Mathematics Institute
University of Warwick

MA469 Project

Finite set theory, arithmetic, and interpretations between them

Daniel Wood
Supervisor: Dr Adam Epstein

March 2009

Abstract

We develop the theories of Peano Arithmetic and Zermelo-Fraenkel set theory minus the Axiom of Infinity with the aim of proving the bi-interpretability result of [8]. We then very briefly discuss the recent result proved in [12]. The novelty is that we take a bottom-up approach to these results, assuming no particular knowledge of formal logic, and that we offer more detail than is usually given in this area.

Contents

1 Introduction **3**

2 Background logic **3**
2.1 Formal languages 3
2.2 Models 7

3 PA and ZF−inf **8**
3.1 Peano arithmetic 8
3.2 ZF−inf 13

4 Interpretations **25**
4.1 Interpretations via the ordinal interpretation 25
4.2 Interpretations: a rigorous definition and a categorical perspective . 26

5 The Ackermann interpretation **28**

6 The inverse Ackermann interpretation **31**

7 Interpretations and bounded formulae: a brief vista **35**

1 Introduction

The aim of this essay is to develop, from the ground up, interpretations between finite set theory and arithmetic. Accordingly, this essay has been written in mind for a fourth-year mathematics student who has little to no training in formal logic, in particular a Warwick MMath student. Knowledge of formal logic will of course be beneficial, and indeed there are some parts of this essay which only those readers with some previous knowledge of logic will properly appreacite, but it will not be essential. We also aim to give more detail than is usually found in treatments of this subject; we shall develop finite set theory particularly rigorously due to its interesting yet accessible subtlties.

In Section 2 we shall cover the basics of formal logic and informally look at models. In Section 3 we will develop the theories of Peano Arithmetic and Zermelo-Fraenkel set theory minus the Axiom of Infinity. We shall develop these theories in their own right, especially the latter, although we shall always have interpretations in mind. In Section 4 we shall develop the theory of interpretations, firstly via an example and then more abstractly, briefly mentioning a categorical approach. In Section 5 we shall go through the Ackermann interpretation of finite set theory in arithmetic. In Section 6 we shall then cover its inverse. Finally in Section 7 we shall very briefly look at a recent result in this area.

2 Background logic

2.1 Formal languages

In this subsection we shall give a brief overview of formal languages and how we shall deal with them. Our coverage really will only be brief and so some understanding of formal logic would be very beneficial, although three years of university mathematics should be enough. A curious and/or confused reader can find help in any one of the many books that have been written on this area; the author is particularly fond of [4] and [2], the latter being more elementary and less mathematical (it is primarily aimed at philosophy students). Also, [13], although not an introduction to formal logic, has some wonderful exposition of formal languages in its third chapter.

A *formal language* is a language with a strictly defined vocabularly and syntax. There are a whole host of different formal languages around in mathematics, but the vast majority have the same underlying structure: each has a set of constant symbols, possibly empty; an arbitrarily large set of variables;[1] a set of finite-arity relation symbols, again possibly empty (although that would be rather boring); the identity/equality symbol '$=$'; the Boolean connectives '$\vee$', '$\wedge$', '$\neg$', '$\rightarrow$', and '$\leftrightarrow$'; parentheses '(' and ')'; the quantifiers '$\forall$' and '$\exists$'; and a collection of rules of inference, i.e. rules that allow one to derive statements from previous statements.[2] We shall also include function symbols, although this isn't strictly necessary, since a function is just a special type of relation.

We shall now cover some terminology and notation. A *formula* of (or in) a given language is a finite string of symbols of the language. So, for example, if R is a

[1] By 'arbitrarily large', we mean that we can always add more variables. This can be done quite easily: simply affix natural numbers or primes: $x_1, x_2, x_3, \ldots$ or $x', x'', x''', \ldots$.

[2] A brief note for any computer scientists, recusrion theorests, or logicians reading this is called for: these rules of inference should be *recursive*. For the uninitiated, this means that we could programme a computer, say in C++, to check whether a given use of a rule is valid. This can be and is indeed done: for example, the programme Fitch that accompanys [2] does precisely this.

binary relation, S is a uninary relation, and x and y are variables of our given language, then

$$\to y \forall R) S(\exists x \tag{1}$$

and

$$\forall x(S(x) \to \exists y R(x, y)) \tag{2}$$

are formulae of our language. Now, hopefully the reader's intuition will have led them to think that (1) doesn't make any sense. This is indeed the case: the formula is not *well-formed*, i.e. it does not follow the syntax. But how do we actually define the syntax? Well, we first need to define *terms* and *atomic formulae*. A *term* is anyone of the following: a constant symbol; a free variable; or $f(x_1, x_2, \ldots, x_n)$, where f is an n-ary function symbol and the x_i are constant symbols or free variables (or a mixture of the two). A formula is called *atomic* iff it is of the form $t_1 = t_2$ or $R(t_1, t_2, \ldots, t_n)$, where R is an n-ary relation symbol and the t_i are terms. We can now define the syntax. We do this inductively: all atomic formulae are well-formed; and if φ and ψ are well-formed, then so are

$$\neg\varphi,\ \varphi \vee \psi,\ \varphi \wedge \psi,\ \varphi \to \psi,\ \varphi \leftrightarrow \psi,\ \exists x\, \varphi, \text{ and } \forall x\, \varphi,$$

where x is any variable.[3] So (2), for example, is well-formed; such a formula is called a *well-formed formula* (or *wff* for short). Since non-well-formed formulae are somewhat pointless to study, from now we shall on use the word 'formula' to mean 'wff', i.e. from now on, all of our formulae will be well-formed.

At this point we shall make a quick note about the order of quantifiers. The order in which quantifiers are placed is *very* important. We illustrate this with the definition of continuity of a function $f\colon \mathbb{R} \to \mathbb{R}$ at $c \in \mathbb{R}$, which should be familiar to the reader:

$$\forall \varepsilon > 0\, \exists \delta > 0\ |x - c| < \delta \to |f(x) - f(c)| < \varepsilon.$$

If we were to switch the order of the quantifiers, the meaning would be quite different:

$$\exists \delta > 0\, \forall \varepsilon > 0\ |x - c| < \delta \to |f(x) - f(c)| < \varepsilon.$$

The first defintion says that for a given $\varepsilon > 0$ we can find a $\delta > 0$ such that the property holds; that is, δ depends on ε. The second definition says that there is a $\delta > 0$ such that the property holds for *all* $\varepsilon > 0$; in particular, δ does not depend on ε.

A variable in a formula that is in the range of a quantifier is said to be *bound*; a variable that is not bound is said to be *free*. So, in the example below, x is bound while y is free:

$$\exists x R(x, y). \tag{3}$$

A formula in which all variables are bound is called a *sentence*. So formula (2) is a sentence, for example. If $\varphi(x_1, x_2, \ldots, x_n)$ is a formula with free variables precisely $x_1, x_2, \ldots, x_n$ and $t_1, t_2, \ldots, t_n$ are terms in our language, then we write

$$\varphi(c_1, c_2, \ldots, c_n)$$

to denote the formula with the free variables $x_1, x_2, \ldots, x_n$ replaced by $t_1, t_2, \ldots, t_n$ (in that order). So, for example, if $\varphi(y)$ denotes formula (3) and c is a constant symbol, then $\varphi(c)$ denotes the sentence

$$\exists x R(x, c).$$

[3] The syntax of a formal language should also be recursive, although this is fairly clear from the way it is defined.

Unless otherwise specified, we write $\varphi(x)$ to denote a formula φ that has *precisely* one free variable.

We make a brief note here about a common notation. Many authors write $\varphi[x/\bar{a}]$ to denote a formula φ with a free variable x and terms $a_1, a_2, \ldots, a_n$ for some n, the latter being called *parameters*. They do this so they can talk about formulae more flexibly, as they can vary the parameters of a given formula, rather than having to talk about a new formula for each set of parameters. We, however, shall not adopt this notation, since for the purposes of this essay such notation is unnecessary and, moreover, the author believes that our notation is slightly clearer, especially to those readers new to formal logic.

All our languages will be *first-order*, which means that quantifiers range, and only range, over variables; they do not range over formulae. The significance of this shall be explained later. The term *first-order* logic refers to the the framework upon which first-order languages are built, i.e. the Booleon connectives, the identity symbol $=$, the quantifiers $\forall$ and $\exists$, a set of variables, and a standard collection of rules of inference, which we shall list shortly.

An *axiom* of a given language is a sentence that is taken to be true,[4] i.e. the axioms of a language are premises from which one can deduce statements using the language's rules of inference.

A *theory* in a language is a consistent set of sentences – by *consistent* we mean that we cannot deduce a contradiction from them, a contradiction being a sentence of the form $\varphi \wedge \neg\varphi$. We shall usually define a theory by specifying a set of axioms, the theory being all sentences that can be derived from the axioms.

We now need to actually state our rules of inderence. There are several equivalent ways of presenting formal proofs in first-order logic, perhaps the three most notable being the Fitch-style or system of natural deduction, as in [2], for example; the proof-tree style, as in [3] and [11], for example; and the sequential calculus, as in [4], for example. We shall adopt the last of these. The sequential calculus ivolves the symbol '$\vdash$', which is read as 'turnstile', 'proves', or 'yields'. For a set of sentences Γ and a sentence φ in a given language, the expression

$$\Gamma \vdash \varphi$$

means that φ can be formally proved from Γ. It is with this notation that we shall state our rules of inference, of which there are 19.[5] They fall into four catgories: structural, Boolean, quantifier, and identity.

Let Γ, Δ be arbitrary sets of sentences; φ, ψ, θ be arbitrary sentences; c be an arbitrary constant symbol; and s, t be arbitrary terms.

1. Structural rules:

   (i) If $\varphi \in \Gamma$, then $\Gamma \vdash \varphi$. (*Assumption*)

   (ii) If $\Gamma \subseteq \Delta$ and $\Gamma \vdash \varphi$, then $\Delta \vdash \varphi$. (*Monotonicity*)

   (iii) If $\Gamma \vdash \Delta$ and $\Delta \vdash \varphi$, then $\Gamma \vdash \varphi$. (*Cut*)

2. Boolean rules:

   (i) If $\Gamma \vdash \varphi \wedge \psi$, then $\Gamma \vdash \varphi$. ($\wedge$-*Elimination*)

[4] I *really* don't want to get into a discussion about truth here, but an argument from an axiom A should be thought of as 'if A, then...'. Whether or not A is actually true (whatever that means) is beside the point: our proofs are conditional.

[5] Perhaps we should call this the *Hardcastle calculus*!

(ii) If $\Gamma \vdash \varphi$ and $\Gamma \vdash \psi$, then $\Gamma \vdash \varphi \wedge \psi$. ($\wedge$-*Introduction*)

(iii) If $\Gamma \vdash \varphi \vee \psi$, $\Gamma \cup \{\varphi\} \vdash \theta$, and $\Gamma \cup \{\psi\} \vdash \theta$, then $\Gamma \vdash \theta$. ($\vee$-*Introduction*)

(iv) If $\Gamma \vdash \varphi$, then $\Gamma \vdash \varphi \vee \psi$. ($\vee$-*Introduction*)

(v) If $\Gamma \vdash \neg\neg\varphi$, then $\Gamma \vdash \varphi$. ($\neg$-*Elimination*)

(vi) If $\Gamma \cup \{\varphi\} \vdash \psi \wedge \neg\psi$, then $\Gamma \vdash \neg\varphi$. ($\neg$-*Introduction*)
(This is often called *proof by contradiction.*)

(vii) If $\Gamma \vdash \varphi$ and $\Gamma \vdash \varphi \to \psi$, then $\Gamma \vdash \psi$. ($\to$-*Elimination*)

(viii) If $\Gamma \cup \{\varphi\} \vdash \psi$, then $\Gamma \vdash \varphi \to \psi$. ($\to$-*Introduction*)

(ix) If $\Gamma \vdash \varphi \leftrightarrow \psi$ and $\Gamma \vdash \varphi$, then $\Gamma \vdash \psi$. ($\leftrightarrow$-*Elimination*)

(x) If $\Gamma \vdash \varphi \to \psi$ and $\Gamma \vdash \psi \to \varphi$, then $\Gamma \vdash \varphi \leftrightarrow \psi$. ($\leftrightarrow$-*Introduction*)

3. Quantifier rules:

(i) If $\Gamma \vdash \forall x\varphi(x)$, then $\Gamma \vdash \varphi(c)$. ($\forall$-*Elimantion*)

(ii) If $\Gamma \vdash \varphi(c)$ and $\varphi(c) \notin \Gamma$, then $\Gamma \vdash \forall x\varphi(x)$. ($\forall$-*Introduction*)

(iii) $\Gamma \vdash \exists x\varphi(x)$, $\Gamma \cup \{\varphi(c)\} \vdash \psi$, and $\varphi(c) \notin \Gamma$, then $\Gamma \vdash \psi$. ($\exists$-*Elimination*)

(iv) $\Gamma \vdash \varphi(c)$, then $\Gamma \vdash \exists x\varphi(x)$. ($\exists$-*Introduction*)

4. Identity rules:

(i) If $\Gamma \vdash s = t$ and $\Gamma \vdash \varphi(s)$, then $\Gamma \vdash \varphi(t)$. (=-*Elimination*)

(ii) $\Gamma \vdash t = t$. (=-*Introduction*)

Let us run through two examples of formal proofs, the first being fairly easy and the second being a little trickier:

Proposition 2.1. $\Gamma \vdash \varphi \wedge \psi$ *iff* $\Gamma \vdash \psi \wedge \varphi$.

Proof.

1.	$\Gamma \vdash \varphi \wedge \psi$	(Premise)
2.	$\Gamma \vdash \varphi$	($\wedge$-Elimation: 1)
3.	$\Gamma \vdash \psi$	($\wedge$-Elimation: 1)
4.	$\Gamma \vdash \psi \wedge \varphi$	($\wedge$-Introduction: 2,3)

We have the other direction by simply running this proof in reverse. □

Proposition 2.2. $\Gamma \vdash \varphi \vee \psi$ *iff* $\Gamma \vdash \neg(\neg\varphi \wedge \neg\psi)$.

Proof.

1.	$\Gamma \vdash \varphi \vee \psi$	(Premise)
2.	$\Gamma \cup \{\neg\varphi \wedge \neg\psi\} \cup \{\varphi\} \vdash \neg\varphi \wedge \neg\psi$	(Assumption)
3.	$\Gamma \cup \{\neg\varphi \wedge \neg\psi\} \cup \{\varphi\} \vdash \neg\varphi$	($\wedge$-Elimination: 2)
4.	$\Gamma \cup \{\neg\varphi \wedge \neg\psi\} \cup \{\varphi\} \vdash \varphi$	(Assumption)
5.	$\Gamma \cup \{\neg\varphi \wedge \neg\psi\} \cup \{\varphi\} \vdash \varphi \wedge \neg\varphi$	($\wedge$-Introduction: 3, 4)
6.	$\Gamma \cup \{\varphi\} \vdash \neg(\neg\varphi \wedge \neg\psi)$	($\neg$-Introduction: 5)
7.	$\Gamma \cup \{\neg\varphi \wedge \neg\psi\} \cup \{\psi\} \vdash \neg\varphi \wedge \psi$	(Assumption)
8.	$\Gamma \cup \{\neg\varphi \wedge \neg\psi\} \cup \{\psi\} \vdash \neg\psi$	($\wedge$-Elimination: 7)
9.	$\Gamma \cup \{\neg\varphi \wedge \neg\psi\} \cup \{\psi\} \vdash \psi$	(Assumption)
10.	$\Gamma \cup \{\neg\varphi \wedge \neg\psi\} \cup \{\psi\} \vdash \psi \wedge \neg\psi$	($\wedge$-Introduction: 8, 9)
11.	$\Gamma \cup \{\psi\} \vdash \neg(\neg\varphi \wedge \neg\psi)$	($\neg$-Introduction: 10)
12.	$\Gamma \vdash \neg(\neg\varphi \wedge \neg\psi)$	($\vee$-Elimination: 1, 6, 11)

1.	$\Gamma \vdash \neg(\neg\varphi \wedge \neg\psi)$	(Premise)
2.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \cup \{\varphi\} \vdash \neg(\varphi \vee \psi)$	(Assumption)
3.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \cup \{\varphi\} \vdash \varphi$	(Assumption)
4.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \cup \{\varphi\} \vdash \varphi \vee \psi$	($\vee$-Introduction: 3)
5.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \cup \{\varphi\} \vdash (\varphi \vee \psi) \wedge \neg(\varphi \vee \psi)$	($\wedge$-Introduction: 2, 4)
6.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \vdash \neg\varphi$	($\neg$-Introduction: 5)
7.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \cup \{\psi\} \vdash \neg(\varphi \vee \psi)$	(Assumption)
8.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \cup \{\psi\} \vdash \psi$	(Assumption)
9.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \cup \{\psi\} \vdash \varphi \vee \psi$	($\vee$-Introduction: 8)
10.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \cup \{\psi\} \vdash (\varphi \vee \psi) \wedge \neg(\varphi \vee \psi)$	($\wedge$-Introduction: 7, 9)
11.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \vdash \neg\psi$	($\neg$-Introduction: 10)
12.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \vdash \neg\varphi \wedge \neg\psi$	($\wedge$-Introduction: 6, 11)
13.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \vdash \neg(\neg\varphi \wedge \neg\psi)$	(Monotonicity: 1)
14.	$\Gamma \cup \{\neg(\varphi \vee \psi)\} \vdash (\neg\varphi \wedge \neg\psi) \wedge \neg(\neg\varphi \wedge \neg\psi)$	($\wedge$-Introduction: 12, 13)
15.	$\Gamma \vdash \neg\neg(\varphi \vee \psi)$	($\neg$-Introduction: 14)
16.	$\Gamma \vdash \varphi \vee \psi$	($\neg$-Elimination: 15)

□

Now, the reader may be worried at this point: 'Surely we don't have to construct such detailed proofs everytime?' Thankfully, we do not. The idea is that we present proofs in such a way that it is clear that we *could*, if we really wanted to, write them in a completely formal way as above, although of we never do of course. Smith summed it up quite nicely:

Sufficient unto the day is the rigour thereof.[6]

Moreover, we shall discuss a result in the next subsection that further justifies such informal proofs, the Soundess and Completeness Theorem of first-order logic.

The last point that we shall cover in this subsection is the word *meta*. We use 'meta' to refer to something *about* a formal theory (or theories), rather than something *in* a formal theory. So, for example, the Soundness and Completeness Theorem that we shall briefly mention in the next subsection is a *meta*theorem, since is it a theorem about formal theories, rather than a theorem in a formal theory. Also, strictly speaking, Propositions 2.1 and 2.2 are metapropositions, since we stated them in English; however, we can easily reformulate them as propositions in the logic: $\Gamma \vdash \varphi \wedge \psi \leftrightarrow \psi \wedge \varphi$ and $\Gamma \vdash \varphi \vee \psi \leftrightarrow \neg(\neg\varphi \wedge \neg\psi)$.

2.2 Models

We will only discuss models informally and briefly, as they are not essential to this essay, although some knowledge of them will aid understanding of some points later on.

We first shall informally define an $\mathcal{L}$-structure. An *$\mathcal{L}$-structure* is a mathemtical object in which we interpret the non-logical symbols of $\mathcal{L}$, [7] i.e. we assign constants, relations, and functions in the structure to each of the constant, relation, and function symbols in $\mathcal{L}$. A *model* of an $\mathcal{L}$-theory T is then an $\mathcal{L}$-structure in which the axioms of T are true. We shall not properly define truth in a model, as it a non-trivial task and fairly involved; the details can be found in any logic textbook, for example [3] or [4]. Instead, we shall give some examples that will hopefully give the reader an intuitive understanding of these concepts.

[6] p. 18 of [13].

[7] The reader should not confuse this with interpreting one theory in another, something that we will be looking at extensively later on in this essay.

Consider the language $\mathcal{L}_G$ of group theory, where the non-logical symbols are 'e' and '$\circ$'. The theory of groups T_G is then the theory in $\mathcal{L}_G$ generated by the usual group axioms. A model of T_G is then simply a group, such as $(\mathbb{Z}/n\mathbb{Z}, +_n)$ or $(\mathbb{R}, \times)$. In the first example we interpret e as 0 and $\circ$ as addition modulo n. In the second example we interpret e as 1 and $\circ$ as multiplication in the reals; of course, we could also make $\mathbb{R}$ into a model of T_G by interpreting e as 0 and $\circ$ as addition. Similarly, we could make $\mathbb{Z}/n\mathbb{Z}$ into a group by interpreting e as 1 and $\circ$ as multiplication modulo n.

Let us consider another example. Let $\mathcal{L}_F$ be the language of fields, with non-logical symbols '0', '1', '+', and '$\times$'. The theory T_F of fields is then the $\mathcal{L}_F$-theory generated by the usual field axioms. A model of T_F would then be a field, such as $\mathbb{C}$ with the natural interpretations of the symbols; or $\mathbb{Z}/p\mathbb{Z}$ for some prime p with 0 and 1 interpreted as 0 and 1 respectively, and + and $\times$ interpreted as addition and multiplication modulo p respectively.

We shall now breifly discuss the Soundness and Completeness Theorem of first-order logic. The proof is quite involved and so we shall only state the theorem.[8] The theorem states that *syntactical* reasoning, i.e. formal proofs, and *semantic* reasoning, i.e deducing results inside models, are the same. More precisely, if we write $\Gamma \models \varphi$ to mean that every model of Γ is also a model of φ,[9] then the Soundness and Completeness Theorem says

$$\Gamma \vdash \varphi \text{ if and only if } \Gamma \models \varphi.$$

As we mentioned at the end of the previous subsection, this theorem further justifies informally reasoning, as it tells us that deducing statements semantically, i.e. by resoning about truth inside an arbitrary model, is equivalent to deducing statements syntactically, i.e. proving them formally. Another important consequence of this theorem is that it justifies the use of counter examples, something that we use in mathematics all the time. So, if we wish to prove that $\Gamma \nvdash \varphi$, i.e. that we cannot deduce φ from Γ, then it suffices to find a model in which Γ is true but φ is false, since $\Gamma \not\models \varphi$ implies $\Gamma \nvdash \varphi$ (where $\Gamma \not\models \varphi$ means that there exists a model in which Γ is true but φ is false).

As said previously, a proper understanding of models is not a prerequisite for reading this essay, although it is helpful. However, model theory is both an interesting and thriving area of research, especially here in the UK, and should the reader wish to investigate the area for his- or herself, the author would recommend [9] as a good place to start.

3 PA and ZF−inf

3.1 Peano arithmetic

Peano Arithmetic, or 'PA' for short, was developed by the Italian mathematician Giuseppe Peano at the end of the 19$^{\text{th}}$ century. There are several ways to formulate PA, all of which are equivalent; our choice of formulation is arbitrary, and is largely based on that in [7]. PA is written in the first-order language of arithmetic, denoted $\mathcal{L}_A$, which of consists of the usual Boolean connectives, quantifiers, and identity symbol; two constant symbols, '0' and '1'; two binary function symbols, '+' and '$\cdot$';

[8] Proofs can be found in all good logic textbooks, such as [3] or [4].

[9] There can be some confusion here: $\models$ is also used to indicate that sentence or set of sentences are ture in given $\mathcal{L}$-structure. So, for example, we might write $(\mathbb{R}, +) \models T_G$. The meaning of $\models$ is usually clear from the context.

and a binary relation, '$<$'. We shall use infix notation for $+$, $\cdot$, and $<$; that is, we shall write '$x+y$', '$x \cdot y$' (or sometimes just 'xy'), and '$x < y$', rather than '$+(x,y)$', '$\cdot(x,y)$', and '$<(x,y)$'. Also, to avoid excessive paremtheses, we let $\cdot$ be dominant over $+$. So, for example, $w \cdot x + y \cdot z$ is short for $(w \cdot x) + (y \cdot z)$. PA is then the $\mathcal{L}_A$-theory based on the following 15 axioms and one axion schema:

(Ax1)	$\forall x \forall y \forall z((x+y)+z = x+(y+z))$	(associativity of $+$)
(Ax2)	$\forall x \forall y(x+y = y+x)$	(commutativity of $+$)
(Ax3)	$\forall x \forall y \forall z((x \cdot y) \cdot z = x \cdot (y \cdot z))$	(associativity of $\cdot$)
(Ax4)	$\forall x \forall y(x \cdot y = y \cdot x)$	(commutativity of $\cdot$)
(Ax5)	$\forall x \forall y \forall z(x \cdot (y+z) = x \cdot y + x \cdot z)$	(distributivity of $\cdot$ over $+$)
(Ax6)	$\forall x(x + 0 = x \wedge x \cdot 0 = 0)$	(0 is the additive identity)
(Ax7)	$\forall x(x \cdot 1 = x)$	(1 is the multiplicative identity)
(Ax8)	$\forall x \forall y \forall z((x < y \wedge y < z) \to x < z)$	(transitivity of $<$)
(Ax9)	$\forall x(\neg x < x)$	($<$ is a strict ordering)
(Ax10)	$\forall x \forall y(x < y \vee y < x \vee x = y)$	($<$ is a total (or linear) ordering)
(Ax11)	$\forall x \forall y \forall z(x < y \to x + z < x + z)$	($<$ is a invariant under addition by a constant)
(Ax12)	$\forall x \forall y \forall z(0 < z \wedge x < y \to x \cdot z < y \cdot z)$	($<$ is a invariant under non-zero multiplication)
(Ax13)	$\forall x \forall y(x < y \to \exists z(x + z = y))$	(subtraction)
(Ax14)	$0 < 1 \wedge \forall x(x > 0 \to (x > 1 \vee x = 1)$	($<$ is a discrete ordering)
(Ax15)	$\forall x(x > 0 \vee x = 0)$	(0 is the least element)

We now come to the axiom schema. Let φ be a $\mathcal{L}_A$-formula with precisely one free variable. Then $\mathrm{I}\varphi$ is defined to be the sentence

$$(\varphi(0) \wedge \forall x(\varphi(x) \to \varphi(x+1)) \to \forall y \varphi(y).$$

The *axiom schema of induction* (or just *induction* for short) is then the collection of all sentences $\mathrm{I}\varphi$ for every $\mathcal{L}_A$-formula φ with precisely one free variable. This is an axiom *schema* because it is a collection of axioms; as we mentioned in the previous subsection, in first-order logic quantifiers only range over variables, so we cannot express the phrase 'for all formulae φ' in a first-order language. Thus PA is not finitely axiomatised, although it is recursively axiomatised: we can programme a computer to check whether a given application of induction is a valid one.

One might initially think that PA precisely describes the natural numbers. However, there are models of PA which are different[10] to $\mathbb{N}$ (see [7]). Such models are called *nonstandard*. We shall not cover them, but it worth the reader's while to be at least aware of their existence.[11]

We can already prove a useful result:

Lemma 3.1. *The z in* (Ax13) *is unique.*

Proof. Suppose that we have z and z' such that $x + z = y$ and $x + z' = y$. Then $x + z = x + z'$. Suppose that $\neg z = z'$. Then, by (Ax9) and (Ax10), $z < z'$ or $z' < z$. Without loss of generality, assume that $z < z'$. Then, by (Ax11), $x + z < x + z'$, which is a contradiction by (Ax9), since $x + z = x + z'$. □

With this lemma in mind, we can rewrite (Ax13) as

$$\forall x \forall y(x < y \to \exists! z(x + z = y)),$$

[10] The technical word is *nonisomorphic*.

[11] For those more familiar with logic: Such models are a consequence of the Incompleteness Theorem; they can also be constructed using the Compactness Theorem.

where $\exists! x R(x)$ is an abbreviation for the formula

$$\exists x(R(x) \wedge \forall y(R(y) \to x = y)).$$

Now is a good point to make a couple of remarks. Firstly, one might reasonably ask why we didn't just specify that the z in (Ax13) be unique in the first place. Well, logically speaking – and we mean this in the meta sense, not the formal one – it is better to make one's axioms as minimal as possible, since this makes the theory generated by the axioms less likely to be contradictory and, from a pragmatic point of view, easier to deal with.[12] Secondly, we should note that we don't include abbreviations as part of the language, also for this last reason. '$\exists!$', despite it's appearance, is not a first-order symbol; it is simply a useful shorthand which is used on the understanding that, if we *really* wanted to, we could do without it. We will be using many such abbreviations in this essay and so it is worth the reader's while to try to understand this point. Let us introduce some more now: '$x \leqslant y$' is abbreviation for $x < y \vee x = y$; '$x \neq y$' is an abbreviation for $\neg x = y$; and '$\exists x < y\, R(x)$' and '$\forall x < y\, R(x)$' are abbreviations for $\exists x(x < y \wedge R(x))$ and $\forall x(x < y \to R(x))$ respectively (and similarly for $\leqslant$).

We will now develop some theoretical machinery in PA that will be invaluable to us later when we deal with interpretations. In particular, we will define functions in PA that are defined recusively, most importantly exponentiation. To do this, we will have to construct recursion in PA. The way to do this is to work through some of the theory that the great mathematician and logician Kurt Gödel developed to prove his famous (and most beautiful) 1931 Incompleteness Theorems.[13] We shall only give a sketch of how one would proceed through such theory, as the proofs involved are very technical and take a *lot* of time to go through; the reader can find the relevant details in [7].

We start by defining a simple function, the *cut-off substraction* of y from x, which is denoted $x \dot{-} y$. Informally, $x \dot{-} y$ is the larger of $x - y$ and 0; formally, $x \dot{-} y = z$ is an abbreviation for the formula

$$(y \leqslant x \wedge x + z = y) \vee (x < y \wedge z = 0). \tag{4}$$

This abbreviation is justified by Lemma 3.1: z is uniquely determined by x and y. Moreover, this function is *total*, i.e. it is defined for all x and y (this is also a consequence of Lemma 3.1). This brings us to a key point: the formula (4) is in fact the *graph* of $x \dot{-} y$. More generally:

Definition 3.2. The *graph* of a function f is the formula $\varphi(x_1, x_2, \ldots, x_n, z)$ (where n is the arity of f) such that

$$f(x_1, x_2, \ldots, x_n,) = z \leftrightarrow \varphi(x_1, x_2, \ldots, x_n, z).$$

Now, $\dot{-}$ isn't actually a function symbol in $\mathcal{L}_A$ and so one might point out that its graph isn't actually defined. However, the following definition and theorem justify its use as a function (and hence the existence of its graph):

Definition 3.3. Let T be a theory in a first-order language $\mathcal{L}$. A theory T^* in a language $\mathcal{L}^*$ is an *extension* of T iff $\mathcal{L} \subseteq \mathcal{L}^*$ and all the axioms of T are also axioms of T^*. T^* is a *conservative* extension of T iff for every $\mathcal{L}$-sentence φ, if $T^* \vdash \varphi$ then $T \vdash \varphi$.

[12] As we will see later, when we start dealing with interpretations between theories, that having as few axioms as possible makes life a lot easier, since it means that there is less to prove.

[13] The author highly recommends [13] should the reader wish to learn these theorems, which the author thinks he/she really should, just like everyone at some point in their life should watch a play by Shakespeare, hear a symphony by Mozart, or see a portrait by Raphael.

Theorem 3.4. *Let T be theory in a first-order language $\mathcal{L}$. Suppose that an $\mathcal{L}$-formula ψ with $n+1$ free variables is such that $T \vdash \forall x_1 \forall x_2 \dots \forall x_n \exists! z \psi(x_1, x_2, \dots, x_n, z)$. If we define $\mathcal{L}^*$ to be the language $\mathcal{L}$ with a new n-ary function symbol f and T^* to be the $\mathcal{L}^*$-theory T with the extra axiom*

$$\forall x_1 \forall x_2 \dots \forall x_n \forall z (f(x_1, x_2, \dots, x_n) = z \leftrightarrow \psi(x_1, x_2, \dots, x_n, z)),$$

then T^* is a conservative extension of T.

Proof. We shall not prove this theorem in this essay due to its length and required background theory. Instead, the reader is referred to [3]. □

This theorem means that we can add $\dot{-}$ to PA as a function symbol without fear of changing PA's strength. We could of course stick to using (4) as an abbreviation, but adding $\dot{-}$ as a function makes life a *lot* easier, since it means we can talk about terms defined from applications of $\dot{-}$, rather than having to stick to using the graph of $\dot{-}$. Both methods are equivalent, in the sense that any theorems produced using the former could be reformulated in terms of the latter, but the former really will save a lot of unnecessary labour, especially when it comes to composing $\dot{-}$ with other functions.

We shall be employing this technique a lot, not just in PA but also to finite set theory later in the essay: whenever we come across a formula of the form of ψ in Theorem 3.4, we shall introduce a function symbol that defines the function in question.

We now move onto slightly harder functions, which are based on the following lemma:

Lemma 3.5 (Euclidean division in PA)**.** $\forall x \forall y (0 < x \to \exists! r \exists! q (y = qx + r \wedge r < x))$.

Proof. We shall first prove existence by induction on y. If $y = 0$ then $q = 0$ and $r = 0$ fit the bill. Now let $y > 0$ and suppose we have $qx + r = y$ and $r < x$ for some r and q. Then $y + 1 = qx + (r + 1)$. $r < x$ and so $r + 1 \leqslant x$ by Lemma below. If $r \leqslant x$ then we are done. If $r = x$, then $y + 1 = qx + x = (q + 1)x + 0$ by (Ax5), (Ax6), and (Ax7) and so we are done because $0 < x$. Thus existence is proved by induction.

We now prove uniqueness. Suppose $y = qx + r$ and $y = q'x + r'$ for some q, q' and $r, r' < x$. Then $qx + r = q'x + r'$. Suppose $\neg q = q'$. Then, by (Ax10), $q < q'$ or $q' < q$; assumme without loss of generality that $q < q'$. Then $qx < q'x$ by (Ax12). Thus $qx + r < q'x + r$ by (Ax11) and so $y < y$, a contradiction by (Ax9). Thus $q = q'$. Suppose $\neg r = r'$. Without loss of generality assume $r < r'$. Then $qx + r < qx + r'$ by (Ax11) and so $y < y$, a contradiction by (Ax9). Thus $r = r'$ and we are done. □

With this lemma in hand we can define the quotient and remainder functions:

Definition 3.6. For $x > 0$, we define the graph of Q to be

$$Q(y/x) = q \leftrightarrow \exists r (y = qx + r \wedge r < x)$$

and the graph of R to be

$$R(y/x) = r \leftrightarrow \exists q (y = qx + r \wedge r < x).$$

As we did with $\dot{-}$, we add Q and R as function symbols.

We adopt Kaye's notation because it makes it clear which number is dividing which: $Q(y/x)$ is far less likely to cause confusion than $Q(y, x)$ (and similarly for R). With the exception of $x = 0$,[14] both Q and R are total.

We shall now sketch how one develops recursion in PA. Firstly, let us consider how we use recursion in everyday mathematics: we define a sequence of numbers by specifying how it starts and then giving a rule for how to get the next number in the sequence from the previous number(s). So, to develop recursion in PA, it is enough to work out how to encode a finite sequence of numbers as a single number in such a way that we can recover any element of the sequence from this number and, crucially, extend the sequence indefinitely. But how do we encode such sequences in PA? Well, let us work out how we might do it in $\mathbb{N}$; the method for PA is the same, except that various points need to be proved quite carefully and functions need to be defined properly (we have in fact done some of the latter). As we said earlier, the precise details can be found in [7].

Consider a finite sequence of numbers $x_0, x_1, \ldots, x_{n-1}$. Let $m = b!$, where $b = max(n, x_0, x_1, \ldots, x_{n-1})$. Then the sequence of numbers

$$m + 1, 2m + 1, 3m + 1, \ldots, nm + 1$$

is pairwise coprime: Suppose that $u \mid im + 1$ and $u \mid jm + 1$ for some $0 < i < j \leqslant n$, $1 \leqslant u$.[15] Then $u \mid (jm + 1) - (im + 1) = (j - i)m$. But $0 < j - i, n \leqslant b$ and so $j - i \mid m$. Now, either $u \mid m$ or $u \mid j - i$ (or both). If $u \mid j - i$, then $u \mid m$ because $j - i \mid m$. Thus in both case we have $u \mid m$, and so $u \mid im$. Thus $u \mid (im + 1) - im = 1$ and so $u = 1$. Now recall the Chinese Remainder Theorem:[16]

Theorem 3.7. *Let R be a Euclidean domain and let $n_1, n_2, \ldots, n_k$ be non-zero coprime elements of R. Then for any $a_1, a_2, \ldots, a_k$ there exists $x \in R$ such that*

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \\ &\vdots \\ x &\equiv a_k \pmod{n_k}. \end{aligned}$$

We can apply this theorem to obtain $a \in \mathbb{N}$ such that

$$a \equiv x_i \pmod{((i + i)m + 1)}$$

for all $i < n$. Then the pair (a, m) encodes the sequence $x_0, x_1, \ldots, x_{n-1}$, since we can obtain any x_i from (a, m):

$$x_i = R\left(\frac{a}{m(i + 1) + 1}\right).$$

We can then encode the pair (a, m) as a single number by setting

$$\langle a, m \rangle = \frac{(a + m)(a + m + 1)}{2} + m.$$

This does indeed encode the pair because $\langle .. , .. \rangle \colon \mathbb{N}^2 \to \mathbb{N}$ is a bijection.

Using this machinery, we can define recursive functions, specifically exponention, 2^x, and summation, $\sum_{y \leqslant x} y$, both of which will be invaluable to us later.

[14] We can't divide by zero!

[15] $x \mid y$ has its usual meaning of x divides y.

[16] This should be familiar to all Warwick mathematics students from MA249 Algebra II.

3.2 ZF−inf

We shall now develop the first-order theory of Zermelo–Fraenkel set theory minus infinity ('ZF−inf' for short). ZF−inf is written in the first-order language of sets, $\mathcal{L}_\in$, whose only non-logical symbol is $\in$, which is binary relation known as *membership*. When we talk about ZF−inf informally, the phrase 'x is a set' simply means '$\exists x$'; in ZF−inf, everything is a set. There are so-called *class–set* theories out there, perhaps most notably the Bernays–Gödel theory of sets (BG),[17] which have two kinds of objects, sets and classes, but we shall be working soley in a set-only set theory. For us, classes will simply be abbreviations; that is, for a class $C = \{x : \varphi(x)\}$, writing $z \in C$ will simply be shorthand for $\varphi(z)$. Note that every set is a class: if x is a set, then $x = \{y : y \in x\}$. A class that is not a set is called a *proper* class.

We will now list the axioms of ZF−inf. We write them informally, i.e. in words, and then write them in $\mathcal{L}_\in$. We shall interspace the axioms with some remarks and some theory. Our formulation is based on that in [6], although it does differ slightly.

Axiom of Extensionality. *If x and y have the same elements, then $x = y$. Formally:*

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

Not that the converse of this statement, i.e. if $x = y$ then x and y have the same elements, is a consequence of the =-Elimination rule.

Axiom of the Empty Set. *There exists a set with no elements. Formally:*

$$\exists x \forall z (\neg z \in x).$$

The use of the definite article in the name of this axiom is justified: this empty set is unique by Extensionality. We shall use the abbreviation $\varnothing$ as shorthand for this set.

Axiom of Pairing. *For all x and y, there exists a set $\{x, y\}$ that precisely contains, x and y. Formally:*

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

The reader is no doubt familiar with the 'curly brackets' or 'braces' notation $\{\ldots\}$; indeed, we have already used it to talk about classes. The symbols '{', '}' are not part of the the language $\mathcal{L}_\in$, although they are very useful as abbreviations. In some instances we can add them to the language without having to worry: for example, just as we used Theorem 3.4 in the previous subsection to add $\dot{-}$ to PA, we can add $\{.., ..\}$ as a binary function symbol to $\mathcal{L}_\in$, its existence justified by Pairing and its functionality justified by Extensionality. We can do the same for the empty set, introducing $\varnothing$ as a 0-place function symbol.

Axiom Schema of Separation. *If φ is a formula with precisely one free variable, then for any x the class $y = \{z \in x : \varphi(x)\}$ exists. Formally:*

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

This is an axiom *schema* for the same reason induction is a schema: we can't quantify over formulae. An interesting aside is that this axiom schema resolves Russell's Paradox; well, it shifts the blame to the universal class, $V := \{x : x = x\}$:

[17] See p. 70 of [6] for the axioms of BG

if V were a set, we could apply Separation to it with the formula $\neg x \in x$ and then run through the usual argument. Thus, in ZF$-$inf, V is not a set. Historically, Russell's Paradox was one of the motivations behind axiomatising set theory, since in naïve set theory we have unrestricted construction of sets. Separation is much weaker: it allows unrestricted construction of *sub*sets, but not of sets in their own right.

Separation also allows us to take intersections over classes: For any nonempty class C, the class

$$\bigcap C := \{z : z \in x \text{ for every } x \in C\}$$

is a set, since we can apply Separation to any set in the class.

Axiom of Union. *For any x there exists $y = \bigcup x$, the set of all elements of members of x. Formally:*

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

As we did with $\{.., ..\}$, we introduce $\bigcup$ as a function symbol. We also introduce the binary function symbol $\cup$, defined by

$$x \cup y = \bigcup \{x, y\}.$$

Axiom of Power Set. *For every x there exists $y = \mathcal{P}(x)$, the set of all subsets of x. Formally:*

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

We introduce $\mathcal{P}(x)$ as a function symbol. We shall also introduce the abbreviation $x \subseteq y$ for the formula

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

Thus we can restate the Aiom of Power Set informally as 'for all x, the set $\mathcal{P}(x) = \{y : y \subseteq x\}$ exists'.

We now come to the negation of the Axiom of Infinity. The Axiom of Infinity states that there exists an *inductive set*; that is, a set S such that $\varnothing \in S$ and $x \in S$ implies $x \cup \{x\} \in S$. Before we state the axiom, we had better check that $x \cup \{x\}$ actually exists. Consider the power set of x. $\{x\} \in \mathcal{P}(x)$ and so by applying Separation to $\mathcal{P}(x)$ with the formula

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

we have that $\{x\}$ is a set. We then have that $x \cup \{x\}$ by the Axioms of Pairing and Union. We can now state the axiom:

The Negation of the Axiom of Infinity. *There does not exist an iductive set. Formally:*

$$\neg \exists x (\varnothing \in x \wedge \forall z (z \in x \rightarrow z \cup \{z\} \in x)).$$

Notice that in the above formula we have used some of the function symbols that we introduced over during this subsection. As a reality check, let us explicitly show that they aren't strictly necessasy. We can write the negation of the Axiom of Infinity purely in $\mathcal{L}_\in$:

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

Hopefully the almost impenetrable nature of this statemtn will convince the reader of the merits of introducing function symbols.

Before we introduce the next axiom, we need to develop some terminology. A formula F with precisely two free variables is said to be *functional* if it has the following property:

$$\forall x \forall y \forall z((F(x,y) \wedge F(x,z)) \rightarrow y = z).$$

The word 'functional' has been chosen for obvious reasons. We shall talk about functions in ZF$-$inf in more detail later, but for the time being we adopt the notation $y = F(x)$ for the formula $F(x,y)$.

Axiom Schema of Replacement. *If F is a functional formula, then for every x the set $y = \{F(w) : w \in x\}$ exists. Formally:*

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

Perhaps the best way to think of this axiom schema is 'the image of a set under a function is a set'.

We now come to the last axiom:

Axiom of Foundation.[18] *Every nonempty set has an $\in$-minimal element. Formally:*

$$\forall x(\neg x = \varnothing \rightarrow \exists y(y \in x \wedge y \cap x = \varnothing)).$$

We shall now run through some theory, which is based on that found in Chapters 1 and 2 of [6]. Firstly, let us define ordered pairs in ZF$-$inf. There are several ways of doing this, but the following is the most popular:

Definition 3.8. (x,y) is an abbreviation for the term $\{\{x\},\{x,y\}\}$.

This does indeed define an ordered pair:

Lemma 3.9. *ZF$-$inf* $\vdash (w,x) = (y,z) \leftrightarrow (w = y \wedge x = z)$.

Proof. $(w = y \wedge x = z) \rightarrow (w,x) = (y,z)$ is simply a consequence of $\wedge$-Elimination.

For the converse, suppose that

$$\{\{w\},\{w,x\}\} = \{\{y\},\{y,z\}\}. \tag{5}$$

First assume that $y = z$. Then $\{\{y\},\{y,z\}\} = \{\{y\},\{y\}\} = \{\{y\}\}$. Thus $\{\{w\},\{w,x\}\} = \{\{y\}\}$. Since the two sets are equal, they must have the same elements and thus $\{w\} = \{w,x\} = \{y\}$, and so $w = x = y = z$ and the lemma holds.

Now asume that $y \neq z$. Suppose that $\{w\} = \{y,z\}$. The two sets are equal and so must have the same elements, so $w = y = z$, a contradiction. So $\{w\} \neq \{y,z\}$. Thus, by 5, we must have $\{w\} = \{y\}$ and so $w = y$. Then, again by 5, we must have $\{w,x\} = \{y,z\}$. But $w = y$ and so $\{w,x\} = \{w,z\}$ and thus $x = z$ and we are done. □

Now that we have ordered pairs, we can talk about ordered n-tuples by setting

$$(x_1, x_2, \ldots, x_n) = ((x_1, x_2, \ldots, x_{n-1}), x_n).$$

This now allows us to construct relations in ZF$-$inf:

Definition 3.10. An *n-ary* relation is a class of ordered n-tuples.

[18] This is also known as the *Axiom of Regularity.*

We can view a relation *symbol* as a relation by considering all n-tuples that satisfy the relation symbol. In ZF$-$inf there is only one relation symbol, $\in$, which we can view as a relation by considering the (proper) class of ordered pairs

$$\{(x, y) : x \in y\}.$$

We are now able to talk properly about functions in ZF$-$inf. So far, we have introduced function symbols (as justified by Theorem 3.4) and seen functional formulae. In general, a *function* f is a binary relation such that $(x, y) \in f$ and $(x, z) \in f$ imply $y = z$. (Notice that we can talk about n-ary functions by letting x be an ordered n-tuple.) So a functional formula F can be viewed as function by considering the class of ordered pairs

$$\{(x, y) : F(x, y)\}.$$

Similarly, we can view function symbols as functions. So, for example, we can view $\mathcal{P}$ as a function by considering the (proper) class of ordered pairs

$$\{(x, y) : \mathcal{P}(x) = y\}.$$

We now introduce some terminology and notation. The *domain* of a function f, denoted $\mathrm{dom}(f)$, is the class

$$\mathrm{dom}(f) = \{x : \exists y\, (x, y) \in f\}.$$

A function whose domain is a proper class is often called a *class function*. Notice that, by Replacement, if $\mathrm{dom}(f)$ is a set then f is a set. The *range* of f, denoted $\mathrm{ran}(f)$, is the class

$$\mathrm{ran}(f) = \{y : \exists x (x, y) \in f\}.$$

We write $f\colon x \to y$ if f $= x$ and $\mathrm{ran}(f) \subseteq y$. *Surjective* and *injective* have their usual definitions. The restriction of f to $z \subseteq \mathrm{dom}(f)$, denoted $f \restriction x$, is the function

$$f \restriction z = \{(x, y) \in f : x \in z\}.$$

We shall shorlty introduce a type of object that will be crucial later on in this essay: ordinals. But first we need some preliminary definitions:

Definition 3.11. A set x is *transitive* if every member of x is a subset of x, that is

$$\forall z (z \in x \to z \subseteq x).$$

At first one might wonder whether such sets exist. Well, quite simple examples exist. We will develop this properly later, but in ZF$-$inf we can build up the natural numbers as follows:

$$0 = \varnothing, 1 = \{0\}, 2 = \{0, 1\}, \ldots, n = \{0, 1, 2, \ldots, n - 1\}, \ldots.$$

These are known as the *von Neumann ordinals*. It is fairly easy to see that all of these sets are transitive, although we will need to develop some more theory to see this rigorously.

Definition 3.12. Let x be a set. A binary relation R is *strict partial ordering* on (or of) x iff

(i) $\neg R(z, z)$ for every $z \in x$ (the ordering is *strict*); and

(ii) for all $u, v, w \in x$, if $R(u, v)$ and $R(v, w)$, then $R(u, w)$ (R is *transitive*[19]).

If, in additon we have

(iii) for every $y, z \in x$ we have $R(x, y)$, $R(y, x)$ or $x = y$,

then R is said to be a *linear* (or *total*) ordering on x.

We shall usually denote orderings by $<$, using infix notation. We use the abbreviation '$x \leqslant y$' as shorthand for $x < y \vee x = y$.

Definition 3.13. Let x be a set linearly ordered by $<$. Let y be a nonempty subset of x. An element $a \in x$ is:

(i) a *greatest* element of y iff $a \in y$ and $\forall z(z \in x \to z \leqslant a)$;

(ii) a *least* element of y iff $a \in y$ and $\forall z(z \in x \to a \leqslant z)$;

(iii) an *upper bound* of y iff $\forall z(z \in x \to z \leqslant a)$;

(iv) a *lower bound* of y iff $\forall z(z \in x \to a \leqslant z)$;

(v) the *supremum* of y iff a is the least upper bound of y, in which case a is denoted '$\sup y$'; and

(vi) the *infimum* of y iff a is the greatest lower bound of y, in which case a is denoted '$\inf y$'.

Definition 3.14. Let R be an linear ordering on a set x. R is said to be a *well-ordering* iff every subset of x has an R-minimal element, i.e.

$$\forall y(y \subseteq x \to \exists w(\neg \exists z(z \in y \wedge R(z, w)))).$$

We can now define what an ordinal is:

Definition 3.15. An *ordinal* is a transitive set that is well-ordered by $\in$.

We shall often denote $\in$ by $<$ when talking about ordinals. We denote the class of all ordinals by 'Ord'. Let us prove some elementary results:

Lemma 3.16.

(i) $0 = \varnothing$ *is an ordinal.*

(ii) If α *is an ordinal and* $\beta \in \alpha$*, then* β *is an ordinal.*

(iii) If $\alpha \neq \beta$ *are ordinals and* $a \subseteq \beta$*, then* $\alpha \in \beta$*.*

(iv) If α*,* β *are ordinals, then either* $\alpha \subseteq \beta$ *or* $\beta \subseteq \alpha$*.*

[19]The reader should not confuse this with the meaning of 'transitive' in Definition 3.11. It is perhaps an unfortunate consequence of the historical developement of this area that the same word has been come to mean two different but yet closely related things. The proof of Lemma 3.16(ii) sheds some light on why the word has developed in this way.

Proof. (i) This is immediate from the definition of an ordinal.

(ii) We need to show that β is transitive and well-ordered by $\in$. Since α is an ordinal, it is transitive and thus $\beta \subseteq \alpha$. Thus β inherits the well-ordering of α. All that is left is to show that β is transitive. Consider some $\gamma \in \beta$. We to show that $\gamma \subseteq \beta$, i.e. $\delta \in \gamma \to \delta \in \beta$. $\beta \subseteq \alpha$ and so $\gamma \in \alpha$. Thus, since α is transitive, $\gamma \subseteq \alpha$. Thus $\delta \in \alpha$. So $\delta, \gamma, \beta \in \alpha$, $\delta \in \gamma$, and $\gamma\beta$. Thus $\delta \in \beta$ because $\in$ is a transitive *relation* on α. Thus $\gamma \subseteq \beta$ and we are done.

(iii) Let γ be the least element of $\beta - \alpha$, where $\beta - \alpha = \{x \in \beta : x \notin \alpha\}$ (which exists by applying Separation to β); we can take the least element because ordinals are *well*-ordered by $\in$. Consider some $\delta \in \gamma$. Then $\delta \in \beta$ because $\gamma \subseteq \beta$. Thus $\delta \in \alpha$ because γ is the *least* element of $\beta - \alpha$. Thus $\{\xi \in \beta : \xi \in \gamma\} \subseteq \alpha$. Now consider some $\sigma \in \alpha$. Then $\sigma \in \beta$ because $\alpha \subseteq \beta$ and so $\sigma \in \gamma$ because γ is the least element of $\beta - \alpha$. Thus $\{\xi \in \beta : \xi \in \gamma\} \subseteq \alpha$ and so $\{\xi \in \beta : \xi \in \gamma\} = \alpha$. Now, clearly $\{\xi \in \beta : \xi \in \gamma\} \subset \gamma$. But what about the converse? Well, suppose we have $z \in \gamma$ such that $z \notin \{\xi \in \beta : \xi \in \gamma\}$. Then, since $z \in \gamma$, we must have $z \notin \beta$. But this is a contradiction, since $\gamma \subseteq \beta$. So $\{\xi \in \beta : \xi \in \gamma\} = \gamma$ and thus $\alpha = \gamma \in \beta$.

(iv) By (ii) and (iii), $\alpha \cap \beta$ is an ordinal. Let $\gamma = \alpha \cap \beta$. Then $\gamma = \alpha$ or $\gamma = \beta$, for otherwise we have $\gamma \in \alpha$ and $\gamma \in \beta$ by (iii), and so $\gamma \in \gamma$, which contradicts $\in$ being a *strict* ordering. □

Now, ZF$-$inf is often called 'finite set theory', and so this would suggest that everything in ZF$-$inf is finite. This is indeed the case, but we shall need to prove this. To do this, we shall run through a series of lemmas and defintions which lead us to this result. The working is quite subtle, and it will require us at times to actually consider an inductive set; this does of course contradict $\neg$inf, but we need to consider such a set in order to properly understand infinity in ZF$-$inf (or rather its lack thereof).

By Lemma 3.16, we have the following results:

(i) $\in$ is a linear ordering of Ord.[20]

(ii) For every ordinal α, $\alpha = \{\beta : \beta \in \alpha\}$.

(iii) If C is a nonempty class of ordinals, then $\bigcap C$ is an ordinal, $\bigcap C \in C$ and $\bigcap C = \inf C$. Thus, by (i), $\in$ is a well-ordering of Ord; that is, every nonempty subclass C of Ord has a least element, namely $\bigcap C$.

(iv) If X is a nonempty set of ordinals, then $\bigcup X$ is an ordinal and $\bigcup = \sup X$.

(v) For every ordinal α, $\alpha \cup \{\alpha\}$ is an ordinal. We define $\alpha + 1$ to be $\alpha \cup \{\alpha\}$.

(vi) Ord is a proper class (otherwise consider $\sup(\text{Ord} + 1)$).

In light of these results, we can make some definitions:

Definition 3.17. An ordinal α of the form $\alpha = \beta + 1$ is called a *successor ordinal*. If α is not a successor ordinal, then $\alpha = \sup\{\beta : \beta \in \alpha\} = \bigcup \alpha$ and we call α a *limit ordinal*. We also consider $\varnothing$ to be a limit ordinal, defining $\sup \varnothing = \varnothing$.

We can now prove *transfinite induction*, or just *induction* for short:

[20] Definition 3.12 can easily be adapted for classes.

Theorem 3.18 (Transfinite induction)**.** *Let C be a class of ordinals such that*

(i) $\varnothing \in C$*;*

(ii) if $\alpha \in C$ *then* $\alpha + 1 \in C$*;*

(iii) if α *is a nonzero limit ordinal and* $\beta \in C$ *for all* $\beta \in \alpha$*, then* $\alpha \in C$*.*

Then $C = \mathrm{Ord}$*.*

Proof. Suppose that the theorem is false. Then, since $\in$ is a well-ordering of Ord, we can consider the least ordinal $\gamma \notin C$. By (i) we have that $\gamma \neq \varnothing$. Suppose that $\gamma = \beta + 1$ for some β. Since γ is the least ordinal not in C, $\beta \in C$. But then by (ii) we have acontradiction. Thus γ must be a limit ordinal. But since γ is the least ordinal not in C, $\beta \in C$ for every $\beta \in \gamma$, and so we have a contradiction by (iii). Thus no such γ can exist. □

As we said earlier, we wish to understand infinity in ZF−inf. To do this, we shall show that there are no limit ordinals in ZF−inf. But to do this, we need to consider inductive sets. We define

$$N = \bigcap \{X : X \text{ is inductive}\}.$$

Clearly such a set does not exist in ZF−inf, but we can consider it in ZF, where we include inf (rather than its negation). We shall now work through a series of lemmas (based on the exercises at the end of Chapter 1 of [6]) twhich show that N behaves like the natural numbers. Let us introduce the following notation: For $n \in N$, let $n + 1 = n \cup \{n\}$, and define $<$ on N by $n < m$ iff $n \in m$. We also reintroduce the notation

$$0 = \varnothing, 1 = \{0\}, 2 = \{0, 1\}, \ldots, n = \{0, 1, 2, \ldots, n - 1\}, \ldots$$

from earlier; these lemmas will lead us to a result that shows that the von Neumann ordinals are in fact all the ordinals in ZF−inf.

Lemma 3.19. *If X is inductive, then the set $\{x \in X : x \subseteq X\}$ is also inductive. Hence N is transitive and for each $n \in N$, $n = \{m \in N : m < n\}$.*

Proof. Let $X' = \{x \in X : x \subseteq X\}$. Clearly $\varnothing \in X'$ (since X is inductive). Consider some $x \in X'$. Since X is inductive, $x \cup \{x\} \in X$. We have $x \subseteq X$ by the definition of X and $\{x\} \subseteq X$ because $a \in X$. Thus $x \cup \{x\} \subseteq X$ and so $x \cup \{x\} \in X'$ and so X' is inductive.

By definition the definition of N, $N' \subseteq N$. Since N is inductive, N' is inductive. Thus, by the definition of N, $N \subseteq N'$. Thus $N = N'$ and so N is transitive.

Let $n \in N$. Then $n \subseteq N$ because N is transitive. Consider some $m \in n$; by definition we have $m < n$. So $n = \{m \in N : m < n\}$. □

Lemma 3.20. *If X is inductive, then the set $\{x \in X : x$ is transitive$\}$ is inductive. Hence every $n \in N$ is transitive.*

Proof. Let $X' = \{x \in X : x$ is transitive$\}$. Clearly $\varnothing \in X'$. Consider some $x \in X'$. Since X is inductive $x \cup \{x\} \in X$. Consider some $z \in x \cup \{x\}$. If $z \in x$, then, since x is transitive, $z \subseteq x$ and so $z \subseteq x \cup \{x\}$. If $z \in \{x\}$, then $z = x$ and so $z \subseteq x \cup \{x\}$. Thus $x \cup \{x\}$ is transitive and so $x \cup \{x\} \in X'$. Therefore X' is inductive

Clearly $N' \subseteq N$. Since N is inductive, N' is inductive. Thus $N' \subseteq N$ and so $N' = N$. Thus every $n \in N$ is transitive. □

Lemma 3.21. *If X is inductive, then the set $\{x \in X : x$ is transitive and $x \notin x\}$ is inductive. Hence $n \notin n$ and $n \neq n+1$ for every $n \in N$.*

Proof. Let $X' = \{x \in X : x$ is transitive and $x \notin x\}$. Clearly $\varnothing \in X'$. Consider some $x \in X'$. Since X is inductive, $x \cup \{x\} \in X$. By the Lemma 3.20 above, to show that X' is inductive it suffices to show that $x \cup \{x\} \notin x \cup \{x\}$. Suppose that $x \cup \{x\} \in x \cup \{x\}$. If $x \cup \{x\} \in x$, then since x is transitive, $x \cup \{x\} \subseteq x$. Thus, since $x \in x \cup \{x\}$, we have $x \in x$, which is a contradiction because $x \in X'$. If $x \cup \{x\} \in \{x\}$, then $x \cup \{x\} = \{x\}$ (since $\{x\}$ has precisely one element). So $\{x\} \subseteq x$ and so $x \in x$, which is again a contradiction. Therefore $x \cup \{x\} \notin x \cup \{x\}$ and so X' is inductive.

Clearly $N = N'$ (the same reasoning as the last to proofs applies). Thus $n \notin n$ for evey $n \in N$. Suppose $n = n+1$ for some $n \in N$. So $n = n \cup \{n\}$ and so $\{n\} \subseteq n$. Thus $n \in n$, a contradiction. Thus $n \neq n+1$ for every $n \in N$. □

Lemma 3.22. *If X is inductive, then $\{x \in X : x$ is transitive and every nonempty $z \subseteq x$ has an $\in$-minimal element$\}$ is inductive. (t is '$\in$-minimal' in z iff there does not exist $s \in z$ such that $s \in t$.)*

Proof. Let $X' = \{x \in X : x$ is transitive and every nonempty $z \subseteq x$ has an $\in$-minimal element$\}$. Clearly $\varnothing \in X'$. Consider some $x \in X'$. By Lemma 3.20 we have that $x \cup \{x\}$ is transitive. $x \notin x$: Suppose $x \in x$ and consider the nonempty subset $\{x\} \subseteq x$; this has no $\in$-minimal element, a contradiction. Consider some nonempty $z \subseteq x \cup \{x\}$. Let $z' = z \cap x$. If $z = z'$, then z has an $\in$-minimal element (since $z \subseteq x$). Now suppose that $z = z' \cap \{x\}$. Let t be an $\in$-minimal element of x. Since x is transitive, $t \subseteq x$. Thus, if $y \in t$, then $y \in y$, which is a contradiction, since $y \notin y$. So $y \notin t$. Therefore t is an $\in$-minimal element of z. Thus $x \cup \{x\} \in X'$ and so X' is inductive. □

Lemma 3.23. *Every nonempty $X \subseteq N$ has an $\in$-minimal element.*

Proof. Consider some $n \in X$. Suppose that $X \cap n = \varnothing$. Then $m \in X$ implies $m \notin n$ and so n is an $\in$-minimal element of X. Now suppose that $X \cap n \neq \varnothing$. By Lemma 3.22, $X \cap n$ has and $\in$-minimal element, u say. Consider some $a \in X$ such thatr $a \in u$. By Lemma 3.20, n is transitive and so $u \subseteq n$. Thus $a \in n$. Hence $a \in X \cap n$, which contradicts u being an $\in$-minimal element of $X \cap n$. Therefore no such a exists and so u is an $\in$-minimal element of X. □

Lemma 3.24. *If X is inductive then so is $\{x \in X : x = \varnothing$ or $x = y \cup \{y\}$ for some $y \in X\}$. Hence every nonzero $n \in N$ is of the form $n = m+1$ for some $m \in N$.*

Proof. Let $X'\{x \in X : x = \varnothing$ *or* $x = y \cup \{y\}$ *for some* $y\}$. Clearly $\varnothing \in X'$. Consider some $x \in X'$. Then $x \cup \{x\} \in X'$ by the definition of X'. Thus X' is inductive. Thus $N = N'$ and we are done. □

With all these lemmas behind us, we can prove induction for N, which we shall need later:

Theorem 3.25 (N-induction)**.** *Let $A \subseteq N$ such that $0 \in A$ and if $n \in A$ then $n+1 \in A$. Then $A = N$.*

Proof. Suppose that $N - A \neq \varnothing$. Then, by Lemma 3.23, $N - A$ has a $\in$-minimal element, u say. By Lemma 3.24, $u = n+1$ for some $n \in N$. Since $n < u$, $n \notin N - A$ (because u is an $\in$-minimal element of $N - A$). Thus, by the hypothesis, $u = n+1 \in A$, which is a contradiction. Thus $N - A = \varnothing$, i.e. $A = N$. □

We are now nearly ready to show that there are no nonzero limit ordinals in ZF−inf. Firstly, notice that if there is a nonzero limit ordinal, then there must be a least nonzero limit ordinal, since $\in$ is well-ordering of Ord. Denote the least limit ordinal, if it exists, by ω. We need another couple of lemmas:

Lemma 3.26. *α is a limit ordinal if and only if for every β, $\beta < \alpha$ implies $\beta+1 < \alpha$.*

Proof. The lemma holds vacuously for $\alpha = 0$, so assume that $\alpha \neq 0$. Let $\beta < \alpha$. Since α is a limit ordinal, $\alpha \neq \beta + 1$. Since $<$ is a linear ordering of Ord, either $\alpha < \beta + 1$ or $\beta + 1 < \alpha$. If $\alpha < \beta + 1$, i.e. $\alpha \in \beta \cup \{\beta\}$, then $\alpha = \beta$ because $\alpha \notin \beta$. This is a contradiction. Therefore $\beta + 1 < \alpha$.

Now suppose that α is an ordinal such that $\beta < \alpha \to \beta + 1 < \alpha$ holds for every β. Then α cannot be a successor ordinal and thus must be a limit ordinal. □

Lemma 3.27. *If a set X is inductive, then $X \cap \mathrm{Ord}$ is inductive. The set N, if it exists, is the least nonzero limit ordinal.*

Proof. Clearly $\varnothing \in X \cap \mathrm{Ord}$. Let $\alpha \in X \cap \mathrm{Ord}$. Since X is inductive, $\alpha \cup \{\alpha\} \in X$. Since α is an ordinal, $\alpha \cup \{\alpha\}$ is an ordinal. Thus $\alpha \cup \{\alpha\} \in X \cap \mathrm{Ord}$ and so $X \cap \mathrm{Ord}$ is inductive.

We have that $N \cap \mathrm{Ord}$ is inductive by our previous work. Thus, by the definition of N, $N = N \cap \mathrm{Ord}$. Since $N \subseteq \mathrm{Ord}$, $<$ is a linear ordering of N; by Lemma 3.23 it is in fact a well-ordering. By Lemma 3.19 we have that N is transitive. Thus N is itself an ordinal.

Since N is inductive, the implication $\alpha < N \to \alpha + 1 < N$ holds for every α. Thus, by Lemma 3.26, N is a limit ordinal. Clearly $N \neq 0$.

Suppose that there exists a nonzero limit ordinal M such that $M \leqslant N$. Then $0 \in M$ and the implication $\alpha < M \to \alpha + 1 < M$ holds for every α by Lemma 3.26. Thus M is inductive and so $N \subseteq M$. But $M \in N$ and so $M \subseteq N$ because N is transitive. Thus $M = N$. □

Definition 3.28. An ordinal n is a *finite ordinal* iff it is not greater than nor equal to any nonzero limit ordinal.[21] A set A is *finite* iff there exists a bijection $A \to n$ for some finite ordinal n. A set is *infinite* iff it is not finite.

Theorem 3.29. *The following are equivalent:*

(i) There exists an inductive set.

(ii) There exists an infinite set.

(iii) ω is a set.

Proof. [(i) → (ii)] Assume there exists an inductive set.

N is a set because the class $\{X : X \text{ is inductive}\}$ is nonempty. Thus, by Exercise 2.3 in Jech, $N = \omega$. We shall now prove that ω is infinite and hence that (ii) holds. To do this we will use N-induction (Theorem 3.25). Clearly there does not exist a bijection $\omega \to 0$. For the induction hypothesis, assume that there does not exist a bijection $\omega \to k$. To prove the case for $k + 1$, we shall suppose the negation and derive a contradiction. So suppose that $f\colon \omega \to k + 1$ is a bijection. Since f is a bijection, there exists a unique $l \in \omega$ such that $f(l) = k + 1$. By Lemma 3.24, either $l = m + 1$ for some $m \in \omega$ or $l = 0$. First suppose $l = m + 1$. Using the Power Set Axiom and the Separation Axiom Schema,[22] define a new function

[21] So $n < \omega$ if ω is a set.

[22] We shall abbreviate 'the Power Set Axiom and the Separation Axiom Schema' to 'PS & S'.

$f' : \omega \to k+1-\{f(0)\}$ by $f'(n) = f(n+1)$. f' is a bijection because f is a bijection. Again using PS & S, define another new function $f'' : \omega \to k$ by

$$f''(n) = \begin{cases} f'(n) & \text{if } n \neq m, \\ f(0) & \text{if } n = m. \end{cases}$$

(Informally: We replace $f'(m) = k+1$ by $f(0)$.) f'' is a bijection because f' is a bijection. This contradicts the induction hypothesis. Now suppose $l = 0$, i.e. $f(0) = k+1$. We will carry out the same procedure as before but with a few minor changes. Using PS & S, define a new function $g : \omega \to k+1-\{f(1)\}$ by

$$g(n) = \begin{cases} f(0) & \text{if } n = 0, \\ f(n+1) & \text{if } n \geqslant 1. \end{cases}$$

g is a bijection because f is a bijection. Again using PS & S, we define another new function $g' : \omega \to k$ by

$$g'(n) = \begin{cases} f(1) & \text{if } n = 0, \\ g(n) & \text{if } n > 0. \end{cases}$$

g' is a bijection because g is a bijection. This contradicts the induction hypothesis. Since both cases ($l = m+1$ for some $m \in \omega$ and $l = 0$) lead to contradiction, our suppostion must be false and thus there does not exist a bijection $\omega \to k+1$. Therefore, by induction, there does not exist a bijection $\omega \to n$ for any $n \in \omega$. Hence ω is infinite.

[(ii) → (iii)] Assume there exists an infinite set.

To prove this implication, we shall suppose that ω is not a set and derive a contradiction. We shall start with some preliminary results, the first of which is not strictly necessay to prove the implication but is nevertheless interesting:

Proposition 3.30. *If ω does not exist, then there does not exist a nonzero limit ordinal.*[23]

Proof. Suppose there exists a nonzero limit ordinal. Then the class $C = \{X : X \text{ is a nonzero limit ordinal}\}$ is nonempty and thus $\bigcap C$ is an ordinal. Consider some $\gamma \in C$. We have $\bigcap C \subseteq \gamma$ by the definition of C. If $\gamma < \bigcap C$, then $\gamma \subseteq \bigcap C$ and thus $\gamma = \bigcap C$. Hence $\bigcap C$ is the least nonzero limit ordinal, i.e. $\bigcap C = \omega$. This is a contradiction. Therefore C must be empty. □

Lemma 3.31. *Let n be a finite ordinal. Then there does not exist a bijection from n onto a proper subset of n.*[24]

Proof. We will prove this result by using transfinite induction. Since n is a finite ordinal, we do not need to worry about the limit step.[25] The empty set does not have any proper subsets so the result holds for 0. For the induction hypothesis, assume that the result holds for some finite ordinal k. To prove the result for $k+1$ we shall assume the negation and derive a contradiction. So suppose that $f : k+1 \to E$ is a

[23] An immediate consequence of this result is that ω is not a set if and only if every ordinal is a finite ordinal.

[24] Notice that we could apply Proposition 3.30 here: since we are assuming that ω is not a set, every ordinal is a finite ordinal and so we could restate the Lemma for all ordinals.

[25] Notice that under the assumption that ω is not a set, Proposition 3.30 makes the limit step in transfinite induction redundant.

bijetion for some proper subset $E \subseteq k+1$. First suppose $k \notin E$. Then $E \subseteq k$ and thus $E - \{f(k)\}$ is a proper subset of k. Hence $f\restriction k$ is a bijection from k to a proper subset of k. This contradicts the induction hypothesis. Now suppose $k \in E$. Since f is a bijection, there exists a unique $l \in k+1$ such that $f(l) = k$. Using PS & S, define a new function $g\colon k \to E - \{k\}$ by

$$g(n) = \begin{cases} f(n) & \text{if } n \neq l, \\ f(k+1) & \text{if } n = l. \end{cases}$$

g is a bijection because f is a bijection. Thus, by the induction hypothesis, $E - \{k\} = k$ (otherwise $E - \{k\}$ would be a proper subset of k). Then $k+1 = E$. This is a contradiction because E is a proper subset of $k+1$. Since both cases ($k \notin E$ and $k \in E$) lead to contradicion, the bijection f cannot exist. Thus the result holds for $k+1$. Therefore, by transfinite induction, we are done. □

Proposition 3.32. *Let A be a set. Suppose that there exist finite ordinals m, n such that $f\colon A \to m$ and $g\colon A \to n$ are bijections. Then $m = n$.*

Proof. If $m \neq n$, then either $m \subseteq n$ or $n \subseteq m$ (Lemma 3.16(iii)). Without loss of generality, assume $m \subseteq n$. Since both f and g are bijections, $f \circ g^{-1}\colon n \to m$ is a bijection. This contradicts the Lemma. Thus m and n must be equal. □

Let X be an infinite set. Using PS & S, define a new set $Y = \{A \subseteq X : A \text{ is finite}\}$. Define a class function $F\colon Y \to Ord$ by $F(A) = n$, where n is such that there exists a bijection $A \to n$. F is well-defined by Propostion 3.32. By the Axiom Schema of Replacement, $F(Y)$ is a set. $\varnothing \in Y$ so $0 \in F(Y)$. Suppose $k \in F(Y)$. Then there exists a set $B \in Y$ such that there exists a bijection $B \to k$. Since X is infinite, $X - B \neq \varnothing$. Consider some $x \in X - B$. Then there exists a bijection $B \cup \{x\} \to k+1$. So $B \cup \{x\} \in Y$ and $k+1 \in F(Y)$. Thus $F(Y)$ is inductive. Therefore, by Lemma 3.27, ω exists (recall the beginning of the proof of (i) → (ii)). This is contradicts our supposition that ω does not exist. Therefore ω must exist.

[(iii) → (i)] Assume that ω is a set.

This implication is refreshingly easy to prove. Since ω is an ordinal, $\varnothing \in \omega$. Since ω is a limit ordinal, we have that, for every α, $\alpha < \omega$ implies $\alpha + 1 < \omega$ (Lemma 3.26). Thus ω is inductive. □

Well, the reader may well wish to sit back and have a cup of tea after all that. What we have shown is that in ZF−inf, the ordinals are precisely the von Neumann ordinals, i.e. Ord = ω (as classes). This makes life a lot easier: for example, we only need to consider N-induction; that is, we can ignore the limit step in transfinite induction.

We shall now develop operations on the (von Neumann) ordinals. To do this, we need to develop recursion in ZF−inf. Usually, i.e. in ZF *with* inf, one developes *transfinite* recursion, but since all ordinals in ZF−inf, we only need a weaker form:

Theorem 3.33 (Recursion)**.** *Let $\beta \in$ Ord and let G be a class function defined on Ord. Then there is a unique class function F defined on Ord such that $F(0) = \beta$ and $F(\alpha + 1) = G(F(\alpha))$ for all $\alpha \in$ Ord.*

Proof. We shall first show that F is defined on all of Ord. Let γ be the least Ord such that F is not defined at γ (such a γ exists because $<$ is a well-ordering of Ord). Then, since we know that F is defined at 0, $\gamma = \delta + 1$ for some $\delta \in$ Ord (recall that

in ZF−inf all ordinals are finite). But then F is defined at δ, and thus it is also defined at γ, since $F(\gamma) = F(\delta + 1) = G(F(\delta))$. This is a contradiction and thus no such γ exists. Thus F is defined on all of Ord.

We now show that F is unique. Suppose that F' is another such function. Let ξ be the least ordinal such that $F'(\xi) \neq F(\xi)$. Since $\xi \neq 0$, $\xi = \varepsilon + 1$ for some $\varepsilon \in$ Ord. But then $F'(\varepsilon) = F(\varepsilon)$ and so $F'(\xi) = F'(\varepsilon + 1) = G(F'(\varepsilon)) = G(F(\varepsilon)) = F(\varepsilon + 1) = F(\xi)$, which is a contradiction and thus no such ξ exists.

Lastly, by the Axiom of Replacement we know that each $F(\alpha)$ is indeed a set. □

We can use recursion to define operations on ordinals:

Definition 3.34. Let $\alpha \in$ Ord. We define addition, multiplication, and exponentiation as follows:

(i) Addition: $\alpha + 0 = \alpha$; $\alpha + (\beta + 1) = (\alpha + \beta) + 1$, for all β.

(ii) Multiplication: $\alpha \cdot 0 = 0$; $\alpha \cdot (\beta + 1) = (\alpha \cdot \beta) + \alpha$, for all β.

(iii) Exponentiation: $\alpha^0 = 1$; $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$, for all β.

To help our understanding, let us explicitly show how we used recusrion to define addition on Ord. In this case, the β from Theorem 3.33 is α and the function G is given by $G(\gamma) = \gamma + 1$ for all $\gamma \in$ Ord.

The operations defined in Definition 3.34 have some basic properties:

Lemma 3.35. *Let $\alpha, \beta, \gamma \in$ Ord be arbitrary. Then:*

(i) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
(ii) $\alpha + \beta = \beta + \alpha$;
(iii) $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$;
(iv) $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$;
(v) $\alpha \cdot \beta = \beta \cdot \alpha$;
(vi) $\alpha \cdot 1 = \alpha$;
(vii) $\alpha < \beta \to \alpha + \gamma < \beta + \gamma$;
(viii) $(\alpha < \beta \wedge 0 < \gamma) \to \alpha \cdot \gamma < \beta \cdot \gamma$; *and*
(ix) $\alpha < \beta \to \exists \gamma\, (\alpha + \gamma = \beta)$.

Proof. These are all proved by induction. Let us explicitly prove (iii), (v), and (ix).

(iii) We shall prove this by induction on γ. The case is trivial for $\gamma = 0$. Suppose that the formula holds for some γ. Then

$$\begin{aligned} \alpha \cdot (\beta + (\gamma + 1)) &= \alpha \cdot ((\beta + \gamma)+) && \text{(definition)} \\ &= \alpha \cdot (\beta + \gamma) + \alpha && \text{(definition)} \\ &= \alpha \cdot \beta + \alpha \cdot \gamma + \alpha && \text{(induction hypothesis)} \\ &= \alpha \cdot \beta + \alpha \cdot (\gamma + 1) && \text{(definition).} \end{aligned}$$

So we are done.

(v) We shall prove this by induction on γ. The case for $\gamma = 0$ is trivial. Suppose that the formula holds for some γ. Then

$$\begin{aligned} (\alpha \cdot \beta) \cdot (\gamma + 1) &= (\alpha \cdot \beta) \cdot \gamma + \alpha \cdot \beta && \text{(definition)} \\ &= \alpha \cdot (\beta \cdot \gamma) + \alpha \cdot \beta && \text{(induction hypothesis)} \\ &= \alpha \cdot (\beta \cdot \gamma + \beta) && \text{(part (iii))} \\ &= \alpha \cdot (\beta \cdot (\gamma + 1)) && \text{(definition).} \end{aligned}$$

So we are done.

(ix) We shall prove this by induction on β. The case for $\beta = 0$ holds vacuously. Suppose that the formula is true for some β. Let $\alpha < \beta + 1$. Then $\alpha \leqslant \beta$. If $\alpha = \beta$, then take $\gamma = 1$. If $\alpha < \beta$, then by the induction hypothesis there exists γ such that $\alpha + \gamma = \beta$. Thus $\alpha + (\gamma + 1) = (\alpha + \gamma) + 1 = \beta + 1$ and we are done. □

We are now ready to move on to the main area of this essay.

4 Interpretations

4.1 Interpretations via the ordinal interpretation

We have now arrived at the main thrust of this essay; all our previous work has been so that we can start talking about interpreting one formal language in another. Informally, an *interpretation* of a theory T_1 in a theory T_2 is a way of talking about T_1 in T_2, in such a way that T_2 can prove the axioms of T_1. We do this by *interpreting* formulae in T_1 by formulae in T_2. We will build up to a rigorous defintion of an interpretation by going through some background theory, using the so-called *ordinal interpretation* as a case study. In Subsection 3.2 we developed the ordinals, which turned out to be the von Neumann ordinals; let us recall their definition:

$$0 = \varnothing, 1 = \{0\}, 2 = \{0, 1\}, \ldots, n = \{0, 1, 2, \ldots, n-1\}, \ldots.$$

By the very way we've labelled them with natural numbers suggests that we can talk about PA in ZF$-$inf. And indeed we can: we have already defined addition and multiplication of ordinals and showed, interpreting the variables as ordinals and $+$, $\cdot$, $<$, 0, 1 naturally, that (Ax1)–(Ax15) and the induction schema from Subsection 3.1 hold for the finite ordinals (see Theorem 3.25 and Lemma 3.35). This brings us to an important point: if we are to interpret a theory T_1 in a theory T_2, we need to be able to restrict the domain of quantification of the interpretation of T_1 in T_2. So, in our example of interpreting PA in ZF$-$inf via the ordinals, we need to be able to restrict the interpretation of PA to the ordinals; we don't want to quantify over arbitrary sets. The way we do this is to introduce a uninary predicate 'Dom' (read 'domain'), specifying that every theory include the axiom $\forall x \mathrm{Dom}(x)$. This doesn't change anything in the theory itself: it simply says that every element is in the domain of the language, which was the case anyway. But given a theory T_1, we can use the interpretation of Dom_{T_1} in T_2 to correctly restrict the domain of quantification. So, in our running example of the ordinal interpretation, we would interpret the predicate $\mathrm{Dom}_{\mathrm{PA}}(x)$ in ZF$-$inf to be $x \in \mathrm{Ord}$. We then define the interpretation of the PA-formula $\forall x \varphi(x)$ in ZF$-$inf to be

$$\forall x(x \in \mathrm{Ord} \to \varphi'(x)),$$

where φ' is the interpretation of φ in ZF$-$inf. Similarly we interpret $\exists x \varphi(x)$ to be

$$\exists x(x \in \mathrm{Ord} \wedge \varphi'(x)).$$

Okay, so we've dealt with quantifiers. But what formulae in general? Well, first we define the interpretations of atomic T_1-formulae in T_2 and then we extend to complex formulae via the natural, logic-preserving way. More specifically, if φ and ψ are T_1-formulae with interpretations φ' and ψ' in T_2 respectively, then we define the interpretations of

$$\neg\varphi, \quad \varphi \vee \psi, \quad \varphi \wedge \psi, \quad \varphi \to \psi, \quad \varphi \leftrightarrow \psi$$

to be

$$\neg\varphi', \;\; \varphi' \vee \psi', \;\; \varphi' \wedge \psi', \;\; \varphi' \to \psi', \;\; \varphi' \leftrightarrow \psi'$$

repectively. So, just as we can define a linear map between vector spaces by specifying how it acts on the elements of a basis and then extending linearly, we can define an interpretation by specifying how it acts on atomic formulae and then 'extending logically'. In this respect it is similar to how we defined syntax. This technique is often called 'induction on (the complexity of) formulae'. We can apply this technique because we specified that both of our theories T_1 and T_2 be first-order, and so they are written in the same underlying logical language. In our example, there are only a few atomic formulae that we have to deal with, namely 0, 1, $x = y$, $x + y$, $x \cdot y$, and $x < y$. These are interpreted as we suggested above:

$$\begin{aligned} 0 &\text{ is interpreted as } \varnothing; \\ 1 &\text{ is interpreted as } \{\varnothing\}; \\ x = y &\text{ is interpreted as } x = y; \\ x + y &\text{ is interpreted as } x +_{\mathrm{o}} y; \text{ and} \\ x \cdot y &\text{ is interpreted as } x \cdot_{\mathrm{o}} y; \end{aligned}$$

where the subscript 'o' in '$+_{\mathrm{o}}$' and '$\cdot_{\mathrm{o}}$' is to emphasise that the operations are on ordinals (the interpretation is so natural that the notations are already the same!). Notice that $=$ in PA is interpreted as $=$ in ZF$-$inf; this will always be the case in this essay, i.e. identity in T_1 will always be interpreted as identity in T_2. Interpretations where this is not the case are studied,[26] but we will not be dealing with them.

We shall move on to the next subsection, where we shall give a much more rigorous definition of an interpretation.

4.2 Interpretations: a rigorous definition and a categorical perspective

In this subsection we shall give a rigorous definition on an interpretation and then build up some more terminology. We shall then quickly review these definitions from a categorical point of view; those readers who are unfamiliar with category theory can happily skip this last part as it is not key to the theme of this essay. Our definitions in this subsection are based on those in [8] and [12]. We shall also adopt their convention in using lower-case Fraktur letters to denote interpretations.

Definition 4.1. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be first-order languages and let T_1 and T_2 be $\mathcal{L}_1$- and $\mathcal{L}_2$-theories respectively. An *interpretation* of T_1 in T_2 is a map $\mathfrak{f}\colon T_1 \to T_2$ given by mapping atomic formulae $\varphi(x_1, x_2, \ldots x_n)$ to $\varphi(x_1, x_2, \ldots x_n)^{\mathfrak{f}}$ in the same free variables and then extending logically to the whole of T_1; that is, for atomic T_1-formulae φ and ψ, we set

(i) $(\neg\varphi)^{\mathfrak{f}}$ to be $\neg(\varphi)^{\mathfrak{f}}$;

(ii) $(\varphi \wedge \psi)^{\mathfrak{f}}$ to be $\varphi^{\mathfrak{f}} \wedge \psi^{\mathfrak{f}}$;

(iii) $(\varphi \vee \psi)^{\mathfrak{f}}$ to be $\varphi^{\mathfrak{f}} \vee \psi^{\mathfrak{f}}$;

(iv) $(\varphi \to \psi)^{\mathfrak{f}}$ to be $\varphi^{\mathfrak{f}} \to \psi^{\mathfrak{f}}$;

(v) $(\varphi \leftrightarrow \psi)^{\mathfrak{f}}$ to be $\varphi^{\mathfrak{f}} \leftrightarrow \psi^{\mathfrak{f}}$;

(vi) $(\forall x \varphi(x))^{\mathfrak{f}}$ to be $\forall x(\mathrm{Dom}(x) \to \varphi(x)^{\mathfrak{f}})$; and

[26] See the cardinal interpretation in [12], for example.

(vii) $(\exists x\varphi(x))^{\mathfrak{f}}$ to be $\exists x(\mathrm{Dom}(x) \wedge \varphi(x)^{\mathfrak{f}})$.

We also specify that $T_2 \vdash \sigma$ for every axiom σ of T_1 and that $T_2 \vdash \exists x \mathrm{Dom}(x)^{\mathfrak{f}}$.

The following propostion is an immediate consequence of this defition:

Proposition 4.2. *Let* $\mathfrak{f}\colon T_1 \to T_2$ *be an interpretation and let* η *be some* T_1*-sentence. If* $T_1 \vdash \eta$ *then* $T_2 \vdash \eta^{\mathfrak{f}}$.

We shall now define what it means for two interpretations to be equivalent:

Definition 4.3. Let $\mathfrak{f}, \mathfrak{g}\colon T_1 \to T_2$ be interpretations. $\mathfrak{f}$ and $\mathfrak{g}$ are *equivalent* iff $T_2 \vdash \forall x(\varphi(x)^{\mathfrak{f}} \leftrightarrow \varphi(x)^{\mathfrak{g}})$ for every T_1-formula φ.

We now define the indentity interpretation and composition of interpretations:

Definition 4.4. The *identity interpretation* of a theory T is the interpretation $1_T\colon T \to T$ defined by setting φ^{1_T} to be φ for all T-formulae φ.

Definition 4.5. Let $\mathfrak{f}\colon T_1 \to T_2$ and $\mathfrak{g}\colon T_2 \to T_3$ be interpretations. We define $\mathfrak{gf}\colon T_1 \to T_3$ by setting $\varphi^{(\mathfrak{gf})}$ to be $(\varphi^{\mathfrak{f}})^{\mathfrak{g}}$.

We shall now define what it means for two interpretations to be inverse to one another:

Definition 4.6. Let $\mathfrak{f}\colon T_1 \to T_2$ and $\mathfrak{g}\colon T_2 \to T_1$ be interpretations. $\mathfrak{f}$ and $\mathfrak{g}$ are *inverse* to one another iff $\mathfrak{gf}$ is equivalent to 1_{T_1} and $\mathfrak{fg}$ is equivalent to 1_{T_2}.

We are now able to develop some terminology regarding interpretability between theories, which we take from [12]:

Definition 4.7. Let T_1 and T_2 be theories. T_1 is *interpretable* in T_2 iff there exists an interpretation $T_1 \to T_2$. T_1 are T_2 *mutually interpretable* iff there exist interpretations $T_1 \to T_2$ and $T_2 \to T_1$. T_1 are T_2 *bi-interpretable* iff there exist inverse interpretations $T_1 \to T_2$ and $T_2 \to T_1$.

Notice that bi-interpretability is a stronger condition than mutually interpretability.

We shall now briefly describe how we can view interpretations in a categorical sense. As we said in the introduction to this subsections, readers not familiar with category theory can happily move on to the next section. We can view theories as a category by regarding theories as objects and interpretations as morphisms, with identity morphisms and composition defined as in Definition 4.2. Two morphisms are defined to be equal in this category iff they are equaivalent in the sense given in Definition 4.3. Associativity is clear:

Lemma 4.8. *Let* $\mathfrak{f}\colon T_1 \to T_2$*,* $\mathfrak{g}\colon T_2 \to T_3$*, and* $\mathfrak{h}\colon T_3 \to T_4$ *be interpretations. Then* $\mathfrak{f}(\mathfrak{gh}) = (\mathfrak{fg})\mathfrak{h}$.

We can now view the defintions given in 4.7 in a categorical sense: T_1 is interpretable in T_2 iff there exists a morphism $T_1 \to T_2$; T_1 are T_2 mutually interpretable iff there exist morphisms $T_1 \to T_2$ and $T_2 \to T_1$; and T_1 are T_2 bi-interpretable iff there exists an isomorphism $T_1 \to T_2$.

5 The Ackermann interpretation

The Ackermann interpretation is an interpretation of ZF−inf in PA. It was first discovered by Wilhelm Ackermann in 1937 ([1]). The idea is beautifully simple: given two numbers m and n, we define m to be a member of n if the m^{th} digit in the binary expansion of n is 1. The fact that this is indeed an interpretation isn't quite so remarkable as it at first might seem: if we wished to programme a computer to work with finite sets, we would probably code sets as a sequence of 1's and 0's, each place in the sequence standing for an object, a 1 indicating that the object is in the set and a 0 indicating that it is not. Before we get into the business of actually showing that the Ackermann interpretation is indeed an interpretation, let us go though some examples informally, working in $\mathbb{N}$, the standard model of PA.

Consider the number 13. In binary, 13 is written as 1101 ($13 = 1 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3$). The 0^{th} digit in this expansion is 1, the 1^{st} is 0, the 2^{nd} is 1, and the 3^{rd} is 1; all higher digits are of course 0. Thus, under the Ackermann interpretation, 0, 2, and 3 are contained in 13. Now consider the number 3. In binary, 3 is expressed as 11 and thus, under the Ackermann interpretation, 0 and 1 are members of 3. One might then ask what the union of 3 and 13 is. Well, we know that it precisely contains 0, 1, 2, and 3, and so in binary it is 1111, which is equal to 15 ($= 1 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3$) in the usual Indo-Arabic notation. But what about power sets? We can calculate these too; let us do this for 13. As we saw, the elements of 13 are 0, 2, and 3. Thus the subsets of 13 are 0 (this is vacuously a subset), $\{0\}$, $\{2\}$, $\{3\}$, $\{0,2\}$, $\{0,3\}$, $\{2,3\}$, and $\{0,2,3\}$; these, under the Ackermann interpretation, are equal to 0, 1, 100, 1000, 101, 1001, 1100, and 1101 respectively, which, in turn, are equal to 0, 1, 4, 8, 5, 9, 12, and 13 respectively as Indo-Arabic numerals. Thus, in binary, the power set of 13 is 11001100110011 (= 13107 as an Indo-Arabic number).

Okay, so we've seen informally how the Ackermann interpretation works. Now we need to get down and show that this is indeed an interpretation of PA in ZF−inf. To do this, we shall use the machinery that we developed in Subsection 3.1 to prove that each of the interpretations of the axioms of ZF−inf are provable in PA. A lot of the following formulae are taken from [5], although our interpretation of the power set is original.

We shall start by saying in PA that the x^{th} digit in the binary expansion of y is 1. We shall do this by introducing a binary function, A (for *A*ckermann),[27] which, informally, is the following:

$$A(x,y) = \begin{cases} 1 & \text{if } x^{\text{th}} \text{ digit in the binary expansion of } y \text{ is } 1 \\ 0 & \text{if } x^{\text{th}} \text{ digit in the binary expansion of } y \text{ is } 0. \end{cases}$$

We can define this in PA by defining its graph:

$$\begin{aligned} A(x,y) = z \leftrightarrow [&((\exists n < y\, \exists m < 2^x (y = 2^{x+1}n + 2^x + m)) \wedge z = 1) \\ &\vee (\neg(\exists n < y\, \exists m < 2^x (y = 2^{x+1}n + 2^x + m)) \wedge z = 0)]. \end{aligned}$$

This isn't *nearly* as bad as it looks: it really is what the reader himslef/herself would write if they sat down and thought about it. With this new function in hand, we can define the Ackermann interpretation formally, which we shall denote $\mathfrak{a}$: ZF–inf $\to$ PA:

$$\begin{aligned} (x \in y)^{\mathfrak{a}} &\text{ is } A(x,y) = 1; \text{ and} \\ \text{Dom}(x)^{\mathfrak{a}} &\text{ is Dom.} \end{aligned}$$

[27] Don't confuse this with the Ackermann function on p. 68 of [7].

This is enough to define $\mathfrak{a}$, since $\in$ and Dom are the only relation symbols in $\mathcal{L}_\in$. Let us now prove our first axiom, Extensionality. Under the Ackermann interpretation, the Axiom of Extensionality is

$$\forall x \forall y (\forall z ((z \in x)^{\mathfrak{a}} \leftrightarrow (z \in y)^{\mathfrak{a}}) \rightarrow x = y),$$

which is

$$\forall x \forall y (\forall z (A(z,x) = 1 \leftrightarrow A(z,y) = 1) \rightarrow x = y).$$

Lemma 5.1. PA $\vdash \forall x \forall y (\forall z (A(z,x) = 1 \leftrightarrow A(z,y) = 1) \rightarrow x = y)$.

Proof. We shall prove this by induction on x. The case $x = 0$ is trivial. Suppose that the formula holds for some $x > 0$. We want to show that the formula holds for $x + 1$. Suppose $A(z, x+1) =\leftrightarrow A(z,y) = 1$ for some arbitrary y and z. Then $A(z,x) = A(z, y \dot{-} 1)$, which implies $x = y \dot{-} 1$ by the induction hypothesis (since z is arbitray) and so $x + 1 = y$. So we are done. □

The interpretation of the Axiom of the Empty Set is easy to prove:

Lemma 5.2. PA $\vdash \exists x \forall y (A(y,x) = 0$.

Proof. Simply take $\varnothing$ to be 0: 0 has no elements because $\neg \exists n < 0$. □

Before we cover any more axioms, we need to define two preliminary functions: $\mu(x)$ is the leftmost postion in the binary expnasion of x to be occupied by a 1, and $k(x) := \sum_{y \leqslant \mu(x)} A(y,x)$. μ is fairly tricky to define in PA, so we will simply use it; the details can be found in [5].

We now come to the Axiom of Pairing. To prove its interpretation in PA, it is enough to find $\{x,y\}^{\mathfrak{a}}$. This is quite straightforward:

$$\{x,y\}^{\mathfrak{a}} = 2^x + (1 \dot{-} \lambda(x,y))2^x,$$

where λ is the function that returns a 1 if x and y are equal and a 0 if they are not equal, which we can define in PA by specifying its graph:

$$\lambda(x,y) = z \leftrightarrow (x = y \wedge z = 1) \vee (\neg x = y \wedge z = 0).$$

We now come to the Power Set Axiom. To define the interpretation of the power set, we will need to define a preliminary function first. Informally, we define $\gamma(x,y)$ by

$$\gamma(x,y) = \begin{cases} 1 & \text{if } x \text{ is a subset of } y \text{ under } \mathfrak{a}, \\ 0 & \text{if } x \text{ is not a subset of } y \text{ under } \mathfrak{a}. \end{cases}$$

But how do we define this in PA? Well, first notice that x is a subset of y under $\mathfrak{a}$ iff

$$\forall z \leqslant \mu x \ (A(z,x) \leqslant A(z,y)).$$

Thus we can define γ in PA by

$$\gamma(x,y) = 1 \dot{-} \sum_{z \leqslant \mu x} (A(z,x) \dot{-} A(z,y)).$$

We can now define $\mathcal{P}(x)^{\mathfrak{a}}$ by

$$\sum_{z \leqslant k(x)} 2^z \gamma(z,x).$$

Let us breifly explain this formula. What we are doing here is going through all the possbile powers of 2, i.e. all the possible elements, and putting a 1 if z is a subset

of x and a 0 if it is not. This is a nice approach and we shall use it for the next three axioms.

We now come to Union. As with the Power Set, we need a preliminary function, which we define informally as

$$\delta(x,y) = \begin{cases} 1 & \text{if } \exists z(A(z,y) = 1 \wedge A(x,z) = 1), \\ 0 & \text{if otherwise.} \end{cases}$$

In other words, $\delta(x,y)$ tells us whether x is in the union of y. We can define δ in PA:

$$\delta(x,y) = (1 \dot{-} (1 \dot{-} \sum_{z \leqslant \mu y} A(x,z)A(z,y)))$$

We can now use this function to define $(\bigcup x)^{\mathfrak{a}}$ by

$$\sum_{z \leqslant \mu x} 2^z \delta(z,x).$$

We now come to Separation, which will require us to define the characteristic function of a formula in PA. Informally, for a formula φ,

$$\chi_\varphi(x) = \begin{cases} 1 & \text{if } \varphi(x), \\ 0 & \text{if } \neg\varphi(x). \end{cases}$$

We define this in PA by specifying its graph:

$$\chi_\varphi(x) = z \leftrightarrow (\varphi(x) \wedge z = 1) \vee (\neg\varphi(x) \wedge z = 0).$$

We can now define $\{z \in x : \varphi(z)\}^{\mathfrak{a}}$ by

$$\sum_{z \leqslant \mu x} 2^z A(z,x)\chi_{\varphi^{\mathfrak{a}}}(z).$$

We now come to the last of the Axioms that we can prove using this method, Replacement. For a functional formula F in ZF$-$inf and a set x, we define $\{F(z) : z \in x\}^{\mathfrak{a}}$ by

$$\sum_{z \leqslant \mu x} A(z,x)2^{F(z)^{\mathfrak{a}}}(1 \dot{-} ((\sum_{y \leqslant z \dot{-} 1} A(y,x)\lambda(F(z)^{\mathfrak{a}}, F(y)^{\mathfrak{a}})) \dot{-} \lambda(z,0))).$$

This really does look awful, so let's go through and informally explain each of the parts of the formula. The first part, $\sum_{z \leqslant \mu x} A(z,x)2^{F(z)^{\mathfrak{a}}}$, goes through and puts in each $F(z)^{\mathfrak{a}}$ as an element. This is not enough though, for the function might not be injective. This is where the next part comes in: For each z, we look at everything less than it, checking if it's in x and whether its image under F is equal to that of z. If the images are equal, then we get a 1, which means we multiply $2^{F(z)^{\mathfrak{a}}}$ by a 0 (since $1 \dot{-} 1 = 0$); otherwise we multiply it by 1. Of course, we need to inculde at least one element for every element in the image of F, and if $z = 0$ then we put it's image in regardless – this is done by the $\dot{-}\lambda(z,0)$ term.

We now have two axioms left: $\neg$Infinity and Foundation. We shall prove the interpretation of the Axiom of Foundation first.

The Axiom of Foundation is

$$\forall x(x \neq \varnothing \to \exists z(z \in x \wedge z \cap x = \varnothing)).$$

So, before we can state the interpretation of Foundation under Ackermann, we'd first better work out the interpretation of intersection. Thankfully, this is straightforward:

$$(x \cap y)^{\mathfrak{a}} = \sum_{z \leqslant \mu(x)+\mu(y)} 2^z A(z,x)A(z,y).$$

So the interpretation of the Axiom of Foundation is

$$\forall x(x > 0 \to \exists z(A(z,x) = 1 \wedge \sum_{z \leqslant \mu(x)+\mu(y)} 2^z A(z,x)A(z,y) = 0)). \tag{6}$$

We can prove this in PA:

Lemma 5.3. PA $\vdash$ (6).

Proof. Take z to be the minimal z such that $A(z,x) = 1$. If $z = 0$ then we are done, so assume that $z > 0$. Then, by Lemma 5.4 below, if $A(y,z) = 1$, then $y < z$ and so $A(y,x) = 0$ because z is minimal. Thus $(z \cap x)^{\mathfrak{a}} = 0$ and we are done. □

Lemma 5.4. PA $\vdash \forall x \forall y(A(x,y) = 1 \to x < y)$.

Proof. Suppose that there exist x and y such that $A(x,y) = 1$ and $y \leqslant x$. Then there exist $n < y$ and $m < 2^x$ such that

$$y = m2^{x+1} + 2^x + n. \tag{7}$$

But $y \leqslant x$ and so $y < 2^y \leqslant 2^x$, which contradicts 7. Thus no such x and y can exist. □

We now only have the interpretation of the negation of the Axiom of Infinity to prove. We can do this by using the ordinal interpretation. Suppose that there was an infinite set, i.e. z such that for every x there exists $y > x$ such that $A(y,z) = 1$. Then the interpretation of z in the ordinals would be an infinite set, contradicting Theorem 3.29. Thus no such set z can exist.

Well, now that we have shown that the Ackermann interpretation is indeed an interpretation of ZF−inf, we shall now consider its inverse.

6 The inverse Ackermann interpretation

We have shown in the last two sections that ZF−inf and PA are *mutually* interpretable; that is, we can interpret PA in ZF−inf (via the ordinals) and we can interpret ZF−inf in PA (via the Ackermann interpretation). But are ZF−inf and PA *bi*-interpretable, i.e. can we find interpretations ZF−inf → PA and PA → ZF−inf that are inverse to one another? Well, the answer is *sort of*. If we wish to find a bi-interpretation, we need to add an extra axiom to ZF−inf, the *Axiom of Transitive Closure* (TC for short); we shall explain what this is shortly. With this new axiom, ZF−inf+TC (that is ZF−inf with the extra axiom TC) and PA are bi-interpretable. We shall abrreviate ZF−inf+TC to ZF−inf*.

Before we actually get to the bi-interpretation, we need to cover some background theory. We start by saying what TC actually is:

Axiom of Transitive Closure. *Every set is contained in a transitive set. Formally:*

$$\forall x \exists y(x \subseteq y \wedge \mathrm{Trans}(y)),$$

where $\mathrm{Trans}(x)$ is the relation that says that x is transitive:

$$\forall z(z \in x \to z \subseteq x).$$

This axiom is stronger than one might have initially thought: We can use it to define the *transitive closure* of x, the smallest transitive set containing x, by taking the intersection of all transitive sets containing x. The transitive closure of x is unique by Extensionality and is denoted $\mathrm{TC}(x)$. Clearly the existence of $\mathrm{TC}(x)$ implies the axiom TC. Thus TC is equivalent to the statement that every set has a transitive closure:

$$\forall x \exists y(y = \mathrm{TC}(x)).$$

Now, the reader may ask why we need to add a new axiom at all: can we not prove TC from the other axioms of ZF$-$inf? The answer is no: there exist models of ZF$-$inf in which TC fails. We shall not offer a proof of this (see [8] for a sketch proof), although we will run through the basic idea. Readers not familiar with more advanced set theory and model theory may wish to skip this explanation and simply take it as read that ZF$-$inf $\nvdash$ TC. The idea is to take a model of ZF$-$inf and from it make a permutation model in which TC fails. Specifically, consider the set of hereditary finite sets V_ω and involute all singletons $\{x \cup \{x\}\}$ of finite ordinals x with x; that is, define a function F that sends x to $\{x \cup \{x\}\}$, $\{x \cup \{x\}\}$ to x, and leaves everthing else. We then define a new membership relation $\in_F$ by $x \in_F y$ iff $x \in F(y)$. It can be shown that $(V_\omega, \in_F) \vDash$ ZF$-$inf but $(V_\omega, \in_F) \vDash \neg$TC; specifically, the empty set has no transitive closure in this model.

We shall now explain why we need TC in the first place. TC allows us to prove something called $\in$*-induction*, which is an extension of transfinite induction to all transitive classes, where a *transitive class* is a class all of whose elements are transitive. Before we state and prove $\in$-induction, we need a lemma:

Lemma 6.1. ZF$-$inf* $\vdash$ *Every nonempty class C has an $\in$-minimal element.*

Proof. Consider some $x \in C$. If $x \cap C = \varnothing$, then x is an $\in$-minimal element and we are done, so assume $x \cap C \neq \varnothing$. Let $y = \mathrm{TC}(x) \cap C$. Since $x \subseteq \mathrm{TC}(x)$, $y \neq \varnothing$. Thus, by the Axiom of Regularity, there exists $z \in y$ such that $z \cap y = \varnothing$. Suppose $z \cap C \neq \varnothing$; let $w \in z \cap C$. Then, since $\mathrm{TC}(x)$ is transitive and $z \in \mathrm{TC}(x)$, $w \in \mathrm{TC}(x)$. Thus $w \in z \cap \mathrm{TC}(x) \cap C = z \cap y$, which is contradiction because $z \cap y = \varnothing$. Thus z is an $\in$-minimal element of C. □

We can now prove $\in$-induction:

Theorem 6.2. *Let φ be an $\mathcal{L}_\in$-formula. Then* ZF$-$inf* $\vdash$

$$\forall x(\forall z \in x(\varphi(z) \to \varphi(x))) \to \forall y \varphi(y).$$

Proof. Consider the class $C = \{x : \neg\varphi(x)\}$. If C is nonempty, then it has an $\in$-minimal element w by Lemma 6.1 above. $w \neq \varnothing$ because the implication

$$\forall z \in \varnothing(\varphi(z) \to \varphi(\varnothing))$$

holds vacuously and so we have $\varphi(\varnothing)$. So consider some $x \in w$. Since w is minimal, we have $\varphi(x)$. But x was chosen arbitrarily and so the hypothesis holds, implying $\varphi(w)$. This is a contradiction and thus C must be empty. □

In fact, the converse holds; that is, $\in$-induction implies TC:

Theorem 6.3. ZF$-$inf $\vdash \in$-induction $\to$ TC.

Proof. Consider some x such that every $z \in x$ has a transitive closure. Then the set

$$\bigcup\{z : \exists y \in x(z = \mathrm{TC}(y))\} \cup x$$

is transtive and has x as a subset. Thus we are done by $\in$-induction. (Recall that the existence of $\mathrm{TC}(x)$ is equivalent to the existence of a transitive set containing x as a subset.) □

Putting Theorems 6.2 and 6.3, we get:

Theorem 6.4. $\mathrm{ZF}-\mathrm{inf} \vdash \in$-induction $\leftrightarrow$ TC

An observant reader will have noticed that we still haven't actually answered the question of why we need TC in the first place. Well, it turns out that $\mathrm{PA} \vdash \mathrm{TC}^{\mathfrak{a}}$ (see [8] for a proof), and so if $\mathfrak{b} : \mathrm{PA} \to \mathrm{ZF}-\mathrm{inf}$ is an inverse to the Ackermann interpretation, then $\mathrm{ZF}-\mathrm{inf} \vdash \mathrm{TC}$ because $\mathrm{TC}^{(\mathfrak{a}\mathfrak{b})} = \mathrm{TC}$, since $\mathfrak{a}$ and $\mathfrak{b}$ are inverse. But this is a contradiction, since $\mathrm{ZF}-\mathrm{inf}$ cannot prove TC. Thus any inverse to the Ackermann interpretation must be from PA to $\mathrm{ZF}-\mathrm{inf}^*$.

Using $\in$-induction, we can develop a new type of recursion, which will be essential to us when we construct the inverse Ackermann interpretation.

Theorem 6.5 ($\in$-recursion)**.** *Let G be a function defined on V and let y be a set. Then there exists a unique function F defined on V such that $F(\varnothing) = y$ and $F(x) = G(F\restriction x)$ for all x.*

Proof. The proof is the same as that of Theorem 3.33, this time using $\in$-induction instead of ordinal induction. □

Notice that for a class of ordinals, ordinal induction and recursion are the same as $\in$-induction and $\in$-recursion.

Now that we have covered all this theory, we can now go about constructing an inverse to the Ackermann interpretation. Our strategy will be to use the ordinal interpretation. We can't use it in its current form because it is not inverse to the Ackermann interpretation: for example, under Ackermann, $\{0, 1\}$ is interpeted as $1 + 1 + 1$ $(= 1 \cdot 2^1 + 1 \cdot 2^0)$ in PA; but under the ordinal interpretation, $1 + 1 + 1$ is interpreted as $\{0, 1, 2\}$ in $\mathrm{ZF}-\mathrm{inf}^*$. However, we can adapt it. The idea is to construct a bijection between the *universe* V, the class of all sets, and the ordinals. We then 'compose' this bijection with the ordinal interpretation to get an inverse to the Ackermann interpretation. Of course we can't just construct any old bijection: we need to do it carefully. The idea is that we construct a bijection that looks like the Ackermann interpretation. If we call our bijection $\mathfrak{p}$, we informally define it recursively by

$$\mathfrak{p}(x) = \sum_{y \in x} 2^{\mathfrak{p}(y)}. \tag{8}$$

We shall sketch a proof later as to why this does in fact lead to an inverse to the Ackermann interpretation.

So far we have been very informal, so we'd better start constructing $\mathfrak{p}$ more rigorously. We follow a lot of the working in [8]. Firstly, we need to define summation over ordinals in $\mathrm{ZF}-\mathrm{inf}^*$; we do this by using ordinal recursion:

Definition 6.6. For a set of ordinals x, define $\hat{\sum}_x$ on Ord as follows: Let $\hat{\sum}_x(0) = 0$ and let

$$\hat{\sum}_x(\alpha + 1) = \begin{cases} \hat{\sum}_x(\alpha) & \text{if } \alpha + 1 \notin x, \\ \hat{\sum}_x(\alpha) + (\alpha + 1) & \text{if } \alpha + 1 \in x. \end{cases}$$

We then define $\sum(x) = \hat{\sum}_x(\bigcup x)$.

This function $\sum(x)$ allows us to sum over all elements of x (using ordinal addition); informally:

$$\sum(x) = \sum_{y \in x} y.$$

We can now define the bijection we want. This time we use $\in$-recursion:

Definition 6.7. Define a class function $\mathfrak{p} \colon V \to \mathrm{Ord}$ by

$$\mathfrak{p}(x) = \sum(\{2^{\mathfrak{p}(y)} : y \in x\}).$$

This is precisely the formal definition of what we defined informally in (8). This is indeed a bijection:

Proposition 6.8. *ZF$-$inf* $\vdash$ $\mathfrak{p}$ is a bijection.*

Proof. First let us prove injectivity. We shall do this by showing that $\mathfrak{p}$ is strictly increasing using $\in$-induction. Suppose that $\mathfrak{p}$ is strictly increasing for all $y \in x$. Now, since all ordinals in ZF$-$inf* are finite, $x = z + 1$ for some $z \in x$. For any ordinal α, $2^\alpha > \alpha$ (this is proved by a simple induction and so $2^{\mathfrak{p}(z)} > \mathfrak{p}(z)$. Thus

$$\begin{aligned}
\mathfrak{p}(x) &= \mathfrak{p}(z+1) \\
&= \sum(\{2^{\mathfrak{p}(y)} : y \in z+1\}) \\
&= \sum(\{2^{\mathfrak{p}(y)} : y \in z\}) + 2^{\mathfrak{p}(z)} \\
&> \mathfrak{p}(z).
\end{aligned}$$

So we are done.

We shall now prove surjectivity by ordinal induction. Suppose that $\mathfrak{p}$ is sujective for all $y < x$. Since all ordinals are finite, $x = z + 1$ for some $z < x$. Then, by the induction hypothesis, there exists some w such that $\mathfrak{p}(w) = z$. Then

$$\begin{aligned}
\mathfrak{p}(w \cup \{0\}) &= \mathfrak{p}(w) + \mathfrak{p}(0) \\
&= y + 1 \\
&= x.
\end{aligned}$$

So we are done. □

We can now define the interpretation $\mathfrak{b} \colon \mathrm{PA} \to \mathrm{ZF} - \mathrm{inf}^*$:

Definition 6.9. We define $\mathfrak{b}$ by setting $\mathrm{Dom}(x)^{\mathfrak{b}}$ to be '$\mathrm{Dom}(x)$'; $(x = y)^{\mathfrak{b}}$ to be '$x = y$'; $(x < y)^{\mathfrak{b}}$ to be '$\mathfrak{p}(x) < \mathfrak{p}(y)$'; $(x + y)^{\mathfrak{b}}$ to be '$\mathfrak{p}(x) + \mathfrak{p}(y)$'; and $(x \cdot y)^{\mathfrak{b}}$ to be '$\mathfrak{p}(x) \cdot \mathfrak{p}(y)$', where the target relation and operations are the usual ordinal ones.

Since the ordinal interpretation is indeed an interpretation, $\mathfrak{b}$ is an interpretation Proposition 6.8. It remains to prove that $\mathfrak{a}$ and $\mathfrak{b}$ are inverse to each other. We shall only sketch the proofs:

Lemma 6.10. $\mathfrak{a}\mathfrak{b} = 1_{\mathrm{PA}}$.

Sketch Proof. Let φ' denote $\varphi^{\mathfrak{a}\mathfrak{b}}$. Clearly $0' = 0$ and $1' = 1$. The other non-logical symbols are then formally proved to be preserved by induction. Informally, the non-logical symbols are preserved because $\mathfrak{p}$ sends x to the set that it encodes under the Ackermann interpretation. □

Lemma 6.11. $\mathfrak{b}\mathfrak{a} = 1_{\mathrm{ZF\text{–}inf}^*}$.

Sketch Proof. Denote $(x \in y)^{\mathfrak{ba}}$ by $x \in' y$. It suffices to show that $\mathrm{ZF}-\mathrm{inf}^* \vdash \forall x \forall y (x \in y \leftrightarrow x \in' y)$. This is done by formally by $\in$-induction. Informally, the implication $x \in y \to x \in' y$ holds because of the way we constructed $\mathfrak{p}$: $\mathfrak{p}(x)$ is effectively the Ackermann epression of x, but in the ordinals rather than in PA. The same goes for the converse, since we can adapt the proof of Extensionality$^{\mathfrak{a}}$ in PA. □

Putting Lemmas 6.10 and 6.11, we get:

Theorem 6.12. $\mathfrak{a}\colon \mathrm{ZF}-\mathrm{inf}^* \to \mathrm{PA}$ *and* $\mathfrak{b}\colon \mathrm{PA} \to \mathrm{ZF}-\mathrm{inf}^*$ *are inverse to one another.*

7 Interpretations and bounded formulae: a brief vista

We shall now *very* briefly talk about the result proved in [12]. We first need to define *bounded formulae*:

Definition 7.1. A PA-formula is *bounded* iff all quantifiers in the formula are of the form $\forall x < y$ or $\exists x < y$, where '$\forall x < y\, \varphi(x)$' and '$\exists x < y\, \varphi(x)$' are abbreviations for '$\forall x(x < y \to \varphi(x))$' and '$\exists x(x < y \wedge \varphi(x))$' respectively. Similarly, a ZF−inf-formula is *bounded* if all the quantifier in the formula are of the form $\forall x \in y$ or $\exists x \in y$, where '$\forall x \in y\, \varphi(x)$' and '$\exists x \in y\, \varphi(x)$' are abbreviations for '$\forall x(x \in y \to \varphi(x)$' and '$\exists x(x \in y \wedge \varphi(x))$' respectively.

Equipped with the notation of a bounded formula, we can define some subsytems of PA and ZF−inf*.

We denote the set of all PA-formulae that are equivalent to a bounded PA-formula by Δ_0.[28] We then define '$\mathrm{I}\Delta_0$' to be the subsystem of PA consisting of the axioms (Ax1)–(Ax15) from Subsection 3.1 and the induction schema restricted to Δ_0-formulae. It turns out that one cannot construct exponentiation as a total function in $\mathrm{I}\Delta_0$, and so we add an axiom 'exp' that states that 2^x is a total function.

We construct a subsystem of ZF−inf* called *Euclidean Arithmetic* ('EA' for short) by restricting the Separation and Replacement Axiom schemata to bounded formulae.[29] We then add another axiom called the *Weak Hierarchy Principle* (or 'WHP' for short), the statement of which we shall not discuss due its prerequisite background theory, to obtain a new system denoted 'EA*'.

Pettigrew then shows that the theories $\mathrm{I}\Delta_0 + \mathrm{exp}$ and EA* are bi-interpretable. This result is interesting because it sheds further light on the respective strengths of systems finite set theory and arithmetic, and thus in turn on the foundations of mathemtics.

Research into interpretations between arithmetic and finite set theory is ongoing as this essay is being written; indeed, [11] will be published later this year. One important outstanding question is the set-theoretical anlogue of $\mathrm{I}\Delta_0$. At present no one has even conjectured an answer.

[28] This notation comes from the *arithmetic hierarchy*, which we will not discuss; the reader is referred to [7].

[29] This was orignally constructed by Mayberry (see [10]); details can also be found in [11].

References

[1] Ackermann, W., 'Die Widerspruchsfreiheit der allgemeinen Mengenlehre', *Mathematische Annalen*, vol. 114: pp. 305–315, 1937.

[2] Barwise, J. and Etchemendy, J., *Language, Proof and Logic*, Stanford: CSLI Publications, 1999, ISBN 1-57586-374-X.

[3] van Dalen, D., *Logic and Structure*, 4th ed., Berlin: Springer-Verlag, 2004, ISBN 3-540-20879-8.

[4] Hedman, S., *A First Course in Logic*, New York: Oxford University Press, 2004, ISBN 0-19-852981-3.

[5] Homolka, V., 'A System of Finite Set Theory Equivalent to Elementary Arithmetic', Ph.D. thesis, University of Bristol, 1983.

[6] Jech, T., *Set Theory*, 3rd ed., Berlin: Springer-Verlag, 2002, ISBN 3-540-44085-2.

[7] Kaye, R., *Models of Peano Arithmetic*, Oxford: Oxford University Press, 1991, ISBN 0-19-853213-X.

[8] Kaye, R. and Wong, T.L., 'On interpretations of arithmetic and set theory', *Notre Dame Journal of Formal Logic*, vol. 48(4): pp. 497–510, 2007.

[9] Marker, D., *Model Theory: An Introduction*, New York: Springer-Verlag, 2002, ISBN 0-387-98760-6.

[10] Mayberry, J.P., *The Foundations of Mathematics in the Theory of Sets*, Cambridge: Cambridge University Press, 2000, ISBN 0-521-77034-3.

[11] Pettigrew, R., 'Natural, Rational, and Real Arithmetic in a Finitary Theory of Finite Sets', Ph.D. thesis, University of Bristol, 2008.

[12] Pettigrew, R., 'On interpretations of bounded arithmetic and bounded set theory', *Notre Dame Journal of Formal Logic*, forthcoming.

[13] Smith, P., *An Introduction to Gödel's Theorems*, Cambridge: Cambridge University Press, 2007, ISBN 978-0-521-67453-9.

Acknowledgements

I would firstly like to thank my supervisor, Adam Epstein. Without his help and encouragement I would never have learnt any serious set theory, let alone written this project. I owe a very great deal to him and I am most grateful.

I would like to thank Richard Pettigrew for all his help. I would like to thank him in particular for his invaluable notes on Homolka's thesis, as well as his most useful and morale-boosting comments shortly before the deadline – in a time of turmoil it is a most welcome thing to hear a calm voice.

I would also like to thank Sebastian Jörn and Josephine Salverda, whose company while writing up this dissertation kept me, well, sane. (Two weeks just isn't enough, is it guys?)

This essay is the culmination of all my work at university, and as such I would like to thank my wonderful parents, Valerie and Anthony. My time at university has not always been smooth, and without their constant love and support I would never have made it this far.