On The Structure of the Set-Theoretic Universe

Dylan Massell

“Mathematics is a study which, when we start from its most familiar portions, may be pursued in either of two opposite directions. The more familiar direction is constructive, towards gradually increasing complexity: from integers to fractions, real numbers, complex numbers; from addition and multiplication to differentiation and integration, and on to higher mathematics. The other direction, which is less familiar, proceeds by analysing, to greater and greater abstractness and logical simplicity; instead of asking what can be defined and deduced from what is assumed to begin with, we ask instead what more general ideas and principles can be found, in terms of which what was our starting point can be defined or deduced.” (Russell, 1930)

I certify that this project report has been written by me, is a record of work carried out by me, and is essentially different from work undertaken for any other purpose or assessment.

Dylan Massell April 27, 2012

Abstract:

There are several reasons for choosing a different set of axioms for a particular theory, ranging from mathematical curiosity (e.g. one wishes to explore the pragmatic differences between an axiom being held as true or not) to, as is more often the case, a metamathematical motivation. This may include, among other reasons, adopting a different set of axioms because they assume different philosophical beliefs about the field of study. Another common reason is because the "overall picture" of the field becomes much more aesthetically pleasing to the intuition with a different set of axioms, or in other words, it makes more sense to use those particular axioms. Set theory is a branch of mathematics which is subdivided into several various axiomatizations, or theories with different axioms. The predominant theory is known as Zermelo-Fraenkel set theory. However, there are countless alternative axioms for set theory one can take, from von Neumann, Bernays, and Gödel's theories which incorporate classes, to Quine's New Foundations theory, and even to Intuitionistic set theory which uses a different logic altogether, and hence a different mathematics, to describe the behavior of sets.

The main intent of this paper is to examine the different axioms of set theory which present different intuitive pictures of the set-theoretic universe, that is, all the objects which set theory as a branch of mathematics aims to study. These different axiomatizations of set theory have been developed by different mathematicians and philosophers over the course of the last century. In particular, we will give particular attention to why, intuitively or philosophically, we should choose to accept one particular axiom system over another. However, many of the various axiom systems, while not strictly philosophically justified, do have a pragmatic mathematical justification; namely the theories they produce are both interesting in their own right as a field of study, and often they can even prove useful in other branches of mathematics.

Contents

I. **Introduction**

Any mathematical, or even logical, theory consists of propositions and definitions. Certain propositions are the theorems, lemmas and corollaries of the theory, and strings of such propositions form a proof of a theorem, if it follows from deductive reasoning. In this manner, the theory is expanded by constructing proofs of theorems from propositions already available. We may also give explicit definitions of various concepts which the propositions express, though these definitions are said to be formally eliminable (Potter, 2004). In other words they are mere shorthand for the formal expressions of the concept given by the prior propositions of the theory. However, it is not possible to construct a theory from nothing, for then it would be about nothing, and thus have no propositions. We must have some starting propositions, or axioms, as well as concepts which have no explicit definition but are taken as given prior to the theory. These undefined notions are the primitive terms of the theory. The axioms of the theory provide a basis from which all the theorems, corollaries and lemmas follow by deductive reasoning. We offer no proof of axioms because they are supposed to be statements which we readily accept intuitively (Blackburn, 2005). From the axioms, the mathematician can deduce theorems and propositions which further characterize the structures and the properties the objects of study take. If we start with different axioms, we produce a different theory of the objects, although often there is the case where two different formulations of an axiom might give the same general principle. Even still it is possible to have different axioms for two different theories, but they both construct a similar picture of the field of study. There may be minor differences, such as notation, but we may still say that these theories are equivalent.

This whole process is thus known as the axiomatic method. All of mathematics uses the axiomatic method to produce theorems over a range of subjects. From geometry, to group theory, arithmetic, to analysis, each branch has axioms concerning the behavior of the objects of study of that particular branch. For example, group theory concerns itself with the notion of groups, geometry with points and lines. The exact manner in which axioms concern themselves with the primitive notions of a given theory is a central meta-mathematical or philosophical concern (Potter, 2004).

Within philosophy of mathematics, the axiomatic method can be approached from two distinct attitudes, namely, realism and formalism (Potter, 2004). These are not strictly the names used for the particular views within all of philosophy of mathematics, but they will serve useful for our discussion. The two philosophies are seen as different responses to the question of how the primitive notions of a theory are treated, and consequently what role do the axioms play in establishing the theory.

The realist answer is that the primitive terms have meaning independent of the theory, and thus the axioms are necessary true statements about those concepts (Potter, 2004). The theory is then precisely all that follows logically from these necessary truths of the primitive notions. Thus we commit ourselves to any ontological claims made by the axioms or propositions derived from them (Potter, 2004). That is to say, if a given proposition states the existence of some object of our theory, we are claiming such an object does indeed exist. A major problem with the realist position is the difficulty in accounting for multiple axiomatizations for a theory, when all axioms are supposed to be true unconditionally, and the theories have conflicting propositions. There is room within the realist perspective to take different attitudes towards the nature of the objects posited to exist.

On the one hand, we have what is known as platonism which states the objects exist independently of the mathematician and his theory; and on the other, constructivism where the objects are mental constructions of the mathematician by way of the theory (Potter, 2004). It is important to note the constructivist attitude does not say the objects are not real, for their existence lies within the meaning of the primitive notions. The theory can be seen as a mental blueprint for the structures constructed by the mathematician.

The other dominant response to the question concerning the meaning of the primitive notions and the axioms of a theory comes from the formalist philosophy. This view holds that the primitive terms have no meaning outside of the theory. Different branches of formalism say more about the exact nature the terms take within the axiomatic system. The axioms therefore are not unconditional truths of these notions; indeed the formalist perspective came about with the development of different axiomatizations of geometry (Potter, 2004).

One attitude under the formalist philosophy is that of implicationism, which is popular with mathematicians of most branches of mathematics. The view states that for any structure which satisfies the axioms, then the theory is a description of that structure and the properties it possesses (Potter, 2004). A simple and clear example of the implicationist attitude can be seen in group theory; where certain axioms are laid down, any structure satisfying them is said to be a group, and the properties it has can be derived from the axioms and supplementary definitions. When applied to set-theory, which is seen as a foundation for all of mathematics due to its ability to provide a model for the natural, rational, and real numbers (and further structures which can be derived from those), we see that an implicationist approach is unsatisfactory.

This is due to the fact that the implicationist approach doesn't allow for any unconditional claims about a theory, thus if applied to a theory which may serve as a foundation for mathematics, we would not be able to make any unconditional claims, or necessary truths, within all of mathematics (Potter, 2004). Consequentially, mathematics would have no substantitive subject matter, and would be but mere derivations of propositions from certain collections of axioms.

There is then pure formalism, which pushes this idea, many would see as an obstacle for implicationism, even further to the claim that mathematics has no semantic meaning at all, but is mere syntactical rules for the manipulation of symbols (Blackburn, 2005). Mathematics can then be seen as a, albeit highly complex, "game played with symbols" (Potter, 2004). This attitude seems too extreme of a response, and there are several objections one can make, the most prominent being an appeal to the direct applicability of mathematics to the empirical world.

Finally, one branch of formalism, which has even been argued to be a close relative of platonism, is postulationism, wherein the primitive notions of our theory take on their meaning from the axioms and propositions derived as theorems and lemmas (Potter, 2004). The theory is then seen as a way of providing implicit definitions of its primitive notions, and moreover, we are allowed to postulate the existence of the objects of the theory with the properties that the axioms and theorems claim they have. There is one obvious objection to this attitude, which is how are we entitled to derive meaning from the axioms and theorems for the primitive notions, and thus postulate the existence of the required objects? While there is very little argument to fully answer this objection, many proponents of the postulationist attitude cite the consistency of the axioms as a necessary criterion, if not also a sufficient one (Potter, 2004). Consistency is the property of a collection of propositions that one cannot prove both one sentence and its negation. Through Gödel's Incompleteness theorem, which states that any consistent theory is not complete, and vice-versa, this yields immediately another problem for the postulationist. The Incompleteness theorem states the existence of a sentence, call it G, which we are inclined to take as true yet the theory neither proves nor refutes G, rendering the theory incomplete. The difficulty for the postulationist is explaining how our intuition that the sentence G is true when G is a proposition involving ultimately only the primitive notions and axioms of our theory (Potter, 2004).

Nonetheless this problem is for the philosopher of mathematics, and need not concern us too greatly here. For in general, postulationism is a stable vantage point for examining set theory, as it places emphasis on the nature of the axioms, and in particular their consistency, and set theory is laden with multiple axiomatizations. Indeed postulationism offers justification for working withinone system of axioms over another, for not only the consistency of the axioms, but for also the implicit definitions they give of the primitive notions. This will be our chief concern in our examination of the two different axiomatizations of Zermelo set theory, and the equivalent theories of von Neumann, Bernays, and Godel, as well as Quine's New Foundations.

II. **Logic**

In constructing a logical theory with axioms and primitive terms, we use as a domain of thought, a *formal* language consisting of logical and non-logical symbols, as well as variables. We will assume a basic familiarity with classical formal logic, but give a short sketch of setting up the logic, first-order predicate calculus, which we will be using. We will make frequent use of the following logical symbols:

$\wedge$: and	$\vee$: or	$\neg$: not
$\exists$: $there\ exists$	$\forall$: $for\ all$	$=$: $equals$
$\rightarrow$: $implies$	$\leftrightarrow$: $if\ and\ only\ if$	

Other symbols in our language are variables and constants. The only non-logical symbol we will use, that we do not define within the course of constructing the various systems, is the relation symbol $\in$. We will take this symbol as primitive, which means we give no formal definition, but instead derive meaning from the axioms we give which determine its behavior. The variables we will be using are $A, B, A', B', \dots, a, b, c, \dots, u, v, \dots, V, V', \dots$ and others which we will introduce at that time. We will also use $x, y, z, \dots, \varphi, \psi, \chi, \dots$ as metavariables; they are not part of our formal object language, but instead are variables in our meta-language. It will be assumed from context what our variables are in each proposition we present. We construct from these symbols *well-formed formulae* by syntactical rules. This ensures that we can systematically infer an objective meaning to the strings constructed as well-formed formulae.

We do so by first stating what the atomic formulae are, and then give rules of how to construct more complicated sentences from those atomic cases. In set theory, we have only two atomic formulae; $x = y$ and $x \in y$, where x, y are terms of our language (Fraenkel, et al., 1973). We may also use the formulae $x \neq y$ and $x \notin y$, though as these are shorthand for $\neg(x = y)$ and $\neg(x \in y)$ respectively, they are not atomic formulae. Terms can either be a free variable, or expressions which denote a particular object. In this paper, we will only use names to denote unique objects, such as $\emptyset$ which is the empty or null set. Thus such names are terms, and we can build up formulae which rely on those terms and other variables to denote some other object. Then by specifying the rules of syntax on the atomic cases we can define the set of well-formed formulae.

Definition: *A string of symbols* φ *is well formed:*

- *If* φ *is atomic*
- *If* φ *is one of the following forms, with* ψ *and* χ *well-formed formulae:* $(\psi \wedge \chi)$, $(\psi \vee \chi)$, $(\psi \rightarrow \chi)$, $(\psi \leftrightarrow \chi)$, *or* $\neg\psi$
- *If* φ *is one of the following forms, with* ψ *well-formed formula, and some variable* x*:* $\exists x(\psi(x))$, $\forall x(\psi(x))$.
- *Nothing else is well formed*

From this definition then, we have a working formal language with which we can formulate propositions and theorems. We will also say that a variable x occurring in a formula φ is free if there are no quantifiers over it (e.g. $\forall x$ *or* $\exists x$); otherwise the variable is bound. A formula in which all the variables are bound is a sentence.

A note on the concept of equality:

We will treat the symbol = as part of the underlying logic, which is first-order predicate calculus. Thus the properties we present below of equality are taken as logical truths (Fraenkel, et al., 1973). Alternatively we could take the symbol as primitive, and thus the following are axioms prior to any of the axiomatizations of set theory we will be discussing. Another alternative is to provide a definition of = from first-order predicate calculus such that the following properties are provable (Fraenkel, et al., 1973). We will take the first attitude, as it makes things simpler, and can easily be applied to all the theories we shall consider. Thus we have for the notion of equality the following:

i) Reflexivity: $\forall x(x = x)$

ii) Symmetry: $x = y \rightarrow y = x$

iii) Transitivity: $(x = y \wedge y = z) \rightarrow x = z$

iv) Substitutivity: *For some formula* $\varphi(x)$; $(\varphi(x) \wedge x = x') \rightarrow \varphi(x')$
(Fraenkel, et al., 1973)

In (iv) we used the phrase "*For some formula* $\varphi(x)$". By this we mean φ is a well-formed formula which depends on the free variable x. We may also say that φ is a condition on x, and if φ contains other free variables we may say those are parameters on the condition.

III. **Naïve Set Theory**

We begin with looking at what is known as "*naïve set theory*". When we think of a collection from an unmathematical, or even an untrained, perspective, we use one of two principles in order make a set which we can refer to; by listing the elements that belong to the set, and by giving a property of all the members which thereby defines the set. For example, we can talk of the set consisting of "Graham Chapman, John Cleese, Terry Gilliam, Eric Idle, Terry Jones, and Michael Palin", or we can refer to the same set by the property of "actors of the Monty Python comedy group". The difference between these two principles is the very same difference between the concept of extension and intension of a concept or property. In naïve set theory, we make these two principles as axioms for the notion of set.

1. **Axiom of Extensionality:**

$$A = B \leftrightarrow \forall x(x \in A \leftrightarrow x \in B)$$

That is, $A = B$ if and only if A has the exact same elements of B.

2. **Axiom Scheme of Comprehension:**

For any formula φ*;*

$$\exists A \forall x\big(x \in A \leftrightarrow \varphi(x)\big)$$

That is, for any formula, or condition, on x, we can *'create'* a set A consisting of the x's such that $\varphi(x)$ holds true.

(Fraenkel, et al., 1973)

Note in the axiom scheme of comprehension, we used again the phrase "*For any formula* φ". The axiom scheme thus states an infinite number of axioms, each filling in a different well-formed formula φ.

The axiom schema of comprehension however leads to paradoxes, and thus any attempt to build a theory with it would prove to be inconsistent. We have already stated why from a postulationist perspective our theory should be consistent; moreover, consistency is important within any theory which uses first-order predicate calculus. This is because within this logic, if we have a proof of both φ and $\neg\varphi$ we can then prove any ψ. If this were the case, our theory is useless in giving us information about the properties and nature of the objects of study.

III.i. The Paradoxes

We have already noted the axiom scheme of comprehension leads to paradoxes. Here we shall state exactly the contradictions which arise from the axiom scheme, rendering naïve set theory inconsistent.

Russell's Paradox:

If we let $\varphi(x)$ be the well-formed formula $x \notin x$ in the axiom schema of comprehension, we get $\exists A \forall x(x \in A \leftrightarrow x \notin x)$. When $x = A$ we have an obvious contradiction; $A \in A \leftrightarrow A \notin A$ (Potter, 2004). What the axiom of comprehension allows us to do is create a set of entities that have a given property (φ). Russell's paradox then is the set of sets which are not members of themselves. Then the contradiction arises when we ask whether that set contains itself. This paradox is the most famous, and there exist other formulations which lead us further to doubt the existence of a universal set, or set of all sets.

Burali-Forti Paradox:

We will give a definition of a well-ordering below, as well as show that every well-ordered set has an associated ordinal number, which themselves are ordered. Considering then the set of all ordinals, as it is well ordered, it has an associated ordinal number, denote it δ, which must be greater than any element of the set. However the set is the set of ordinal numbers, so it must contain δ, a contradiction (Aken, 1986).

Cantor's Paradox:

This paradox gives another reason to deny the existence of the set of all sets. The paradox runs as follows. If we consider the set of all sets, denote it U, and its power set, or the set of all subsets of it, $\mathcal{P}(U)$ and compare their sizes, we find a contradiction. For the power set of any set is bigger than the set itself, by Cantor's Theorem; however the set of all sets U contains all sets, so it contains $\mathcal{P}(U)$ as a member (Blackburn, 2005).

III.ii. How to respond to the paradoxes

We have several ways in which to respond to these three paradoxes. The most prominent response is to relinquish the axiom scheme of comprehension, and provide multiple other axioms to supplement a weaker form of the axiom scheme. This is especially the case in Zermelo-Fraenkel set theory, where we have the axiom scheme of separation (or subsets), which allows us to construct a set of objects that satisfy some property φ if we restrict ourselves within a given set we already know to exist (Potter, 2004). Thus, what the common response to the paradox of Russell is to admit, as a theorem in most cases, that there is no universal set, or set of all sets. Likewise for the Burali-Forti paradox, there is no set of all ordinals. The ways in which the paradoxes are handled are dependent on the axioms we choose to start with, as well as if we add any other concepts into our theory.

We will first look at constructing Zermelo-Fraenkel set theory from the principle of *limitation of size,* and then contrast it with Potter-Scott set theory, which is equivalent to Zermelo set theory, but is built up in a more intuitive manner using *the iterative hierarchy* motivation. We will also briefly sketch the systems of von Neumann, Gödel, and Bernays which involves the concept of classes, and the related system of Quine which informally uses the notion of types. These different axiomatizations are, for a large part, equivalent to one another. However each presents a different motivation for the axioms they give, and thus a different intuitive picture of the set-theoretic universe. The traditional formulation of the axioms of Zermelo-Fraenkel set theory are motivated by ensuring the paradoxes are avoided, and thus the justification for the particular axioms given are they are what is needed to make a fully functional set theory in which we can do mathematics. This way of justifying axioms is termed regressive, that is the axioms give us as much of our intuitions as possible, and there is a mathematical field that can be studied which follows from the axioms (Potter, 2004). It is contrasted with an intuitive justification, where we provide axioms such that the primitive notions become sufficiently clear, and the axioms can be taken as true, given their consistency (Potter, 2004). This later form of justification is more in accordance with the postulationist attitude, however, the regressive strategy is the default way for a mathematician, who does not give much concern for metamathematical issues, and just wants to get on with the actual deduction of theorems.

IV. **Constructing Zermelo-Fraenkel Set Theory**

IV.i. Limitation of Size

The motivation behind this principle is to stick as close to our naive system, in which it was only the axiom of comprehension that proved inconsistent. It did so because it allowed us to create arbitrarily large sets (Fraenkel, et al., 1973). It follows to reason that we should avoid the paradoxes if we are careful with respect to the size of the sets which we create. This metamathematical dogma is what is known as the principle of *limitation of size*, wherein we do not create sets of too large a size compared to than ones we already have (Fraenkel, et al., 1973). The traditional axioms of Zermelo set theory are various operations we can use to construct larger sets from whatever starting ones we are given.

We start with the following;

Axiom of Pairing:

For any a and b, there exists the set $c = \{a, b\}$ which contains exactly just a and b.

$$\forall a \forall b \exists c \forall x \big(x \in c \leftrightarrow (x = a \vee x = b)\big)$$

Axiom of Union:

For any set a there exists the set $b = \bigcup a$ which contains exactly the members of a.

$$\forall a \exists b \forall x \big(x \in b \leftrightarrow \exists y (y \in a \wedge x \in y)\big)$$

Axiom of Powerset:

For any set a there exists the set $b = \mathcal{P}(a)$ which contains exactly all the subsets of a.

$$\forall a \exists b \forall x (x \in b \leftrightarrow x \subseteq a)$$

Axiom Schema of Separation:

For any set a and formula $\varphi(x)$ where x is free in φ, there exists the set $b = \{x : x \in a \wedge \varphi(x)\}$ which contains exactly all members of a such that $\varphi(x)$.

$$\forall a \exists b \forall x \Big(x \in b \leftrightarrow \big(x \in a \wedge \varphi(x)\big)\Big)$$

(Parsons, 1974)

The existence of null-set follows from letting $\varphi(x)$ be the well formed formula $x \neq x$, so that b has no elements (Fraenkel, et al., 1973). From the axiom of extensionality, it is easily seen that the null set is unique, and from separation, it is a subset of every set.

These axioms alone however will not allow us to form arbitrary infinite sets, although we do have an infinite number of objects which we can construct an infinite chain as sets of sets of sets of... and so on where each set is finite (Fraenkel, et al., 1973). From this though, we can give models for the natural and rational numbers. However, if we are interested in analysis, and in order to construct the reals, we need proper infinite sets which are not hereditarily finite (Fraenkel, et al., 1973). We therefore need the following axiom.

Axiom of Infinity:

There exists a set $\mathbb{N}$*, with the following properties:*

(i) $\emptyset \in \mathbb{N}$

(ii) If $x \in \mathbb{N}$ *then also* $\{x\} \in \mathbb{N}$

$$\exists \mathbb{N}[\emptyset \in \mathbb{N} \land \forall x(x \in \mathbb{N} \rightarrow \{x\} \in \mathbb{N})]$$

(Parsons, 1974)

Why can we be allowed to construct such a set? Is it not in conflict with our guiding principle, limitation of size? The idea is that we can construct an infinite number of objects, such as $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots$ (Fraenkel, et al., 1973). It seems quite intuitive we can form a set of these objects, especially by way of formalizing it as above so that it is finitely expressible (Fraenkel, et al., 1973). By way of notation, we have alluded to the fact that this axiom is able to provide a definition of the natural numbers, with $0 = \emptyset$, $1 = \{\emptyset\}$, $2 = \{\{\emptyset\}\}$, and so on. This gives us Peano's axioms of arithmetic, if we take $\{x\}$ to be the successor of any number x (Fraenkel, et al., 1973). This way of identifying the natural numbers within set theory is not unique however, and we can have different formulations of the axiom of infinity that prefer one model of the natural numbers over another.

Furthermore, we can construct relations and functions from the axioms given thus far. This would then allow us to create models for the rational and real numbers as well.

Relations

From the axiom of pairing, we have for any sets a and b the set $\{a, b\}$. From this, we wish to define the notion of an *ordered pair* (a, b) such that $(a, b) = (c, d) \leftrightarrow (a = c) \wedge (b = d)$. We have then the following

Definition: $(a, b) = \{\{a\}, \{a, b\}\}$

From this we can define *triples*, $(a, b, c) = ((a, b), c)$, and more generally, *n-tuples*, $(a_1, \dots, a_{n+1}) = ((a_1, \dots, a_n), a_{n+1})$ (Parsons, 1974). From these ordered objects, we can define relations. We first consider a general *n-ary relation* r before restricting ourselves to the primary discussion of *binary relations.* We say a set where each member is an n-tuple is an *n-ary relation* (Parsons, 1974). Thus a *binary relation*, henceforth simply a *relation*, is a set with every member is an ordered pair. We use the convention of writing $x\ r\ y$ or $r(x, y)$ (the latter more so if r has arity greater than 2) rather than $(x, y) \in r$ (Fraenkel, et al., 1973). It is simple to then define the domain and image of a relation. We give here the definition of the Cartesian product of two sets, so that we can see how a relation can be formed on a single set.

Definition: *The Cartesian product of two sets* a *and* b*, is given by*

$$a \times b = \{(x, y) : x \in a \wedge y \in b\}$$

Definition: *A relation* $r \subseteq a \times b$ *is a relation between* a *and* b*. A relation* $r \subseteq a \times a$*, i.e. a relation between a set* a *and itself, is called a relation on* a*.*

Definition: *An ordered pair* (a, r) *is a structure if* r *is a relation on* a*.*

Definition: *A relation* r *on* a *is*

i) *Reflexive on* a *if* $\forall x \in a(x\ r\ x)$
ii) *Irreflexive on* a *if* $\forall x \in a\big(\neg(x\ r\ x)\big)$
iii) *Transitive if* $\forall x, y, z \in a((x\ r\ y \wedge y\ r\ z) \rightarrow x\ r\ z)$
iv) *Symmetric if* $\forall x, y \in a(x\ r\ y \rightarrow y\ r\ x)$

v) *Antisymmetric if* $\forall x, y \in a\big((x\ r\ y \wedge y\ r\ x) \to x = y\big)$

(Parsons, 1974)

We can now define functions in terms of binary relations on sets.

Definition: *A binary relation f betweens sets a and b is called a function if* $\forall x \in a(\exists y \in b(x\ f\ y) \wedge [\exists z \in b(x\ f\ z) \to (z = y)])$. *The conventional notation for a function f which maps x to y is* $f(x) = y$.

(Fraenkel, et al., 1973)

Definition: *A function f from a to b is surjective if* $\forall y \in b\big(\exists x \in a(f(x) = y)\big)$. *A function f is injective if* $\forall x, x' \in a\big(x \neq x' \to f(x) \neq f(x')\big)$. *A function is a bijection, if it is both surjective and injective.*

(Potter, 2004)

We will later use two notions which are central to developing set theory, and we give their definitions here, though we will not use them till later.

Definition: *Two sets A, B are said to be equinumerous if there is a function f from A to B which is a bijection.*

Definition: *An isomorphism between structures (A, r) and (A', r') is a bijection between A and A' where* $\forall x, y \in A\big(x\ r\ y \leftrightarrow f(x)\ r'\ f(y)\big)$.

(Potter, 2004)

With this toolset, we can also introduce another axiom. The principle of limitation of size only restricts us to the size of each set, but we would like to be able, it seems intuitively reasonable, to have as many sets as we wish of any given size, if we already have one of that size. Therefore we introduce the following axiom scheme;

Axiom Scheme of Replacement:

For any set a, and any functional condition $f(x, y)$ on a, then there exists a set $b = \{y : f(x) = y\}$, which contains exactly those elements y for which $f(x) = y$ is true for some $x \in a$.

$$\forall a \left[\forall u \forall v \forall w (u \in a \wedge \varphi(u, v) \wedge \varphi(u, w) \to v = w) \to \exists b \forall y \Big(y \in b \leftrightarrow \exists x\big(x \in a \wedge \varphi(x, y)\big)\Big)\right]$$

In other words, for any set a we can take a function $f(x, y)$ from the elements $x \in a$ to other elements y and form a set of the output of $f(x, y)$ (Fraenkel, et al., 1973). That is to say, if the domain of a function is a set, then its range or image also forms a set (Fraenkel, et al., 1973).

The axiom scheme of replacement was not one of the original axioms of Zermelo set theory, but instead was first stated by Fraenkel (Potter, 2004). Thus adding it to the other axioms produces Zermelo-Fraenkel set theory. The justification for the addition of this axiom is regressive in nature, as it provides a simpler manner to develop a theory of ordinals, and other notions of set theory of much higher complexity (Potter, 2004). The scheme of replacement simplifies things as it unifies all the extra axioms one would need to develop those notions without the scheme.

Orderings

Definition: *A relation* $<$ *over a set* A *is said to be a partial ordering of* A *if* $<$ *is irreflexive and transitive. It is total ordering if* $\forall x, y \in A[(x < y) \vee (x = y) \vee (y < x)]$.

(Parsons, 1974)

We also make the following definitions on relations:

Definition: *For a partially ordered set* B *with partial ordering* $<$, *and non-empty subset* A, *and element* $a \in B$;

i) a *is a maximal element of* A *if* $a \in A \wedge \left(\forall x \in A(\neg a < x)\right)$

ii) a *is a minimal element of* A *if* $a \in A \wedge \left(\forall x \in A(\neg x < a)\right)$

We can also define the greatest and least elements, as well as upper and lower bounds, and the supremum and infimum of a set necessary to define a Dedekind cut to define the real numbers from the rational numbers (Parsons, 1974). As we are not interested in the formal construction of these structures, we omit the definitions. Instead we make the important definition of a well-ordering.

Definition: *A relation* r *on a set* A *is well-founded if every non-empty subset of* A *has an* r*-minimal element. We say for a partial ordering* $<$ *that it is a well-ordering on* A *if it is well-founded (alternatively, we say the set* $(A, <)$ *is well-ordered).*

(Potter, 2004)

A small corollary of this definition is if a partial ordering is a well-ordering, then it is a total ordering. From the concept of orderings, we can form what is known as the ordinal numbers. There is an entire theory of ordinal numbers, and here we give only the simple definition and a few resulting properties.

Ordinals

If we take the notion of membership, $\in$ as a relation, we can make a very natural definition of ordinal numbers.

Definition: *If a set A is well-ordered by $\in$ and every member of A is a subset of A, then it is an ordinal. That is if the following all hold:*

i) $\forall x, y((x \in y \wedge y \in A) \rightarrow x \in A)$*, every member of a member of A is itself a member. This is equivalent to the condition that every member of A is a subset of A.*
ii) $\forall x \in A(x \notin x)$*, irreflexitivity for the $\in$ relation to be a partial ordering*
iii) $\forall x, y, z \in A((x \in y \wedge y \in z) \rightarrow x \in z)$*, transitivity again for the $\in$ relation to be a partial ordering*
iv) $\forall x \subseteq A[x \neq \emptyset \rightarrow \exists y \in x(y \cap x = \emptyset)]$*, every non-empty subset of A has an $\in$-minimal element.*

(Fraenkel, et al., 1973)

We have the following propositions concerning the behavior of ordinals as easy consequences of this definition. We use α, β, γ as variables over ordinals, and $\alpha < \beta$ for $\alpha \in \beta$, and $\alpha \leq \beta$ for $(\alpha < \beta) \vee (\alpha = \beta)$.

1. $\forall \alpha(\alpha \notin \alpha)$
2. $\forall \alpha, \beta, \gamma[(\alpha < \beta) \wedge (\beta < \gamma) \rightarrow (\alpha < \gamma)]$
3. $x \in \alpha \rightarrow x \textit{ is an ordinal}$. This shows that $\forall \alpha[\alpha = \{\beta : \beta < \alpha\}]$, or every ordinal is the set of all ordinals less than it.
4. $\forall \alpha, \beta[(\alpha < \beta) \vee (\alpha = \beta) \vee (\beta < \alpha)]$
5. For some condition $\varphi(x)$, if there is some α such that $\varphi(x)$, then there is a β such that $\varphi(x)$ and there is no $\gamma < \beta$ such that $\varphi(x)$.
6. For every well-ordered set (a, r) there is a unique α such that (a, r) is isomorphic to $(\alpha, \in_\alpha)$ where $\in_\alpha$ is the relation given by the ordered pairs (β, γ) where $\beta < \gamma < \alpha$.

(Fraenkel, et al., 1973)

We can also define functions on ordinals by what is known as transfinite induction. We do this by first making the following definition.

Definition: *For a function f and set A such that $f(x)$ is defined for every $x \in A$, we define* $f|A = \{(x, f(x) : x \in A\}$*.* (Fraenkel, et al., 1973)

Then we can find a functional condition defined on ordinals for any function g defined over sets such that $\forall\alpha[f(\alpha) = g(f|\alpha)]$. As $f|\alpha = \{(\beta, f(\beta)) : \forall\beta < \alpha\}$, we have the following:

$$f(\alpha) = g(\{(\beta, f(\beta)) : \forall\beta < \alpha\})$$

Hence, to define a function of some ordinal α we must define it for all ordinals $\beta < \alpha$. (Fraenkel, et al., 1973)

By definition by transfinite recursion, we can develop an axiom which is seen as not fundamental to set theory being the foundational field for all of mathematics. If we left out any of the other axioms above, we could not fully develop set theory, and other branches of mathematics would be handicapped (Fraenkel, et al., 1973). The axiom however does offer a clearer picture of the set theoretic universe, and we will see it directly aids in the development of the iterative conception of set theory. We first define a function on the ordinals.

Definition: *The function* R *on all ordinals is defined by transfinite induction as follows;*

$$R(\alpha) = \bigcup_{\beta<\alpha} \mathcal{P}(R(\beta))$$

(Fraenkel, et al., 1973)

This function defines the iterative hierarchy from within the framework of the traditional axiomatization of Zermelo-Fraenkel set theory. We can think of the set $R(\alpha)$ for each ordinal α as a 'level' containing the subsets of the 'level' below (Fraenkel, et al., 1973). We will see that the $R(\alpha)$ in this development of Zermelo-Fraenkel set theory correspond to the V_α in Scott-Potter set theory. We will focus on developing the later as a more intuitive perspective of set theory in the next section. For now however, we only list a few theorems which follow for $R(\alpha)$:

Proposition: *For every ordinal* α, $\forall x, y\left((x \in y \wedge y \in R(\alpha)) \to x \in R(\alpha)\right)$

Proposition: *If* $\beta < \alpha$ *then* $R(\beta) \in R(\alpha)$ *and* $R(\beta) \subseteq R(\alpha)$

Proposition: $x \in R(\alpha) \leftrightarrow x \subseteq R(\alpha)$

(Fraenkel, et al., 1973)

We here give a definition so that we can formulate a final axiom to our theory. Also, the definition which follows defines the *well-foundedness* of a set A, which is a slightly distinct notion from the *well-foundedness* of a relation given above. We will see how the two notions are linked.

Definition: *A set A is well-founded if $A \in R(\alpha)$ for some ordinal α. The rank $\rho(A)$ of a well-founded set A is the least ordinal β such that $A \subseteq R(\beta)$.* (Potter, 2004)

We thus have the following immediate propositions:

Proposition: *For a well-founded set A*; $\rho(A) \leq \alpha \leftrightarrow A \subseteq R(\alpha)$

Proposition: *For a well-founded set A*; $\rho(A) < \alpha \leftrightarrow A \in R(\alpha)$

Proposition: *For a well-founded set A*; $B \in A \rightarrow B$ *is well-founded* $\wedge\ \rho(B) < \rho(A)$

Proposition: *If all members of a set A are well-founded, then A is well-founded.*

(Fraenkel, et al., 1973)

We now can give the **Axiom of Foundation:** *Every set is well-founded.*

We can equivalently say that *every non-empty set has an* $\in$*-minimal element, that is* $\forall A(A \neq \emptyset \rightarrow \exists x \in A(A \cap x = \emptyset))$ (Potter, 2004). Thus this axiom states that for all sets, the $\in$ relation is well-founded.

It is important to note that the axiom of foundation is considered by many as strictly unnecessary (Potter, 2004). For we can still construct ordinals, as well as cardinal numbers, and other mathematical structures without direct use of the axiom. Indeed, the only reason to refute the axiom of foundation is for greater generality, such that there can be sets which are not well-founded. This however is unnecessary as "no field of set theory or mathematics is in any need of sets which are not well-founded."(p.88) (Fraenkel, et al., 1973).

We will leave discussion of cardinal numbers for within the framework of the iterative hierarchy, as they are easier to define and work with given that set up. However this is not to say they cannot be formulated in the traditional manner, in fact most of the development of the theory of cardinal numbers was done within this axiomatization of Zermelo-Fraenkel set theory (Fraenkel, et al., 1973).

IV.ii. The Iterative Conception

Another motivation quite different from the principle of limitation of size that guided us above is the iterative conception of sets. This conception forms what is known as the cumulative hierarchy. We have already seen one manner in which it can be presented through the axioms of Zermelo-Fraenkel set theory. However there have been developments of the iterative conception from a more intuitive point of view by Scott, Potter, and Boolos. Each of these developments use slightly different axioms, but the theories developed are largely equivalent. We will be focusing on presenting the conception which Michael Potter gives in his book *Set Theory and its Philosophy* (2004).

The manner in which the iterative conception is formed, as its name suggests, is by a recursive definition, whereby we use stages to construct 'levels' that form a hierarchy (Boolos, 1971). We start with the lowest level V_0 containing all the individuals, or the null set (Boolos, 1971).[1] The next stage we take all the subsets of the first level thus $V_1 = \mathcal{P}(V_0)$ (Boolos, 1971). We continue in this manner such that for each ordinal α we have $V_{\alpha+1} = \mathcal{P}(V_\alpha)$ and for a limit ordinal γ we have $V_\gamma = \bigcup_{\alpha<\gamma} V_\alpha$ (Boolos, 1971). Thus the hierarchy is cumulative, once a set is formed at one level, it is a member of all sequent levels (Potter, 2004).

Formal Construction:

We will follow Michael Potter's construction of the cumulative iterative hierarchy. We begin with some definitions and lemmas.

Definition: *For a formula $\varphi(x)$, we may abbreviate 'the collection of all x such that $\varphi(x)$' by* $\{x: \varphi(x)\}$.

Lemma: *If* $a = \{x: \varphi(x)\}$ *exists, then* $\forall x(x \in a \leftrightarrow \varphi(x))$ *, for some formula* $\varphi(x)$.

Lemma: *For some formulae $\varphi(x)$ and $\psi(x)$, we have*

$$\forall x\big(\varphi(x) \leftrightarrow \psi(x)\big) \rightarrow \{x: \varphi(x)\} = \{x: \psi(x)\}$$

(Potter, 2004)

[1] Michael Potter allowed for individuals in his construction of the iterative hierarchy in order to allow set theory to serve as a container for any other theory of any particular field. That is, given some theory T, we can embed it within our set theory. We will leave out the precise formulation of how it is done, and we will assume there is the possibility of having some individuals. However this assumption is not necessary and for a mathematician who wishes to cut everything which is not needed, it is easy to drop, and ignore all points when we say *x is an individual*, replacing it with the null set. Thus, the first level would be the null set.

Potter here introduces Russell's paradox, prior to any axioms, and as a theorem we must accept.

Russell's Paradox: $\{x: x \notin x\}$ *does not exist.*

Definition: *If* $b = \{x: x \in b\}$ *then* b *is a collection.*

Lemma: *For some formula* $\varphi(x)$*, if* $\{x: \varphi(x)\}$ *exists, then it is a collection.*

Lemma: *For some formula* $\varphi(x)$*, we have*

$$\exists a \forall x\big(x \in a \leftrightarrow \varphi(x)\big) \leftrightarrow \{x: \varphi(x)\}$$

(Potter, 2004)

We can then produce the Extensionality axiom above as a theorem:

Extensionality Principle:

$$\forall x(x \in a \leftrightarrow x \in b) \rightarrow a = b$$

Proof: From the lemma above, we have $\forall x(x \in a \leftrightarrow x \in b) \rightarrow \{x: x \in a\} = \{x: x \in b\}$, where $\varphi(x) = x \in a$ and $\psi(x) = x \in b$. Thus $\{x: x \in a\} = \{x: x \in b\} \leftrightarrow a = b$ as required. (Potter, 2004)

At this point, there have been no axioms, thus no commitments to any collection's ontology; that is, its existence. We make a few more definitions concerning operations on collections:

Definition: *We write* $a \subseteq b$ *as an abbreviation for the formula* $\forall x(x \in a \rightarrow x \in b)$*, and* $a \subset b$ *for* $(a \subseteq b) \wedge (a \neq b)$*.*

Definition: $\emptyset = \{x: x \neq x\}$

Definition: $a \setminus b = \{x: x \in a \wedge x \notin b\}$

Definition: $\mathcal{P}(a) = \{b: b \subseteq a\}$

Definition: $\bigcap a = \{x: \forall b \in a(x \in b)\}$

Definition: $\bigcup a = \{x: \exists b \in a(x \in b)\}$

(Potter, 2004)

To begin the level construction, we need a few definitions, before we introduce any axioms.

Definition: *The accumulation of* a *is given by the following*

$$acc(a) = \{x: x \text{ is an individual} \vee \exists b \in a(x \in b \vee x \subseteq b)\}$$

Definition: $\mathcal{V}$ *is called a history if* $\forall V \in \mathcal{V}\big(V = acc(\mathcal{V} \cap V)\big)$

Definition: *The accumulation of a history is called a level.*

Definition: *A set,or a grounded collection, is some subcollection of a level.*

Axiom Scheme of Separation:

For a formula $\varphi(x)$*, the following is an axiom:*

$$\forall V(\{x \in V: \varphi(x)\}\ exists)$$

(Potter, 2004)

From which we are able to give another version of Russell's paradox: *There is no set of all sets*.

This is shown from the following proposition and proof.

Proposition: *There is no set* b *such that* $\forall a(a \subseteq b \rightarrow a \in b)$*.*
Proof: Let there be such a set b. Then $\exists V(b \subseteq V)$. And let $a = \{x \in b: x \notin x\}$. Thus we have $a = \{x \in V: x \in a \wedge x \notin x\}$, which exists by the axiom scheme of separation. Clearly, $a \subseteq b$, but if $a \in b$ then $a \in a \leftrightarrow a \notin a$, a contradiction. (Potter, 2004)

Thus if the set of all sets existed, it would contain as members all subsets of it, which we have just proven impossible. Potter in fact uses this result to prove some facts about the behavior of levels.

Definition: *A collection* a *is transitive if* $\forall b \in a(\forall x \in b(x \in a))$*.*
Proposition: *Every level is transitive, and hence* $a \in V \rightarrow a \subseteq V$

Proposition: $V = acc(\{V': V' \in V\})$

From which we can say if $V_1 \in V_2$ then V_1 *is lower than* V_2 and any level is the accumulation of all lower levels (Potter, 2004). Thus "the hierarchy of levels is cumulative: if an object belongs to a particular level, then it belongs to all subsequent levels." (p 45) (Potter, 2004).

Proposition: *For any two levels* V_1, V_2 *we have the following*

$$(V_1 \in V_2) \vee (V_2 \in V_1) \vee (V_1 = V_2)$$

Proposition: *For any level* V*, we have* $V \notin V$*.*

Proposition: $V \subseteq V' \leftrightarrow (V \in V' \vee V = V')$

Proposition: $V \subseteq V' \vee V' \subseteq V$

Proposition: $V \subset V' \leftrightarrow V \in V'$

(Potter, 2004)

We can also prove a few facts about sets.

Proposition: *For some formula* $\varphi(x)$*,* $\{x: \varphi(x)\}$ *is a set if and only if there is a level* V *such that* $\forall x(\varphi(x) \rightarrow x \in V)$*.*

Definition: *For a set* a*, the lowest level* V *such that* $a \subseteq V$ *is called the birthday of* a *and denoted* $\mathbb{V}(a)$*.*

Proposition: *If* a *is a set then* $a \notin a$*.*

Proof: $a \subseteq \mathbb{V}(a)$ from the definition above. If $a \in a$, then $a \in \mathbb{V}(a)$. But as we have from above, $V = acc(\{V' : V' \in V\})$ and thus $\mathbb{V}(a) = acc(\{V' : V' \in \mathbb{V}(a)\})$, therefore we have $\exists V \in \mathbb{V}(a)(a \subseteq V)$, a contradiction of the definition of $\mathbb{V}(a)$.

Proposition: *For some formula* $\varphi(x)$ *and set* a*,* $\{x \in a : \varphi(x)\}$ *is a set.*

(Potter, 2004)

This proposition is equivalent to the axiom scheme of separation of Zermelo set theory we constructed in the previous section. We also have the formulation of the axiom of foundation as a proposition which can be proved from what has been given. We can begin to see how with one axiom in the iterative conception we can derive several axioms of traditional Zermelo set theory. Below we also have the axioms of union and powerset as propositions.

Foundation Principle: $a \neq \emptyset \rightarrow \exists x \in a(x \text{ is an individual} \vee x \cap a = \emptyset)$

Proposition: *If* a *is a nonempty set of sets, then* $\bigcap a$ *is a set.*

Proposition: *If* a, b *are sets, then* $a \cap b$ *is a set.*

Proposition: *If* a, b *are sets, then* $a \setminus b$ *is a set.*

Proposition: *If* a *is a set, then* $\bigcup a$ *is a set.*

Proposition: *If* a, b *are sets, then* $a \cup b$ *is a set.*

(Potter, 2004)

As the axiom scheme of separation quantifies over all levels, in order to posit sets' existence, all instances of it will be true even if there are no levels (Potter, 2004). Potter thus uses a temporary axiom; *there is at least one level*, to start things off and to prove the nullset $\emptyset$ is a set. The axiom of infinity Potter uses covers the existence of one level which ensures the initial level, as well as infinitely others (Potter, 2004). We shall continue thus first giving definitions to postulate the axiom of creation, from which we can work our way towards the axiom of infinity.

Definition: *The level above (or after) a level* V *is the lowest level* V' *such that* $V \in V'$*.*

Proposition: *If* V' *is the level after* V *then* $\forall x(x \in V') \leftrightarrow (x \text{ is an individual} \vee x \subseteq V)$*.*

Axiom of creation: *For each level* V *there is a level* V' *such that* $V \in V'$.

Proposition: *For any set* a*, the power set* $\mathcal{P}(a)$ *is a set.*

Proposition: *For any set* a*, we also have the set* $\{a\}$. *Thus for any two objects (sets or individuals)* x, y *the set* $\{x, y\} = \{x\} \cup \{y\}$ *exists.*

(Potter, 2004)

The axiom of creation and the scheme of separation however still has yet to provide us with an actual level we can work with, thus we introduce the last axiom to the iterative hierarchy.

Definition: *A limit level is a level* V *that is neither the initial level nor the level above any other level. Thus* V *is a limit level iff* $\forall x \in V(\exists V' \in V(x \in V'))$.

Axiom of Infinity: *There exists at least one limit level.*

Definition: *Let the lowest limit level be denoted* V_ω.

Proposition: *The history of* V_ω *is infinite.*

(Potter, 2004)

From these three axioms, the separation scheme, creation, and infinity, we can prove all the standard mathematical equipment, such as ordered pairs, cross products, relations, functions, from which we can create models of the natural, rational, and real numbers, that the axioms of Zermelo set theory (without the replacement scheme) affords us.

Cardinals

Definition: *For any set* A*, the set* $\langle X : X \text{ and } A \text{ are equinumerous} \rangle$ *is called the cardinality of* A *and denoted* $|A|$. *A cardinal number is the cardinality of some set.*

Theorem Hume's principle*:* $|A| = |B|$ *iff* A *and* B *are equinumerous.*

(Potter, 2004)

This is the guiding principle in the development of the theory of cardinals. We only give a brief sketch of their definition to show how Cantor's paradox is proven. In the following, we will use $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \ldots$ to denote cardinal numbers.

Definition: *Let* $\mathfrak{a} = |A|$ *and* $\mathfrak{b} = |B|$*, then* $\mathfrak{a} \leq \mathfrak{b}$ *iff there is an injection from* A *to* B. *We let* $\mathfrak{a} < b$ *to be* $\mathfrak{a} \leq \mathfrak{b} \wedge \mathfrak{a} \neq \mathfrak{b}$

Bernstein's Equinumerosity Theorem: *If there exists an injection f from A to B, and an injection g from B to A, then there exists a bijection between A and B.*

Proposition: *The relation $\leq$ partially orders the cardinals as we have the following:*

i) $\mathfrak{a} \leq \mathfrak{a}$
ii) $(\mathfrak{a} \leq \mathfrak{b} \wedge \mathfrak{b} \leq \mathfrak{c}) \rightarrow \mathfrak{a} \leq \mathfrak{c}$
iii) $(\mathfrak{a} \leq \mathfrak{b} \wedge \mathfrak{b} \leq \mathfrak{a}) \rightarrow \mathfrak{a} = \mathfrak{b}$

Proposition: *For a formula φ, the set $B = \{\mathfrak{a}: \varphi(\mathfrak{a})\}$ exists iff $\exists \mathfrak{c}(\varphi(\mathfrak{a}) \rightarrow \mathfrak{a} \leq \mathfrak{c})$.*

Cantor's Theorem: *If A is a set, then there are injective functions from A to $\mathcal{P}(A)$, but there are no surjective functions from A to $\mathcal{P}(A)$.*

Corollary: *For every cardinal $\mathfrak{a}$ there is a cardinal $\mathfrak{a}'$ such that $\mathfrak{a} < \mathfrak{a}'$.*

Proposition: *There is no set of all cardinal numbers.*

(Potter, 2004)

This then is Cantor's Paradox, but stated as a theorem we accept readily from what we have defined and stated as axioms. We can still however go on to derive all the properties of finite and infinite cardinal numbers, including their arithmetic and functions over them. Instead we turn here again to ordinals, and how they are formulated within the iterative hierarchy.

Ordinals

Definition: *For any structure (A, r), the set $\langle (X, z): (X, z)\ and\ (A, r)\ are\ isomorphic \rangle$ is called the order type of (A, r) and denoted $ord(A, r)$. An ordinal number is the order type of a well-ordered set.* (Potter, 2004)

In fact this definition of ordinals is largely equivalent to the definition we presented above. We can prove the following for ordinal numbers $\alpha, \beta, and\ \gamma$.

Proposition:

i) $(\alpha \leq \beta \wedge \beta \leq \gamma)$
ii) $\alpha \leq \alpha$
iii) $(\alpha \leq \beta \wedge \beta \leq \alpha) \rightarrow \alpha = \beta$
iv) $\alpha \leq \beta \vee \beta \leq \alpha$

(Potter, 2004)

We can then go on to show everything we have seen about ordinals, including their arithmetic, and how to define a function over them by transfinite induction. One result we will state here is the **Burali-Forti's Paradox:** $\{\alpha : \alpha is\ an\ ordinal\}$ *does not exist.* Therefore the ordinals are not totally ordered by $\leq$ (Menzel, 1986). There is debate about why the ordinals form an inaccessible set, see Menzel's essay *On the Iterative Explanation of Set* for a more detailed account of this.

IV.iii. Axiom of Choice:

Definition: *A choice function is a function* f *such that* $\forall A \in dom[f](f(A) \in A)$.

Axiom of Choice: *For every set* $\mathcal{A}$ *of disjoint nonempty sets there exists a set* C *such that* $\forall A \in \mathcal{A}$ *the set* $C \cap A$ *has exactly one member.*

(Parsons, 1974)

This is equivalent to saying that every set has a choice function. The axiom of choice can be used to prove Zermelo's well-ordering principle, which is '*every set is well-orderable'* (Potter, 2004).

The axiom of choice is independent of all the other axioms, and can be formulated in numerous ways. It can be added to either construction of Zermelo-Fraenkel set theory, as both the limitation of size and iterative hierarchy produce the same theories, whichever construction one uses. The main justification for the axiom of choice is regressive, it allows us to prove a lot of important results in various branches of mathematics (Potter, 2004). There is not much intuitive justification for it as it is difficult to see exactly how it further defines the primitive notion of membership.

V. Theories with classes and types

In the iterative conception above, Potter began with the informal notion of a *collection*, though his theory did not treat explicitly them as objects. There is a closely related system, that of von Neumann, Gödel, and Bernays, which admits *classes* in addition to sets (Fraenkel, et al., 1973). The systems of these different mathematicians do differ slightly, but here we present them as a single theory with the following axioms which determine the role of classes. We use the variables $A, B, C, A', B', C', \ldots$ to range over classes.

Axiom of Extensionality over Classes: $\forall x(x \in A \leftrightarrow x \in B) \rightarrow A = B$

Definition: *A class x is a set if $\exists A(x \in A)$, where A is a class. A proper class is a class which is not a set.*

Axiom of Predicative Comprehension for Classes: *For some formula $\varphi(x)$ which only quantifies over sets, $\exists A = \{x: \varphi(x)\}$, where A is the class containing as members exactly those sets x such that $\varphi(x)$ holds.*

(Holmes, 2012)

The Russell paradox then simply becomes a class, specifically the proper class $\{x: x \notin x\}$. We also have the following axiom:

Axiom of Limitation of Size: *A class C is proper iff there is a class bijection between C and the universe* (Holmes, 2012).

However, these axioms are not in themselves sufficient to create the sets and classes which we work with in Zermelo-Fraenkel set theory. But it is a simple matter of importing those axioms into our system with classes (Fraenkel, et al., 1973). The axioms of extensionality and choice from ZF are not needed, however (Holmes, 2012). It seems that classes are integral part of set theory, even if one is working in a system which does not admit them as mathematical objects. The concept of a class is a useful tool in the metamathematics to gain a clearer intuitive picture of the set-theoretic universe.

V.i. Quines NF theory

Quine's answer to Russell's paradox and the others was to restrict the variables in formulae so that we cannot have statements such as $x \notin x$. This was done by first restricting variables to be of certain types, indexed by the natural numbers as a typographical shorthand (Holmes, 2012). The point is we do not need a theory of natural numbers to use them as an indexing set (Holmes, 2012). This is similar to using the ordinals to index the levels in the iterative conception, apart from the fact that we define the ordinals in terms of the levels.

Thus variables of type 0 range over individuals, type 1 range over sets of type 0 objects, and in general type $n + 1$ variables range over sets of type n objects (Holmes, 2012). Using this type restriction of variables, we can say a formula is *stratified* if all the variables which occur in the formula are assigned a certain type (Fraenkel, et al., 1973). Therefore the atomic cases of well formed formula are $x^n = y^n$ and $x^n \in y^{n+1}$, where the superscript denotes the variable type (Holmes, 2012). Quine's goal was that this restriction would allow for the naïve axioms to follow without inconsistencies, but that ultimately we could drop the type restrictions in the formalization (Holmes, 2012). Thus the axioms which Quine presented in his paper *New Foundations for Mathematics*, known as the 'New Foundations' system (NF) are simply those of naïve set theory:

Axiom of Extensionality: $\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y)$
Axiom Schema of Stratified Comprehension: $\exists y \forall x (x \in y \leftrightarrow \varphi(x)$ *where* $\varphi(x)$ *is a stratified formula in which* y *does not occur free.*

(Holmes, 2012)

Quine held that the type restrictions on variables were not necessary in the formalization of the system, as for any formula φ which is stratified, we can associate with it the formula φ^+ where all the indices of type occurring in φ are raised by 1 (Holmes, 2012). It can be shown that any proof of φ can be shown to also prove φ^+, and thus Quine conjectured $\varphi \equiv \varphi^+$ for all formula φ (Holmes, 2012). From this, we can drop the type distinction of the variables in the formulae of the system in the formal notation, while still holding as an informal principle that we only admit stratified formulae into our system. We thus avoid Russell's paradox, as the set $\{x: x \notin x\}$ is not given from the schema of stratified comprehension (Holmes, 2012). We do however get other sets which Zermelo set theory cannot have, for example the universal set $V = \{x: x = x\}$ (Holmes, 2012).

Furthermore, in Quine's NF, while no contradictions follow from these axioms, other anomalies do occur. They are not contradictory, but mere drawbacks when compared to the standard Zermelo-Fraenkel set theory. For example, as the universal set exists in NF, it is a set which is equinumerous with its power-set (Holmes, 2012). Moreover, the axiom of choice fails in NF. However, this does not affect NF from being able to provide a model for the natural numbers, and thus, with a few extensions, serve as a foundation for mathematics (Fraenkel, et al., 1973).

VI. **Conclusion**

Mathematician or not, we all have a rough idea of what the real number line looks line. We can imagine an infinite line stretching in both directions, with tick marks equally spaced by a given unit all centered from the mark 0. The natural numbers are even easier to conceive of. A tape of blocks labeled from 0 to 1 to 2 and so on forever. Moving from one to the next is just adding one, or in Peano's formulation simply the successor of each number, with 0 not being the successor to any. Even the rational numbers can be seen loosely in this way, with all the blocks divided up an infinite amount of times such that between any two rational numbers, there is another we can find greater than the first, and less than the second. When we come to sets however, the picture seems less clear, especially using the traditional axioms of Zermelo-Fraenkel set theory. The axioms seem quite *ad hoc* in nature, given merely to avoid paradox rather than to be statements which give implicit meaning to the notion of set, as a postulationist attitude seems the most appropriate to take. The iterative hierarchy however affords us with an intuitive picture of how sets can be made, and offers a relatively clear implicit definition of what it means for an entity to be a set. We start with the set of all things, and can form any subset of them. From which we can construct sets of sets, and so on, as we go up the hierarchy, we create more subsets from what we already have, which can then be treated as elements of higher rank sets.

Even still, as Fraenkel states, "as we see it, there exists a multi-dimensional continuum of possible attitudes" (p.154) (Fraenkel, et al., 1973) we can take in response to the paradoxes of naïve set theory. There are countless more set theories constructed from different axioms. These range from the conservative ways of finding ways to restrict the axioms of naïve set theory so that they are consistent, to developing an entirely different mathematics (Holmes, 2012). Intuitionistic set theory does such by relying on intuitionistic logic. This is different from classical logic by denying the law of excluded middle, which states for all models for any formula $\varphi(x)$ we have $\varphi(x) \vee \neg\varphi(x)$ (Iemhoff, 2009). Thus as a consequence, we do not have the following for any formula $\varphi(x)$;

$\neg\neg\varphi(x) \rightarrow \varphi(x)$. As such we cannot prove the existence of some object, or the truth of some condition on an object, by supposing the converse true and deriving a contradiction (Iemhoff, 2009). The development of set theory then becomes much more complex.

However, it seems from the presentation of the iterative conception given, this is unnecessary if we are interested in developing set theory as a foundation for mathematics, or even merely a large branch of mathematics with numerous sub-theories. From the theory of ordinals, to cardinals, and higher systems of infinity, set theory is a rich field of mathematics, with numerous ways to ground it in axioms. In order to make sense of which system we wish to take, we can appeal to meta-mathematical and philosophical criteria. Though as the every-day mathematician is largely unconcerned with those issues, we can choose axioms from a simple pragmatic view, and investigate the consequences of a particular axiom system for no motivation other than mathematical curiosity.

Bibliography

Aken, J. V., 1986. Axioms for the Set-Theoretic Hierarchy. *The Journal of Symbolic Logic,* 51(4), pp. 992-1004.

Blackburn, S., 2005. *Oxford Dictionary of Philosophy.* Oxford, UK: Oxford University Press.

Boolos, G., 1971. The Iterative Conception of Set. *The Journal of Philosophy,* 68(8), pp. 215-231.

Fraenkel, A. A., Bar-Hillel, Y. & Levy, A., 1973. *Foundations of Set Theory.* 2nd ed. Amsterdam: North-Holland Publishing Company.

Holmes, R. M., 2012. *Alternative Axiomatic Set Theories.* [Online]
Available at: http://plato.stanford.edu/archives/sum2012/entries/settheory-alternative/
[Accessed 27 April 2012].

Iemhoff, R., 2009. *Intuitionism in the Philosophy of Mathematics.* [Online]
Available at: http://plato.stanford.edu/archives/win2009/entries/intuitionism/
[Accessed 27 April 2012].

Menzel, C., 1986. On the Iterative Explanation of the Paradoxes. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition,* 49(1), pp. 37-61.

Parsons, C., 1974. Sets and Classes. *Noûs,* 8(1), pp. 1-12.

Potter, M., 2004. *Set Theory and its Philosophy: A Critical Introduction.* New York: Oxford University Press.

Russell, B., 1930. *Introduction to Mathematical Philosophy.* 2nd ed. London: George Allen & Unwin, Ltd..