For any automorphism θ in any $\mathfrak{G} \in \mathfrak{F}$, $\theta(s_n) = s_{\theta(n)}$. This is classical equality: they are the same set. (This proof is omitted in [SF96]).*

Proof. Using definition 4.39,

$$x \in \theta(s_n) \leftrightarrow x = \langle \theta(p), \hat{k} \rangle \text{ (with } p(n) = \langle P, N \rangle, k \in P)$$

Substituting q for $\theta(p)$, we get the equivalent statement

$$x \in \theta(s_n) \leftrightarrow x = \langle q, \hat{k} \rangle \text{ (with } \theta^{-1}(q)(n) = \langle P, N \rangle, k \in P).$$

But using definition 4.42, this is equivalent to

$$x \in \theta(s_n) \leftrightarrow x = \langle q, \hat{k} \rangle \text{ (with } q(\theta(n)) = \langle P, N \rangle, k \in P).$$

And from definition 4.39, this is what it means to be in $s_{\theta(n)}$. □

Lemma 4.45. *For $\theta \in \mathfrak{G}$ for some $\mathfrak{G} \in \mathfrak{F}$, and for any world $p \in G$, and any $n, k \in \omega$, we have*

$$p \vDash (\hat{k} \; \varepsilon \; s_n) \text{ if and only if } p \vDash (\hat{k} \; \varepsilon \; \theta(s_n)).$$

Proof. First, we observe that for any $\psi \in \mathfrak{G}$ for some $\mathfrak{G} \in \mathfrak{F}$, we have

$$p \vDash (\hat{k} \; \varepsilon \; s_n) \text{ if and only if } \psi(p) \vDash (\hat{k} \; \varepsilon \; s_{\psi(n)}).$$

This is a consequence of theorem 4.29 and lemma 4.45.

Taking the case where $\psi = \theta^{-1}$:

$$p \vDash (\hat{k} \; \varepsilon \; s_n) \text{ if and only if } \theta^{-1}(p) \vDash (\hat{k} \; \varepsilon \; s_{\theta^{-1}(n)}).$$

Since $\theta \in \mathfrak{G}_{n+1}$ and $\mathfrak{G}_{n+1}$ is a group, $\theta^{-1} \in \mathfrak{G}_{n+1}$, and it must be the identity on all numbers less than $n + 1$, so we can substitute on the right-hand side:

$$p \vDash (\hat{k} \; \varepsilon \; s_n) \text{ if and only if } \theta^{-1}(p) \vDash (\hat{k} \; \varepsilon \; s_n).$$

Applying theorem 4.29 to the right-hand side, we get:

$$p \vDash (\hat{k} \ \varepsilon \ s_n) \text{ if and only if } p \vDash \theta(\hat{k} \ \varepsilon \ s_n).$$

We apply the definition of $\theta(\phi)$ for formula ϕ to get:

$$p \vDash (\hat{k} \ \varepsilon \ s_n) \text{ if and only if } p \vDash (\hat{k} \ \varepsilon \ \theta(s_n)).$$

$\square$

Theorem 4.46. *For each $n \in \omega$, s_n is $\mathfrak{F}$-invariant.*

Proof. s_n was given in definition 4.39 as

$$s_n = \{\langle p, \hat{k} \rangle \mid p \in G, p(n) = \langle P_n, N_n \rangle \text{ and } k \in P_n\}.$$

Fix n for the rest of the discussion. Clearly each $\hat{k}$ is $\mathfrak{F}$-invariant, so we just need to show that condition 1 of definition 4.34 holds, i.e. that s_n is $\mathfrak{G}$-invariant for some $\mathfrak{G}$ in the automorphism sequence $\mathfrak{F}$. It will be $\mathfrak{G}_{n+1}$-invariant. For any world p and any natural number k we can use lemma 4.45 to get

$$p \vDash (\hat{k} \ \varepsilon \ s_n) \text{ if and only if } p \vDash (\hat{k} \ \varepsilon \ \theta(s_n)).$$

Since each element of s_n (as a classical set in **M**) is of the form $\langle p, \hat{k} \rangle$, we can conclude from the statement above and the definition of ε that in **M** we have $\forall x (x \in s_n \leftrightarrow x \in \theta(s_n))$. As classical sets, s_n and $\theta(s_n)$ are identical. As modal sets, they are equal at every world. The permutation θ has no effect on s_n. From this it easily follows that $p \vDash [\![s_n \approx \theta(s_n)]\!]$ at every world p.

$\square$

Theorem 4.47. *For each $n \in \omega$, t_n is $\mathfrak{F}$-invariant. (This proof does not appear in [SF96]).*

Proof. As in the previous theorem, we will show that for any world p, and any $k, n \in \omega$,

$$p \vDash s_k \ \varepsilon \ t_n \text{ if and only if } p \vDash s_k \ \varepsilon \ \theta(t_n).$$

We know that $p \vDash s_k \ \varepsilon \ t_n$ if and only if $k \in I_n$. Let θ be any automorphism in any $\mathfrak{G} \in \mathfrak{F}$. Then $\theta^{-1} \in \mathfrak{G}$ also. By definition of θ, we know $\theta^{-1}(k) \in I_n$, so

$$p \vDash s_k \ \varepsilon \ t_n \text{ if and only if } p \vDash s_{\theta^{-1}(k)} \ \varepsilon \ t_n.$$

By definition 4.39, every world sees the same immediate members of t_n, so we have

$$p \vDash s_k \; \varepsilon \; t_n \text{ if and only if } \theta^{-1}(p) \vDash s_{\theta^{-1}(k)} \; \varepsilon \; t_n.$$

By theorem 4.29 this is equivalent to

$$p \vDash s_k \; \varepsilon \; t_n \text{ if and only if } p \vDash \theta(s_{\theta^{-1}(k)}) \; \varepsilon \; \theta(t_n).$$

And by lemma 4.44, this is in turn equivalent to

$$p \vDash s_k \; \varepsilon \; t_n \text{ if and only if } p \vDash s_k \; \varepsilon \; \theta(t_n).$$

Invariance follows easily as in the previous theorem, and again we have classical identity: in $\mathbf{M}$, $t_n = \theta(t_n)$. $\square$

As for T, its invariance follows easily from that of t_n. And again, $T = \theta(T)$.

So $s_n \in D_{\mathfrak{F}}$ and $t_n \in D_{\mathfrak{F}}$ for all n, and $T \in D_{\mathfrak{F}}$. Now, finally, we show that $[\![AC]\!]$ is not valid in $\mathbb{M}_{\mathfrak{F}}$. In fact, it fails at every possible world. In order to show this, we need a classical formula ϕ that expresses some form of the Axiom. We choose form 2 from theorem 2.1, which says roughly that for every set a there is a function F that chooses one element of each element of a. We avoid referring to F as a "choice function" since this term has been reserved for a different form of the Axiom. Suppose we have the formula ϕ. It will be in the form $\forall y \exists x (AC(x,y))$ where $AC(x,y)$ is a formula that says "x chooses one element of each element of y". If we show that for some y there is no x that can satisfy the formula, we will be asserting the failure of the Axiom of Choice. So how do we define $AC(x,y)$? The details are not really necessary. Using some of the simple formulas listed in theorem 2.19, we could define:

$$\mathrm{AC}(x,y) = \mathrm{Function}(x) \wedge \forall z (z \in \mathrm{Dom}(x) \leftrightarrow z \in y \wedge z \neq \emptyset) \wedge (\forall z \in y)[z \neq \emptyset \rightarrow (\exists w \in z)(\langle z, w \rangle \in x)].$$

Theorem 4.48. *Let* T, s_n, *and* t_n *be as given in definition 4.39. Then for any set* $F \in D_{\mathfrak{F}}$, $[\![\neg AC(F,T)]\!]$ *is valid in* $\mathbb{M}_{\mathfrak{F}}$.

Proof. We prove the result by way of contradiction. In this proof we must be careful to restrict our formulas to the correct language when interpreting them in the submodel $\mathbb{M}_{\mathfrak{F}}$. We still allow all of our defined symbols in $\mathbb{M}$. When we need to show equality in the submodel, we use $\approx$ equality in $\mathbb{M}$ and apply theorem 4.38.

Suppose the result is false, i.e. there is a world p and a function $F \in D_{\mathfrak{F}}$ that chooses an element from each element of T at p, i.e. $p \vDash_{\mathfrak{F}} [\![AC(F,T)]\!]$. So $p \vDash_{\mathfrak{F}} [\![\mathrm{Function}(F)]\!]$ and $p \vDash_{\mathfrak{F}} [\![\forall z(z \in$

$\mathrm{Dom}(F) \leftrightarrow z \in T \wedge z \neq \emptyset)]$ and $p \Vdash_{\mathfrak{F}} [(\forall z \in T)[z \neq \emptyset \rightarrow (\exists w \in z)(\langle z, w \rangle \in F)]]$. By item 1 of lemma 4.6, these statements are also true in any worlds accessible from p in the model $\mathbb{M}_{\mathfrak{F}}$.

Now, since $F \in D_{\mathfrak{F}}$, F is $\mathfrak{F}$-invariant and thus $\mathfrak{G}_n$-invariant for some n. We fix n at this value for the rest of the discussion. $p \Vdash_{\mathfrak{F}} [t_n \in T]$, so p must also think that F chooses an element of t_n, i.e. $p \Vdash_{\mathfrak{F}} [(\exists w \in t_n)(\langle t_n, w \rangle \in F)]$. That implies that for some world p' accessible from p, and some element a of the domain $D_{\mathfrak{F}}$, $p' \Vdash_{\mathfrak{F}} [a \in t_n \wedge \langle t_n, a \rangle \in F]$. Since $p' \Vdash_{\mathfrak{F}} [a \in t_n]$ and the models $\mathbb{M}$ and $\mathbb{M}_{\mathfrak{F}}$ agree on atomic statments, we have $p' \Vdash [a \in t_n]$, i.e. for some $b \in D$, $p' \Vdash [a \approx b \wedge b \, \varepsilon \, t_n]$. This means that for some world p'' accessible from p' we have $p'' \Vdash b \, \varepsilon \, t_n$. So b must be s_k for some $k \in I_n$, by definition 4.39, and we have $p' \Vdash [a \approx s_k]$ and by theorem 4.38, $p' \Vdash_{\mathfrak{F}} [a = s_k]$. Thus by substitutivity of equality, $p' \Vdash_{\mathfrak{F}} [s_k \in t_n \wedge \langle t_n, s_k \rangle \in F]$.

We now turn our attention away from p' to find a world q that thinks some other s_m ($m \neq k$) is the element of t_n chosen by F. Then we show that p' and q are compatible, that is, there is a world r accessible from both of them. At r, we have an easy contradiction. The role of q will be played by $\theta(p')$, and we now turn to the definition of θ.

By the definition of the possible worlds of our model, and the I-partition of the natural numbers, we know there exists a number m such that $m \neq k$ and $m \in I_n$ (I_n is infinite), and such that $p'(m) = \langle \emptyset, \emptyset \rangle$ (a possible world must take this value at all but a finite number of naturals). Now, let θ be the permutation of ω that simply swaps m and k, and is the identity on all others. By definition of I_n, both k and m are greater than or equal to n, so θ is the identity on all numbers below n. Also, θ only moves numbers in the same I-partition, and it is the identity on all but a finite number of naturals. Thus $\theta \in \mathfrak{G}_n$.

We already know that $p' \Vdash_{\mathfrak{F}} [\langle t_n, s_k \rangle \in F]$. Applying theorem 4.29, we get $\theta(p') \Vdash_{\mathfrak{F}} [\langle \theta(t_n), \theta(s_k) \rangle \in \theta(F)]$. We also know that t_n and F are $\mathfrak{G}_n$-invariant, so using theorem 4.38, we have $[\theta(t_n) = t_n]$ and $[\theta(F) = F]$ at every world of $\mathbb{M}_{\mathfrak{F}}$ and thus $\theta(p') \Vdash_{\mathfrak{F}} [\langle t_n, \theta(s_k) \rangle \in F]$.

Note that we cannot replace $\theta(s_k)$ with s_k since we just know that s_k is $\mathfrak{G}_{k+1}$-invariant (by theorem 4.46) and $k + 1 > n$. In fact, $\theta(s_k) = s_{\theta(k)}$ by theorem 4.44, so $\theta(s_k) = s_m$ and by lemma 4.40, s_k and $\theta(s_k)$ are two different sets since $k \neq m$. This is important to the rest of the argument. (Strictly speaking, lemma 4.40 only gives the result using $\approx$ equality in $\mathbb{M}$, not $=$ equality in $\mathbb{M}_{\mathfrak{F}}$, but this follows easily using theorem 4.38). Now we have $\theta(p') \Vdash_{\mathfrak{F}} [\langle t_n, s_m \rangle \in F]$

So p' thinks that s_k is the distinguished element of t_n and $\theta(p')$ thinks it is s_m. But $s_k \neq s_m$. This in itself is not a contradiction. Different worlds can force incompatible statements, but if they have a common accessible world, that world will contradict itself and this is impossible. In this case,

from the definition of θ we can conclude that there is a world r accessible from both p' and $\theta(p')$, which gives us the contradiction we need.

We know that p' and $\theta(p')$ are the same at all natural numbers except k and m, since these are the only numbers moved by θ. We also know that $p'(m) = \langle \emptyset, \emptyset \rangle$ (m was chosen to make this true) and thus that $\theta(p')(k) = \langle \emptyset, \emptyset \rangle$ (since $\theta(p')(k) = p'(\theta^{-1}(k)) = p'(m)$). Now, we choose a world r that agrees with p' and $\theta(p')$ on all naturals except k and m. Let $r(k) = p'(k)$ and $r(m) = \theta(p')(m)$. By the definition of the set of possible worlds and accessibility relation in $\mathbb{M}$ (and thus in $\mathbb{M}_{\mathfrak{F}}$, we can see that r is accessible from both p' and $\theta(p')$.

Now, r asserts a number of statements which are not consistent with each other. It asserts that $[\![\langle t_n, s_k \rangle \in F]\!]$ and $[\![\langle t_n, s_m \rangle \in F]\!]$. It also says $[\![s_k \neq s_m]\!]$. Finally, it says $[\![\text{Function}(F)]\!]$. But a function must be single-valued, so we must abandon the hypothesis that a set F exists that chooses one element of each element of T.   □

Since we now have a model that contains a set T for which the formula $[\![AC(F, T)]\!]$ fails for all F, we cannot possibly have the validity of the S4 translate of the Axiom of Choice—$[\![\forall x \exists y AC(y, x)]\!]$—in this model (or even its truth at a single world), and the plan given at the beginning of this section is complete: the negation of AC is consistent with ZF. Together with the result from section 3, we have the full independence of the Axiom.

Theorem 4.48 was given in a different form by Paul Cohen in [Coh66], though he obtained the result in 1963. At that time he also used forcing to prove the independence of another set theoretical statement, the Generalized Continuum Hypothesis. This is the well-known statement that the cardinality of the power set of x is the next cardinal after the cardinality of x. Many other independence results have been established using forcing since 1963.

The standard forcing technique is not modal. Possible worlds are referred to as *forcing conditions*. "p is accessible from q" is usually expressed as "p strengthens q". And what we have called the domain of our modal model is referred to as the class of *names*. Statements that we made using the machinery of modal logic (such as $p \vDash \Box \Diamond (x \in y)$) are made outside the system (so "for each forcing condition q that strengthens p, there is a condition r that strengthens q such that $p \vDash x \in y$"). All work is done within a classical model of ZF.

Other approaches to forcing exist, most notably the method of Boolean-valued models, due to Dana Scott and Robert Solovay (see [Bel85]). In this approach, a partially-ordered set of conditions is created, a domain is constructed (much like our $\mathcal{D}$), and truth values are assigned to the formulas from a complete Boolean algebra rather than the standard {true, false}. The notion of forcing

can then be captured by statements about the Boolean algebra, just as we captured it by modal statements about possible worlds. The result is the same.

The modal approach of Smullyan and Fitting cleverly omits some of the requirements of the standard approach that aren't needed to establish independence but are needed to construct models in which certain formulas are true or false. The modal approach establishes independence but doesn't settle the truth or falsity of each formula in the language of set theory. The standard approach does this by restricting the set of forcing conditions (possible worlds) in such a way that a formula is never forced true by one condition and false by another, and every formula is eventually forced either true or false by some condition. In this way a complete classical model is described while the freedom to force statements (as we did when creating sets of accessible worlds) is maintained.

Cohen's contribution to set theory was significant. To a large extent, his forcing method was new and unexpected, and it caused a period of high activity in the field in the 1960's and 1970's. Some work had been done on independence; as mentioned above, Fraenkel had developed a method of constructing permutation models in which the Axiom of Choice fails over a universe that contains urlements. This method is described in detail in [Jec73], and its influence can be seen on the proof of the full independence of AC by Cohen. Also, the Axiom had been proven independent of a version of set theory with no urelements but where the Axiom of Foundation was dropped (see [Moo82, p.302]). After these various partial independence results, as Moore points out, "it would have astonished set theorists if someone had shown that the Axiom of Choice was really a theorem of ZF." So the result was not surprising, but the method and its widespread applicability were.

An account of the discovery of forcing, by Cohen himself, can be found in [Coh02]. There are also historical summaries in [Kun80] and [SF96] of the development of the technique after Cohen's original publication, including Scott and Solovay's reformulation of the forcing technique using Boolean-valued models. Today it is an essential tool in set theory. As Dana Scott points out, "...there is simply no comparison whatsoever in the sophistication of our knowledge about models for set theory today as contrasted to the pre-Cohen era." (Foreword of [Bel85])

Reviewing the search for these independence proofs from 1904 to 1963 provides a fascinating glimpse into the evolution of the foundations of mathematics in the 20th century. So the Axiom of Choice is independent of ZF set theory: what does that mean for mathematics? For many mathematicians outside the realm of logic and set theory, it may not mean anything. But it is an important philosophical question, and the independence proofs allow it to be asked in a meaningful and non-trivial way. And, as we have discussed above, the methods of proof inspired by it have

forever enriched mathematics itself.

5   Appendix: Properties of the S4 translate

The following are elementary properties of the S4 translate, as introduced in defnition 4.4 in section 4.1, and are given here without proof. They are listed in an appendix for reference and to avoid cluttering proofs.

Throughout this section, let p be a world of some S4-ZF model, ϕ and ψ be formulas, and R be the accessibility relation between possible worlds of the model.

1. If $p \vDash [\![\phi]\!]$ and pRp' then $p' \vDash [\![\phi]\!]$.

2. If $p \nvDash [\![\phi]\!]$ then for some p' such that pRp', $p' \vDash [\![\neg\phi]\!]$.

3. If $p \vDash [\![\neg\phi]\!]$ then $p \nvDash [\![\phi]\!]$.

4. $p \vDash [\![\phi \wedge \psi]\!]$ if and only if $p \vDash [\![\phi]\!]$ and $p \vDash [\![\psi]\!]$.

5. If $p \vDash [\![\phi]\!]$ or $p \vDash [\![\psi]\!]$ then $p \vDash [\![\phi \vee \psi]\!]$.

6. If $p \vDash [\![\phi \rightarrow \psi]\!]$ and $p \vDash [\![\phi]\!]$, then $p \vDash [\![\psi]\!]$.

7. If $p \nvDash [\![\phi \rightarrow \psi]\!]$ then there is some world p such that pRp' and such that $p' \vDash [\![\phi]\!]$ and $p' \vDash [\![\neg\psi]\!]$.

8. If $p \vDash [\![\phi(x)]\!]$ for some $x \in D$, then $p \vDash [\![\exists x(\phi(x))]\!]$.

9. If $p \vDash [\![\exists x(\phi(x))]\!]$ then for some world p' such that pRp', there is some $x \in D$ such that $p' \vDash [\![\phi(x)]\!]$.

10. $p \vDash [\![\forall x(\phi(x))]\!]$ if and only if for all $x \in D$, $p \vDash [\![\phi(x)]\!]$.

References

[BdRV01] P. Blackburn, M. de Rijke, and Y. Venema. *Modal logic.* Cambridge University Press, Cambridge, 2001.

[Bel85] J. L. Bell. *Boolean-Valued models and independence proofs in set theory, 2nd edition.* Oxford University Press, Oxford, 1985.

[Coh63] P. J. Cohen. The independence of the continuum hypothesis, I. *Proc. Nat. Acad. Sci. U.S.A.*, 50:105–110, 1963.

[Coh66] P. J. Cohen. *Set Theory and the Continuum Hypothesis.* Benjamin, New York, 1966.

[Coh02] P. J. Cohen. The discovery of forcing. *Rocky Mountain Journal of Mathematics*, 32:1071–1100, 2002.

[End77] H. Enderton. *Elements of Set Theory.* Academic Press, Los Angeles, 1977.

[End87] H. Enderton. *Set Theory,* volume 12 of *Encyclopedia of physical science and technology,* pages 545–553. Academic Press, Los Angeles, 1987. Also available at http://www.math.ucla.edu/~hbe/.

[Fra58] A.A. Fraenkel. *Foundations of Set Theory.* North-Holland Publishing Company, Amsterdam, 1958.

[Göd40] K. Gödel. *The consistency of the axiom of choice and the generalized continuum-hypothesis with the axioms of set theory.* Princeton University Press, Princeton, N.J., 1940.

[Jec73] T. Jech. *The Axiom of Choice.* North-Holland Publishing Company, Amsterdam, 1973.

[Jec03] T. Jech. *Set Theory.* Springer, Berlin, 2003.

[Kun80] K. Kunen. *Set Theory: An Introduction to Independence Proofs.* North-Holland Publishing Company, Amsterdam, 1980.

[Lev65] A. Levy. A hierarchy of formulas in set theory. *Memoirs of the American Mathematical Society,* 57, 1965.

[Moo82] G. Moore. *Zermelo's Axiom of Choice: Its origins, development, and influence.* Springer-Verlag, New York, 1982.

[RR63]   H. Rubin and J. Rubin. *Equivalents of the Axiom of Choice.* North-Holland Publishing Company, Amsterdam, 1963.

[SF96]   R. Smullyan and M. Fitting. *Set Theory and the Continuum Problem.* Oxford Science Publications, Oxford, 1996. See also errata at http://comet.lehman.cuny.edu/fitting/errata/errata.html.

[TZ70]   G. Takeuti and W.M. Zaring. *Introduction to Axiomatic Set Theory.* Springer-Verlag, New York, NY, 1970.

[Zer04] E. Zermelo. Proof that every set can be well-ordered (reprinted in van heijenoort 1967). *Mathematische Annalen*, 59, 1904.

[Zer08] E. Zermelo. A new proof of the possibility of a well-ordering (reprinted in van heijenoort 1967). *Mathematische Annalen*, 59, 1908.