The Axiom of Choice

—

MA213 Second Year Essay

0511028

1 The axioms of set theory

The reader is probably familiar with the basic concepts of set theory. Indeed, they permeate through the whole of mathematics, from analysis — where sets form the domains and codomains of functions — to algebra — where many different algebraic structures can be considered as sets equipped with various additional properties.[1] What is perhaps less well-known is the axiomatic formulation of set theory, developed in the late nineteenth and early twentieth centuries by individuals including Cantor, Peano, Zermelo, Fraenkel, Skolem, Russell and Bernays.[2]

In 1901 the English philosopher Bertrand Russell proposed his now famous paradox:

Let M be the set of all sets that do not contain themselves. That is,

$$M = \{A \mid A \notin A\}.$$

Is M itself an element of M?

Suppose $M \in M$. Then M must satisfy the condition that $M \notin M$, contradicting our assumption. Similarly, if we assume that $M \notin M$ then M fulfills the condition of not belonging to itself and hence $M \in M$, another contradiction. In other words, $M \in M$ implies $M \notin M$ and vice versa. Clearly this is absurd!

How do we get around such paradoxes?[3] The problem ultimately lies in what we allow to be sets. In some sense, collections such as "the set of all sets that don't contain themselves", or indeed the "set of all sets", are simply too big to be allowed to be sets. There are essentially two ways to progress, both of which are based on restricting the usage of the term "set". One option is to allow such entities to exist, but not as sets; instead we introduce the additional

[1]For example, a group is a set G together with an associative binary operation on G, an identity element and an inverse for every element of G.

[2]Georg Cantor (1845–1918), Giuseppe Peano (1958–1932), Ernst Zermelo (1871–1953), Adolf Fraenkel (1891–1965), Thoralf Skolem (1887–1963), Bertrand Russell (1872–1970), Paul Bernays (1888–1977).

[3]Russell's paradox is not the only paradox arising in so-called "naive set theory", but is one of the more straightforward to describe. Another example, found in Cameron [**1**], relies on the notion of a "game", by which we mean an activity for two players in which the players take it in turns to do some sort of "move". A game is said to be "well-founded" if it always ends after a finite number of moves. As an example, consider the game where the first player picks an arbitrary positive whole number and the players then take it in turns to pick a positive whole number less than the previous one; this game is clearly well-founded. By virtue of its "draw by repetition" rule, the game of chess is also well-founded. Now consider the following game, the "hypergame": the first player picks any well-founded game; then the players (starting with player two) play this game until its completion. Since each of the games that can be picked in the first move is well-founded and the hypergame only lasts for only one move more, the hypergame is clearly well-founded too, isn't it? But if so, then player one could pick the hypergame on his first move, and then player two could pick the hypergame, and then player one could pick it, etc. *ad infinitum*, contradicting the fact that the hypergame is well-founded.

notion of *proper classes*, for example the "class of all sets", which may not be members of any other class.[4] Alternatively we can lay out rules (called axioms[5]) for how to construct sets from the fundamental concept of the existence of an empty set $\varnothing$, a set with no members; anything that can be constructed from application of these rules is deemed to be a set, while collections such as the collection of all sets are never reached and are therefore not sets. It is easy to see that Russell's paradox can no longer arise in either of these systems.

In this essay we will be concerned with the second of these options, known as Zermelo–Fraenkel set theory. The initial axioms were laid out by the German mathematician Ernst Zermelo in the first decade of the twentieth century and later added to independently by Adolf Fraenkel and Thoralf Skolem. Depending on which version you look at there are between nine and twelve axioms[6], the first of which seem almost trivial. For example, the first axiom, the Axiom of Extensionality, states that[7]

Two sets are equal if and only if they contain the same elements. Formally,

$$(\forall x)(x \in A \iff x \in B) \iff A = B.$$

For anyone with a reasonable level of familiarity with mathematics, this seems obvious. However, once the basic rules for creating pairs, unions and power sets, etc. are listed we are left with some rather more troublesome — and controversial — axioms, none more so than the so-called Axiom of Choice.

2 The Axiom of Choice

There are many different ways to formulate the axiom of choice. The first definition we give is in terms of choice functions, in which it can be stated very concisely.

Axiom of Choice. *Any collection of non-empty sets has a choice function.*

What is a choice function, exactly? Intuitively it is a function that somehow "chooses" a unique element from each set. More precisely, if C is a collection of non-empty sets, then a choice function on C is a function $f\colon C \to \bigcup C$ such that for every $X \in C$ we have

$$f(X) \in X.$$

Here the notation $\bigcup C$, known as the union set of C, refers to the set of all members of members of C. Note that in the case where $C = \{X_i \mid i \in I\}$, $\bigcup C$ is just another way of writing $\bigcup_{i \in I} X_i$. If C *is* indexed by a set I, i.e. if $C = \{X_i \mid i \in I\}$, then an alternative way of describing a choice function is as a function $f\colon I \to \bigcup_{i \in I} X_i$ such that for each $i \in I$ we have

$$f(i) \in X_i.$$

The equivalence of these definitions follows from the fact that in the case where C is indexed there is a natural bijection between the index set and C. The details are omitted.

Another commonly cited definition of the axiom is

[4] This is the approach using taken in the branch of mathematics known as category theory, in which mathematical objects, such as sets, groups and topological spaces are considered together with their structure-preserving functions as "categories" and the relationships between different categories are studied.

[5] These are not axioms in the sense that Euclid used the term, i.e. self-evident truths. Rather they are assumptions on which we base a theory; whether their truth is absolute in any sense is irrelevant.

[6] For a full list, see e.g. Suppes [**9**] or Skolem [**7**].

[7] In fact, from a very technical point of view, this axiom only tells us about the "if" direction; the other implication follows from the definition of the equality symbol and its property of substitution.

Given any collection of non-empty sets there exists a set containing precisely one element from each set.

This form of the axiom gives perhaps the clearest idea of what the axiom of choice actually allows us to do, in a "practical" sense: namely, the axiom allows us to make infinitely many choices simultaneously, which otherwise would not be possible. (There are exceptions; see below.)

2.1 Situations where the Axiom of Choice is not needed

There are several situations where the axiom of choice is not actually needed. For example, if each of the sets in the collection C is a singleton set[8] then there is no choice involved: for each set we simply pick its only element as our representative of that set. If the number of sets in the collection is finite then we can prove the existence of a choice function by induction on the number of sets.[9] Furthermore, if every set in the collection has some sort of structure (if each set is well-ordered, for example. See Section 3.2.) then it is possible to pick an element without resorting to the axiom of choice. For instance, if C is a collection of groups we can simply take the identity element of each group as our representative; if each element is a set of natural numbers we can take the smallest element each time. If each set is a pair of shoes the axiom of choice is not required — we simply pick the left shoe; but if we are dealing with a collection of pairs of indistinguishable socks then it is absolutely necessary![10]

The reason we need the axiom of choice when the collection of sets is infinite is because the process of picking one element from each set in turn will never terminate. For a finite collection of non-empty sets the process will be complete after a finite period of time, but in the infinite case we need the axiom of choice to allow us to effectively make infinitely many choices at once. The website [**5**] contains a particularly clear example of this distinction.

If we do not want to allow the axiom of choice in its full form, there are several weaker forms of the axiom that can help to prove some, but not all, of the results that the axiom of choice implies. Probably the most common of these is the so-called Axiom of Countable Choice. Predictably, this states that

Any countable collection of non-empty sets has a choice function.

The axiom of countable choice allows us to prove that any countable union of countable sets is itself countable, something which is very useful in many areas of mathematics, including topology.

2.2 An alternative formulation of the Axiom of Choice

Although the definition we have given in terms of choice functions is perhaps the most intuitive with regards to how the axiom gets its name there are many alternative formulations found in the literature, of which the following is one of the most common.

The Cartesian product of any collection of non-empty sets is non-empty. That is, if the sets $X_i \neq \varnothing$ for each i in some index set I then

$$\prod_{i \in I} X_i \neq \varnothing.$$

[8] A singleton set is a set with precisely one member.

[9] See Jech [**3**].

[10] This example is attributed to Bertrand Russell.

We will now show how this definition is equivalent to the definition given in the previous section as it is an interesting example of how concisely a complex idea can be stateful with the right terminology. The first thing that needs to be determined is precisely what is meant by the Cartesian product of an arbitrary collection of sets; for instance, what does an element of such a product look like? Recall the definition of the Cartesian product of two sets, encountered in elementary set theory.

Definition 1. Let A and B be sets. The *Cartesian product* of A and B, written $A \times B$, is defined to be the set of ordered pairs[11] (a, b) such that $a \in A$ and $b \in B$. In symbols

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}.$$

This definition can easily be extended to include the Cartesian product of three sets A, B, C, where an element of $A \times B \times C$ is an ordered triple[12] (a, b, c) such that $a \in A, b \in B$ and $c \in C$. Indeed this can be further extended to include the Cartesian product of any finite number of sets $X_1, \ldots, X_n$ for a natural number n in the obvious way; hence, we can describe $\prod X_i$ in the case where I is some finite subset of the natural numbers. For simplicity we will assume that $I = \{1, \ldots, n\}$ for some $n \in \mathbb{N}$. A typical member of this product is shown below.

$$\prod_{i \in I} X_i = \prod_{i=1}^{n} X_i \ni (x_1, x_2, \ldots, x_n)$$

where the term on the right-hand side is an ordered n-tuple, whose ith element is a member of the set X_i for each $i \in \{1, 2, \ldots, n\}$. It is understood that $\prod X_i$ is precisely the set of all such terms. Notice that this n-tuple can be interpreted as a choice function from $\{1, 2, \ldots, n\}$ to the union of the X_i, selecting for each i an element of X_i.

This brings us to the definition of $\prod X_i$ when I is an infinite set. First we assume that I is countable, i.e. that I has the cardinality of the natural numbers. Then, without loss of generality, we may assume that $I = \mathbb{N}$, and extrapolating from the previous definition we suggest the form of a generic element of the product:

$$\prod_{i \in I} X_i = \prod_{i=1}^{\infty} X_i \ni (x_1, x_2, \ldots).$$

What are we to make of the entity on the right-hand side of this equation? It is simply an infinite sequence indexed by the natural numbers satisfying the condition that the ith term of the sequence is contained in the ith set, namely X_i — as in the previous paragraph we may interpret this as a function that for each $i \in \mathbb{N}$ selects an element from the corresponding set X_i. This idea helps us to propose a definition of the Cartesian product of any collection of sets X_i, even when the indexing set I may be uncountably infinite. Without this insight it would be impossible to attach any sort of meaning to the expression $\prod X_i$ in the general case.

Definition 2. Let I be a set and consider the collection of sets $\{X_i \mid i \in I\}$. The *Cartesian product* of the X_i, written $\prod_{i \in I} X_i$, is defined

$$\prod_{i \in I} X_i = \{f \colon I \to \bigcup_{i \in I} X_i \mid f(i) \in X_i \text{ for each } i \in I\}.$$

[11] We use the notation (x, y) for an ordered pair whose first element is x and whose second element is y; many authors use angled brackets, i.e. $\langle x, y\rangle$. Note that $(x, y) = \{\{x\}, \{x, y\}\}$. It is easy to check that $(x_1, y_1) = (x_2, y_2)$ if and only if $x_1 = x_2$ and $y_1 = y_2$.

[12] Formally the ordered triple (a, b, c) can be constructed as the ordered pair $((a, b), c)$.

The elements of the Cartesian product are precisely choice functions as defined in the previous section. Thus, given any collection of non-empty sets X_i, saying that the Cartesian product is non-empty amounts to saying that there exists a choice function for the collection. The converse is also clearly true. Hence we have shown our alternative formulation to be equivalent to our original statement of the axiom of choice.

3 Other results equivalent to the Axiom of Choice

The axiom of choice, obscure as it may seem, is actually equivalent to many different results. That is, there are results that are both implied by, and imply the axiom of choice. Two of the most famous are Zorn's Lemma and the Well-Ordering Principle. We state these and outline parts of the proof of their equivalence.

3.1 Zorn's Lemma

One result famously equivalent to the axiom of choice is Zorn's lemma (sometimes known as the Kuratowski–Zorn Lemma). It was discovered by the German mathematician Max Zorn[13] in 1935, independently of Kuratowski[14] who had proven it thirteen years earlier. There is at least one well-known (amongst mathematicians, at least) joke about it:

What's yellow and equivalent to the axiom of choice? — Zorn's Lemon.

Before stating Zorn's lemma we must first introduce some terms. There is considerable variation in the literature over the precise meaning of some of the following terminology, so we start with some definitions. Recall that a relation on a set X is just a subset of the Cartesian product $X \times X$. It is conventional to write xRy instead of $(x, y) \in R$ to indicate that the relation R holds between two elements x, y in the set.[15] An ordering is just a particular type of relation on a set; a function is another.

Definition 3. A *partial order* on a set X is a relation $\leqslant$ satisfying the following conditions:

- $\leqslant$ is *antisymmetric*, i.e. if $x \leqslant y$ and $y \leqslant x$ then $x = y$;
- $\leqslant$ is *transitive*, i.e. if $x \leqslant y$ and $y \leqslant z$ then $x \leqslant z$.

We call a pair $(X, \leqslant)$ of a set X and a partial order $\leqslant$ on X a *partially ordered set*. We will use the phrase "X is partially ordered (with respect to $\leqslant$)" to mean that $(X, \leqslant)$ is a partially ordered set.

Although the symbol $\leqslant$ has been used in our definition of a partial order, it should be noted that the familiar "is less than" on real numbers, also denoted by $\leqslant$, is only one example of a partial order; another possibility is the $\subseteq$ relation on a set of sets and there are many more. Sometimes an additional property of a partial order is stated: that $\leqslant$ is *reflexive*, i.e. $x \leqslant x$ for all $x \in X$, but this is unnecessary as it follows from transitivity and antisymmetry by letting $z = x$.

[13] Max Zorn (1906–1993).

[14] Kazimierz Kuratowski (1896–1980).

[15] It is important to be aware that a relation need not correspond to any obvious pattern. While some relations exhibit familiar properties such as transitivity or symmetry, a relation can essentially be an arbitrary subset of the Cartesian product.

Definition 4. A *total order*[16] on a set X is a partial order $\leqslant$ which satisfies the additional property that $\leqslant$ is *strongly connected*, i.e. either $x \leqslant y$ or $y \leqslant x$ for all $x, y \in X$. In other words, all elements in X are comparable.

The pair $(X, \leqslant)$ of a set X and a total order $\leqslant$ on X is called a *totally ordered set.* We will use the phrase "X is totally ordered (with respect to $\leqslant$)" to mean that $(X, \leqslant)$ is a totally ordered set.

These definitions are based on those of Suppes [**9**]. Another approach is to define a *strict* partial order $<$ on a set X as a relation that is asymmetric ($x < y$ implies $y \not< x$) and transitive. Then a total order is a strict partial order that is connected (either $x < y$ or $y < x$ for all $x, y \in X$, except perhaps when $x = y$). In this case, the so-called "law of trichotomy" then holds, that is, for all $x, y \in X$ exactly one of the following is true: $x < y$, $y < x$ or $x = y$. This corresponds to the statement in the above definition about comparability in totally ordered sets.

Definition 5. A *chain*[17] in a partially ordered set $(X, \leqslant)$ is a subset Y of X such that Y is totally ordered with respect to the restriction of $\leqslant$ to Y.

Definition 6. An *upper bound* for a subset Y of a partially ordered set $(X, \leqslant)$ is an element $x^* \in X$ such that $y \leqslant x^*$ for all $y \in Y$. Note that x^* is not required to be in Y. A *lower bound* is an element $x_* \in X$ (not necessarily in Y) such that $x_* \leqslant y$ for all $y \in Y$.

Definition 7. A *maximal element* in a partially ordered set $(X, \leqslant)$ is an element $x_{\max} \in X$ such that for all $x \in X$ if $x_{\max} \leqslant x$ then $x_{\max} = x$. A *minimal element* is defined similarly.

One might expect the definition of a maximal element to be something more like "$x_{\max}$ is a maximal element if for all $x \in X$ we have $x \leqslant x_{\max}$", but this is the definition of a different concept, that of a *greatest element.* (The corresponding definition of a *least element* is analagous.) A greatest element is always a maximal element and the converse holds if X is totally ordered, but not in general. For example, consider the set $X = \{x, y, z\}$ and the relation $\preccurlyeq$ on X defined by $x \preccurlyeq x$, $y \preccurlyeq y$, $z \preccurlyeq z$, $x \preccurlyeq z$. It is easy to see that $\preccurlyeq$ defines a partial order on X. Now z is certainly a maximal element, because no other element is greater than it with respect to $\preccurlyeq$, but it is not a greatest element, for we do not know that $y \preccurlyeq z$.

We now have all the terminology we need to state Zorn's lemma. Now that we have defined all the above terms it can be written very concisely.

Zorn's Lemma. *Let $(X, \leqslant)$ be a partially ordered set in which each chain has an upper bound. Then X has a maximal element.*

It was mentioned earlier that Zorn's lemma is in fact equivalent to the axiom of choice. Many of the proofs of these equivalences are complicated and require the use of advanced techniques based around extending the principle of induction to the "transfinite".[18]

Nevertheless, we outline the proof of the following theorem.

Theorem 1. *Zorn's lemma implies the axiom of choice.*

(The following outline is based on Machover [**4**], p. 86.) The idea of the proof is to consider the set of "partial choice functions" on a collection C of non-empty sets, i.e. functions which are

[16] A total order is often called a "linear order".

[17] Sometimes the term "chain" is used exclusively to refer to a totally ordered subset of the partially ordered set $(\mathcal{P}(X), \subseteq)$ for a set X. (The notation $\mathcal{P}(X)$ refers to the power set of X, the set of all subsets of X.) Historically this was what was used first, but this essay uses a generalisation of it.

[18] More information can be found in most books on axiomatic set theory, including Jech [**3**], Suppes [**9**] or Machover [**4**].

choice functions on some subset of C. For instance, for any set $X \in C$ and $x \in X$ the function f defined on C with $f(X) = x$ is a partial choice function. We consider the set F of partial choice functions as a partially ordered set with respect to the subset relation.

We then show that $(F, \subseteq)$ satisfies the conditions of Zorn's lemma (the details are omitted) and apply it, showing that $(F, \subseteq)$ has a maximal element g. We then argue that g must be a choice function on the whole of C, for if not then it is possible to extend this g (a partial choice function) by considering one of the sets for which it is not a choice function, contradicting the fact that g is maximal. Hence a choice function exists.

3.2 The Well-Ordering Principle

We first give a definition of what it means for a set to be well-ordered.

Definition 8. A *well-order*[19] on a set X is a total order $\leqslant$ such that every subset of X has a least element with respect to $\leqslant$, provided it is non-empty.

We call a pair $(X, \leqslant)$ of a set X and a well-order $\leqslant$ on X a *well-ordered set*. We will use the phrase "X is well-ordered (with respect to $\leqslant$)" to mean that $(X, \leqslant)$ is a well-ordered set.

This definition may seem so familiar from undergraduate analysis that it appears obvious, but there is a very important difference between this and the completeness property of the real numbers, which states that every non-empty subset of the reals that is bounded above has a *least upper bound*, or even the (equivalent) property that every non-empty set that is bounded below has a greatest lower bound. It does not take much work to show that the set $\{x \in \mathbb{R} \mid x > 0\}$ has 0 as its greatest lower bound; the same cannot be said of finding a least element in this set.

Which sets can be well-ordered? The natural numbers themselves are well-ordered with respect to the usual $\leqslant$ ("less than or equal to") ordering. (The negative whole numbers can be well-ordered too, this time by the $\geqslant$ ("greater than or equal to") relation — in this case a "least element" with respect to the well-order is actually a larger number in the usual sense.) Hence, it is clear that any finite set can be well-ordered: we define a (total) order on the set by assigning a natural number to each element and then in each non-empty subset pick the element with the smallest corresponding natural number as our least element with respect to this order.

The following lemma is useful for showing that certain familiar sets can be well-ordered.

Lemma 1. *If $(X, \leqslant)$ is a well-ordered set and there is a bijection $\phi\colon X \to Y$ then Y can be well-ordered.*

Proof. Define a total order[20] $\leqslant'$ on Y by the rule

$$\phi(x_1) \leqslant' \phi(x_2) \text{ if and only if } x_1 \leqslant x_2.$$

We now show that every non-empty subset of Y has a least element. Let $Z \subseteq Y$ be non-empty. Since there is a bijection between X and Y, every element of Z is of the form $z = \phi(x)$ for exactly one $x \in X$. Consider the set

$$\phi^{-1}(Z) = \{x \in X \mid \phi(x) \in Z\} \subseteq X,$$

which is non-empty following our assumption that $Z \neq \varnothing$. As X is well-ordered (with respect to $\leqslant$) $\phi^{-1}(Z)$ has a least element, x_0. Then the element $\phi(x_0)$ is a least element in Z, since the fact that $x \leqslant x_0$ for every $x \in \phi^{-1}(Z)$ implies that $z \leqslant' \phi(x_0)$ for every $z \in Z$. Hence $(Y, \leqslant')$ is a well-ordered set. □

[19] This ungrammatical terminology is a back-formation of "well-ordered set".

[20] That $\leqslant'$ satisfies the conditions of antisymmetry and transitivity required for a total order follow directly from the fact that $\leqslant$ is a total order.

Note that the order relation on Y is not the same as the order relation on X. For example, if $X = \mathbb{N} = \{1, 2, 3, \dots\}$ with the "less than or equal to" ordering then the ordering induced on $Y = \{-1, -2, -3, \dots\}$ by the bijection $\phi(n) = -n$ is not the "less than or equal to" ordering, but the "greater than or equal to" ordering.

We can use the conclusion of Lemma 1 to see that various other familiar sets can be well-ordered. For example, since the integers and the rational numbers are countably infinite, they too can be well-ordered although the actual orderings look a little bit different. Define $\phi\colon \mathbb{N}\cup\{0\} \to \mathbb{Z}$ by

$$\phi(n) = \begin{cases} n & \text{if } n \text{ is odd} \\ -n & \text{if } n \text{ is even} \end{cases}$$

then the induced ordering is

$$0 \leqslant 1 \leqslant -1 \leqslant 2 \leqslant -2 \leqslant 3 \leqslant \cdots .$$

Similarly, a possible well-ordering on the positive rational numbers could be

$$0 \leqslant 1 \leqslant \tfrac{1}{2} \leqslant 2 \leqslant 3 \leqslant \tfrac{1}{3} \leqslant \tfrac{1}{4} \leqslant \tfrac{2}{3} \leqslant \tfrac{3}{2} \leqslant 4 \leqslant 5 \leqslant \tfrac{1}{5} \leqslant \cdots ,$$

based on the well-known bijection between the naturals (excluding zero[21]) and the positive rationals as summarised in the following table:

1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\cdots$
2		$\frac{2}{3}$		$\frac{2}{5}$		$\cdots$
3	$\frac{3}{2}$		$\frac{3}{4}$	$\frac{3}{5}$		$\cdots$
4		$\frac{4}{3}$		$\frac{4}{5}$		$\cdots$
5	$\frac{5}{2}$	$\frac{5}{3}$	$\frac{5}{4}$		$\frac{5}{6}$	$\cdots$
6				$\frac{6}{5}$		$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

This is all very well and good (no pun intended) but so far we have only shown that countable sets can be well-ordered. What if X is uncountably infinite, like the set of real numbers? Is it possible to define an order on X such that every non-empty subset of X has a least element with respect to that order? The answer is by no means obvious, but it turns out that we can. The catch? We need the axiom of choice, of course!

Well-Ordering Principle. *Every set can be well-ordered.*

As stated previously, it turns out that the well-ordering principle is actually equivalent to the axiom of choice. As for Zorn's lemma, the proof that the axiom of choice implies the well-ordering principle is difficult and requires use of a process called "transfinite recursion" and the manipulation of ordinal numbers, which is beyond the scope of this essay.[22] However, the reverse implication is straightforward:

Theorem 2. *The well-ordering principle implies the axiom of choice*

Proof. Let C be a collection of non-empty sets. Well-order the union set $\bigcup C$. Define $f\colon C \to \bigcup C$ as follows: for each set $X \in C$, let $f(X)$ be the least element of X (which we can do since X is a subset of the union set of C). Then f is a choice function on C. $\square$

[21] Here I take $\mathbb{N}$ to be the set $\{1, 2, 3, \dots\}$ of positive integers. When zero is to be included I will write $\mathbb{N} \cup \{0\}$.
[22] For a proof see Roitman [**6**] or Machover [**4**].

4 Some results from the Axiom of Choice or its equivalents

There are many results — too many to list here — that follow from the axiom of choice, either directly or via one of its equivalents. They arise in all branches of mathematics from set theory to analysis, from algebra to measure theory to topology. We give three examples.

4.1 Every ring has a maximal ideal

First recall the definitions of a ring and an ideal.

Definition 9. A *ring* is set R together with two binary operations $+, \cdot \colon R \times R \to R$ (addition and multiplication) such that the following conditions hold[23]:

- R is an abelian group with respect to the operation $+$;
- For all $x, y, z \in R$ we have $x \cdot (y \cdot z) = (x \cdot y) \cdot z$;
- For all $x, y, z \in R$ we have $(x + y) \cdot z = x \cdot z + y \cdot z$ and $x \cdot (y + z) = x \cdot y + x \cdot z$;
- There is an element $1_R \in R$ such that for every $x \in R$ we have $x \cdot 1_R = 1_R \cdot x = x$.

There is some discrepancy regarding the definition of a ring; some authors prefer to refer to the structure defined above as a *ring with identity* to allow rings where the fourth condition above is not satisfied. In what follows a ring will be taken to have an identity element 1_R.

Definition 10. An *ideal* of a ring R is a set I such that:

- I is a subgroup of R with respect to the operation $+$;
- For every $x \in R$ we have $xI \subseteq I$;[24]
- For every $x \in R$ we have $Ix \subseteq I$.

Furthermore, we say that I is a *maximal ideal* of R if I is an ideal of R and whenever J is an ideal of R such that $I \subseteq J$ we have either $J = I$ or $J = R$.

The following theorem relies on Zorn's lemma.

Theorem 3. *Every ring has a maximal ideal.*

Proof. (Based on Cameron [**1**], p. 120.) Let R be a ring and let $\mathcal{I}$ denote the set of all proper ideals of R, i.e. ideals of R which are not equal to R itself. This is a partially ordered set with the inclusion relation $\subseteq$. We now verify the hypotheses of Zorn's lemma.

Let $\mathcal{C}$ be a chain in $\mathcal{I}$. There are two possibilities: either $\mathcal{C}$ is empty or it is not. In the first case the zero ring $\{0_R\}$ is a proper ideal of R and is trivially an upper bound for $\mathcal{C}$ (since there are no elements of $\mathcal{C}$ to cause a problem). So suppose that $\mathcal{C}$ is non-empty. We show that the set $I = \bigcup \mathcal{C}$ is an element of $\mathcal{I}$ and is an upper bound for $\mathcal{C}$.

- I is a subgroup of R with respect to addition — Take $x_1, x_2 \in I$. Then $x_1 \in I_1$ and $x_2 \in I_2$ for some $I_1, I_2 \in \mathcal{C}$. Since $\mathcal{C}$ is a chain, we can assume, without loss of generality, that $I_1 \subseteq I_2$. Hence, by virtue of I_2 being an ideal of R, we have $x_1 + x_2 \in I_2 \subseteq I$.
- For every $x \in R$ we have $xI, Ix \subseteq I$ — Let y be an element of some ideal $I' \in \mathcal{C}$. As I' is an ideal we have immediately that $xI', I'x \subseteq I' \subseteq I$.

[23]Formally the ring is the ordered triple $(R, +, \cdot)$, but I shall refer to "the ring R", as is common practice.
[24]The notation xI is used to denote the set $\{x \cdot y \mid y \in I\}$.

- I is a proper ideal — Any ideal containing the identity 1_R must be the whole ring.[25] Since all the ideals in $\mathcal{C}$ are proper, none contains 1_R and so neither does their union I, which is therefore proper.

Hence $(\mathcal{I}, \subseteq)$ is a partially ordered set in which each chain has an upper bound and applying Zorn's lemma we see that $\mathcal{I}$ contains a maximal element, which is therefore a maximal ideal of R. □

4.2 Non-measurable sets

We first introduce the concept of a measure.

Definition 11. The *Lebesgue measure*[26] of a subset of the real line is function μ which satisfies the following:

- If A is a bounded set of real numbers then $\mu(A)$ (if it exists) is a non-negative real number;
- If $A \subseteq B$ then $\mu(A) \leqslant \mu(B)$ (if the measures exist);
- The measure of an interval is its length, e.g. $\mu([a, b]) = b - a$;
- For pairwise disjoint sets X_n the measure is "countably additive", in the sense that

$$\mu\left(\bigcup_{n\in\mathbb{N}} X_n\right) = \sum_{n\in\mathbb{N}} \mu(X_n);$$

- The measure of a set (if it exists) is unaffected by translation, i.e. $\mu(A + b) = \mu(A)$ where, for a given real number b, $A + b$ is the set $\{a + b \mid a \in A\}$, which we call a *translate* of A.

It might seem like every subset of the real line should be measurable — and indeed nearly every set of real numbers encountered in undergraduate analysis is — but the following theorem proves that this is not the case. Crucially, this result relies on the axiom of choice.

Theorem 4. *There exist non-measurable bounded sets of real numbers.*

Proof. (Based on Cameron [**1**], p. 122.) Define a relation on the interval $[0, 1]$ by the rule that $x \sim y$ if $y - x \in \mathbb{Q}$. This is clearly an equivalence relation.[27] Denote by C the partition of $[0, 1]$ corresponding to this relation, i.e. the set of equivalence classes. Now, by the axiom of choice, it is possible to form a set S containing exactly one element from each of these mutually disjoint sets.

For a given $q \in [-1, 1] \cap \mathbb{Q}$ (a rational number between -1 and 1) denote by S_q the translate $S + q$, as defined above. These sets are pairwise disjoint. For given distinct rational numbers $q_1, q_2 \in [-1, 1]$, suppose that $(S + q_1) \cap (S + q_2) \neq \varnothing$; then there is an $x = s_1 + q_1 = s_2 + q_2$ for $s_1, s_2 \in S$ and this implies that $s_1 - s_2 = q_2 - q_1 \in \mathbb{Q}$ and hence that $s_1 \sim s_2$, contradicting the fact that S contains a unique representative from each equivalence class. Clearly each S_q is contained in the interval $[-1, 2]$. Also, each $x \in [0, 1]$ is equivalent to exactly one element $s \in S$ giving $x - s \in [-1, 1] \cap \mathbb{Q}$ and so $x \in S_{x-s}$. Thus,

$$[0, 1] \subseteq S^* = \bigcup_{q\in[-1,1]\cap\mathbb{Q}} S_q \subseteq [-1, 2].$$

---

[25] For then the second (or third) condition of Definition 10 implies that $x = x \cdot 1_R\ (= 1_R \cdot x) \in I$ for every $x \in R$.

[26] Henri Lebesgue (1875–1941).

[27] Recall that an equivalence relation is a relation that is reflexive, symmetric and transitive.

We will show that our original set S is non-measurable by way of a contradiction.

Suppose that S is measurable and let $\mu(S) = c$. By translation invariance, $\mu(S_q) = c$ for every $q \in [-1, 1] \cap \mathbb{Q}$. If $c = 0$ then, since S^* is a countable union of sets all with measure zero, we have $\mu(S^*) = 0$, which contradicts the fact that $[0, 1] \subseteq S^*$ (and hence that $\mu(S^*) \geqslant 1$). However, if $c > 0$ then $\mu(S^*)$ must be infinite, contradicting the fact that S^* is bounded.

Hence, we must conclude that S is non-measurable. Moreover, it is clear from the construction that S is by no means unique in this regard. □

Non-measurable sets are the basis of many unintuitive results in mathematics. One of the most famous is the somewhat notorious Banach–Tarski paradox, which states that it possible to slice up the unit sphere into a finite number of pieces (five!) and reassemble them, using only rigid motions, to form two unit spheres. Naturally, the pieces are non-measurable sets. A proof can be found in Wagon [**10**].

4.3 Every vector space has a basis

Another important result that requires the axiom of choice (either via Zorn's lemma or the well-ordering principle) is that every vector space has a basis. We will not prove this here as the proof requires the manipulation of ordinal numbers, a proper description of which is beyond the scope of this essay.

Theorem 5. *Every vector space has a basis.*

There is a proof in Cameron [**1**]. The basic idea is to well-order the vector space and construct a sequence of vectors via an algorithm. Then, by manipulating ordinal numbers, this sequence is shown to be linearly independent and to span the vector space; hence a basis exists.

5 Consistency and Independence of the Axiom of Choice

We have seen that the axiom of choice leads to some surprising results: that the reals can be well-ordered, that all vector spaces can be described as sets of linear combinations of some fixed basis vectors and, not least, the existence of non-measurable sets and the Banach–Tarski paradox! Because of its highly unintuitive nature and its bizarre consequences there has naturally been much concern over the validity of the axiom; does it really deserve its place in the axioms of set theory? Perhaps — one might suggest — it would be safer to assume the negation of the axiom of choice, i.e. that there exist sets of non-empty sets without a choice function, where it is impossible to simultaneously pick exactly one representative from each member set. Of course, those collections of sets where there is definite rule about which element to pick or those where the number of sets is finite, as mentioned in Section 2.1, will still admit choice functions, since the axiom of choice is not required to prove this; perhaps those sets of non-empty sets where no choice function is possible would be strange enough objects themselves for the lack of one to seem reasonable.

However, if we disallow the axiom of choice and accept its negation we find we run into more unseen problems. Many familiar intuitive results in mathematics become false. For example, without the axiom of choice it is impossible to prove that the union of countably many countable sets is countable, something which is often taken for granted.[28] Without this result it is possible that the real line, an uncountably infinite set, is a countable union of countable subsets of

[28]Strictly speaking, as mentioned in Section 2.1 it is possible to prove this with the slightly weaker axiom of countable choice.

real numbers; this is pretty hard to imagine! If our main reason for dismissing the axiom of choice is the unintuitiveness of some of its consequences then there is little grounds for accepting its negation either. While it may be unsatisfactory to some that the axiom of choice implies, completely non-constructively[29], that every vector space has a basis, for example, the existence of a vector space that can't be described in terms of a set of basis vectors is to many people (the author included) even worse!

So where does the axiom of choice stand in the world of mathematical logic and foundations? Do we accept it or its negation; is it even possible to pick one over the other? The answer, in short, is that it doesn't matter. There are models[30] of set theory where the axiom of choice is assumed and ones where its negation is assumed. This is because the axiom of choice is both consistent with and independent of the other axioms of set theory.

The first of these means that if the axiom of choice is assumed along with the other axioms of set theory then there are no contradictions, i.e. it is not possible to prove something and also its negation. This was proved by Kurt Gödel[31] in 1940. Later, in 1963, the American mathematician Paul Cohen[32] proved that the same is true of the negation of the axiom of choice. Together these results mean that the axiom of choice is logically independent of the rest of set theory and therefore it is impossible to prove that the axiom of choice is true or that it is false. For the proofs the interested reader should refer to Crossley [**2**].

Because of this logical independence, the axiom of choice is sometimes said to be "undecidable". Nowadays, however, most mathematicians choose to accept the axiom, despite its consequences. But whether or not you want to, the choice is yours.

References

1. Peter J. Cameron. *Sets, Logic and Categories.* Springer, 1999.

2. J. Crossley. *What is Mathematical Logic?* Dover, 1970.

3. Thomas Jech. *Set Theory.* Academic Press, 1978.

4. Moshé Machover. *Set Theory, Logic and their limitations.* CUP, 1996.

5. Joseph R. Mileti. The Axiom of Choice and Zorn's Lemma, 2006. URL `http://www.math.uchicago.edu/ mileti/museum/choice.html`.

6. Judith Roitman. *Introduction to Modern Set Theory.* John Wiley & Sons, 1990.

7. Thoralf A. Skolem. *Abstract Set Theory.* Notre Dame, 1962.

8. Ian Stewart and David Tall. *The Foundations of Mathematics.* OUP, 1977.

9. Patrick Suppes. *Axiomatic Set Theory.* D. Van Nostrand Company, 1960.

10. Stan Wagon. *The Banach–Tarski Paradox.* CUP, 1985.

[29] The non-constructive nature of the axiom of choice is another reason why it is so controversial. The axiom simply tells us that for any colleciton of non-empty sets a choice function must exist without any indication of how such a function is to be defined and there are schools of thought where this is not an acceptable assumption.

[30] The formal definition of a model can be found in Crossley [**2**]. Essentially, a model of a set of a set of axioms is a set where the axioms work, i.e. where they are consistent. For instance, the natural numbers model the Peano axioms. (For more information see Stewart and Tall [**8**].)

[31] Kurt Gödel (1906–1978).

[32] Paul Cohen (1934–2007).