Set theory

Christian Wüthrich Spring 2022

Remark. I claim no originality in the following content, and sometimes not even for the particular formulations used. There are only so many ways in which one can express mathematical definitions grammatically and with full rigour. In order not to break the flow of the presentation, however, I have not given precise credit where it is due, but only summarily in the references.

1 What is a set?

It is actually already a very hard question to state exactly how a set differs from the (mereological) fusion of its elements, and how it differs from other types of collections. We ignore these subtle, yet important, questions, at least for now and hope that the concept of a 'set' will emerge through the following work. So let's just dive in.

Definition 1 (Set)**.** *"A* set *is any collection into a whole of definite well-distinguished objects of our intuition or our thought." (Georg Cantor, 1895 in* Lexikon der Mathematik*)*

Definition 2 (Set, alternative definition)**.** *"A* set *is a many, which can be thought of as one, i.e., a totality of definite elements that can be combined into a whole by a law."*

The identity criterion for sets is rather straightforward and demands (and is usually stated, in a weaker form, as an axiom in axiomatic set theory, e.g., as an axiom of 'extensionality') that two sets are identical just in case they have exactly the same elements (or 'members').

1.1 Specifying a set

There are two ways to specify sets: either we do it *extensionally* and use curly brackets between which all its elements are listed, or we do it *intensionally* by giving some sort of constructive rule, which allows us to uniquely determine what its elements are, in which case we say that the rule or condition 'generates' the set.

We can thus extensionally define a set S by stating $S = \{\text{Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune}\}$. More abstractly, $S = \{a, b, c\}$ says that S is a set whose elements are a, b, and c.[1] Alternatively, we could specify the set intensionally by stating that S is the set of all planets in our solar system. Abstractly, if $F(x)$ is a determinate property or condition, we can say that the set S consists of all objects x which exemplify property $F(x)$ or satisfy condition $F(x)$. In our notation:

$$S = \{x : F(x)\}.$$

The planetary example can thus be written as $S = \{x : x \text{ is a planet in our solar system}\}$.

The *naive comprehension* or *existence axiom* states that every genuine condition or property $F(x)$ does in fact generate a set via above procedure. As we will see shortly, this axiom spells trouble.

1.2 Elementary relations

The *membership relation*, denoted by the infix $\in$, obtains between two sets or between an 'atom' and a set just in case one is an element of the other. Apart from identity, this is the only non-logical symbol which enters set theory primitively. In the previous case, we thus write, for example, $a \in S$. Since d is not an element of S, we write $d \notin S$ or $\neg(d \in S)$.

Another key relation is the *subset relation*, denoted by the infix $\subseteq$.

[1] **Important note:** I often distinguish between a set and its elements, as is common in introductory texts, by using uppercase Latin letters for the former and lowercase Latin letters for the latter. But this is unprincipled. I should, as advanced texts usually do, just use one or the other.

Definition 3 (Subset)**.** *A set S is a subset of another set T just in case every element of S is also an element of T. Symbolically, $S \subseteq T :\leftrightarrow \forall a(a \in S \rightarrow a \in T)$.*

Definition 4 (Proper subset)**.** *S is a* proper subset *of T, symbolically $S \subset T$, just in case that $S \subseteq T$ and $T \nsubseteq S$, i.e., every element of S is an element of T, but not every element of T is an element of S.*

In order to give an exemplary list of all subsets of a given set, we need another definition:

Definition 5 (Empty set)**.** *The* empty set *is the (unique) set which contains no elements. Symbolically, the empty set is denoted by $\emptyset$.*

Why should we think that there is such a thing as an empty set? Apart from its great practical utility in mathematics, it helps in formalizations of the intuitive difference between, as Raymond Smullyan (1992) says, an empty theatre and no theatre at all. Nota bene: no curly brackets are used for the empty set, except in the alternative notation of $\{\}$.

With this in place, we can now state the following example:

Example 1. *For a set $S = \{a, b, c\}$, the following sets are subsets: $\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}$, $\{a, b, c\}$ and $\emptyset$.*

It is important to note that for any set S, S and $\emptyset$ are always among its subsets.

Let us define another concept while we are at it (not a relation):

Definition 6 (Power set)**.** *The* power set $\mathfrak{P}(S)$ *of the set S is a set whose elements are exactly the subsets of S. Symbolically, $\mathfrak{P}(S) := \{T : T \subseteq S\}$.*

Example 2. *The power set of $S = \{a, b, c\}$ is $\mathfrak{P}(S) = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}, \emptyset\}$.*

For a finite set S (more on those below) of cardinality n, the cardinality of $\mathfrak{P}(S)$ is 2^n, hence the name.

Here are a few more important relations:

Definition 7 (Relative complement)**.** *The* relative complement *$S \backslash T$ of T in S, or the* set-theoretic difference *of S and T, is the set of all elements in S, but not in T. Symbolically, $S \backslash T := \{x : x \in S$ and $x \notin T\}$.*

Definition 8 (Intersection)**.** *The* intersection *of two sets S and T is the set consisting of those elements which are contained in both S and T. Symbolically, $S \cap T := \{x : x \in S$ and $x \in T\}$.*

Definition 9 (Union)**.** *The* union *of two sets S and T is the set consisting of those elements which are contained in either S or T. Symbolically, $S \cup T := \{x : x \in S$ or $x \in T\}$.*

It is straightforward to generalize the definitions of intersection and union to an arbitrary number of sets. In this case, the relation is symbolized not by an infix, but by a prefix, like so:

Definition 10. $\bigcap S := \{x : \forall T \in S(x \in T)\}$.

Definition 11. $\bigcup S := \{x : \exists T \in S(x \in T)\}$.

A bit more terminology.

Definition 12. *Two sets are* disjoint *just in case their intersection is the empty set. A set of sets is* pairwise disjoint *if every pair of sets is disjoint, i.e., if no object belongs to more than one of the sets.*

1.3 Some special and specially important sets

Arguably the most important one of those—the empty set—, we have already encountered. Another special class of sets is that of *singleton sets*, i.e. of sets containing just one element such as the singleton set containing Aristotle, {Aristotle}. Importantly, Aristotle is different from {Aristotle}. Some sets are important because they have additional algebraic structure which makes them very important in mathematics. Here are some famous examples of sets of numbers with such additional algebraic structure (from `http://en.wikipedia.org/wiki/Set_(mathematics)#Special_sets`):

- $\mathbb{P}$, the set of all *primes*: $\mathbb{P} = \{2, 3, 5, 7, 11, 13, 17, ...\}$.
- $\mathbb{N}$, the set of all *natural numbers*: $\mathbb{N} = \{0, 1, 2, 3, ...\}$.
- $\mathbb{Z}$, the set of all *integers*: $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$.
- $\mathbb{Q}$, the set of all *rational numbers*: $\mathbb{Q} = \{a/b : a, b \in \mathbb{Z}, b \neq 0\}$. Since all integers z can be expressed by the fraction $z/1$, it follows that $\mathbb{Z} \subset \mathbb{Q}$.
- $\mathbb{R}$, the set of all *real numbers*. Apart from all the rational numbers in $\mathbb{Q}$, $\mathbb{R}$ contains 'irrational' numbers such as π, e, and $\sqrt{2}$ (and others that cannot be defined).
- $\mathbb{C}$, the set of all *complex numbers*: $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$.
- $\mathbb{H}$, the set of all *quaternions*: $\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\}$.

It is an interesting fact that $\mathbb{P} \subset \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C} \subset \mathbb{H}$. Since these are *proper* subset relations, does this mean that the sets get larger and larger as we move from left to right? The answer to this question is not evident, since these sets are all 'infinite', in a sense yet to be defined. As it turns out, this is a rather deep question in the foundations of mathematics. Interestingly, it only turns on a seemingly simple activity that you have all learned a long time ago: *counting*.

2 Counting all the way to infinity

In order to answer the question in the previous paragraph, and in order to conceptualize infinity in a mathematically rigorous sense, we need to understand what it means to say that two sets S and T are of the 'same size', or are 'equinumerous', or have 'the same cardinality', or have the same 'power'. Usually, we mean that the elements of the two sets can be matched in a one-one fashion [see blackboard]. This means that there exists a map $f : S \to T$ which is one-to-one and onto, or 'bijective' as mathematicians say. This map (or 'function') f is a non-ambiguous rule which associates with each element $x \in S$ an element $f(x) \in T$.

Bijectivity or, equivalently, the relationship of one-to-one and onto conjoined is defined as follows [see also blackboard]:

Definition 13 (Bijective map)**.** *A map $f : S \to T$ is* bijective *just in case it satisfies the following two conditions:*

1. *For any elements $x_1, x_2 \in S$, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$ ('one-to-one').*
2. *For all $y \in T$, there exists an $x \in S$ such that $y = f(x)$ ('onto').*

If there exists a bijective map between two sets, we say that its elements can be brought into a *one-one correspondence*. Thus, two sets S and T are the *same size*, or have the *same cardinality*, or are *equinumerous* iff there exists a bijective map $f : S \to T$. By setting $T = \mathbb{N}$, this definition allows us to count the elements of a set and to say exactly how many elements a given set has.

Definition 14. *For a given $n \in \mathbb{N}$, a set has exactly n elements if its elements can be brought into one-one correspondence with the set of natural numbers from 1 to n.*

It immediately follows that the above defined set of all planets in our solar system has eight elements and is thus equinumerous to the set of chemical elements up to oxygen, the set of all days of Hanukkah, the set of apparitions appearing to Macbeth in Act 4, Scene 1 of *Macbeth*, and the set of women appearing in François Ozon's *8 femmes*.

2.1 Finite and infinite sets

From these definitions, one can then easily articulate a rigorous definition of when sets are 'finite' or 'infinite'.

Definition 15 (Finite set)**.** *A set S is* finite *iff there exists an $n \in \mathbb{N}$ such that S has n elements. Equivalently, S is finite just in case if it cannot be brought into one-one correspondence with a proper subset of itself.*

Definition 16 (Infinite set)**.** *A set S is* infinite *iff it is not finite, i.e., if there exists no $n \in \mathbb{N}$ such that S has n elements. Equivalently, S is infinite just in case it can be brought into a one-one correspondence with a proper subset of itself.*

Exercise 1. *Suppose that S is an infinite set. Prove that by removing one single element, the set remains infinite. Hint: use the first part of the definitions of 'finite' and 'infinite'.*

Exercise 2. *Show that the set $\mathbb{N}$ is infinite by stating a proper subset $S \subset \mathbb{N}$ and a bijective map $f : \mathbb{N} \to S$.*

2.2 Grasping the infinite: Hilbert's Hotel

Infinity is a strange animal. In order to appreciate this, let us engage in a little thought experiment called 'Hilbert's Hotel', after David Hilbert, which beautifully illustrates this strangeness. It will also serve as a propaedeutic for Georg Cantor's diagonalization proof. What follows in this section and the next is closely based on Smullyan (1992, Ch. 18).

Suppose you have a hotel of one hundred rooms. There is exactly one guest in each of the rooms. Late at night, a traveller arrives and requests a room for the night. Can she be accommodated? The problem is that neither the traveller nor any of the guests are prepared to share rooms, which renders her accommodation impossible since one cannot put 101 people into a one-one correspondence with 100 rooms.

But this is very different in *Hilbert's Hotel*, which has an infinite number of rooms. If it is equally fully booked by an infinite number of lonely, non-sharing guests, the manager can accommodate the late arrival, as long as everyone is prepared to move rooms.

Exercise 3. *How can this be achieved?*

But just as everyone went back to sleep after having accommodated the traveller, a bus full of new guests arrives, all seeking a bed for the night. The problem is that it's a very large bus; in fact, the bus contains infinitely many new guests! The surprising fact is that all of them can be accommodated as long every original guest is prepared to move to a new room once.

Exercise 4. *How?*

Hilbert's Hotel, as it turns out, it just one hotel of *Hilbert's Chain* of infinitely many hotels with infinitely many rooms each. Suppose that all hotels of Hilbert's Chain are fully booked with exactly one guest in each room of each hotel. One day, the management decides to remodel all but one of the chain's hotels. In order to loose no business, they hatch a plan of how to accommodate all the guests in all the hotels which close down in the one hotel which will remain open during the remodelling of the others.

Exercise 5. *Is this possible? If so, how? If not, why not?*

Exercise 6. *Which of the two infinities—the Chain's total number of rooms prior to and during the remodelling—, if any, is larger?*

These examples show some of the strange properties of infinite sets. For instance, Hilbert's Hotel and Chain show that (and how) there exist bijective maps between infinite sets and some of their proper subsets. I have stated this above as part of the definition of what it is for a set to be infinite. In fact, however, this can be shown as a *theorem* given just the first part of the definitions of finite and infinite sets. As strange as this seems, it is not really paradoxical. And it has certainly been known, at least for special cases, for quite a while. E.g., Galileo has pointed out that the natural numbers and their squares can be brought into one-one correspondence by the bijective map $f(n) = n^2$.

Despite his advocacy of *actual* infinities—as opposed to the Aristotelian doctrine that there only exist *potential* infinities—, particularly as concerning the number of individuals in our world, Gottfried Leibniz did not countenance actual infinities of *numbers*, because, as he claimed, it is inconsistent with the axiom that the whole is greater than any proper subpart. But cases such as Hilbert's Hotel show that this is not true at least in a very general class of theories. This can also be seen from another simple thought experiment [see blackboard].

Thus, even though there is a sense in which Hilbert's Chain contains more rooms than the original Hilbert's Hotel—in the sense that it contains all the rooms of the latter and some more—the number of its rooms is not *numerically* larger than the number of rooms in the original hotel. Above, we stated that two sets are equinumerous, or are numerically the same, if one can find a one-one correspondence between them. This is still the case for infinite sets, but we have to tread carefully here. It is not the case for infinite sets that a set S is 'numerically smaller than' T (or S has 'fewer elements than' T) if one can bring S into a one-one correspondence with a proper subset of T. The problem with this characterization is that it can be the case that S stands in such a correspondence with a proper subset of T *and* T stands in such a correspondence with a proper subset of S.

In order to see this, consider the set O ($\subset \mathbb{N}$) of odd numbers and the set E ($\subset \mathbb{N}$) of even numbers. It is easy to find a bijective map between O and E, e.g. (where $n \in \mathbb{N}$):

$$\begin{array}{ccccccc} 1, & 3, & 5, & 7, & 9, & \ldots & 2n+1\ldots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & & \updownarrow \\ 0, & 2, & 4, & 6, & 8, & \ldots & \ldots 2n \ldots \end{array}$$

This means that O and E are equinumerous. However, it is also rather straightforward to find a bijection between O and a proper subset of E, as follows:

$$\begin{array}{ccccccc} 1, & 3, & 5, & 7, & 9, & \ldots & 2n+1\ldots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & & \updownarrow \\ 2, & 4, & 6, & 8, & 10, & \ldots & \ldots 2n+2\ldots \end{array}$$

At the same time, however, we can do the same for E and a proper subset of O:

$$\begin{array}{ccccccc} 0, & 2, & 4, & 6, & 8, & \ldots & 2n\ldots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & & \updownarrow \\ 3, & 5, & 7, & 9, & 11, & \ldots & \ldots 2n+3\ldots \end{array}$$

So this would suggest then that on this definition, O would be numerically smaller, equal, and larger than E! This shows that the definition must be refined for infinite sets:

Definition 17. *A set S is* smaller *than another set T, or T is* larger *than S iff the following two conditions are met:*

1. *There exists a bijective map between S and a proper subset of T;*
2. *T cannot be brought into a one-one correspondence with a proper subset of S.*

This definition works also for infinite sets and reduces to the intuitive characterization given above for finite sets, since condition (2) is automatically satisfied for finite sets if condition (1) is.

In order to formulate a theory of the infinite, one needs to answer the following question: are any two arbitrary infinite sets necessarily equinumerous, or can they have different sizes? Georg Cantor (1845-1918), who made seminal contributions to set theory, answered this question in his *theory of transfinite numbers*. This theory met with fierce resistance even from some of the greatest minds of the time, such as Leopold Kronecker and Henri Poincaré, and later from Hermann Weyl, L E J Brouwer, and Ludwig Wittgenstein. Before I give you the answer, or you read on, pause and do the following

Exercise 7. *What do you think, are all infinite sets the same size or different?*

2.3 Cantor's heaven

I obviously don't know what your answer to Exercise 7 was, but Cantor first suspected that any two infinite sets would be of the same size. He spent twelve years trying to prove this when he found a counterexample. This means that there exist different kinds of infinities. In fact, as we will see, there exist infinitely many!

Finite sets are all 'countable' in the sense that we can count how many elements they contain. But this concept can be extended so as to include some infinite sets as 'countable':

Definition 18 (Countable set)**.** *A set is called* countable *just in case it has the same cardinality as a subset of $\mathbb{N}$. Equivalently, a set S is countable iff there exists an injective map from S to $\mathbb{N}$.*

Exercise 8. *Convince yourself that there exist infinite sets which are countable. By the way, such as set is called* countably infinite *or* denumerable*—or simply* countable *of course. An infinite set, which is not countable, is called* nondenumerable*.*

It is clear from Definition 18 that $\mathbb{N}$ is a countably infinite set. Thus, Cantor's question becomes whether every infinite set is countable. Cantor's strategy was to study infinite sets which seemed too large to be denumerable; yet, he found a trick to count them.

Let's start with another thought experiment. Suppose I write a natural number on a piece of paper and your task is to guess it. It is straightforward to devise a strategy which will succeed in doing this in a finite number of steps. Let's make your task a little bit harder: this time I write an integer number on a piece of paper.

Exercise 9. *Can you formulate a strategy which guarantees your success in a finite number of steps? This is essentially the same task as solving Exercise 4 on Hilbert's Hotel.*

Because this is also possible, it means that $\mathbb{N}$ and $\mathbb{Z}$ have the same cardinality (which makes $\mathbb{Z}$ denumerable), despite initial appearances that $\mathbb{Z}$ might be twice as large as $\mathbb{N}$ (modulo 0).

Now the next task is clearly more difficult. Suppose I write *two* natural numbers on a piece of paper and your task is to guess both of them at the same step. Only if your answer is the *pair* of numbers identical to the one I wrote will the task be solved.

Exercise 10. *Do you think there is a strategy which guarantees success in a finite number of steps? In other words, the question is whether pairs of natural numbers are denumerable, just as natural or integer numbers are. If so, provide the strategy. If not, why not?*

Now let's make it a bit harder yet. Suppose that now you don't only have to guess the two numbers, but also the order in which they are written down.

Exercise 11. *Devise a strategy to solve this in a finite number of steps.*

As you get a knack on devising counting strategies, you should be able to answer the following question:

Exercise 12. *Is the set $\mathbb{Q}^+$ of positive rational numbers denumerable?*

The answer to this question is part of what incited the strong reactions against Cantor's theory of transfinite numbers mentioned above.

Now to another, yet harder, problem. Suppose I write down a finite set of natural numbers. You do not know either *how many* numbers I wrote down or which is the largest number among those on the paper. Do you think there is a strategy to solve this problem in finitely many steps?

Exercise 13. *There is, as every* finite *set of natural numbers is denumerable. But how can we denumerate the set of finite sets of natural numbers?*

This leaves open the question regarding the set of all sets of natural numbers—finite *and infinite*. It was Cantor's great discovery to show that *this* set was nondenumerable. Before we prove this, let us consider the cardinality of the real numbers $\mathbb{R}$ and—as a warm-up exercise—prove that $\mathbb{R}$ is nondenumerable.

Exercise 14. *Prove that $\mathbb{R}$ is nondenumerable. If you haven't seen this before, you will probably not find the answer. But don't despair—it took Cantor a long time to realize this himself and we will go over this carefully in class! Proving this will introduce you to Cantor's ingenious 'diagonalization technique'.*

Let's return to the main item of business, to prove Cantor's Theorem. First, consider the following fact:

Fact 1. *The set of all sets of natural numbers is nondenumerable. In other words, the power set $\mathfrak{P}(\mathbb{N})$ of $\mathbb{N}$ is nondenumerable.*

Smullyan sketches the proof as follows. Suppose you have a book with countably infinitely many pages, numbered consecutively. In order to show that the above set is not denumerable it suffices to show that there exists a set of natural numbers which cannot be listed on any page of this book.

Exercise 15. *Your task now is to describe this set which cannot be given on any page of the book.*

Please note that this essentially makes use of the diagonalization technique. [blackboard]

Exercise 16. *The set of* finite *sets of natural numbers, as we have seen, is denumerable. What about the set of infinite sets of natural numbers—is it denumerable?*

We have thus shown that there is no one-one correspondence between $\mathfrak{P}(\mathbb{N})$ and $\mathbb{N}$ (or, *a fortiori*, any subset of $\mathbb{N}$). But can we conclude that $\mathfrak{P}(\mathbb{N})$ is thus of larger cardinality than $\mathbb{N}$? According to Definition 17, we also need to show that there exists a bijective map between $\mathbb{N}$ and a proper subset of $\mathfrak{P}(\mathbb{N})$.

Exercise 17. *Show this. Hint: we have actually already found such a map.*

The cardinality $\mathfrak{P}(\mathbb{N})$ is thus larger than that of $\mathbb{N}$. This means that the infinity of the number of sets of natural numbers is larger than the infinity of the natural numbers themselves. In other words, there are infinities of different 'sizes', or different *cardinalities*. This immediately leads to the question of whether there are sets even larger than $\mathfrak{P}(\mathbb{N})$. As Cantor famously found, the answer to this question is affirmative. That $\mathfrak{P}(\mathbb{N})$ is larger than $\mathbb{N}$ is only a special case of his famous theorem:

Fact 2 (Cantor's Theorem)**.** *For every set A, the set $\mathfrak{P}(A)$ of all subsets of A is larger than A.*

The proof of the theorem is essentially similar to our answer to Exercise 15. The proof idea is illustrated by Smullyan as follows. Imagine a universe in which each set of inhabitants forms a club. The inhabitants decide that each club gets named after an inhabitant of the universe such that no two clubs bear the name of the same inhabitant and such that each inhabitant lends its name to exactly one club. To this end, it is not necessary that the inhabitant is a member of the club which is named after him. It is clear that this project fails in a universe with a finite number of inhabitants, for n inhabitants, the number of clubs is 2^n. Fortunately, it turns out that this universe has infinitely many inhabitants, which is why no one sees a reason why the plans couldn't be realized. However, every scheme that is attempted fails.

Exercise 18. *Why is it impossible to find such a naming scheme, and how does this relate to Cantor's Theorem?*

[Proof of Cantor's Theorem on the blackboard]. As a corollary, this immediately entails that there are sets larger than $\mathfrak{P}(\mathbb{N})$, viz. $\mathfrak{P}(\mathfrak{P}(\mathbb{N}))$. In fact, it entails that there are infinities of infinitely many differing sizes, as this construction can always be repeated. In other words, there are larger sets for *every* set, and hence there are infinitely many sizes or cardinalities of infinite sets. This hierarchy of infinite cardinal numbers is often referred to as 'Cantor's heaven'.

3 Paradoxes and the foundations of set theory

Exercise 19. *A warmup exercise for the set-theoretic paradoxes: the quarter and the penny (from Smullyan, p. 234ff).*

3.1 The paradox of the universal set and Russell's paradox

Continuing to follow Smullyan's presentation, let us see how all of this concerns the foundations of set theory. In 1897, Cantor discovered a paradox, now often called the 'paradox of the universal set'. Suppose U is the set of *all* sets (the *universal* set), i.e. $U := \{x : x \text{ is a set}\}$. According to Cantor's Theorem (Fact 2), there is always a larger set, viz. $\mathfrak{P}(U)$. But could it be that there is a set larger than the set of all sets? Surely, $\mathfrak{P}(U)$ must be a subset of U, since U contains *all* sets. How can a subset of a set be larger than the set itself? It can't, and that's the paradox.

When Bertrand Russell studied Cantor's proof, searching for errors, in 1901, he found what is now called *Russell's paradox* and is similar to Cantor's (and was also found by Zermelo in 1900). Informally, the idea behind it is as follows. Call those sets which are not elements of themselves *normal*. The set of all books surely is normal, since it is itself not a book. A set is *abnormal* if it does contain itself. An example would be the set of all things which are not books. Clearly, the set of all non-books is a not a book and hence an element of itself.

Exercise 20. *Now consider the set R of all normal sets. Is R normal or not? Show that it cannot be determined whether R is normal or abnormal.* Hint: *Make sure to derive a contraction for all possibilities.*

This is Russell's paradox. A more formal version is obtained if we remind ourselves that the following is an axiom of naive set theory as formulated by Cantor[2]

Axiom 1 (Naive comprehension)**.** $\exists y \forall x (x \in y \Leftrightarrow F(x))$ *for any predicate F in a free variable x.*

In other words, for any predicate there exists a set which contains all and only those objects which exemplify the property denoted by the predicate. Consider the property of not containing itself.

[2] I say 'remind' because we encountered this above, stated informally. I mentioned that it will get us in trouble; Russell's paradox is that trouble.

According to the naive comprehension (or existence) Axiom 1, there exists a set containing all and only objects not containing themselves. If we thus substitute $x \notin x$ for $F(x)$ and apply existential and universal instantiation, we get

$$z \in z \Leftrightarrow z \notin z. \tag{1}$$

This is a contradiction and we have Russell's paradox again. Russell's paradox is a simplification of Cantor's paradox in that it does not rely on the concept of cardinality.

In 1919, Russell proposed a simple analogue of this paradox in the form of a barber who lives in a small village, allegedly in Sicily. He says of himself: "I shave all and only the men in my village who do not shave themselves." Thus, the barber doesn't shave any village man who shaves himself and every man in the village who does not shave himself is shaved by the barber. Does he shave himself? If he does, he shaves somebody (namely himself) who shaves himself, in violation of the rule that he shaves nobody who shaves himself. If he doesn't, then he is a village man who doesn't shave himself and is thus shaved by him, in accordance to the rule that he shaves all men in the village who don't shave themselves. So we have a contradiction again.

Exercise 21. *So, does he shave himself or not? How would you resolve the paradox?*

Similarly to the resolution of the barber paradox, one can see that the assumption of Axiom 1 leads to inconsistency, as evidenced by Russell's paradox, and must thus be discarded. Axiom 1 also leads to Cantor's version of the paradox if we assume that the relevant property is that of being a set. If we do, the set of all sets also exists. On the one hand, this set can be arbitrarily large; on the other hand, by Cantor's Theorem, for each set there exists a larger set. But this is the inconsistency again. The error consists in assuming that there is such a set in the first place.

Any solution to these paradoxes must block trouble-making sets such as U and R, while at the same time allow for sufficiently rich mathematics, e.g., so as to still allow Cantor's heaven. The rejection of Axiom 1, however, challenged the very foundations of mathematics at the time and nourished the suspicion that mathematics may not be reconcilable with logic.

3.2 Consequences for the foundations of mathematics

At the time, the most worked-out framework for mathematics was the one by Gottlob Frege. Frege's system was designed to realize his vision of deriving all of mathematics from logic and set theory. We will have a closer look at the philosophy of his 'logicism' later in the term. For now, it suffices to note that apart from a few axioms of logic, he used only one axiom concerning set, viz. essentially the '(naive) comprehension axiom', Axiom 1. Frege managed to derive all sets required to do mathematics at the time from this single assumption (plus the logic). The first order of business was to get the empty set. One can obtain that by stating a property, which is not exemplified, such as 'being dissimilar from oneself' or 'being non-identical with itself'. Since this gives us an admissible predicate, we know by Axiom 1 that there exists the set of all objects with the property it denotes. But since there are no such objects, the set is empty. Next, take two arbitrary objects x and y. There is the property of 'being either identical to x or to y', which gives us, again via Axiom 1, the set $\{x, y\}$ with exactly the elements x and y and none else. This is still the case if x and y happen to be identical. In this case, we simply have the singleton set $\{x\}$.

Thus we have the empty set $\emptyset$—and thus an object—and a constructive rule to produce more sets. To get the construction going, build the set of exactly those objects which are identical to $\emptyset$. There is only one such object, and we thus have the singleton set $\{\emptyset\}$, consisting only of the empty set. Note that $\emptyset$ and $\{\emptyset\}$ are different sets—the latter is a set of one element, the former has no elements. We can reapply that same step to obtain $\{\{\emptyset\}\}$, the set whose only element is the set whose only element is the empty set. We continue doing this to get $\{\{\{\emptyset\}\}\}, \{\{\{\{\emptyset\}\}\}\}, \ldots$. We thus obtain an infinitude of sets which can serve as natural numbers. In fact, this is how Ernst Zermelo introduced the natural numbers: he named the empty set '0', $\{\emptyset\}$ '1', $\{1\}$—which is another name for $\{\{\emptyset\}\}$—'2', etc. This procedure gives us all the natural numbers, which according to Axiom 1 also form a set—the set of natural numbers $\mathbb{N}$.

Given a set S, one can further say that there is the property of 'being a subset of S'. According to Axiom 1, there consequently exists a set of all subsets of S, i.e., what we called the *power set* $\mathfrak{P}(S)$ of S. Consider the property of 'being an element of at least one element of S'. Hence there exists a set of all elements of all elements of S, called the *set union* $\mathfrak{U}(S)$ of S. For instance, if S is the set of all clubs, then $\mathfrak{U}(S)$ is the set of all club members.

So far so good, as it seems as if we can get all the sets which matter in mathematics in this manner. But there is one problem with all of this: the theory is inconsistent! Sanctioned by Axiom 1, we can assume the existence of the set of all normal sets, which leads us directly to the contradiction known as Russell's paradox. Similarly for the set of all sets, which leads to Cantor's version of the contradiction. Despite this inconsistency, Frege's work was soon recognized (e.g. by Russell) to contain the seeds of a revolution in the foundations of logic and mathematics. One of the reasons for this was because the inconsistency could actually be eradicated. One of the ways to do that (but not the only one) was to ditch Axiom 1 and to replace it with what became known as the *Zermelo-Fraenkel axiomatization* of set theory. Before we delve into that, let's briefly look at Russell's own solution.

3.3 Russell's theory of types

The general idea of Russell's *theory of types* was the recognition that sets are something fundamentally different from elements of sets. Russell constructed a hierarchy of *types*:

Level 0. individuals $a, b, c, ...$ (don't have set-theoretic elements)

Level 1. sets whose elements live on level 0: $\{a\}, \{a, b\}, ...$

Level 2. sets whose elements live on level 0 or 1: $\{a\}, \{a, \{b, c\}\}, ...$

Level 3. etc.

It is important to insist that only those sets exist, which exist at a level. How does this solve the paradoxes?

(a) At what level is the universal set U? $U \in U \Rightarrow$ cannot exist on any level and thus doesn't exist.

(b) At which level do we find $R := \{x : x \text{ is a set } \wedge\ x \notin x\}$? R cannot be an element of itself and thus doesn't exist either.

While this perfectly solves the paradoxes, the overall system proposed by Russell and Alfred North Whitehead in their monumental *Principia Mathematica* (1910, 1912, 1913) was generally too complicated for many to adopt it, even though many notational innovations remain with us today. And there was another problem: the theory of types required an axiom of infinity, postulating infinitely many individuals at level 0, for otherwise each level consists only of finitely many objects.

Exercise 22. *Why is this a problem for Russell's logicistic program? We may have to get back to this one when we look at logicism.*

4 Zermelo-Fraenkel set theory

Before I write down all the axioms in their full glory, let's approach the axiomatization due to Ernst Zermelo (1871-1953), with later improvements by Adolf (later: Abraham) Fraenkel (1891-1965), more informally, again following Smullyan. The Zermelo-Fraenkel axiomatization is now the most widely used axiomatic theory of sets.

4.1 Zermelo set theory

Zermelo started out by replacing the axiom of '*unrestricted* comprehension'—our Axiom 1—by what is called the *axiom schema of specification* ('Aussonderungsaxiom') or of *restricted comprehension*, because it does not permit every class as set.[3] As compared to Russell's theory of types, which achieves the elimination of the contradictions by restricting the syntax of admissible predicates quite significantly, Zermelo's path didn't restrict the syntax, but instead restricted 'comprehension' such that it was no longer the case that all classes qualified as sets. Zermelo showed a quarter century later that his axiom (scheme) of restricted comprehension was a theorem of set theory as improved by Fraenkel by the introduction of the axiom schema of replacement ('Ersetzungsaxiom') and was thus obsolete as an independent assumption. But we are getting ahead of ourselves.

Informally stated, Zermelo's axiom schema of specification is as follows:

Axiom 2 (Axiom schema of specification)**.** *Given an arbitrary property as well as an arbitrary set S, then the set of all elements of the set S which exemplify the property exists.*

Because of the need for a prior set S, we can no longer speak of a set of *all* x with a certain property. Instead, we can speak of all $x \in S$ with this property. No contradiction can be derived from assuming Axiom 2 as far as we know. And it suffices for the purposes of mathematical practice: when mathematicians speak of the 'set of all numbers', or the 'set of all points in a place', etc, they talk about objects and sets whose existence was ordained before.

It should also be clear that this dissolves Russell's paradox, as the set of *all* normal sets cannot be constructed anymore on the basis of Axiom 2. However, for a given set S, we can create the set T of all normal elements of the set S (i.e., the set T of all those sets in S which do not contain themselves). This does not lead to a paradox, since T is not—cannot be—an *element* of S, even though it is a *subset* of it.

To say that a set T is an *element* of a set S means that S is a collection of objects, one of which is T. But to say that T is a *subset* of S means that all elements of T are also elements of S, but this does not entail that T itself is one of these elements. Suppose that S is the set of all humans on Earth and T is the set of human Earthlings who are left-handed. Surely, all left-handed humans are also humans. Hence $T \subset S$. But the *set* of all left-handed humans is not itself a human (even though all its elements are), and hence $T \notin S$.

Exercise 23. *Why is the set B of all normal elements of A not itself an element of A, even though it is certainly a subset of A?*

Exercise 24. *Show that the universal set U cannot exist under the assumption of Axiom 2. This resolves Cantor's paradox of the universal set.*

As a result of giving up Axiom 1, Zermelo could no longer construct the sets needed for mathematics as we did in §3.2 and had to introduce specific axioms to guarantee the existence of $\emptyset, \{x, y\}, \mathfrak{P}(S)$, and $\mathfrak{U}(S)$. Similarly, the existence of the natural numbers $\mathbb{N}$ had to be underwritten by a separate axiom, the *Axiom of Infinity*. Zermelo (1908) postulates the following axioms for set theory:[4]

Axiom 3 (Extensionality (Axiom der Bestimmtheit))**.** *"If every element of a set M is also an element of N and vice versa... then $M \equiv N$. Briefly, every set is determined by its elements."*

Axiom 4 (Elementary sets (Axiom der Elementarmengen))**.** *"There exists a set, the null set, $\emptyset$, that contains no element at all. If a is any object of the domain, there exists a set $\{a\}$ containing a and only a as element. If a and b are any two objects of the domain, there always exists a set $\{a, b\}$ containing as elements a and b but no object x distinct from them both."*

[3] This axiom, or axiom *schema*, actually has a few more names in the literature.

[4] Except for a few typographical corrections, all English translations are from Wikipedia (`http://en.wikipedia.org/wiki/Zermelo_set_theory`). This translation is not very faithful to the original, but commits no gross mistake regarding content. The typographical changes have been made both to improve textual accuracy and continuity with the present text.

Axiom 5 (Separation (Axiom der Aussonderung)). *"Whenever the propositional function $F(x)$ is definite for all elements of a set M, M possesses a subset M_F containing as elements precisely those elements x of M for which $F(x)$ is true."*

This axiom is essentially Axiom 2 above. It is ultimately responsible for the elimination of the paradoxes as it suffices to prove the following important theorem:

Fact 3. *Every set M possesses at least one subset M_0 that is not an element of M.*

Proof (as given by Zermelo 1908, 265). "For each element x of M, it is determinate whether $x \in x$ or not; this possibility $x \in x$ is in itself not excluded by our axioms. If M_0 is the subset of M which contains, by [Axiom 5], all those elements of M for which $x \notin x$, then M_0 cannot be an element of M. For either $M_0 \in M_0$ or not. In the first case M_0 contained an element $x = M_0$, for which $x \in x$, against the definition of M_0. Hence, it is certainly not the case that $M_0 \in M_0$, and if M_0 were an element of M, it would have to also be an element of M_0, which was just excluded." This means that the reductio assumption $M_0 \in M$ is false, proving the theorem.

Zermelo concludes from Fact 3 that not all objects x in the universal domain $\mathfrak{B}$ can be elements of one and the same set, which means, in other words, that the domain $\mathfrak{B}$ itself is not a set. This precludes the universal set and "disposes the 'Russellian antinomy' as far as we are concerned." (ibid.)

Axiom 6 (Power set (Axiom der Potenzmenge)). *"To every set T there corresponds a set $\mathfrak{P}(T)$, the power set of T, that contains as elements precisely all subsets of T."*

Axiom 7 (Union (Axiom der Vereinigung)). *"To every set T there corresponds a set $\mathfrak{U}(T)$, the union of T, that contains as elements precisely all elements of the elements of T."*

Axiom 8 (Choice (Axiom der Auswahl)). *"If T is a set whose elements all are sets that are different from $\emptyset$ and mutually disjoint, its union $\mathfrak{U}(T)$ includes at least one subset S_1 having one and only one element in common with each element of T."*

This is the famous and controversial Axiom of Choice, which Zermelo considers an "unobjectionable logical principle". Zermelo states that one can also express this axiom by saying that it is always possible to choose a particular element $m, n, r, ...$ from each element $M, N, R, ...$ of T and to collect them to a set S_1.

Axiom 9 (Infinity (Axiom des Unendlichen)). *"There exists in the domain at least one set Z that contains the null set as an element and is so constituted that to each of its elements a there corresponds a further element of the form $\{a\}$, in other words, that with each of its elements a it also contains the corresponding set $\{a\}$ as element."*

Axiom 9 establishes the existence of infinitely many sets. Axioms 3-9 constitute the axiomatic basis of *Zermelo set theory*, or the *theory* ZC.

4.2 The well-ordering theorem and the Banach-Tarski paradox

Assuming the Axiom of Choice (Axiom 8), Zermelo was able to prove the so-called *well-ordering theorem* (also called 'Zermelo's theorem'), which asserts that every set can be well-ordered.

Definition 19. *A set S is* well-ordered *by a strict total order just in case every non-empty subset of S has a least element under the ordering.*

Orders (total, strict total, partial, etc) are binary relations defined on some set, the 'domain'.

Definition 20 (Total order). *A set S is* totally ordered *under a binary relation, here denoted by the infix $\leq$, just in case for all $a, b, c \in S$, the following three conditions hold:*

1. *If $a \leq b$ and $b \leq a$, then $a = b$ (antisymmetry);*

2. *If* $a \leq b$ *and* $b \leq c$*, then* $a \leq c$ *(transitivity);*

3. $a \leq b$ *or* $b \leq a$ *(totality).*

Exercise 25. *Show that* totality *in Definition 20 entails* reflexivity*, i.e.,* $\forall a \in S, a \leq a$.

Definition 21 (Partial order)**.** *A set* S *is* partially ordered *under a binary relation, denoted by the infix* $\leq$*, just in case for all* $a, b, c \in S$*, the following three conditions hold:*

1. *If* $a \leq b$ *and* $b \leq a$*, then* $a = b$ *(antisymmetry);*

2. *If* $a \leq b$ *and* $b \leq c$*, then* $a \leq c$ *(transitivity);*

3. $a \leq a$ *(reflexivity).*

Note that because of the result established in Exercise 25 any totally ordered set is also partially ordered, but not vice versa. We can now define the remaining two undefined concepts used in Definition 19:

Definition 22. *The* least element *of a subset* S *of a partially ordered set* T *is an element of* S *which is smaller than or equal to any other element of* S*, where 'smaller than' and 'equal to' is defined in terms of the partial order of* T*. The* greatest element *is defined dually.*

Definition 23 (Strict total order)**.** *For every total order* $\leq$*, there exists an associated asymmetric relation, denoted by the infix* $<$ *and called a* strict total order*, which can be defined in two equivalent ways:*

- *Either* $a < b$ *iff* $a \leq b$ *and* $a \neq b$*;*
- *or* $a < b$ *iff not* $b \leq a$ *(making* $<$ *the inverse of the complement of* $\leq$*).*

Exercise 26. *An* asymmetric relation R *is a binary relation defined on a domain* S *such that* $(\forall a, b \in S, Rab \rightarrow \neg Rba)$*. Show that the asymmetry of a relation entails that it is irreflexive, i.e.* $\forall a \in S, \neg Raa$.

It turns out that the well-ordering theorem is equivalent to the Axiom of Choice, as is 'Zorn's lemma', in the sense that in first-order logic either one of the two conjoined with the (remaining) axioms of Zermelo-Fraenkel set theory (see §4.3 below) entails the other.[5]

Fact 4 (Well-ordering theorem (Zermelo's theorem))**.** *For every set* S*, there exists a well-ordering with domain* S.

For the proof (which is not very hard), cf. e.g. Halmos 1960.

One of the most intriguing paradoxes of modern mathematics arises as a consequence of the well-ordering theorem: the 'Banach-Tarski paradox'. This paradox was first stated by Stefan Banach and Alfred Tarski (1924), who showed how it is possible to cut a solid 3-dimensional ball into finitely many pieces and to reassemble the pieces into two solid balls of the same size as the original ball. The proof of this crucially depends on assuming the Axiom of Choice (Axiom 8) or the well-ordering theorem. Please read the article by Robert French if you're interested. You are also invited to watch this video for an illustration of how to use the Banach-Tarski paradox to multiply oranges: `http://www.youtube.com/watch?v=uFvokQUHh08`.

[5] This is not true in second-order logic, where the well-ordering theorem is strictly stronger than the Axiom of Choice.

4.3 Zermelo-Fraenkel set theory

For most purposes, the theory ZC introduced in §4.1 suffices to satisfy the demands on a set theory. However, it does not allow the construction of ordinal numbers (cf. Moore 1990) and Axiom 5 (Separation) insists on 'definite' functions (or formulae) without specifying what is meant by that. Fraenkel (and Thoralf Skolem) thus proposed a slightly modified theory, adding an axiom scheme of replacement and an axiom of regularity, while freeing some of the other axioms from redundancy. This resulted in what is today the most commonly used set of axioms for a set theory, the *Zermelo-Fraenkel axioms*.

Starting out from an informal notion of 'classes', which look just as sets naively construed prior to the advent of axiomatic set theories. All sets as characterized by the axioms to follow are classes, but there are some classes which are not sets (so-called 'proper classes'), most notoriously the *universal class* $V = \{x : x = x\}$. So like most authors, let's presuppose a non-empty domain of discourse as part of our semantics of the first-order logic in which the following set theory is axiomatized. This assumption is harmless, and certainly underwritten by Axiom 15 (Infinity) below, which entails the existence of a set.

In the formal statements of the axioms (ignore them if you want), we will use the notion of *formulae*, which are constructed from the 'atomic' formulae $a \in b$ and $a = b$ by means of the usual logical connectives $\phi \wedge \psi$ (conjunction), $\phi \vee \psi$ (disjunction), $\neg\phi$ (negation), $\phi \to \psi$ (implication), $\phi \leftrightarrow \psi$ (equivalence) and quantifiers $\forall x\phi$ (universal) and $\exists x\phi$ (existential). This means that fundamentally, we introduce '$\in$' and '$=$' as the only non-logical symbols. If a formula has free variables, we often write $\phi(x_1, ..., x_n)$ and mean that the free variables are among the x_i's. Anyway, if you want to ignore the formal statement of the axioms below and focus on the informal statements which always precede the formal ones, I think you can safely forget about everything I said in this paragraph.

Here are the Zermelo-Fraenkel axioms, given both informally and formally:[6]

Axiom 10 (Extensionality)**.** *If S and T have the same elements, then $S = T$. In other words,*

$$\forall S \forall T[\forall x(x \in S \leftrightarrow x \in T) \to S = T].$$

It should be noted that the converse, viz. if $S = T$, then $\forall x(x \in S \Leftrightarrow x \in T)$, is an axiom of predicate calculus. The axiom simply expressed the idea that a set is determined by its 'extension', i.e. by its elements.

Axiom 11 (Pairing)**.** *For any a and b, there exists a set $\{a, b\}$ that contains exactly a and b. In other words,*

$$\forall a \forall b \exists z \forall x(x \in z \leftrightarrow x = a \vee x = b),$$

or, more simply, there exists a set z whose only members are a and b (which themselves may be sets).

Given Axiom 10 (Extensionality), the set z is guaranteed to be unique, which justifies our notation $\{a, b\}$. You may be surprised that we start by stating an axiom about sets of *two* members, rather than just one. But really, we have just included those sets: the *singleton* $\{a\}$ is defined as the set $\{a\} := \{a, a\}$. At this point, we can also define an *ordered pair* $\langle a, b\rangle$ of elements or sets a and b by

$$\langle a, b\rangle := \{\{a\}, \{a, b\}\}.$$

Exercise 27. *Show that this definition satisfies the intuitive requirement on ordered pairs that $\langle a, b\rangle = \langle c, d\rangle$ iff $a = c \wedge b = d$.*

Axiom 12 (Schema of Separation)**.** *If P is a property (with parameter p), then for any S and p, there exists a set $T = \{x \in S : P(x, p)\}$ that contains all those $x \in S$ which exemplify property P. In other words, for a formula $\phi(x, p)$, and for any S and p, there exists a set $T = \{x \in S; \phi(x, p)\}$:*

$$\forall S \forall p \exists T \forall x(x \in T \leftrightarrow x \in S \wedge \phi(x, p)).$$

[6] I consulted various sources for this, in an attempt to get a list as standard as possible. Among them are Hajnal and Hamburger 1999, Jech 2003, Tiles 1989, and `http://en.wikipedia.org/wiki/Zermelo-Fraenkel_set_theory`.

This is called an 'axiom *schema*' rather than just an 'axiom' because it is strictly speaking a distinct axiom for each formula $\phi(x, p)$. Again by Axiom 10 (Extensionality), the set T is unique. As a consequence of Axiom 12 (Schema of Separation), we can introduce set intersection and difference, as follows:

$$S \cap T := \{x \in S : x \in T\} \text{ and } S \setminus T := \{x \in S : x \notin T\},$$

respectively. Compare this with Definitions 7 and 8 above. Moreover, it follows that the empty class $\emptyset = \{x : x \neq x\}$ is a set—our beloved empty set as originally introduced in Definition 5. The empty set exists and is unique just in case at least one set exists. But existence of at least one set is guaranteed by Axiom 15 (Infinity).

An important consequence of Axiom 12 (Schema of Separation) is that the universal class V as defined above is a proper class (and hence not a set), for otherwise, $S = \{x \in V : x \notin x\}$ would also be a set.

Axiom 13 (Union)**.** *For any S, there exists a set $T = \cup S$, the union of all elements of S. In other words, for any S, there exists a set $T = \cup S$:*

$$\forall S \exists T \forall x [x \in T \leftrightarrow \exists z (z \in S \wedge x \in z)].$$

As a result of this axiom, for every S, there exists a *unique* set

$$T = \{x : \exists z \in S (x \in z)\} =: \cup\{z : z \in S\} = \cup S,$$

interpreted as the unique set T whose elements are exactly the elements of the elements of a given set S. Jointly with Axiom 11, this axiom entails that for any two sets S, T, there is a unique set $S \cup T := \cup\{S, T\}$ (which can be generalized to any number of sets). A corresponding 'axiom of intersection' is quite unnecessary, as we were able to define it above, simply using Axiom 12 (Schema of Separation).

Axiom 14 (Power Set)**.** *For any S, there exists a set $T = \mathfrak{P}(S)$, the set of all subsets of S. In other words, for any S, there exists a set $T = \mathfrak{P}(S)$:*

$$\forall S \exists T \forall x (x \in T \leftrightarrow x \subseteq S).$$

At this point of setting up the axioms, we can now introduce quite a bit of useful stuff. For instance, one can use the concept of power sets to define the Cartesian product of two sets, define n-ary relations, fields, functions, etc. I am not going to do this, except to say how to define the Cartesian product. If you are interested to see how this can be used to define relations and functions, please see my handout on structure, available here: `https://wuthrich.net/teaching/2010_246/246HandoutStructureMath_2010.pdf`.

Definition 24 (Cartesian product)**.** *The Cartesian product $S \times T$ of two sets S and T is the set of all ordered pairs such that the first element is an element of S, and the second one of T:*

$$S \times T := \{\langle x, y\rangle : x \in S \wedge y \in T\}.$$

Strictly speaking, we have not yet established that $S \times T$ is a set and that it exists. Noting that $x, y \in S \cup T$ and $\{x\}, \{x, y\} \in \mathfrak{P}(S \cup T)$, we recognize that

$$\langle x, y\rangle = \{\{x\}, \{x, y\}\} \in \mathfrak{P}(\mathfrak{P}(S \cup T)).$$

This now allows us to recognize that the Cartesian product of two sets is a set itself:

$$S \times T \subseteq \mathfrak{P}(\mathfrak{P}(S \cup T)).$$

Axiom 15 (Infinity)**.** *There exists an infinite set.*

We will not state this more formally. Although possible, it would involve introducing new notions we will not otherwise need. Note that this is the only axiom making an existential assertion; courtesy of this axiom, we are guaranteed that there are sets.

Axiom 16 (Schema of Replacement)**.** *If a class F is a function, then for any S, there exists a set $T = F(S) = \{F(x) : x \in S\}$. In other words, for each formula $\phi(x, y, p)$, the following formula is an Axiom (of Replacement):*

$$\forall x \forall y \forall z[\phi(x, y, p) \wedge \phi(x, z, p) \rightarrow y = z] \rightarrow \forall S \exists T \forall y[y \in T \leftrightarrow (\exists x \in S)\phi(x, y, p)].$$

Don't worry if the formal statement becomes too unwieldy for your taste. But just so you know, there are more unwieldy formulations out there! Generally, the exact formulation of the axiom statements varies from source to source. This is not normally a problem, since different formulations can be equivalent. One different but equivalent formulation of Axiom 16 (Schema of Replacement) is that if a class F is a function and $\mathrm{dom}(F)$ is a set, then $\mathrm{codom}(F)$ is also a set.[7]

Axiom 17 (Regularity)**.** *Every non-empty set S contains an element T which is disjoint from S. In other words,*

$$\forall S[\exists T(T \in S) \rightarrow \exists T(T \in S \wedge \neg \exists X(X \in S \wedge X \in T))].$$

There is an important consequence of this axiom: no set is an element of itself. Suppose we have a set S and apply Axiom 17 (Regularity) to $\{S\}$, which is the singleton set of S whose existence is guaranteed by Axiom 11 (Pairing). By Axiom 17 (Regularity), there must be an element of $\{S\}$ which is disjoint from $\{S\}$. But the only element of $\{S\}$ is S, of course, and thus it must be that S is disjoint from $\{S\}$. Since $S \in \{S\}$, it follows that $S \notin S$, by the definition of disjoint sets (Definition 12).

Axiom 17 (Regularity) plays a lesser role in mathematics, as most results still obtain without it. It does play a role in establishing well-ordering results and results concerning ordinal numbers.

Axiom 18 (Choice)**.** *Every family of non-empty sets has a choice function. In other words, for a set S of non-empty elements, there exists a function f from S to the union of all elements of S—the 'choice function'—such that for all $T \in S$, $f(T) \in T$.*

This statement of the axiom of choice (AC) is equivalent to the one given in Axiom 8 (Choice), where each choice function is replaced by its codomain. For finite sets, Axiom 18 (Choice) is quite unnecessary, as it follows as a theorem from Axioms 10–17. This is not the case for infinite sets.

There are a dozen or so statements used across mathematics which are demonstrably equivalent to the AC. The first one to note is Fact 4, the well-ordering or Zermelo's theorem. There are other statements in set theory, such as Tarski's and König's theorems, which are equivalent to the AC. Apart from statements in abstract algebra, functional analysis, topology, and mathematical logic, which are equivalent to the AC, a famous lemma from order theory is as well:

Fact 5 (Zorn's lemma)**.** *Every non-empty partially ordered set in which every totally ordered subset has an upper bound contains at least one 'maximal element', i.e. an element that is not smaller (as defined by the ordering relation) than any other element of the subset.*

Note that I have labelled Zorn's lemma a 'fact', which of course it is given the AC. For a list of statements equivalent to the AC, consult `http://en.wikipedia.org/wiki/Axiom_of_choice`.

The theory resulting from Axioms 10–18 is denoted 'ZFC' or 'ZF', depending on whether Axiom 18 (Choice) is also assumed or not. For some time, it was an open question not only whether ZFC was consistent, but also whether the AC was indeed independent of Axioms 10–17 in the sense that neither it nor its negation could be derived from the other axioms. Assuming the consistency of ZF, Kurt Gödel and Paul Cohen settled this question, Gödel proving that the negation of the AC is not a theorem of Axioms 10–17 and that hence ZFC is consistent, and Cohen showing that the AC is not theorem of Axioms 10–17 by establishing that $\mathsf{ZF}\neg\mathsf{C}$, i.e. ZF conjoined with the negation of the AC, is consistent.

[7] I am sure most of you have encountered the terms 'domain' and 'codomain' (or 'range') of a function. What you may not know is that these concepts can be generalized to relations, such that for instance for a binary relation R, the *domain* of R is defined as the set $\mathrm{dom}(R) := \{x : \exists y \langle x, y\rangle \in R\}$ and the *codomain* of R is defined as the set $\mathrm{codom}(R) := \{y : \exists x \langle x, y\rangle \in R\}$.

5 The continuum hypothesis

One important question remains: given that the set $\mathfrak{P}(\mathbb{N})$ of all subsets of natural numbers is larger than the set $\mathbb{N}$ of natural numbers itself, does there exist a set S such that it is larger than $\mathbb{N}$, but smaller than $\mathfrak{P}(\mathbb{N})$? To ask differently, is there a set whose cardinality lies between that of $\mathbb{N}$ and $\mathfrak{P}(\mathbb{N})$, or is $\mathfrak{P}(\mathbb{N})$ the next largest set after $\mathbb{N}$?

Cantor asked the same question, but the answer is not known to this day. He conjectured that $\mathfrak{P}(\mathbb{N})$ is indeed the next largest set after $\mathbb{N}$, and this conjecture is known as the *continuum hypothesis* (CH). The reason for calling this conjecture the *continuum* hypothesis is because one can bring the set $\mathfrak{P}(\mathbb{N})$ into a one-one correspondence to the set of points on a infinite line, which is sometimes called the *continuum*. Thus, $\mathfrak{P}(\mathbb{N})$ has the same size as the continuum. The question then is whether there exist sets whose size is larger than $\mathbb{N}$ but smaller than the continuum. More generally, Cantor also hypothesized that there does not exist, for any infinite set S, a set whose size lies between that of S and that of $\mathfrak{P}(S)$. This is called the *generalized continuum hypothesis*, or simply *GCH*. Some people, including Smullyan (who again served as basis for the informal part of this section), believe that to proof the GCH or its negation is the biggest outstanding problem in all of mathematics.

What the status of the CH is depends on your larger point of view about the status and function of mathematical theories. Cantor himself believed that the CH was true and expended significant intellectual energies to prove it. Formalists do not consider it to be true or false simpliciter, but make it a question of which axiomatic system is used. In some of these, the CH is true, while in others, it is false. Both the conjunction $\mathsf{ZFC}\wedge\mathrm{GCH}$ of ZFC and the GCH and the conjunction $\mathsf{ZFC}\wedge\neg\mathrm{GCH}$ of ZFC and the negation of the GCH are consistent—assuming that ZFC itself is consistent, which is what most mathematicians believe. This is because the GCH can be shown to be independent of ZFC. This independence was established, again, by Gödel and Cohen. Gödel showed in 1940 that the GCH cannot be *disproved* in ZFC, and Cohen showed in 1963 that the GCH cannot be *proved* in ZFC either. Together, these results mean that the GCH is independent of ZFC, similarly to the infamous Fifth Axiom's independence of the rest of Euclid's axioms. One thing to note, however, is that $\mathsf{ZF}\wedge\mathrm{GCH}$ entails the AC. This means that, while both independent of ZF, the GCH is a strictly stronger claim than the AC because, to repeat, the entailment does not hold in the other direction.

On the other hand, realists/platonists insist that the CH is either true or false—i.e., there is a fact of the matter whether or not the CH is true—and that we simply don't know which. The reason for our ignorance, they continue, is not because the question is in principle unanswerable, but simply because we don't yet know enough about sets. The platonist Gödel predicted that the correct or *true* axiomatic system for set theory, once found, will show that the CH (and the GCH) is false.

In order to just make a connection to what we could go on to study in more detail, let me state the CH more concisely and more formally. The cardinality of $\mathbb{N}$ is usually denoted $\aleph_0$ ('aleph-naught'). Since it can be shown that the cardinality of $\mathbb{R}$ is the same as that of $\mathfrak{P}(\mathbb{N})$ (although we didn't show this here, cf. Jech (2003, 37) or the appendix 'Basic set theory' to Jech (2002)), it is $2^{\aleph_0}$. As the AC guarantees that there exists a smallest cardinal number $\aleph_1$ larger than $\aleph_0$, the CH can be expressed very compactly:

Hypothesis 1 (Continuum)**.**

$$2^{\aleph_0} = \aleph_1.$$

Exercise 28. *Prove the continuum hypothesis.*

This can be generalized for any 'ordinal' α:

Hypothesis 2 (Generalized continuum)**.**

$$2^{\aleph_\alpha} = \aleph_{\alpha+1}.$$

At this point, I could go on about transfinite mathematics, ordinals and cardinals, and the wonderful world of the alephs, the Löwenheim-Skolem theorem and a theorem due to Gödel, but we will not cover this in this course. If you are interested, please read Chapter 10-12 of Moore 1990.

References

[1] Stefan Banach and Alfred Tarski, 'Sur la décomposition des ensembles de points en parties respectivement congruentes', *Fundamenta Mathematicae* **6** (1924): 244-277.

[2] John Earman, Handout 'Sets' for his class *Paradoxes*, Fall 2001, University of Pittsburgh.

[3] Herbert B Enderton, *Elements of Set Theory*, Academic Press (1977).

[4] Robert M French, 'The Banach-Tarski Theorem', *The Mathematical Intelligencer* **10** (1988): 21-28.

[5] Kurt Gödel, *The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis with the Axioms of Set Theory*, Princeton University Press (1940).

[6] András Hajnal and Peter Hamburger, *Set Theory*, Cambridge University Press (1999).

[7] Paul Halmos, *Naive Set Theory*, Litton Educational (1960).

[8] Karel Hrbacek and Thomas Jech, *Introduction to Set Theory*, Marcel Dekker (^{2}1984).

[9] Thomas Jech, *Set Theory*, Springer (2003).

[10] Thomas Jech, 'Set Theory', in Edward N. Zalta (ed.), *The Stanford Online Encyclopedia of Philosophy*, URL `http://plato.stanford.edu/entries/set-theory/` (2002).

[11] David Lewis, *Parts of Classes*, Blackwell (1991).

[12] A W Moore, *The Infinite*, Routledge (1990).

[13] Michael Potter, *Set Theory and Its Philosophy*, Oxford University Press (2004).

[14] Raymond Smullyan. *Satan, Cantor, and Infinity and Other Mind-Boggling Puzzles*, Knopf (1992).

[15] Robert R Stoll, *Set Theory and Logic*, W H Freeman and Company (1963).

[16] Patrick Suppes, *Axiomatic Set Theory*, Dover (1972).

[17] Mary Tiles, *The Philosophy of Set Theory: An Historical Introduction to Cantor's Paradise*, Dover (1989).

[18] Ernst Zermelo, 'Untersuchungen über die Grundlagen der Mengenlehre I', *Mathematische Annalen* **65** (1908): 261-281.

Additional resources used:

- Planet math entries (http://planetmath.org/):
  - `http://planetmath.org/encyclopedia/SetTheory.html`
  - `http://planetmath.org/encyclopedia/ZermeloFraenkelAxioms.html`
- Wikipedia entries:
  - `http://en.wikipedia.org/wiki/Set_(mathematics)`
  - `http://en.wikipedia.org/wiki/Russell's_paradox`
  - `http://en.wikipedia.org/wiki/Zermelo_set_theory`
  - `http://en.wikipedia.org/wiki/Well-ordering_theorem`
  - `http://en.wikipedia.org/wiki/Banach-Tarski_paradox`
  - `http://en.wikipedia.org/wiki/Zermelo-Fraenkel_set_theory`
  - `http://en.wikipedia.org/wiki/Axiom_of_choice`
- Wolfram MathWorld entries (`http://mathworld.wolfram.com/`):
  - `http://mathworld.wolfram.com/Zermelo-FraenkelAxioms.html`