Chapter 3

Sets

No one shall expel us from the Paradise that Cantor has created.

— David Hilbert, *On the Infinite, Math. Ann. 95*

Set theory was invented by Georg Cantor in a remarkable series of six seminal papers. In this chapter, we will learn notation, terminology, and basic operations of set theory. We will follow the axiomatic method by Zermelo with the amendments suggested by Fraenkel and Skolem, so that we can avoid the consistency problems of "naive set theory". The axioms of this set theory are commonly known as ZFC, where the acronym honors Zermelo with Z and Fraenkel with F, and the letter C indicates the inclusion of the axiom of choice. This axiom schema forms the basis of most areas of modern mathematics.

3.1 Background and Motivation

A set is a collection of distinct objects. The objects contained in a set are called its elements. For instance,

$$\{1, 2, 3\}$$

is the set containing three elements, namely 1, 2, and 3. Another set is $\{A, B, C, D, E\}$ that contains the five elements A, B, C, D, and E.

Sets were implicitly used by mathematicians since antiquity. However, they were not seriously studied in their own right before Cantor conceived his theory of sets. Set theory quickly flourished and soon became an established mathematical area. Nowadays, sets are commonly used as a foundation for most mathematical theories.

Sets can have a finite number of elements or an infinite number of elements. One of the key insights by Cantor was that infinite sets can come in different sizes (or, in technical terms, cardinalities). One of the many consequences of this fact is a simple proof for the existence of functions on the natural numbers that cannot be computed. So Cantor's insight into the infinite has tangible benefits to the theory of computation!

The pioneering work by Cantor was not without some flaws. He did not delineate well enough what can and cannot constitute a set, so his theory lead to contradictions. Fortunately, it is possible to fix these flaws, and the subsequent sections detail the improved version of set theory that was developed by Zermelo, Fraenkel and others.

For those who are familiar with Cantor's set theory, we point out some apparent differences. In ZFC, every element is a set itself – so there are no so-called urelements that are not sets themselves. This allows for a more coherent development of the theory, but the difference is not essential. The unrestricted set comprehension of Cantor's set theory is replaced by a more restricted version, which eliminates troublesome antinomies.

3.2 Fundamental Concepts

The central concept of mathematics is a **set**. A set is an abstraction of a container. The objects contained in it are called its **elements**. We write

$$a \in A$$

to denote that a is an element of the set A. If a is not an element of A, then we write $a \notin A$. Put differently, $a \notin A$ is the negation of $a \in A$.

The first axiom asserts that there exists a set that does not contain any elements.

> **S1.** There exists a set that does not have any elements.

We call it the **empty set** and denote it by $\varnothing$, or by $\{\}$. The membership relation $a \in \varnothing$ is always false. The empty set occurs more often than you might think. Its role in set theory is similar to the role of 0 in the integers.

The second axiom asserts that the membership relation completely determines the equality of sets.

> **S2.** Two sets are **equal** if and only if they have the same elements.

In the Zermelo-Fraenkel set theory, every element is a set. Our third axiom schema asserts that we can form a finite set given its elements.

> **S3.** Let $a_1, \dots, a_n$ be sets. Then there exists a set S such that $x \in S$ if and only if $x = a_1$ or $x = a_2$ or $\cdots$ or $x = a_n$.

The set S is commonly denoted by listing its elements in curly braces,

$$S = \{a_1, a_2, \dots, a_n\}.$$

For example, $\{a, b, c\}$ is the set containing the elements a, b, c, and no other elements.

It should be stressed that listing elements more than once does not change the set. For instance, $\{a, a, b\}$ denotes the same set as $\{a, b\}$. We drive this point home in the next example.

Example 3.1. The four sets

$$S_1 = \{\varnothing\}, S_2 = \{\varnothing, \varnothing\}, S_3 = \{\varnothing, \varnothing, \varnothing\}, S_4 = \{\varnothing, \varnothing, \varnothing, \varnothing\}$$

all contain the empty set $\varnothing$ as an element. Since none of them contain any other element, they actually all specify the same set! So, despite appearances, the sets S_1, S_2, S_3, and S_4 each contain a single element. Listing an element multiple times is simply redundant.

Although we have only learned a few axioms of ZFC set theory, they already allow us to specify many different sets. Starting from the empty set and the set builder notation, we can form many different sets.

We call a set $\{a\}$ with one element a **singleton set**. This terminology is helpful in the next example.

Example 3.2. The empty set $\varnothing$ and the singleton set $\{\varnothing\}$ containing the empty set as an element are different sets. Indeed, the empty set $\varnothing$ does not contain any element, but the singleton set $\{\varnothing\}$ contains one element, namely the empty set. It follows from **S2** that $\varnothing$ is not equal to $\{\varnothing\}$.

By the axiom **S2**, the order of the elements does not matter, so $\{a, b, c\}$ denotes the same set as $\{b, c, a\}$.

Example 3.3. We can form a set S containing as elements the empty set $\varnothing$, the singleton set $\{\varnothing\}$, and the set $\{\varnothing, \{\varnothing\}\}$ containing two elements. Apparently, all three elements of this set are different, as the first contains no element, the second contains one element, and the third contains two elements. So even if we do not repeat elements, there are already six different ways to list the elements of S, namely

$$\begin{aligned}
S_1 &= \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}, & S_2 &= \{\varnothing, \{\varnothing, \{\varnothing\}\}, \{\varnothing\}\}, \\
S_3 &= \{\{\varnothing\}, \varnothing, \{\varnothing, \{\varnothing\}\}\}, & S_4 &= \{\{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \varnothing\}, \\
S_5 &= \{\{\varnothing, \{\varnothing\}\}, \varnothing, \{\varnothing\}\}, & S_6 &= \{\{\varnothing, \{\varnothing\}\}, \{\varnothing\}, \varnothing\}.
\end{aligned}$$

Each set $S_1, S_2, \ldots, S_6$ is equal to S, as the order of elements does not matter by axiom **S2**.

If we form sets starting with the empty set and set builder notations, we can quickly end up with a bewildering number of curly braces. This can be a bit tedious to decipher. Therefore, it is customary to introduce abbreviations. For instance, the next example reformulates the previous example with abbreviations, which dramatically increase the readability.

Example 3.4. If we denote the empty set $\varnothing$ as 0, the singleton set $\{\varnothing\}$ as 1, and the set $\{\varnothing, \{\varnothing\}\}$ with two elements as 2, then the sets from the previous example can be stated in the more readable form $S_1 = \{0, 1, 2\}$, $S_2 = \{0, 2, 1\}$, $S_3 = \{1, 0, 2\}$, $S_4 = \{1, 2, 0\}$, $S_5 = \{2, 0, 1\}$, and $S_6 = \{2, 1, 0\}$.

The idea of defining a nonnegative integer n by a set with n elements goes back to von Neumann. We will explore his construction in depth later on. One can apply the same principle to other objects. The next example briefly sketches one possible set-theoretic representation of ASCII characters.

Example 3.5. If we want to reason about sets of ASCII characters, then we can use von Neumann's idea and represent the upper case `A` with ASCII code 65 by a set with 65 elements, the lower case `a` with ASCII code 97 by a set with 97 elements, and so on. This allows us to form sets of characters such as $\{\mathtt{A}, \mathtt{B}, \mathtt{C}\}$.

A set A is called a **subset** of a set B if and only if all elements of A are also elements of B. We write $A \subseteq B$ to denote that A is a subset of B. If A is a subset of B, but not equal to B, then we call A a **proper subset** of B. We write $A \subsetneq B$ to denote that A is a proper subset of B.

Example 3.6. If $A = \{1, 2\}$, $B = \{2, 3\}$ and $C = \{1, 2, 3\}$, then $A \subseteq C$ and $B \subseteq C$, so A and B are subsets of C. In fact, they are both proper subsets of C. However, A is not a subset of B, as 1 is an element of A that is not contained in B. Furthermore, B is not a subset of A, as 3 is an element of B that is not contained in A.

A convenient way to prove that two sets A and B are equal is to show that $A \subseteq B$ and $B \subseteq A$ holds. The next proposition establishes this simple fact.

Proposition 3.7. *If A and B are two sets such that $A \subseteq B$ and $B \subseteq A$, then $A = B$.*

Proof. Since $A \subseteq B$, $x \in A$ implies $x \in B$. Since $B \subseteq A$, $x \in B$ implies $x \in A$. Therefore, $x \in A$ if and only if $x \in B$, which means that $A = B$. □

Every set A has itself as a subset, since the definition of a subset does not require that a subset is a proper subset. The next proposition shows that every set has the empty set as a subset. This is obvious if you have a good grounding in logic, but perhaps a bit surprising for a novice.

Proposition 3.8. *The empty set is a subset of each set.*

Proof. Seeking a contradiction, we assume that there exists a set A such that the empty set $\varnothing$ is not a subset of A. This means that there must exist an element $x \in \varnothing$ such that $x \notin A$. However, this is absurd since the empty set does not contain any element x. □

The most common way to specify a subset of a set A is by defining a property $S(x)$ that an element x of A may or may not have.

S4. Let $S(x)$ be a property of x. For a set A, there exists a set B such that $x \in B$ if and only if x is an element of A that has property $S(x)$.

The set B is commonly denoted by

$$\{x \in A \mid S(x)\}.$$

We need to be able to specify the property $S(x)$ using a formula in the first-order language of set theory that has only x as a free variable. Furthermore, the formula $S(x)$ cannot reference the set A, so that circular definitions are avoided. In other words, $S(x)$ is constructed using the equality $=$ and the membership relation $\in$, the logical operators $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$, as well as the quantifiers $\forall$ and $\exists$ using the rules **F1**–**F4** of Section 2.8.

The main point of the construction is that a subset of A is obtained by filtering out all elements x that do not satisfy $S(x)$. What remains is evidently a subset of A, and it contains precisely the elements x of A that do satisfy the property $S(x)$.

The next example illustrates this principle.

Example 3.9. Let A denote the set

$$A = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}\}$$

with four distinct elements. The set A contains the empty set $\varnothing$, the singleton set $\{\varnothing\}$, the set $\{\varnothing, \{\varnothing\}\}$ with two elements, and the set $\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$ with three elements. If $S(x)$ is the property $(\{\varnothing\} \in x)$ then

$$\{x \in A \mid S(x)\} = \{x \in A \mid \{\varnothing\} \in x\} = \{\{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}\}$$

is the subset of A containing as elements $\{\varnothing, \{\varnothing\}\}$ and $\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$. Both elements evidently contain $\{\varnothing\}$. Missing from this subset of A are the elements $\varnothing$ and $\{\varnothing\}$ of A that do not contain $\{\varnothing\}$ as an element.[1]

One does not need to express the property $S(x)$ in such a formal way. It is sufficient to specify the property $S(x)$ unambiguously in a way that can be expressed – in principle – by a formal statement. The next example illustrate this point.

Example 3.10. Let A denote the set

$$A = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}\}$$

with four distinct elements. If $S(x)$ is the property that (x has an even number of elements), then

$$\{x \in A \mid S(x)\} = \{\varnothing, \{\varnothing, \{\varnothing\}\}\}$$

[1]Indeed, the empty set $\varnothing$ does not contain any elements, and the set $\{\varnothing\}$ contain just one element, namely $\varnothing$, but this element is different from $\{\varnothing\}$.

(a) If $n \in S$, then there exists some nonnegative integer $m \leqslant n$ such that $n = f(m)$. We define a new function

$$g(i) = \begin{cases} f(i) & \text{for all } i \in \{0,\ldots,n-1\} \text{ such that } i \neq m, \\ f(n) & \text{if } i = m \text{ and } m \neq n. \end{cases}$$

Then $g\colon \{0,\ldots,n-1\} \to S - \{n\}$ is an injective map, contradicting $P(n)$.

(b) If $n \notin S$, then the restriction $f \upharpoonright n$ maps $n = \{0,\ldots,n-1\}$ onto the proper subset $S - \{f(n)\}$, contradicting $P(n)$.

Therefore, $P(n+1)$ is true. We can conclude by induction that the claim holds for all nonnegative integers n. □

Remark 3.49. Dedekind realized that for any infinite set A one can find an injective map onto a proper subset of A. For example, the map $s\colon \mathbf{N}_0 \to \mathbf{N}_0 \backslash \{0\}$ given by $s(n) = n + 1$ is injective and maps $\mathbf{N}_0$ to the proper subset of positive integers. It follows from the pigeonhole principle and Dedekind's observation that a set S is infinite if and only if one can find an injective map from S onto a proper subset.

A set A satisfying $|A| \leqslant |\mathbf{N}_0|$ is called a **countable set**. In other words, it suffices to find an injective function from A into $\mathbf{N}_0$ to prove that A is countable. Alternatively, we can prove that A is countable if we can find a surjection from $\mathbf{N}_0$ onto A, see Proposition 3.41. A set A that is not countable is called **uncountable**.

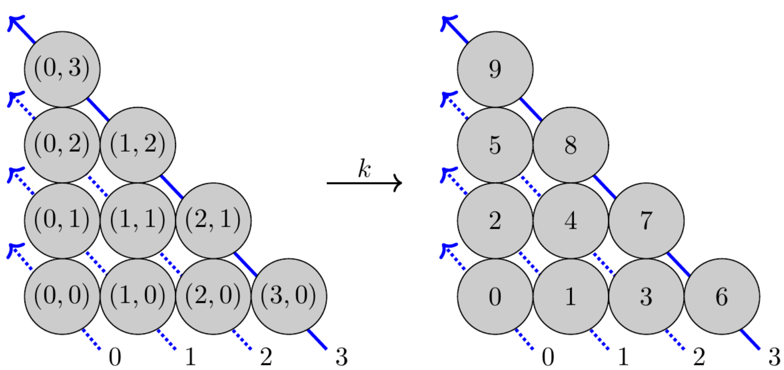

Figure 3.7: A pair of nonnegative integers (x, y) lying on the nth diagonal, where $n = x+y$, is mapped to the value $T_n + y$. For example, $(2, 1)$ lies on the 3rd diagonal. The number of elements on the previous three diagonals is $T_3 = 1 + 2 + 3$. So $(2, 1)$ is mapped to the value $T_3 + 1 = 6 + 1 = 7$.

Proposition 3.50. *Suppose that A and B are countable sets. Then $A \times B$ is a countable set.*

Proof. By assumption, there exist injective maps $f\colon A \to \mathbf{N}_0$ and $g\colon B \to \mathbf{N}_0$. Therefore, the function $h\colon A \times B \to \mathbf{N}_0 \times \mathbf{N}_0$ given by $h(x, y) = (f(x), g(y))$

is an injective function. Therefore, it suffices to show that the set $\mathbf{N}_0 \times \mathbf{N}_0$ is countable.

Let us follow Cantor and define a function $k\colon \mathbf{N}_0 \times \mathbf{N}_0 \to \mathbf{N}_0$ by

$$k(x, y) = \frac{1}{2}(x + y + 1)(x + y) + y. \tag{3.2}$$

The first few values of $k(x, y)$ are illustrated in Figure 3.7.

Note that $k(x, y) = T_n + y$, where $n = x + y$ designates the diagonal to which (x, y) belongs and $T_n = \sum_{k=1}^{n} k$ is the n-th triangular number, which counts the number of elements on all previous diagonals. The function $k(x, y)$ is bounded by the triangular numbers,

$$T_{x+y} \leqslant k(x, y) = T_{x+y} + y < T_{x+y} + x + y + 1 = T_{x+y+1}.$$

We will now show that k is an injective function. Suppose that $k(x, y) = k(a, b)$. Seeking a contradiction, let us assume that $x + y \neq a + b$. Without loss of generality, we may assume that $x + y < a + b$. By the above bound for $k(x, y)$, we get

$$k(x, y) = T_{x+y} + y < T_{a+b} \leqslant k(a, b),$$

contradicting the equality $k(x, y) = k(a, b)$. Therefore, $x + y = a + b$.

It follows that $k(x, y) = T_{x+y} + y = T_{x+y} + b = k(a, b)$, so $y = b$, whence $x = a$. Therefore, $(x, y) = (a, b)$ and it follows that k is injective. ❑

We encourage you to explore further properties of Cantor's function k, see Exercise 3.62.

Proposition 3.51. *Let $S = \{A_i \mid i \in I\}$ be a family of countable sets A_i such that I is a nonempty countable set. Then $\bigcup S$ is a countable set.*

Proof. The claim holds if all sets A_i are empty, so we may assume that not all sets A_i are empty. Let $I_0 = \{i \in I \mid A_i \neq \varnothing\}$ be the indices of all nonempty A_i. Since I is countable, the subset I_0 is countable as well. Therefore, there exists a surjection $s\colon \mathbf{N}_0 \to I_0$, so we can form the set family $T = \{A_{s(n)} \mid n \in \mathbf{N}_0\}$, which satisfies $\bigcup T = \bigcup S$.

Since $A_{s(n)}$ is a nonempty set, there exists a surjective map $f_n\colon \mathbf{N}_0 \to A_{s(n)}$. Therefore, the function $f\colon \mathbf{N}_0 \times \mathbf{N}_0 \to \bigcup S$ given by $f(m, n) = f_n(m)$ is a surjective map onto $\bigcup S = \bigcup T$. If we compose this map with the inverse of Cantors map $k^{-1}\colon \mathbf{N}_0 \to \mathbf{N}_0 \times \mathbf{N}_0$, we get a surjective map $f \circ k^{-1}\colon \mathbf{N}_0 \to \bigcup S$. Therefore, $\bigcup S$ is countable. ❑

Let A and B be sets. We write $|A| < |B|$ if and only if there exists an injective function from A into B, but no bijective function from A onto B.

Proposition 3.52 (Cantor)**.** *Let A be a set. There does not exist any surjection from A onto $P(A)$.*

Proof. Given any function $f\colon A \to P(A)$, we can construct the set

$$S = \{x \in A \mid x \notin f(x)\}.$$

Seeking a contradiction, let us assume that there exists an element $a \in A$ such that $f(a) = S$. Then we have $a \in f(a)$ if and only if $a \notin f(a)$, which is a contradiction. Thus, f is not surjective. $\square$

Corollary 3.53. *Let A be a set. Then $|A| < |P(A)|$.*

Proof. The function $f\colon A \to P(A)$ given by $f(x) = \{x\}$ is injective, so $|A| \leqslant |P(A)|$. By the previous proposition, there cannot exist any surjective map from A onto $P(A)$. Therefore, $|A| < |P(A)|$. $\square$

A remarkable consequence of the previous corollary is that one can construct sets of larger and larger cardinality.

Let us denote by $2^{\mathbf{N}_0}$ the set of all functions from the nonnegative integers to the set $\{0, 1\}$ with two elements.

Proposition 3.54. *The power set $P(\mathbf{N}_0)$ and the set $2^{\mathbf{N}_0}$ are equipotent,*

$$|2^{\mathbf{N}_0}| = |P(\mathbf{N}_0)|.$$

Proof. Define a function $f\colon P(\mathbf{N}_0) \to 2^{\mathbf{N}_0}$ by

$$f(A) = \chi_A$$

where χ_A is the characteristic function of A defined by

$$\chi_A(n) = \begin{cases} 1 & n \in A \\ 0 & n \notin A \end{cases}$$

Then f is a bijection. Indeed, if two characteristic functions χ_A and χ_B are the same, then $A = B$; thus, f is injective. The function f is surjective, since every function in $2^{\mathbf{N}_0}$ is a characteristic function of its set of arguments yielding the value 1. $\square$

Consider functions that takes a nonnegative integer as an argument and have a single bit as a value. It might seem that one can compute any such function with a suitable computer program given enough time. However, nothing could be further from the truth. We will now sketch the proof that for most functions in $2^{\mathbf{N}_0}$ one cannot write a program that will compute it.

Choose a programming language L of your preference that is expressive enough to compute functions. For instance, the `while` language or the λ-calculus are primitive choices. We call a function f in $2^{\mathbf{N}_0}$ computable in L if and only if there exists a program that for each input $n \in N_0$ will compute $f(n)$.

Theorem 3.55. *There exist functions in $2^{\mathbf{N}_0}$ that cannot be computed in L.*

Proof. Let A_k denote the set of all programs in L that have k characters. Then A_k is a finite set. The set of all programs that can be expressed in L is given $\bigcup S$, where $S = \{A_k \mid k \in N_0\}$. By Proposition 3.51, $\bigcup S$ is a countable set.

By contrast, the set $2^{\mathbf{N}_0}$ of characteristic functions is not countable, since $|2^{\mathbf{N}_0}| = |P(\mathbf{N}_0)| > |\mathbf{N}_0|$ by Corollary 3.53. This proves the claim. $\square$

One should note how crude our argument was. We simply listed all possible programs that can be written in the programming language L. Since this set is countable, the subset of programs computing a function with a nonnegative integer as the input and a single bit as the output is countable as well. Since the set of characteristic functions is not countable, there are many functions that we cannot compute. Courses on computability will explore many examples of interesting functions that cannot be computed.

EXERCISES

3.57. Show that the sets $A = \{x \in \mathbf{R} \mid 1 \leqslant x \leqslant 4\}$ and $B = \{x \in \mathbf{R} \mid 2 \leqslant x \leqslant 7\}$ have the same cardinality.

3.58. Let (a) $A_0 = \{x \in \mathbf{R} \mid 0 < x < 1\}$, (b) $A_1 = \{x \in \mathbf{R} \mid 0 \leqslant x < 1\}$, (c) $A_2 = \{x \in \mathbf{R} \mid 0 < x \leqslant 1\}$, (d) $A_3 = \{x \in \mathbf{R} \mid 0 \leqslant x \leqslant 1\}$. Show that all four sets have the same cardinality using Proposition 3.47 (b).

3.59. Let $A = \{x \in \mathbf{R} \mid 0 < x < 1\}$ and $B = \mathbf{R}$. Show that $|A| = |B|$ by giving an explicit bijection from A onto B.

3.60. Show that $|\mathbf{N}_0| = |\mathbf{Z}|$ by explicitly giving a bijective function from the set $\mathbf{N}_0$ of nonnegative integers onto the set $\mathbf{Z}$ of integers.

3.61. Suppose that A and B are sets such that $|A| = |B|$. Show that $|P(A)| = |P(B)|$.

3.62. Show that Cantor's function $k\colon \mathbf{N}_0 \times \mathbf{N}_0 \to \mathbf{N}_0$ defined in equation (3.2) is surjective. This means that k is a bijective function.

3.63. Show that $|P(\mathbf{N}_0)| = |P(\mathbf{N}_0 \times \mathbf{N}_0)|$.

3.64. Show that the following sets are countable:
(a) The set of nonnegative even integers.
(b) The set of prime numbers.
(c) The set of (positive and negative) integers.

3.65. Show that a subset of a countable set is countable.

3.10 Notes

Set theory was conceived by Georg Cantor [12] in a series of seminal papers. A brief informal treatment of set theory can be found in a delightful book by Kaplansky [46].

We have given a brief overview of Zermelo-Fraenkel set theory. The two books by Halmos [34] and Hrbacek and Jech [37] are excellent introductions to Zermelo-Fraenkel set theory that are recommended for further reading. Our debt to the books by Halmos and Hrbacek and Jech should be obvious. We also consulted the books by Deiser [20], Ebbinghaus [23], Enderton [26], and Schindler [68] in the preparation of this chapter.

Another axiomatic set theory was conceived by von Neumann, Bernays, and Gödel. This theory contains proper classes in addition to sets and this can be convenient at times. We recommend the book by Smullyan and Fitting [73] for further information.

For a more advanced treatment of set theory including a thorough discussion of forcing, see Jech [42] and Kunen [56].