The Mathematics of Infinity

PURE AND APPLIED MATHEMATICS

A Wiley Series of Texts, Monographs, and Tracts

A complete list of the titles in this series appears at the end of this volume.

The Mathematics of Infinity

A Guide to Great Ideas

Second Edition

Theodore G. Faticoni
Department of Mathematics
Fordham University
Bronx, NY

WILEY
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Faticoni, Theodore G. (Theodore Gerard), 1954–
The mathematics of infinity : a guide to great ideas / Theodore G. Faticoni. — 2nd ed.
p. cm. — (Pure and applied mathematics)
Includes bibliographical references and index.
ISBN 978-1-118-20448-1
1. Cardinal numbers. 2. Set theory. 3. Infinite. I. Title.
QA248.F29 2012
511.3'22—dc23 2011041439

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To *Professor Elliot Wolk* who taught me Set Theory.

Contents

Preface for the Second Edition

The most primitive of herdsman used a pouch of stones to keep track of the number of sheep he had in the field. As each sheep would enter the field, the herdsman would place a stone in a pile. As the sheep would leave the field, the herdsman would place the stones back into the pouch. If there were stones left on the ground, then some sheep were missing. If there were no stones left, and no sheep left then all was well with the herd. And if there were no more stones but there were more sheep, then somehow the herdsman had picked up an ewe or two.

This correspondence between pouch stones and sheep is one of the most primitive forms of counting known. In today's language, this is known as a one-to-one correspondence, or a bijection between pouch stones and sheep. This kind of counting is continued today when we make an attendance sheet. Each name on the sheet corresponds to exactly one child in the class, and we know some child is missing if he or she does not respond to his or her name. A more important correspondence is found in the grocery store. There we associate a certain number called a price with each item we put in our cart. The items in the cart correspond to a number called the total price of the cart. When we compare our receipt with the objects in the cart, we are imitating the sheep herdsman's pouch stones.

Believe it or not, mathematicians count like the primitive herdsmen. The number 1 is all sets that match up in an exact manner to the set $\{\bullet\}$. Thus, we say that $\mathrm{card}(\{\bullet\}) = 1$, and we say that $\mathrm{card}(\{*\}) = 1$. The number 1 becomes all that we associate with one element. We use the convenient symbol 1 to denote all possible sets that match up perfectly with $\{\bullet\}$. The symbol 1 is convenient because it is what we have been taught all these years. The number 2 is defined to be all of those sets that match up perfectly with $\{\bullet, *\}$.

$$\mathrm{card}(\{\bullet, *\}) = 2.$$

This is 2 because we define it that way. It agrees with our training. It represents all possible sets that match up exactly with the set $\{\bullet, *\}$. This is exactly what you have been taught.

Next up is what we mean by *matches up perfectly*. This is the bijection we alluded to earlier. Sets A and B are called *equivalent* if there is a bijection between them. That is, they match up perfectly. In other words, there is a way of matching up elements between A and B, called a *function* or bijection

$$f : A \longrightarrow B$$

such that

1. different elements of A are mapped to different elements of B, and

2. each element of B is associated with some element of A.

For finite sets, this bijection can be drawn as a picture. Let $A = \{a_1, a_2, a_3\}$ and let $B = \{b_1, b_2, b_3\}$. Then one bijection between A and B is

$$\begin{cases} a_1 & \longmapsto & b_1 \\ a_2 & \longmapsto & b_2 \\ a_3 & \longmapsto & b_3 \end{cases}$$

which matches A up with B in an exact manner. Here is another such bijection

$$\begin{cases} a_1 & \longmapsto & b_3 \\ a_2 & \longmapsto & b_2 \\ a_3 & \longmapsto & b_1 \end{cases}$$

between A and B. You see, the bijection you choose does not have to respect the subscripts. These mappings are bijections because as you can see the elements a_k are sent to different elements b_ℓ. Also each element in B is associated with an element in A. That is exactly how mathematicians count elements in sets.

An impressive extension of this idea is that we can count *infinite sets* in the same manner, but you must use different symbols to denote card(A). We let

$$\text{card}(A) = \text{ the } \textit{cardinality} \text{ of } A,$$

which is simply all sets B such that A is equivalent to B. That is, card(A) is all those sets B for which there exists a bijection $f : A \longrightarrow B$. Hence $B \in \text{card}(A)$ or $\text{card}(A) = \text{card}(B)$ exactly when there is a function $f : A \longrightarrow B$ such that

1. different elements of A are mapped to different elements of B, and

2. each element of B is associated with some element of A.

Notice that the definition of *bijection* has not changed.

Since these sets are infinite we need a new symbol to denote card(A) of infinite sets. It is traditional to use the Hebrew letter *aleph*

$$\aleph$$

to denote infinite cardinals. Let

$$\begin{aligned} \mathbb{N} &= \{0, 1, 2, 3, \ldots\}, \\ \mathbb{R} &= \{x \mid x \text{ is a real number}\}. \end{aligned}$$

So $\mathbb{N}$ is the set of whole, nonnegative numbers, and $\mathbb{R}$ is the set of all real numbers. These would be decimal expansions like 1.414 and 3.14159. Then we write

$$\text{card}(\mathbb{N}) = \aleph_0$$

and we say *aleph naught.* It is quite a surprising mathematical (universal) truth that there is a cardinal $\aleph_1$ such that

$$\aleph_0 < \aleph_1.$$

Indeed there is an infinite chain of infinite cardinals

$$\boxed{\aleph_0 < \aleph_1 < \aleph_2 < \aleph_3 < \cdots.}$$

We will have a chance to expand on this idea in the later chapters of this book.

This second edition of *The Mathematics of Infinity: A Guide to Great Ideas* contains some new ideas about mathematics as logic. We begin with the *binary logic* that we all learn at an earlier age.

It is well known that most statements P and Q can have logical states True or False. This is the basis for most legal conversations or in scientific argument or in mathematics in general. In fact, some years ago (up to 1920) several logicians tried to derive most of mathematics from binary logic. Four long and laborious volumes on their research were written. The best effort achieved a proof that $1 + 1 = 2$ after almost 2000 pages of logical symbolism. Shortly thereafter a German logician/mathematician Curt Gödel (circa 1930) proved that this line of research was mathematically impossible, as no effort from logic could deduce all of mathematics. There would always be a mathematical statement that had been missed, or the authors would have made a mistake.

We approach binary logic in a more traditional manner. We introduce the operations *and, or, not, implies* on statements in the way Aristotle must have defined them some 2600 years ago. We describe how statements *lingually and logically* combine under these operations, and we significantly reduce the importance of the more recently used charts of symbols P, Q, T, and F. We feel the chart has its place in the binary design of a computer and not as a form of conversation or argument between people.

Thus, we define *not* P so that it changes the logical state from one state into another. This agrees with the modern chart. We say that P *and* Q is True precisely when both P and Q are True. Implied in this is that P *and* Q is False in case either statement is False. The same lingual manner is applied to defining P *or* Q, which is False precisely when both P and Q are False. The implication $P \Rightarrow Q$ is False precisely when a Truth implies a Falsehood.

The most compelling use of $P \Rightarrow Q$ is to form the classic arguments. So Truth with a correct argument leads to Truth, and a False premise P can lead us to either a Truth or a False conclusion Q. For this reason, any arguer that proceeds from a False premise cannot decide the Truth of his conclusions. We will avoid such arguments, but they do lead us to some fun as we investigate the

Epimenides Paradox (which is no paradox at all) and the World's Hardest Logic Puzzle (which we dispatch in a couple of lines).

We provide an elementary investigation of the logical state of the Liar's Paradox *This statement is False.* Some have concluded that the Liar's Paradox is *always* False, which is a statement no serious mathematician would make. For example, in the conversation *All statements are False, This statement is False*, the Liar's Paradox is True, but in the conversation *All statements are True, This statement is False*, the Liar's Paradox is False. We prove that the logical state of the Liar's Paradox is more like our cultural Walrus than Aristotle, our ancient Lord of Logic.

The last results in the book are extensions of Gödel's Theorem showing that if $\mathbf{C}$ is a set of True statements from some logical system, then there is some statement Q that is True over $\mathbf{C}$ but not deducible from $\mathbf{C}$. This is used to prove that there can be no theory of everything for any logical intellectual endeavor.

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

Chapter 1

Logic

The ideal writing style ascribed to by mathematicians is that in writing mathematics, *less is more.* If we can convey the exact idea of a concept with 5 words instead of 10, then we will use 5. Thus, we will use the statement *Cardinal numbers form a well-ordered collection* over the wordier statement *The well-ordered property is enjoyed by the collection of cardinal numbers.* The second statement is mathematically correct, but it is more than we need to convey the idea.

I have tried to practice this ideal while writing the mathematics in this book. The only exceptions to this ideal are made on the basis of decisions on the educational value of sentence structure, the anecdotal comments, or discussions of this sort that occur between mathematical discourse. Sometimes it is good to sacrifice some mathematical austerity in the interest of getting an important point across to the reader. As the reader will clearly see, this economy of words in mathematical writings is not exercised in the text of a discussion. Discussions and intermediate anecdotes contain examples and illustrations that are the only tools we have to illustrate a concept. Since I have sacrificed a good bit of mathematical rigor in favor of clarity, examples and illustrations are necessary if I am to get some subtle ideas across to the reader. This form of personalized writing style is unavoidable when discussing advanced ideas from mathematics in the popular press.

We have a bit of a mountain to climb in this book, so please be patient. Perhaps you can sit down in an overstuffed chair or at a table and open the book. Maybe you have a pencil and paper handy. That's a good idea. Some of these topics need to be diagrammed.

And certainly you have a cup of beverage, coffee would be my choice. Now turn on that lamp overhead and blend in that final inspiring ingredient: cream in your coffee. Good luck.

1.1 Axiomatic Method

The Axiomatic Method is how mathematicians apply logic. It is how we advance from one topic to the next, and so this is how future generations will discover more sophisticated forms of mathematics. The section will be brief, but it is how the mathematics in each successive chapter is treated.

Axioms are mathematical statements that we assume are True. We do not prove axioms, they come to us as statements whose Truth we do not deduce. The use of axioms first comes to us from the Greek slave Euclid circa 300 BC in his book *The Elements*. *The Elements* begins by stating five axioms and five postulates to be taken as primitive Truths. By assuming these 10 statements, Euclid was building a foundation on which logic would be used to deduce the mathematics in the reminder of *The Elements*. Today, the method of applying logic to a small set of primitive Truths is called the *axiomatic method*. It is the way mathematics has been practiced for the last 2300 years. It has lasted essentially unchanged since Euclid wrote it, including the many editions printed in the various lands in which *The Elements* was read and studied. It is how mathematics will progress to find larger thoughts using today's theorems.

For example, Euclid defines a *right angle* as the bisection of a line, and then he assumes that two right angles are equal. Today, we would say that it is obvious that any two right angles are equal, and so it was with Euclid's contemporaries. You might even suggest that you can prove it, but when you do you are assuming that any two lines represent an angle of the same measure, π. You have assumed what you wanted to prove. Euclid did not have angular measure, so he could not talk about π radians, but he knew what he was assuming. Thus the fourth postulate of *The Elements* assumes that any two right angles have equal measure.

A more subtle axiom is the fifth postulate, today called the *parallel postulate.* This was an attempt to describe the interior angles of two lines cut by a transversal. After thousands of years of investigation the parallel postulate has evolved into the equivalent form that we know today. It states that *through a point P not on a line L there is a unique line L' that is parallel to L.* Today's plane geometry is based on this parallel postulate. It is an interesting topic for further reading that one can change the parallel postulate into two different parallel postulates, and that each has its own use.

1.2 Tabular Logic

Formal logic is the logic used in Computer Science to design and construct the guts of your computer and its central processing chip. You have used this logic every time you analyzed regions in a Venn diagram.

And then there is Aristotle's logic. This is *the* logic used by rational men to form rational arguments. This logic is used to form arguments to prove that something is, or to prove that something is not. While we will examine formal logic, we are most interested in Aristotle's logic. Before we use Aristotle's logic to construct arguments we will introduce elementary or primitive logical statements.

The statements P, Q, and R are variables. They represent all statements from the language we are speaking. They do not exclude values unless we state so. Thus, P represents something simple like *The sky is blue*, or $1 \neq 0$, or something more complicated like

> *The sum of the squares of the lengths of the legs in a right triangle is the square of the length of the hypothenuse.*

Even that last statement is a possible value for P.

Aristotle's Logic begins with these statements and combines them using the elementary logical operations *not, and, or.* There might be other logical operations but they can be expressed as combinations of these three. The logical state of a statement formed by using P and Q is determined by the entries in a few tables.

The operation *not* simply changes the logical state of P from one logical state into the other. In tabular form, *not* can be described by the following.

P	$not P$
T	F
F	T

In the first column of this table, we are considering all possible logical states for P. True T and False F are all of the logical states that P can achieve in this book. (We consider only binary logic here.) The second column is the logical state of the statement $not P$. We should be clear about this. We begin with a table and *define* what *not* means. That defined meaning reflects exactly what you have used *not* for in your life. We do not begin with a word *not* and then try to make up a table for it. We have tried to define a logical operation here, and that is what the table does.

Notice that the table for *not* does just what we first stated *not* will do. It takes a logical states for P and changes it from True to False, or from False to True. Follow the logic for P today.

1. $P =$ *The sky is blue* is a True statement.

2. $not P =$ *The sky is not blue* is a False statement.

Of course, if it is a slate gray sky, or if we are on Mars, then *The sky is blue* will be a False statement. This is what you mean when you say to someone *What is the color of the sky in your Universe?* You are asking for the logical state of the statement *The sky is blue.* You are asking that individual for a logical foundation from which the two of you can intelligently converse.

Our operation *not* has a familiar property that comes from that early English class you had. Given a statement P, then $not\, not P$ has the same logical states as P. That is, a double negative does not change the logical state of P. In terms of a table, we have

P	$not P$	$not\, not P$
T	F	T
F	T	F

The first and third columns of the table show that P and $not\, not P$ have the same logical state. If P is True, then $not\, not P$ is True, and if P is False, then $not\, not P$ is False. Notice that the first way we described the double negative, the table we gave for it, and the last couple of lines all describe the same operation. From input to ultimate output, *not not* does not change the logical state of the input statement. It does change P, though, doesn't it. If we let $P =$ *The sky is blue* then one reading of *not not* P is *It is False that the sky is not blue.* This last statement will have the same logical state as P, but it is awkward in its presentation. We will avoid the double negative whenever possible but we will find that at times we are forced to deal with it.

The operation *and* combines two statements P and Q and makes a compound statement P *and* Q. The statement P *and* Q is True exactly when both P and Q are True. So, of course, the other possible combinations of T and F for P and Q yield logical states False. Thus, if one or more of P, Q is False then the statement P *and* Q is False. If we let $P =$ *The sky is blue* and if we let $Q =$ *I am human* then today P *and* Q is a True statement. *The sky is blue and I am human* is a True statement on the day this is written. But if it is a slate gray sky, then *The sky is blue and I am human* is a False statement. If I come from Mars then *The sky is blue and I am human* is a False statement. If I am writing this on Mars in January of 1900 then *The sky is blue and I am human* is a False statement because both $P =$ *The sky is blue* and $Q =$ *I am human* are False statements.

The table of values T, F for *and* will make the above discussion short and mechanical. That table is

P	Q	P *and* Q
T	T	T
T	F	F
F	T	F
F	F	F

In other words, P *and* Q is True exactly when both P and Q are True. In any other situation, P *and* Q is False.

The first two columns of the above table gives us all of the possible pairs of logical states T, F for P and Q, and in the third column we read the corresponding logical states for the compound statement P *and* Q. Notice that P *and* Q is a True statement exactly when both P and Q are True. Otherwise, P *and* Q is a False statement. This is an effective shorthand since once we know that

P *and* Q is True exactly when both P and Q are True, then the logical states in the rest of the table fall into place.

Let us see how the logic of our discussion proceeds. It is elementary, but it also shows us what the undercurrent of our thought process is.

1. Let $P =$ *The sky is blue* and let $Q =$ *I am human.*

2. P is True, and Q is True.

3. P *and* Q is then *True*.

4. Thus *The sky is blue and I am human* is True.

You may have skipped all of those thoughts but this list of thoughts fills in all of those nagging details about the logic of the compound statement. Actually, this linear discussion shows how logic is part of the structure of your language. We just don't think in that much detail, now do we?

The third operation is the *or* operation. This operation takes two statements P and Q and assigns a True logical state to P *or* Q when at least one of them is True. Or to put it another way, P *or* Q is False exactly when both P and Q are False. The rest of the cases T, F for P and Q yield a True statement P *or* Q.

So if we let $P =$ *The sky is blue*, and if we let $Q =$ *I am human*, then P *or* Q is a True statement. That is, *The sky is blue or I am human* is a True statement. That is because Q is True. The logical state of P in this case does not matter. If it is a slate gray sky, then *The sky is blue or I am human* is still a True statement. The True statement *I am human* makes the compound statement *The sky is blue or I am human* a Truth. But if I am writing this on Mars in January of 1900, then *The sky is blue and I am human* is a False statement because both $P =$*The sky is blue* and $Q =$*I am human* are False statements.

The table for the operation *or* will again make the above discussion short and mechanical.

P	Q	P *or* Q
T	T	T
T	F	T
F	T	T
F	F	F

In other words, $P\ or\ Q$ is False exactly when both P and Q are False. The other values T in the third column of the table are then forced.

The first two columns give us all possible pairs of logical states for the statements P and Q. The third column gives us the logical states of $P\ or\ Q$ that correspond to the first two columns of the table. Notice that the logical states in column three show that $P\ or\ Q$ is False exactly when both P and Q are False.

The next way to combine statements P and Q we will call *implication.* We write

$$\boxed{P \Rightarrow Q}$$

when we want to say that P *implies* Q. In its simplest form, $P \Rightarrow Q$ is False exactly when P is True and Q is False. In every other instance the statement $P \Rightarrow Q$ is True. Its tabular description follows from this verbal description.

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

From the table defining *implication*, we see several important properties. The first row of the table for $P \Rightarrow Q$ gives us the most important argument in mathematics, that of deductive reasoning. The first row shows us that if we make no mistake, that is, if $P \Rightarrow Q$ is True, then the Truth of P implies the Truth of Q. Thus, if P is True and if we make no mistakes, that is, if $P \Rightarrow Q$ is True, then Q is True. We can find new Truths from old Truths in this way.

The last row shows us that if we make no mistakes in our argument, then a Falsehood Q comes from a Falsehood P. Hence, if $P \Rightarrow Q$ is True and if Q is False, then P is False. We will often work with this argument. It is called the *indirect proof.*

The first line of the table allows us to deduce the second line. The Truth of P implies the Truth of Q if we make no mistakes in our argument $P \Rightarrow Q$. Therefore, if P is True and if Q is False, then we made a mistake somewhere, and $P \Rightarrow Q$ is False. If you have deduced a Falsehood from a Truth then you have made a mistake. That is what the False logical state of $P \Rightarrow Q$ stands for: a mistake. Thus, the implication $P \Rightarrow 1{=}0$ is False if P is True. The conclusion $1 = 0$ is False, so the implication is a Falsehood.

There is another interesting possibility for $P \Rightarrow Q$. A Falsehood P will imply anything. If we begin our argument with a False premise P, then subsequent deductions Q do not possess a predictable logical state. These deductions Q can be either True or False. If P is a Falsehood then $P \Rightarrow Q$ is True no matter what the logical states of Q is. Thus, if P is False, you can deduce that $1 = 0$, that *All opinions are valid*, and that there is a Universal Set. But, as we will prove later, each of these is a Falsehood. The conclusion drawn will have no logical weight whatsoever because your premise P was False

After all, we can deduce that there are no prime numbers if we assume that $1 = 0$, but of course the conclusion is False. The argument goes like this. Assume that $1 = 0$. Then $1 + 1 = 0 + 0$ and so $2 = 0$. In this manner, we can prove that $n = 0$ for each $n \in \mathbb{N}$. That is correct. From the premise $1 = 0$ we can prove that there are no other natural numbers but 0. Since 0 is not a prime number, we have proved that there are no prime numbers. This is the kind of foolishness we can arrive at by proceeding from a False premise. However, the steps in our argument were all True, so that the implication $1{=}0 \Rightarrow$ *there are no prime numbers* is a True statement. Think about that for awhile.

Exercise 1.2.1 Let P, Q, and R be statements. Make Truth Tables for the compound statements in the following exercises.

1. $\textit{not}\,(P\ \textit{or}\ Q)$

2. $\textit{not}\,(P\ \textit{and}\ Q)$

3. $P\ \textit{and}\ (Q\ \textit{and}\ R)$

4. $P\ \textit{or}\ (Q\ \textit{or}\ R)$

5. $((\textit{not}P)\ \textit{or}\ Q)\ \textit{and}\ (P\ \textit{or}\ (\textit{not}Q))$

1.3 Tautology

A *tautology* is a logical statement R that is always True. When its table is established, the output logical states, those values in the rightmost column, are all T. Let us examine a few tautologies.

Consider $P\, or\, (not\, P)$. We can see that this is tautological by observing that by the table for the *not* operation, either P or $not\, P$ is a Truth. That is, one of P and $not\, P$ is True. Examining the table for *or* shows us that $P\, or\, (not\, P)$ is then True. In its tabular form, we have

P	$not\, P$	$P\, or\, (not\, P)$
T	F	T
F	T	T

Thus, $P\, or\, not\, P$ is a tautology. $P\, or\, not\, P$ is a True statement given any logical state for P. In other words, as the table suggests, no matter which statement is used for P, the output $P\, or\, (not\, P)$ is a True statement. For example, *The sky is blue or the sky is not blue* is True, as is $1 = 0\ or\ 1 \neq 0$. Also, the statement *There is a Universal Set or there is no Universal Set* is True. We may not know what a Universal Set is, but we know that the statement *There is a Universal Set or there is no Universal Set* is True. That is, it either is or it is not.

In the same way, the statement $P\, and\, (not\, P)$ is False because by definition of *not*, the statements P and $not\, P$ have different logical states. For instance, if P is True, then $not\, P$ is False. Then by the the definition of *and*, $P\, and\, (not\, P)$ is False. The next table shows all of this in tabular form.

P	$not\, P$	$P\, and\, (not\, P)$
T	F	F
F	T	F

For example, let X be a any statement. Given the statement $P =$ *X is valid* then $not\, P =$ *X is not valid.* Hence, the statement

$$X \text{ is valid and } X \text{ is not valid}$$

is a False statement. We will encounter this kind of Falsehood often as this chapter moves along.

A more common logical comparison of statements is the following. Let X and Y be statements. We say that X and Y are *logically equivalent* iff X and Y have the same logical states. In terms of a Truth Table, the two right-hand columns in the table for X and Y have the same sequence of T's and F's. We saw above that P and $not\, not P$ have the same logical values, so

> The statements P and $not\, not P$ are logically equivalent

A more complex example of logically equivalent statements is formed from an implication and its *contrapositive.* The contrapositive of the implication $P \Rightarrow Q$ is the implication

$$\boxed{notQ \Rightarrow notP.}$$

Notice the reversal in the roles of P and Q. The Truth Table of the contrapositive of $P \Rightarrow Q$ is given below.

P	Q	$notP$	$notQ$	$notQ \Rightarrow notP$
T	T	F	F	T
T	F	F	T	F
F	T	T	F	T
F	F	T	T	T

Let us show that the implication and its contrapositive are logically equivalent. We use a sizable Truth Table.

P	Q	$notQ$	$notP$	$P \Rightarrow Q$	$notQ \Rightarrow notP$
T	T	F	F	T	T
T	F	T	F	F	F
F	T	F	T	T	T
F	F	T	T	T	T

Notice that the two right most columns have exactly the same entries in the same order. Thus, $P \Rightarrow Q$ and $notQ \Rightarrow notP$ are logically equivalent. They are not the *same* statement. They differ in their sentence structure. They differ in the way they are written in English. This is important. Logically equivalent statements do not have to look alike at all.

Consider the statement

Let $a > 0$ be a natural number. If a^2 is odd, then a is odd.

Its contrapositive is True when the implication is True and its contrapositive is False when the implication is False. That contrapositive is

Let $a > 0$ be a natural number. If a is even, then a^2 is even.

These two statements are logically equivalent even though their statements are different.

Something more symbolic is the logical expression

$$(not P)\ or\ Q.$$

It is a simple combination of the operations *not* and *or*, and yet we will see that it is logically equivalent to a familiar statement.

To begin, we argue verbally. The statement $(not P)\ or\ Q$ is a False statement exactly when the twostatements $not P$ and Q are False. This occurs when P is True and Q is False, and only when P and Q are in these logical states. For any other logical states, (not P) or Q is a True statement. There is a coincidence here. The implication $P \Rightarrow Q$ is False only when P is True and Q is False. Given any other logical states for P and Q, $P \Rightarrow Q$ is True. Hence (not P) or Q and $P \Rightarrow Q$ have the same logical states. They are then logically equivalent. The relevant Truth Table looks like this.

P	Q	$not P$	(not P) or Q	$P \Rightarrow Q$
T	T	F	T	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

We conclude that

(not P) or Q and $P \Rightarrow Q$ are logically equivalent.

At this point, we will abandon the use of Truth Tables. One can use them to discuss the other logical tautologies that we will bring out presently, but we feel that in our present setting, they are Baroque. The reader should feel free to translate our lingual discussion into a tabular one. The exercise will do you good.

Some logical statements are combinations of two or more smaller statements. Let P, Q, and R be statements. Different ways to combine and manipulate these statements are from the following tautologies.

associative law	$P \textit{ and } (Q \textit{ and } R)$	$=$	$(P \textit{ and } Q) \textit{ and } R$
distributive law	$P \textit{ and } (Q \textit{ or } R)$	$=$	$(P \textit{ and } Q) \textit{ or } (P \textit{ and } R)$
distributive law	$P \textit{ or } (Q \textit{ and } R)$	$=$	$(P \textit{ or } Q) \textit{ and } (P \textit{ or } R)$
the biconditional	$P \Leftrightarrow Q$	$=$	$(P \Rightarrow Q) \textit{ and } (Q \Rightarrow P)$

The associative law, for example, states that there is no reason to use parentheses in conjuncted statements using the *and* operation. The statement

The sky is blue and the grass is green and I am human

is unambiguous in the calculation of its logical state. The distributive laws simply give us reasons to replace commas with parentheses. For example,

Either the sky is blue, or the grass is green and I am human

can be rewritten as

The sky is blue or the grass is green,
and the sky is blue or I am human

without changing the logical state of the compound statement.

The *biconditional* is a short way of writing that P implies Q, and also that Q implies P. The biconditional $P \Leftrightarrow Q$ is read P *if and only if* Q. This means *If* P *then* Q and *If* Q *then* P. It is common to write

$$\boxed{P \text{ iff } Q}$$

for $P \Leftrightarrow Q$.

For instance, we will prove that

$$1 + 1 = 2 \text{ iff } 1 + 1 + 1 = 3.$$

Proof: We must first prove that $1 + 1 = 2 \Rightarrow 1 + 1 + 1 = 3$. This is done by beginning with $1 + 1 = 2$, and then adding 1 to each side, which yields $1 + 1 + 1 = 2 + 1 = 3$.

Conversely, begin with $1 + 1 + 1 = 3$, and then subtract 1 from both sides, from which we have $1 + 1 = 1 + 1 + 1 - 1 = 3 - 1 = 2$. Thus, $1 + 1 = 2$, which is what we had to prove.

Therefore, having proved the implication $1+1 = 2 \Rightarrow 1+1+1 = 3$ and the implication $1 + 1 + 1 = 3 \Rightarrow 1 + 1 = 2$, we conclude that $1 + 1 = 2$ iff $1 + 1 + 1 = 3$. This completes the argument.

Let us prove another biconditional statement.

Let $a \in \mathbb{N}$. Then $a < 2a$ iff $a \neq 0$.

Proof: We must prove that *if* $a < 2a$ *then* $a \neq 0$ and that *if* $a \neq 0$ *then* $a < 2a$. Begin by proving *if* $a < 2a$ *then* $a \neq 0$. Suppose that $a < 2a$. Then, allowing that we all know the arithmetic, $0 = a - a < 2a - a = a$. Thus $a \neq 0$, which proves this half of the larger proof.

Conversely, we must prove *if* $a \neq 0$ *then* $a < 2a$. Begin with $a \neq 0$. Then $a > 0$, so that $2a = a + a > 0 + a = a$. Thus $a < 2a$ is proved. Therefore, by having proved both *if* $a < 2a$ *then* $a \neq 0$ and *if* $a \neq 0$ *then* $a < 2a$, we conclude that $a \neq 0$ iff $a < 2a$. This completes the larger proof.

There will be much more to say about the biconditional statement in the later pages of this book.

Another tautology that we will find useful in the rest of this book is called *DeMorgan's Laws for Logic*. It shows how the *not* operation combines with the *and* and the *or* operations. Symbolically, DeMorgan's Laws for Logic look like this.

DeMorgan's Law 1.3.1 *Let P and Q be statements. Then*

1. $\text{not}(P \,\text{and}\, Q) = (\text{not}\, P) \,\text{or}\, (\text{not}\, Q)$

2. $\text{not}(P \,\text{or}\, Q) = (\text{not}\, P) \,\text{and}\, (\text{not}\, Q)$

Note the change from *and* to *or* and from *or* to *and* in each of DeMorgan's Laws. Thus

Either the sky is not blue, or the grass is not green.

has the same logical state as

It is False that the sky is blue and the grass is green.

Furthermore,

The sky is not blue and the grass is not green.

has the same logical state as

It is False that the sky is blue or that the grass is green.

Exercise 1.3.2 Let P, Q, and R be statements.

1. Show that $P \,\textit{and}\, (\textit{not} Q)$ is logically equivalent to $\textit{not}((\textit{not} P) \,\textit{or}\, Q)$.

2. Show that $P \,\textit{or}\, (\textit{not} Q)$ is logically equivalent to $\textit{not}((\textit{not} P) \,\textit{and}\, Q)$.

3. Find a shorter statement that is logically equivalent to $\textit{not}(\textit{not}((\textit{not} P) \,\textit{or}\, Q)) \,\textit{or}\, R)$.

4. Give a highly detailed proof that *If* $a = ab$ *then* $b = 1$.

5. Give a highly detailed proof that *If* $x^2 - 1 = 0$ *then* $x \in \{-1, 1\}$.

1.4 Logical Strategies

The first logical observation is that there is a statement that is always False, no matter what the logical value of P is. The statement $P \,\textit{and}\, (\textit{not} P)$ is a Falsehood. It is False F all of the time. We discussed this in the previous section.

Example 1.4.1 1. Let P be the statement *The sky is blue.* Then *the sky is blue and the sky is not blue* is a Falsehood.

2. Let P be the statement *There is a mountain.* Then the statement *There is a mountain and there is no mountain* is a Falsehood.

3. Let P be the statement *This is True.* Then *This is True and this is not True* is a Falsehood.

We continue our discussion with *implications.* Let P and Q be statements. It is good to know that $P \Rightarrow Q$ is True or False, but it is better to know how one uses the implication $P \Rightarrow Q$ to argue correctly. For example, let us reexamine the first line of the Truth Table for $P \Rightarrow Q$. That first row is T, T, T. It can be read as follows. If we start with a True premise P, and if we make no mistakes in our argument, then our conclusion Q is True. You might expect this, since every deductive argument is based on it.

Example 1.4.2 1. Here is a Greek classic. Begin with the statement P, *Socrates is a man.* It is True that *If he is a man then he is mortal.* We conclude Q that *Socrates is mortal* is True. I know that P is True, and that $P \Rightarrow Q$ is True. From this I deduce that Q is True.

2. Let P be the statement *I see the sky on Earth* (on a sunny day). It is True that *If I see the sky on Earth then the sky is blue.* Therefore I conclude Q that *The sky I see is blue.* I have assumed P and observed that $P \Rightarrow Q$ is True. From this I deduce that Q is True.

Example 1.4.3 1. Let P be the True statement *The sky is not blue.* Argue correctly as follows. If the sky is not blue, then we are not on Earth. Conclude the statement Q that *We are not on Earth* is a True statement.

2. Something more advanced goes like this. Let $a, b \in \mathbb{N}$ be nonzero. We prove that $a \geq ab$.

Proof: We will write one True implication after another to form our argument. We will then conclude that $a < ab$. Begin with a True statement. Since $a, b \in \mathbb{N}$ are nonzero, $a, b \geq 1$. Since $a \geq 1$, multiplying $1 \leq b$ by a implies that $a = a \cdot 1 \leq ab$. Thus, $a \leq ab$ is a True statement, which is what we had to prove.

A more subtle argument shows us the strength of logic when arguments are concatenated. Consider the statements

$$\begin{aligned}&P\\&P \Rightarrow Q\\&Q \Rightarrow R\\&R\end{aligned}$$

and assume that P is a True statement, that $P \Rightarrow Q$ is a True statement, and that $Q \Rightarrow R$ is a True statement. Make no assumptions about R. Argue as follows.

From the Truth of P and the Truth Table of $P \Rightarrow Q$, we deduce that Q is a True statement. Then Q and $Q \Rightarrow R$ are True, so we may deduce that R is a True statement.

We began with the Truth of P. The Truth of the statements $P \Rightarrow Q$ and $Q \Rightarrow R$ allows us to make a correct argument $P \Rightarrow Q$ *and then* $Q \Rightarrow R$. We then deduce the Truth of R. This justifies our use of this compound argument in our deliberations.

More generally, we have shown that if your argument begins with a True statement P, and if, as above, the compound argument consists of one True implication after another, then you have formed a longer True argument. You then deduce the Truth of the last statement R in your argument.

Example 1.4.4 This example shows how the above discussion can be applied to longer arguments.

a) Assume that $10 < 2^{10}$. (This is the statement P.)

b) Because $10 < 2^{10}$, we see that $2 \cdot 10 < 2 \cdot 2^{10}$. (This is $P \Rightarrow Q$.)

c) Because $20 = 2 \cdot 10$, we have $11 < 2 \cdot 10$. (This is $Q \Rightarrow R$.)

d) Because $2 \cdot 10 < 2 \cdot 2^{10}$, it follows that $2 \cdot 10 < 2^{1+10} = 2^{11}$. (This is $R \Rightarrow S$.)

e) The conclusions in lines c) and d) combine to show that $11 < 2 \cdot 10 < 2^{11}$.

We conclude that $11 < 2^{11}$.

We will not use this level of detail in future arguments.

Exercise 1.4.5 Find the error in the following arguments. These errors make the arguments False.

1. (a) This amoeba lives on Earth.
   (b) All life on Earth has hair.
   (c) Thus, this amoeba has hair.
2. (a) Assume that $x \in \mathbb{R}$ satisfies $x^2 - x = 0$.
   (b) If $x^2 - x = 0$ then $x^2 = x$.
   (c) $x^2/x = x$ and $x/x = 1$.
   (d) Thus, $x = 1$.
3. (a) I swim in the sea.
   (b) If I am a fish, then I swim in the sea.
   (c) I swim in the sea, so I am a fish.
   (d) Thus, I am a fish.

1.5 Implications From Implications

The implication $P \Rightarrow Q$ comes with what is called the *converse*. *The converse of* $P \Rightarrow Q$ *is* $Q \Rightarrow P$. Let us write down the Truth Table for $Q \Rightarrow P$ and compare it with $P \Rightarrow Q$.

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

P	Q	$Q \Rightarrow P$
T	T	T
T	F	T
F	T	F
F	F	T

As you can see, the implication and the converse do not have the same Truth Table. The value of the implication $P \Rightarrow Q$ in the third row is T, while the logical value of the implication $Q \Rightarrow P$ in the third row is F. Thus, the converse implication $Q \Rightarrow P$ can be False even when $P \Rightarrow Q$ is True. For this reason, the converse cannot be assumed to be a True statement even when the original implication is True. Therefore, we must be careful to avoid the classic error of using the converse of an implication. Some examples will help.

Example 1.5.1 1. Let P be *The sky is blue*, and let Q be *The world is flat*. Then $P \Rightarrow Q$ is *If the sky is blue then the world is* *flat.* This is a False statement. Its converse is $Q \Rightarrow P$. *If the world is flat then the sky is blue.* The Truth of this converse is established by looking at line 3 in the Truth Table for $\Rightarrow$.

2. The converse of the implication *If today is Monday then my schedule is clear* is *If my schedule is clear then today is Monday.* The implication may be True, but the converse is False if my schedule is clear on Sunday.

Suppose that we have an implication $P \Rightarrow Q$. If the implication is True and if Q is False, then line 4 of the Truth Table for $P \Rightarrow Q$ shows us that P is False. This leads us to an important implication known as the *contrapositive.*

$$notQ \Rightarrow notP$$

We discussed the contrapositive beginning on page 18. The implication $P \Rightarrow Q$ and its contrapositive $notQ \Rightarrow notP$ are logically equivalent statements.

Example 1.5.2 1. Let P be *My GPS is working* and let Q be *I am not lost.* The implication is *If my GPS is working then I am not lost.* Its contrapositive is *If I am lost then my GPS is not working.* Notice that both the implication and its contrapositive are True, provided that I always use my GPS.

2. Let P be *My spell-check program is running*, and let Q be *I misspell all the time.* The implication *If my spell-check program is running then I do not misspell all the time* has contrapositive *If I misspell all the time then my spell-check program is not running.* Notice that both the implication and its contrapositive are True.

3. The implication *If the sky is not blue then this is not Earth* has as contrapositive *If this is Earth then the sky is blue.* Once again the implication and its contrapositive are True.

The next form of argument is meant to demonstrate how to use a counterexample as a means of proof. The idea is as follows. If I tell you that *All colors are blue*, then you can show that this is a Falsehood by producing some color that is not blue. One nonblue color would be red. The existence of red shows that *All colors are blue* is False. You just proved *Some colors are not blue.*

This approach to proof is called *proof by contradiction.* They all proceed in the same way. You are given a statement that claims either *All of something has a property* or *Nothing has a property.* To show that these two statements are Falsehoods, and to prove a statement about something having a property, you produce a counterexample to the claim. Some examples will help.

Example 1.5.3 1. *All numbers are even* is claimed. You produce the number 3, which is not even. This 3 is a counterexample to the claim that *all* numbers are even. Thus *All numbers are even* is False. You have thus proved that *Some numbers are not even.*

2. *All people are Truth sayers* is claimed. That is, it is claimed that every person tells the Truth all of the time. You react by producing a known False statement and then speaking it. You speak *1 = 0*, thus uttering a False statement. The claim is proved False. You have proved that *Someone will sometimes speak Falsehoods.*

3. *All statements are False* is claimed. You state that *The number line has no end.* This Truth is a counterexample to the claim. You have shown that *Some statements are True.*

4. *Nothing is interesting* is claimed. The claim is that nothing in this world is interesting at all. You say, *I find that the lack of interesting facts in this world is interesting,* thus showing that something is interesting. With this counterexample to the claim, you have shown that *Somethings are interesting.*

Exercise 1.5.4 Prove these statements using the proof by counter example.

1. All math professors are male.

2. All numbers are positive.

3. If $a \leq a^2$ then $1 < a$.

4. $\tan(\theta) \leq 1$ for all $\theta \in (-\pi/2, \pi/2)$.

5. All men are liars.

6. All men are Truth sayers.

7. Left alone things do not change.

1.6 Universal Quantifiers

Statements about *All* or *Nothing* of something are called *universal quantifiers.* We will discuss proofs of universal quantifiers in this section. In so doing, we will show that a certain kind of statement is always False. These statements are common when speaking about logical ideas. We will show that these Falsehoods not only look alike but they also have similar proofs.

A logical strategy that comes from the bottom row

P	Q	$P \Rightarrow Q$
F	F	T

of the Truth Table we used to define the implication $P \Rightarrow Q$ is given as follows. *Suppose that you make no mistakes in your argument, but that you conclude that Q is False. Then your premise P is also a Falsehood.* To employ this proof, you begin by assuming that P is True. Then you make a careful argument with no mistakes. Once you have come upon a Falsehood, Q, you can conclude that P is False. We will continually use this strategy in this section.

The *universal* and *existential quantifiers* are logical statements used to define the logical value of infinitely many statements without writing down infinitely many statements. Let $P(t)$ be a statement that is defined in terms of a parameter t. For example, $P(n)$ might be $n < n+1$, or $P(x)$ might be *If $x \neq 0$ then $\frac{1}{x}$ is defined.* The *universal quantifier* allows us to discuss the logical values of $P(n)$ and $P(x)$ all at once. It is the infinite counterpart of the *and* operation. Suppose that $t \in I$ for some index set I. Then

$$\boxed{\forall t \in I, P(t) \text{ is read } \textit{for all } t \textit{ in } I,\ P(t) \textit{ is True.}}$$

You may delete the *is True* as the language permits. The universal quantifier $\forall t \in I, P(t)$ is True only when $P(t)$ is True for every $t \in I$. Thus, $\forall n \in \mathbb{N}, n < n+1$ is a True universal quantifier since $n < n+1$ is True for all $n \in \mathbb{N}$.

The *existential quantifier* is the infinite equivalent of the *or* operation. Given a statement $P(t)$ with parameter $t \in I$ then

$\exists t \in I, P(t)$ is read *there exists a t in I such that P(t) is True.*

You may drop *is True* as the language permits. The existential quantifier $\exists t \in I, P(t)$ is True when there is some $t \in I$ such that $P(t)$ is True. (The words $P(t)$ *is True for some* t mean that $P(t)$ is True for at least one value t and at most all t.)

The *negation* of $\forall t \in I,\ P(t)$ is argued as follows. It is False that $P(t)$ is True for all $t \in I$, so $P(t)$ is False for some $t \in I$. Put succinctly,

$\exists t \in I,\ not P(t)$ is the logical negation of $\forall t \in I,\ P(t)$.

Thus, the negation of a universal quantifier is an existential quantifier. For instance, $\forall x \in \mathbb{R}, \sqrt{x}$ *is real* is a False statement because $\sqrt{-1}$ is not real. The logical negation of $\forall x \in \mathbb{R}, \sqrt{x}$ *is real* is $\exists x \in \mathbb{R},\ \sqrt{x}$ *is not real.*

The logical negation of $\exists t \in I,\ P(t)$ is rationalized as follows. Suppose it is False that $\exists t \in I,\ P(t)$. Then no statement $P(t)$ is True, or equivalently $P(t)$ is not True for *any* $t \in I$. Given $t \in I$ then $P(t)$ is False, so that *not* $P(t)$ is True for each $t \in I$. Symbolically we would write as follows.

$\forall t \in I,\ not P(t)$ is the logical negation of $\exists t \in I,\ P(t)$.

Example 1.6.1 1. Let $P(n)$ be the statement $2n = 2^n$. Since $2 \cdot 2 = 2^2$, the existential statement $\exists n \in \mathbb{N},\ 2n = 2^n$ is a True statement, while $2 \cdot 3 \neq 2^3$ shows that $\forall n \in \mathbb{N},\ 2n = 2^n$ is False.

2. Fix a continuous function f of a real variable on the closed bounded interval $[a, b]$, and let F denote an arbitrary function on $[a, b]$. If $P(F)$ is the statement $\dfrac{dF}{dx} = f$, then $\exists F, P(F)$ is True by The Fundamental Theorem of Calculus. The function $F(x) = \int_a^x f(t)\, dt$ has derivative $f(x)$.

Exercise 1.6.2 In each of the following exercises, negate the False quantifier.

1. $\forall x \in \mathbb{R},\ x^2 > 0$

2. $\exists x, y, z \in \mathbb{N},\ x^3 + y^3 = z^3$

3. $\exists$ continuous function f on $[0, 1]$, $\frac{d}{dx}\int_0^1 f(t)dt \neq f(x)$.

4. $\exists$ a set $\mathbf{C}$, $\mathbf{C}$ contains all sets.

5. $\forall$ opinion P, P is valid.

1.7 Fun With Language and Logic

Here are some examples that will illustrate how deduction is used in mathematics. To understand these examples, we must first examine how to diagram a sentence.

Let Y be some quality of statements. So Y could be, but is not restricted to, the values *True, valid, known, recursive, hard to read, impossible to understand.* Let Q be the statement *This statement has quality* Y. We would say that Q and *This statement has quality* Y are the same. Furthermore, since *This statement* refers to Q, Q is the same as *This statement* Q *has quality* Y, which is also the same as Q *has quality* Y. The above discussion is easily translated into the following table, each of whose statements are the same.

1. Q

2. *This statement has quality* Y

3. *This statement* Q *has quality* Y

4. Q *has quality* Y

Some clever arguments can be formed by interchanging these equivalent ways to write Q.

Example 1.7.1 This problem goes back to ancient Greece, sometime before 300 BC. Epimenides is a philosopher from the island of Crete. He walks into a room of Greek scholars and cries out, *All*

Cretans are liars. Then he says, *I am lying.* The Greek scholars proceed to determine the logical value of *I am lying*, deciding that *I am lying* is both a lie and the Truth, an intolerable situation in those days.

We will give a modern approach to the Epimenides Paradox that explains the paradox. I have hired a Cretan public speaker to speak the quotations in this discussion. Assume that *All Cretans are liars.* Consider the statement Q: *This statement when spoken by a Cretan is not a lie.* Since our Cretan spoke Q, Epimenides' declaration states that Q is a lie. But Q states that *This statement* Q *is not a lie*, so that Q is not a lie. But the statement R: Q *is a lie* has logical negation *not*R: Q *is not a lie.* Recall that R *and* (*not*R) is a Falsehood from page 9. We have argued to a Falsehood, so our premise is False. That is, *All Cretan's are liars* is a Falsehood.

Let us review how we argued. We assumed that the statement P : *All Cretans are liars* is True. We then made no errors in our implication, and we deduced the two statements R: Q *is a lie* and its logical negation *not*R: *Q is not lie.* Since this conclusion of ours R *and* (*not*R) is then False, we began with a False premise. That is, *All Cretans are liars* is a Falsehood.

Returning to the ancient paradox, we proceed from the False premise *All Cretans are liars.* Thus, we can deduce many things, but we have no means of showing that the conclusions are True. Specifically, we claim that we have deduced that *I am lying* is paradox being neither True nor False. But the Truth is that we cannot identify the logical state of *"I am lyin" is a paradox.* Beginning with a Falsehood the way we did makes any analysis of the logical state of *"I am lying" is a paradox* impossible. This illustrates just how badly facts can be confused when we proceed from a False premise.

Example 1.7.2 Suppose we assume that *All is known.* That is, if there is a fact to be learned, then it is already known. Consider the statement Q : *This fact is not known.* By our assumption, Q is known. By reading Q we see that *The fact* Q *is not known*, so that Q is not known. We have deduced a statement R: Q *is known* and its logical negation *not*R: Q *is not known*, which we combine to form the Falsehood R *and* (*not*R) as on page 9. We conclude that we began with a False premise, so that *All is known* is False.

Suppose that in some future time we have deduced a body of knowledge B. Then B represents all of the known facts, which by the above paragraph cannot be all facts. Specifically, knowledge is neither finite nor bounded. Consequently, there will always be some fact to be discovered. Hence, unless we willingly stop work, researchers and scholars cannot work themselves out of a job. There will always be some unknown fact to be researched.

Example 1.7.3 Suppose we assume that *All opinions are valid.* Consider the statement Q: *This opinion is not valid.* Since Q is an opinion, our supposition implies that Q is valid. But Q itself states that this opinion Q is not valid. Thus, we have deduced R: Q *is valid* and its logical negation *not*R: Q *is not valid.* When combined we have deduced the Falsehood R *and* (*not*R). Thus, our premise *All opinions are valid* is False. This can be interpreted as stating that not every opinion advanced is worth listening to.

Example 1.7.4 The above examples illustrate a general form of statement and argument. Let Y be one of the properties in the set {True, known, valid, complicated, assumed, hard to understand, worth listening to}. Suppose we assume that *All self-referential statements are of type* Y. Consider the statement Q: *This statement is not of type* Y. Because Q is a self-referential statement, it is True that Q *is of type* Y. But Q itself states that the statement Q *is not of type* Y. Thus, we have deduced R: Q *is of type* Y and its logical negation *not*R: Q *is not of type* Y. The Falsehood R *and* (*not*R) shows us that *All statements are* Y is False.

Example 1.7.5 A *perfect logician* is a person who knows all of logic. Let us decide the logical state of *There is a perfect logician* PL. Assume that there is a perfect logician PL. We begin our argument with the statement Q: *This statement of logic is not known to* PL. Then this logical statement, Q, is not known to PL. Furthermore, by assumption, PL is a perfect logician, so Q is known to PL. The contradiction is that we have deduced Q *is not known to* PL and its logical negation Q *is known to* PL. Thus *There is a perfect logician* PL is a Falsehood, and hence there can be no perfect logicians.

Let us apply the previous example.

Example 1.7.6 At the time of this writing, *The World's Hardest Logic Puzzle* has been a popular stop for those who surf the web. The puzzle begins with 200 perfect logicians on an island, 100 of them are blue eyed, and 100 of them are brown eyed. Once a perfect logician has determined his eye color through the use of logic alone, he will leave the island. The problem is to determine when all of these perfect logicians have left the island. That's it. That is all that we assume in this version of the puzzle. There are some internet versions of this puzzle that include much more detail than this version, but they and their solutions follow from our solution given below. In other words, once we solve this problem then we can solve any other version of it.

The solution is that the problem begins by assuming that there are 200 perfect logicians, while we have proved that there are *no* perfect logicians. Thus the premise of the problem is False. You can therefore deduce anything you want, but you have no way of knowing which deduction is True. Thus, you might deduce that 200 people leave the island Friday, you might deduce that 200 people leave the island instantly, or you might deduce that seven of them never leave the island. But we cannot know the Truth of these conclusions as we proceed from a False premise.

This kind of indirect argument will appear often in the succeeding chapters. The readers should familiarize themselves with it.

The Q statements used above are examples of *self-referential statements*. These are statements that refer to themselves. One fun example of a self-referential statement is the following story that in the end refers to itself.

Mythology 1.7.7 There once was a girl who liked to travel from town to town, telling this story about herself. One day, while traveling in the dense forest, she entered a small village in a small clearing. She told them that she was hungry and tired, and then asked if she could exchange a telling of her story for some food and a place to sleep. But the villagers knew that only evil came from the dense forest, so they threw garbage at her, and chased her in large numbers. She was so overcome by these people that she stumbled and fell into a great blaze just outside the village. There she went up in a dark cloud of smoke. This is always how her story ended, though, with her death in a fiery place. It seems that the myth and the miss had this end in common.

We say that a statement P is a *paradox* if it is neither True nor False. While it is True that *I am lying* is a paradox in the context of the conversation *I am a Cretan, All Cretans are Liars, I am lying*, what follows is a statement that is a paradox without being contained in such a conversation. That statement is called the *Liar's Paradox* or *Satan's Statement.*

Satan's Statement: This statement is False.

In the context of the conversation

All statements are False
This statement is False,

Satan's Statement is True. In the context of the conversation

The next statement is True
This statement is False

Satan's Statement is a Falsehood. These examples contrast the commonly held erroneous belief that Satan's Statement is False in every conversation that contains it. In Section 9.2 we will give a correct proof that shows that Satan's statement is neither True nor False, thus making it a True paradox.

Exercise 1.7.8 Using Q statements, prove that the statements 1 through 5 below are Falsehoods. Ignore simpler arguments if they should exist.

1. All statements are valid.

2. All men are liars.

3. Each statement is False.

4. All that is written is known.

5. All of Algebra in known.

6. Define: A perfect physicist is one who knows all of physics. Show that *There is a perfect physicist* is a Falsehood.

7. Define: A perfect mathematician is one who knows all of mathematics. Show that *There is a perfect mathematician* is a Falsehood.

Chapter 2

Sets

Logic is our alphabet, and sets will be the words in our language of mathematics. Some foundation has to be laid down, though, before any discussion in this language can proceed. Our foundation is *Set Theory and its functions*, and we will discuss these notions in the next two chapters.

There is one assumption that we will use implicitly and explicitly throughout this book. Our underlying assumption is that all mathematical objects considered in this book are from a Mathematical Universe in which these objects exist. Thus, when we say that $\aleph$ is a cardinal, it is to be understood that this cardinal lives in a Mathematical Universe and that we can examine it there. This Mathematical Universe is a classical idea attributed to the Greek philosopher and mathematician Plato. Therefore in making our universal assumption, we are following in a good classic tradition. The intent here is clear.

Our definition of *Set* requires us to know when an object is given. We will write *given* x if we wish to examine an object in our assumed Mathematical Universe. As you can see, our universal assumption goes to work right away. Whatever else you might believe, let us agree that this our assumed Mathematical Universe exists and that we can study the elements in it. Said assumption will not change as we work our way through this book.

2.1 Elements and Predicates

Mathematical statements are explicitly or implicitly about elements in *sets*. One of the highlights of twentieth century mathematics was to show that all of mathematics can be derived from the basic concepts in *Set Theory*. There is also strong empirical evidence that a firm grounding in Set Theory gives the reader the mathematical experience necessary to understand the advanced abstract ideas covered in later chapters.

We begin our introduction to modern mathematics with some work in *Set Theory*. However, the Elementary Set Theory that we will cover comes with a price. It can be terribly dry and uninspiring. We ask the reader to please be patient and work through the unions, intersections, and complements contained in this chapter. Your work will pay large dividends in our later discussions.

Definition 2.1.1 *A* set *is a collection A of objects called* elements *that satisfy the property*

Given x then either x is an element in A or x is not an element in A.

We will use the standard notation $x \in A$ when we want to say that *x is an element of A*. The notation $x \notin A$ is used when we want to say that *x is not an element of A*.

Observe that the defining condition of sets applies to objects that are *given* or that *exist* in our Mathematical Universe, and to nothing else. That must seem strange. But that little bit of philosophy must be addressed elsewhere. In this text, we will assume that everything we consider exists in our assumed Mathematical Universe. The use of the words *Given x* simply gives me a way of pointing out where that x is. This is perhaps the last time in this book that you will have the opportunity to think about this, so spend some time with it, and then read on.

Some examples of sets are in order. Consider the simplest possible set. The *empty set* is the collection that contains no elements. It is denoted by

$$\boxed{\emptyset.}$$

The empty set is indeed a set since given x then $x \notin \emptyset$. It arises in a surprising number of places in this book. For example, the set of all living people who will celebrate their -2nd birthday today is empty, as is the set of places in the universe whose temperature is measured at *absolute zero.* Chemists and physicists define absolute zero as that tempertaure at which all atomic motion stops, while Heisenberg's Uncertainty Principle tells us that this cannot be measured. The set of dogs with a PhD in mathematics is, at the time of this writing, also empty.

The simplest of sets in this book are listed according to their elements. Illustrations of how this is done include

1. $\{\bullet\}$, a set with the one element $\bullet$
2. $\{+,=,:\}$ the set whose elements are the symbols $+,=,:$
3. $\{\text{Lori}, \text{Christian}, \text{Marcus}, \text{Gwendolyn}\}$.

The set of *natural numbers*

$$\boxed{\mathbb{N} = \{0, 1, 2, 3, \ldots\}}$$

is given as an *implied list.* You might have called the elements of $\mathbb{N}$ *whole numbers* in the past, but the common professional designation for $x \in \mathbb{N}$ is that x is a *natural number.* The set of *integers*

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$

is also an implied list.

An *implied list* is an infinite list of elements. An ellipsis $\cdots$ is usually used to denote the fact that the elements in an implied list continue in the same manner or under the same pattern. It is hoped that each reader sees the same pattern when they get to $\cdots$. Implied lists can be trouble like that. While we would agree that

$$\{2, 4 \ldots\}$$

begins the set of even positive numbers (so that the next listed element is 6), we might also see that $\{2, 4, \cdots\}$ begins the powers of 2, so that the next element on the implied list would be $2^3 = 8$.

$$\{2, 4, 8, \ldots\}.$$

If we begin a set with numbers

$$2, 3, 5, x,$$

then what might the *next natural number* x be? Ponder this awhile before you read on.

You might answer that $5 = 2 + 3$ so $x = 3 + 5 = 8$. You might also expand on this formula and say that x is the sum of *all* of the previous numbers, $x = 2 + 3 + 5 = 10$. You might also see that $5 = 2 \cdot 3 - 1$ so that $x = 3 \cdot 5 - 1 = 14$. Or you might say that $2, 3, 5$ are the first three prime numbers so that $x = 7$, the fourth prime number.

In fact, there are infinitely values x that can be filled in the implied list while still correctly claiming to be the next logical element on the list. Just come up with an arithmetic relationship between $2, 3, 5$ that I have not used to this point, and you have a new value for x. You are invited to try to find some of these undiscovered values. Be sure to explain your claim that x is the *next natural number* on the list.

To avoid this kind of profusion of *next natural number*, we will be explicit in the pattern used in an implied list. As we read further, there will be no confusion as to the next value in a list. It is just interesting to know that this many numbers can claim to be *next.*

Sometimes the best way to denote a set is to use a *predicate* to describe the elements. A predicate is a description in words and symbols. For example, *blue* and *a temperature on Earth* are predicates.

Let P be any predicate. The symbols

$$A = \{x \mid x \text{ satisfies } P\}$$

are read *A is the set of elements x such that x satisfies P*. The symbol $\mid$ is read as *such that* or *with the property that.* For example, we would find it most difficult to give a complete list of the elements in the set $\{x \mid x \text{ is a person on the Earth}\}$. It is better by far to use the predicate to describe the set than to list the seven billion people on the Earth in 2011.

Implied lists are often better written as predicates.

$$\begin{aligned}
\mathbb{N} &= \{x \mid x \text{ is a natural number}\}, \\
\mathbb{Z} &= \{x \mid x \text{ is an integer}\}, \\
\mathbb{E} &= \{x \mid x \text{ is an even natural number}\}, \\
\mathbb{P} &= \{x \mid x \text{ is a prime number}\}.
\end{aligned}$$

We could not define the following sets of numbers without the use of the predicate form. Let

$$\boxed{\begin{aligned}
\mathbb{R} &= \{x \mid x \text{ is a real number}\}, \\
\mathbb{Q} &= \left\{\frac{n}{m} \mid m \neq 0 \text{ and } n, m \in \mathbb{Z}\right\}.
\end{aligned}}$$

$\mathbb{R}$ is called the set of *real numbers*, and $\mathbb{Q}$ is called the set of *rational numbers*. While we would have great difficulty in trying to list the elements in $\mathbb{R}$, (a mathematical impossibility that will be proved in Theorem 5.3.1), we can make an implied list of the rational numbers $\mathbb{Q}$. To illustrate this point, let us make an implied list of the set

$$\mathbb{Q}^+ = \left\{\frac{n}{m} \mid m \neq 0 \text{ and } n, m \in \mathbb{N}\right\}$$

of *positive rational numbers*. We begin the implied list by listing those rational numbers with numerator 1. The next row will be those rationals with numerator 2, and then I hope the pattern is clear. We have thus constructed the following implied list rational numbers.

$$
\begin{array}{ccccc}
\frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\
\frac{2}{1} & \frac{2}{2} & \frac{2}{3} & \frac{2}{4} & \cdots \\
\frac{3}{1} & \frac{3}{2} & \frac{3}{3} & \frac{3}{4} & \cdots \\
\frac{4}{1} & \frac{4}{2} & \frac{4}{3} & \frac{4}{4} & \cdots \\
\vdots & \vdots & \vdots & \vdots &
\end{array}
\tag{2.1}
$$

Observe that the rows of this infinite rectangular array list the positive rational numbers by increasing the denominator of the fraction and holding the numerator fixed. There certainly are repeated values on this implied list, aren't there? We could clean that up by deleting the rational numbers that are not in reduced form. The result is the implied list below.

$$
\begin{array}{ccccc}
\frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\
\frac{2}{1} & & \frac{2}{3} & & \cdots \\
\frac{3}{1} & \frac{3}{2} & & \frac{3}{4} & \cdots \\
\frac{4}{1} & & \frac{4}{3} & & \cdots \\
\vdots & \vdots & \vdots & \vdots &
\end{array}
\tag{2.2}
$$

Note that among others, the fractions

$$\frac{2}{2}, \frac{3}{3}, \frac{2}{4}, \frac{4}{2}$$

have been deleted from our list.

We will return to this list several times in this book. You might ask: Why is he doing that? Why would he present a list like that over and over again? The answer is that by referring to this list several times I hope to pique your curiosity. This list will be a friendlier mathematical object when next you see it.

Given a general mathematical collection X, like a set, there are several ways that mathematicians will examine X. We will define operations between X and sets like it. We will consider things Y in X that have the properties possessed by X. We will also compare sets X and Y by considering functions between them. In this section, we will consider operations on sets and subsets of sets.

We operate on sets in the following ways. Notice the use of the logic we learned in Chapter 1.

Definition 2.1.2 *Let A and B be sets.*

1. $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$.

2. $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$. *Our use of the word* or *is the* inclusive or, *so that $x \in A \cup B$ means at least* one *of the three inclusions*
$$x \in A, \quad x \in B, \quad x \in A \cap B.$$

Because this book is being directed at people with at least a high school education, the assumption is that you are familiar with the operations $\cap$ and $\cup$ for sets. For this reason, we will make a minimal number of concrete examples of what these operations can do.

Example 2.1.3 1. Let $A = \{\bullet, :\}$ and $B = \{\bullet, *\}$. Then $A \cap B = \{\bullet\}$, $A \cup B = \{\bullet, :, *\}$.

2. Let $A = \{x \in \mathbb{Z} \mid x \text{ is even}\}$ and let $B = \{x \in \mathbb{Z} \mid x \text{ is odd}\}$. Then $A \cup B = \mathbb{Z}$ and $A \cap B = \emptyset$.

Example 2.1.4 1. A natural number $p \geq 2$ is a *prime number* if it is divisible only by 1 and itself. Thus $2, 3, 5, 7$ are the first four prime numbers, while 9 and 111 are not prime.

To those who would ask why we do not include 1 as a prime number, let me say that a definition of prime number must encompass properties of primes that cannot be discussed in detail in a text like this. So just allow me to say that 1 is not a prime because a field must have 2 or more elements in it. A good reference for this fact is [10].

2. A natural number x is *composite* if it is the product of two or more primes. For instance, 4, 9, 12, and 111 are composite numbers, as are even numbers other than 2. But even though $2 = 1 \cdot 2$, 2 is not composite.

3. One can visualize prime numbers as follows. A positive natural number p is a prime number if given a set of p dots, $\bullet$, then the only rectangle that can formed from the given dots is a line of dots.

For example, 2 and 3 are prime because the only arrays that can be formed with 2 or 3 dots are

$$\bullet\bullet \quad \text{and} \quad \bullet\bullet\bullet$$

while 4 and 6 are not primes since we can draw *rectangles*

$$\begin{matrix}\bullet\bullet\\ \bullet\bullet\end{matrix} \qquad \begin{matrix}\bullet\bullet\bullet\\ \bullet\bullet\bullet\end{matrix}$$

One dot $\bullet$ does not denote a prime since two dots are required to form a line.

Example 2.1.5 If we let $\mathbb{E}$ denote the set of even numbers, and if $\mathbb{P}$ denotes the set of prime numbers, then $\mathbb{E} \cap \mathbb{P} = \{2\}$ since 2 is the only even prime. Also, $\mathbb{E} \cup \mathbb{P} = \{x \mid x \text{ is even or } x \text{ is prime}\}$. Some but not all of the elements in $\mathbb{E} \cup \mathbb{P}$ are

$$2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 101, 112.$$

The definition of *subset* represents a common technique in mathematics. Once we have defined an object X that satisfies some properties, then we can define a similar subobject in X that satisfies the properties of X. If we consider these subobjects as slices of X, then the subobjects of X can be used to study X.

Definition 2.1.6 *Let A and B be sets. We say that A is* a subset *of B if*

given $x \in A$ then we can show that $x \in B$.

We write

$$\boxed{A \subset B}$$

if A is a subset *of B.*

For instance, $\{3\} \subset \{1,2,3,4\}$ because 3 is an element of $\{1,2,3,4\}$, $\{2,3\} \subset \mathbb{P}$ because 2 and 3 are primes, and the set of children on Earth is a subset of the set of people on the Earth.

If $A \subset B$, then we also say that A *is contained in* B. The containments

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$$

are clear since each $x \in \mathbb{Z}$ can be written as $x = \dfrac{x}{1}$, and each fraction is a real number. Other containments are the set of your family contained in the set of people on Earth. The land making up the country of Peru is contained in the land that makes up the South American continent. College professors form a set that is a subset of the set of educated people on Earth, and the ages of people on this Earth in years forms a subset of the natural numbers $\mathbb{N}$.

A word of warning. When you are writing proofs, please do not take shortcuts. Use as much pencil lead, pen ink, and blank paper as possible when proving theorems. Fill in the details. Don't be afraid to use some paper. The trees that were used to manufacture that paper have since departed. You are not saving trees by using less paper in your mathematical deliberations. Get a feel for the flow of your argument. Let the mathematics guide your argument. It is unlikely that we can find a logically consistent and convincing proof by writing down an argument with many gaps. Your hard work will be rewarded when you can follow some of the subtler arguments presented in later chapters. For now, to give you a running start in your reading, we present our arguments in all their baroque detail.

The identities in the next theorem are not hard. By including the proofs, we are improving the reader's intuition of what constitutes a proof. This could be the first mathematics you have ever met that did not reduce to a formula or that did not require you to plug in some kind of numbers.

I therefore recommend that you carefully read the following theorem as it consists of elementary examples. For instance, to show that $A \subset B$, we start with an element $x \in A$ and then show that $x \in B$. There is no formula that will produce the proof, and there is no shorter way to do this.

Theorem 2.1.7 *Let A, B, and C be sets.*

1. $A \subset A$.

2. $A \cap B \subset A$.

3. *If $A \subset B$ and if $B \subset C$ then $A \subset C$.*

Proof: 1. Proofs like this are mathematical double-talk. $A \subset A$ because given $x \in A$, then $x \in A$.

2. As with every proof that attempts to prove $X \subset Y$, we start with an element of X and show that it is an element of Y. So let $x \in A \cap B$. By the definition of $A \cap B$, $x \in A$ and $x \in B$. Specifically, $x \in A$. Then $A \cap B \subset A$ by the definition of subset.

3. Suppose that $A \subset B$ and that $B \subset C$. Let $x \in A$. We must show that $x \in C$. Using the definition of subset, $x \in A$ and $A \subset B$ implies that $x \in B$. Also, $x \in B$ and $B \subset C$ implies that $x \in C$. We have just proved that

$$x \in A \text{ implies that } x \in C$$

so we can conclude that $A \subset C$. This completes the proof.

Admittedly we have included a great deal of detail in these proofs. We do this to make a point. When you construct a proof, you are trying to tell the reader what you are thinking, so in your argument, every reason for the given steps must be explained. Since the reader does not know your mind, even the simplest of arguments can be confusing to them. Furthermore, the ideas flow one to the next and *every statement in the proof is accompanied by a reason.* This flow helps the reader follow the argument. A jumble of loosely connected statements will not form a convincing argument.

The next type of argument, called a proof by contradiction, can be kicked around and explored. It is the first time since Chapter 1 you have read such an argument. Let us review how a proof by contradiction proceeds.

Suppose that we want to prove that P is a True statement and that we wish to prove that Q is True. The *proof by contradiction* begins with the assumption that P is a False statement. We then follow our logical noses until we arrive at a mathematical mistake called a *contradiction*. This contradiction shows us that we began our proof with a False statement. That is, it is False that P is False. Therefore, P must be a True statement.

The rest of this book contains many proofs by contradiction, so try to work your way through this one.

Theorem 2.1.8 *Let A be a set. Then $\emptyset \subset A$.*

Proof: Suppose for the sake of contradiction that $\emptyset \subset A$ is a False statement. Then $\emptyset \not\subset A$. Consider for a moment what it means when $B \not\subset A$. It means that some element of B is not in A. Since $\emptyset \not\subset A$, there is some element $x \in \emptyset$ such that $x \notin A$. But $x \in \emptyset$ contradicts the fact that $\emptyset$ has no elements. This is the desired contradiction, the mathematical mistake. We conclude that our initial assumption $\emptyset \not\subset A$ is False. Therefore, $\emptyset \subset A$, and this completes the proof.

2.2 Equality

Have you ever wondered what it means for objects x and y to be equal, $x = y$? If x and y are numbers then we have an intuitive understanding of what it means for $x = y$. As an illustration of the weakness in your intuition, would you have recognized the following equations on your own?

$$\begin{aligned} \frac{1}{2} &= .5 \\ &= .4\overline{999} \\ &= \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \cdots . \end{aligned}$$

Equality is not always such a simple concept. For our purposes we will accept that *if we are given objects x and y, there is an understanding of the identification $x = y$.*

The equality of sets is not intuitive at all. It is mathematically definite and mathematically rigorous. What we mean by this is that if we are given sets A and B, there is a precise definition of equality, $A = B$, of A and B.

Definition 2.2.1 *Let A and B be sets. We say that A* equals *B if $A \subset B$ and $B \subset A$. The traditional notation is*

$$\boxed{A = B.}$$

Thus, to prove that $A = B$, we must first prove that $A \subset B$ and then that $B \subset A$. There are no shortcuts in this kind of proof. There are two things to show, and we must be prepared to prove them. Some examples will show you what we mean.

Example 2.2.2 If we let

$$\mathbb{F} = \text{the set of finite or repeating decimal numbers}$$

then we can prove that

$$\mathbb{Q} = \mathbb{F}$$

as follows. To prove that $\mathbb{Q} \subset \mathbb{F}$, let $x \in \mathbb{Q}$. Then x is a fraction and your training in converting fractions into decimals makes x a finite or a repeating decimal. Thus, $\mathbb{Q} \subset \mathbb{F}$.

Conversely, we show that $\mathbb{F} \subset \mathbb{Q}$. Let $x \in \mathbb{F}$. You may remember some training in arithmetic that shows you how to convert the finite or repeating decimal x into a fraction. We will not pursue that arithmetic here. Then $x \in \mathbb{Q}$ and we conclude that $\mathbb{F} = \mathbb{Q}$.

Example 2.2.3 Let

$$A = \{x \mid x \text{ is a solution to } x^2 - 1 = 0\}.$$

We will show that

$$A = \{-1, 1\}.$$

The numbers -1 and 1 are solutions to $x^2 - 1 = 0$ since

$$(-1)^2 - 1 = 0 \text{ and } 1^2 - 1 = 0.$$

It follows that $\{-1, 1\} \subset A$. Conversely, let us suppose that $x \in A$. Then x is a real number such that $x^2 - 1 = 0$. Our high school algebra gives us the steps used to solve the equation $x^2 - 1 = 0$.

$$\begin{aligned} x^2 - 1 &= 0, \\ (x-1)(x+1) &= 0, \\ x - 1 = 0 \qquad & x + 1 = 0, \\ x = 1 \qquad & x = -1. \end{aligned}$$

Then $A \subset \{-1, 1\}$, from which we conclude that $\{-1, 1\} = A$.

Other examples come from the operations $\cup$ and $\cap$ on sets A, B, and C.

Theorem 2.2.4 *Let A, B, and C be sets.*

1. $A \cup B = B \cup A$.

2. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Proof: 1. We begin by showing that $A \cup B \subset B \cup A$. The reader will note the interplay between the definitions of $\cup$ and $\subset$. Because of its elementary nature, the proof is going to sound like mathematical doublespeak.

Let $x \in A \cup B$. By definition of $\cup$, $x \in A$ or $x \in B$, so that $x \in B$ or $x \in A$. By the definition of $\cup$, $x \in B \cup A$, and hence, $A \cup B \subset B \cup A$.

Conversely, we must show that $B \cup A \subset A \cup A$. Let $x \in B \cup A$. Then $x \in B$ or $x \in A$ so that $x \in A$ or $x \in B$. Thus, by definition of $\cup$, $x \in A \cup B$, and hence, $B \cup A \subset A \cup B$.

Therefore, $A \cup B = B \cup A$.

2. The proof is an exercise in the language that interprets commas and conjunctions. It is left as an exercise for the reader.

Now that was deadly dull. You might complain that the whole idea of the proof was to say that P or Q is the same as saying Q or P, and you'd be right. This is a tedious example of what it means to prove that two sets are equal. But the essence of the proof is the important thing. The introduction of all that detail is an attempt to get you to fill in more detail when you think and write.

People tend to make intellectual leaps when they argue. I suppose that the person arguing feels that he can fill in the detail later or perhaps the arguer feels that the detail is unimportant. When original arguments avoid the details, they are usually hard to follow and they are often incorrect. Therefore, at this stage of our mathematical development, we cannot allow ourselves the luxury of gaps in our train of thought when we argue.

Think of a mathematical argument as a maze in which you want to get from point A to point B. There are many paths we can choose. Because this is a maze, we cannot lift our pencils off the paper, thus ignoring the boundaries of the paths, and then drop the pencil point on B. That would be contrary to the spirit of playing with mazes. Many paths from A lead to a blind alley. We cannot find B down these paths. However, we can learn something from these blind alleys. They indicate the paths that we should avoid in trying to get to B. Perhaps there is some general principle that we can learn from these blind alleys. Once learned, they inevitably lead us to B along a logical path.

Remark 2.2.5 Let us apply the maze analogy to the construction of a proof. We know what we have to start with (the hypotheses), and we know where we want to finish (the conclusion). Try repeatedly to construct a logical path from the hypotheses to the desired conclusion. Don't give up. Keep at it. Refine your argument but do not leave out essential details.

Mathematical proofs do not come cheaply. They require a good deal of hard work. Eventually, we will find that path from the hypotheses to the conclusion. Then, if we apply what we have learned about the problem, we will rewrite the proof and perhaps find an elegant or beautiful path through the logical maze from the point of origin represented by the hypotheses to the ending point represented by the conclusion.

We want to investigate what it means when an element x *is not in* a set A, $x \notin A$. This is also a good opportunity to stretch our logical experience with the word *not*.

Let me introduce a new notion here. Sometimes we will find that several sets A, B, and perhaps C are contained in one larger set $\mathcal{U}$. This set $\mathcal{U}$ plays the role of the universe for A, B, C in that all operations on A, B, and C are relative to $\mathcal{U}$. For this reason, $\mathcal{U}$ is called a *universal set.* For example, the set of real numbers $\mathbb{R}$ is a universal set for the natural numbers $\mathbb{N}$ and the rational numbers $\mathbb{Q}$. On the other hand, $\mathbb{N}$ is a universal set for $\emptyset$, $\{1\}$, and $\{2, 4, 6, \cdots\}$.

Definition 2.2.6 *Let A and B be sets in some universal set $\mathcal{U}$. Then*

$$\boxed{A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}}$$

and

$$\boxed{A' \;=\; \{x \,|\, x \in \mathcal{U} \text{ and } x \notin A\}.}$$

We say that A' is the complement of A.

Notice the use of logic in defining these sets. Intuitively, to form $A \setminus B$, we take the elements of A and we throw out the elements of B. To form A', we take those elements in the universal set $\mathcal{U}$ that are not in A.

Some examples will help us see what we are talking about.

Example 2.2.7 Let us begin by letting $A = \{a, b, c, d, e\}$, $B = \{c, d, e, f, g\}$, and $C = \{a, b, c, d, e, f, g, h\}$.

1. Because $A \setminus B$ is the set of elements in A that are not in B, we have

$$A \setminus B = \{a, b\}$$

as the reader can easily verify.

2. Since $A \subset C$,

$$A \setminus C = \emptyset.$$

This may come as somewhat of a surprise. We will work on this example by removing one element at a time. I do not recommend this approach in general. We throw out a since $a \in A$ and $a \in C$. We throw out b since $b \in A$ and $b \in C$. We throw out c since $c \in A$ and $c \in C$. Keep going until you run out of elements in A to be tested. Then $A \setminus C$ has no elements, or in set notation, $A \setminus C = \emptyset$.

3. Let us find $C \setminus B$. Write down the elements of C that are not in B, thus deleting the elements of C that are in B. The result is

$$C \setminus B = \{a, b, h\}.$$

Example 2.2.8 Let $\mathcal{P}$ be the set of people on Earth, and let A be the set of male people. Then $\mathcal{P} \setminus A$ is the set of female people. This should be enough examples of a finite nature.

Example 2.2.9 1. Recall that $\mathbb{E}$ = the set of even integers, and let $\mathcal{U} = \mathbb{N}$. Then $\mathbb{E}'$ is the set of integers that are not even. That is, $\mathbb{E}'$ is the set of odd numbers. If $\mathbb{T}$ is the set of natural numbers that are divisible by 3, then $\mathbb{T}'$ is the set of natural numbers not divisible by 3. That is, $\mathbb{T}'$ is the set of natural numbers that have a nonzero remainder when divided by 3.

2. Recall that $\mathbb{P}$ = the set of prime numbers, and let $\mathcal{U} = \mathbb{N}$. Because a prime cannot be properly factored, and while composite numbers can be properly factored, $\mathbb{P}'$ is the set whose elements are 0 and 1, and the composite numbers. A number > 1 not in $\mathbb{P}$ but in $\mathbb{N}$ is divisible by at least two numbers a, $b \neq 1$. Then such a number is a composite number, and hence, the elements of $\mathbb{P}'$ are 0, 1, and the composite numbers.

Example 2.2.10 Let $\mathbb{Q}^-$ be the set of negative rational numbers,

$$\begin{aligned}\mathbb{Q}^- &= \left\{-\frac{n}{m} \,\middle|\, n, m \neq 0 \in \mathbb{N}\right\} \\ &= \{-q \mid q \in \mathbb{Q}^+\}.\end{aligned}$$

Let $\mathcal{U} = \mathbb{Q}$ and let $A = \mathbb{Q}^- \cup \{0\}$. Then A' is the set $\mathbb{Q}^+$ of positive rational numbers.

Another example is found by setting $\mathcal{U} = \{x \in \mathbb{R} \mid x \geq 0\}$ = the set of *nonnegative real numbers*. Then $\mathbb{Q}^+ \subset \mathcal{U}$ and $(\mathbb{Q}^+)'$ is the set of *nonnegative irrational numbers.*

Here's a challenge for the reader. Without peeking below, try to describe the symbols $x \in (A')'$ using words. This might help you see the power of this symbolism over the language in dealing with double negatives. As you work on this exercise, note the interplay amongst the symbols x, A, $()'$, and the word *not.*

The next theorem is another example of a *proof by contradiction.*

Theorem 2.2.11 *Assume that some universal set $\mathcal{U}$ contains the set A. Then $A \cap A' = \emptyset$.*

Proof: We begin by assuming for the sake of contradiction that $A \cap A' \neq \emptyset$. We now seek that mathematical mistake, that contradiction. Then $A \cap A'$ must contain an element, say, $x \in A \cap A'$. By the definition of $\cap$, $x \in A$ and $x \in A'$ or equivalently $x \in A$ and $x \notin A$. This conclusion contradicts the fact that A is a set. (Reread Definition 2.1.1 to see that if A is a set, then $x \in A$ or $x \notin A$ but not both.) It then follows that our initial assumption is False. That is, $A \cap A' = \emptyset$, which completes the proof.

Example 2.2.12 At this point, I hope you have convinced yourself that

$$(A')' = A$$

for each set A. Here is how I hope you argued. Let $x \in A$. By the definition of A', $x \notin A'$, and hence x is in the complement of A'. The complement of A' is $(A')'$, so that $x \in (A')'$, and so $A \subset (A')'$.

On the other hand, suppose that $x \in (A')'$. Then by definition of complement, $x \notin A'$. Either $x \in A$ or $x \notin A$. Assume to the contrary that $x \notin A$. Then $x \in A'$, contrary to $x \notin A'$. Consequently, $x \in A$, and hence, $(A')' \subset A$. Therefore, $A = (A')'$, which completes the proof.

These two proofs were instructive exercises in mathematical double talk. The level of detail I used was strictly an educational device. No professional includes the attention to minutiae that we have given here. However, in an introductory work like the present one, it is necessary for us to observe this obsessive level of detail in a proof. This level of detail allows you very little freedom of thought when you read it, and that is the key. By giving this level of detail, you and I are certain to be thinking the same thing, thus giving me knowledge of how much further I can push the level of understanding.

To hone the reader's writing skills, we present a series of elementary exercises. The reader is encouraged to try these exercises and to practice the completeness and clarity of a rational argument that accompanies a well-constructed proof.

Exercise 2.2.13 Let A, B, and C be sets contained in some larger universal set $\mathcal{U}$.

1. $A \cap B \subset A \cup B$.

2. $A \cap B = B \cap A$.

3. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

4. $A \cup A' = \mathcal{U}$.

5. **DeMorgan's Law**

   (a) $(A \cap B)' = A' \cup B'$.

(b) $(A \cup B)' = A' \cap B'$.

6. $(A \setminus B) \cap B = \emptyset$.
7. $(A \setminus B) \cap (B \setminus A) = \emptyset$.
8. $(A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$.

2.3 Cartesian Products

The next operation on sets allows us to make a larger dimensional set from a number of smaller dimensional sets.

Definition 2.3.1 *Let A and B be sets. The* Cartesian product *of A and B is the set of pairs*

$$\boxed{A \times B \quad = \quad \{(x,y) \mid x \in A \text{ and } y \in B\}.}$$

To form $A \times B$, we take the elements of A and B and pair them up. For example, if we let $A = \{a,b,c\}$ and let $B = \{x,y\}$ then

$$A \times B = \{(a,x),(b,x),(c,x),(a,y),(b,y),(c,y)\}.$$

It is natural to picture $A \times B$ as a lattice of points in two dimensions.

$$\begin{array}{ccccc} & & & & (c,y) \\ y & — & \bullet & \bullet & \bullet \\ & & & (b,x) & \\ x & — & \bullet & \bullet & \bullet \\ & & | & | & | \\ & & a & b & c \end{array} \tag{2.3}$$

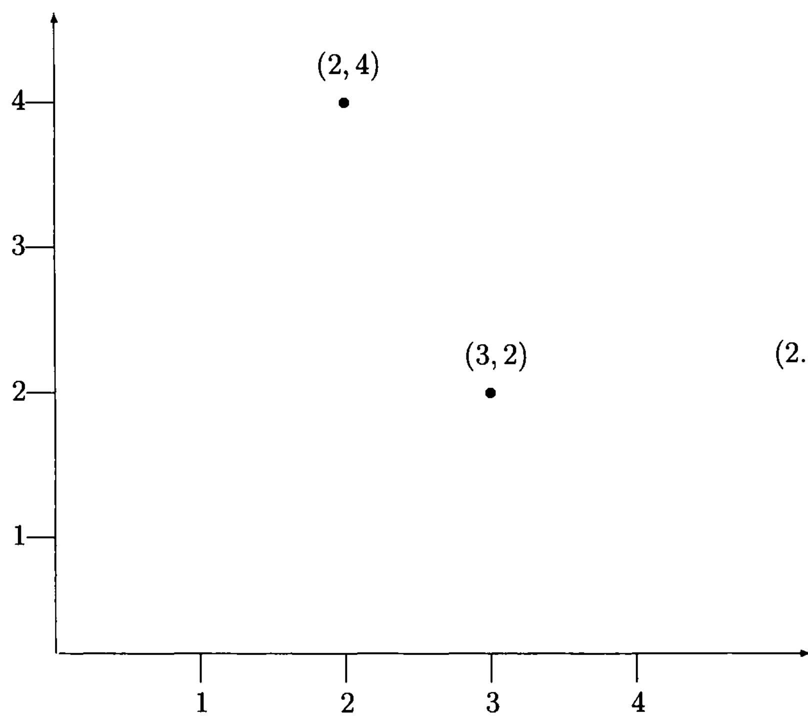
The *Cartesian plane*

$$\boxed{\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}}$$

is the plane equipped with x and y axes with which you are familiar. This is the plane on which you graphed points, lines, and parabolas in an early high school Algebra course.

The points in $\mathbb{R}^2$ are written as (x, y), and they specify a set of directions to be followed in locating the physical point (x, y) in $\mathbb{R}^2$. These directions work like this. Beginning at the origin $(0, 0)$, move x units in the horizontal, or x, direction. This is followed by a movement of y units in the vertical, or y, direction. You are now at the point (x, y).

For instance, the point labeled by $(2, 4)$ is found by moving 2 units in the x direction and then 4 units in the y direction. Here are a few more examples.

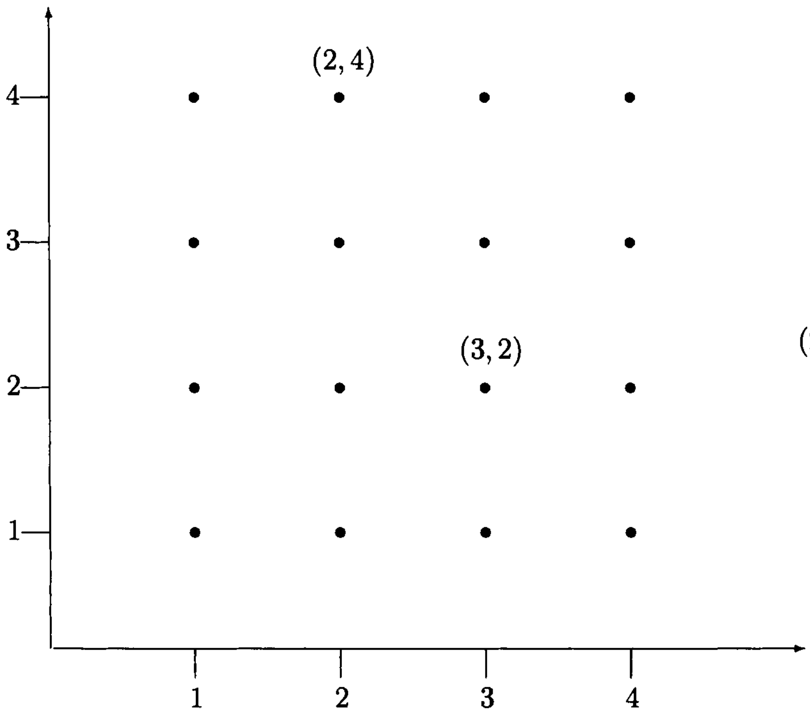


(2.4)

The Cartesian Plane $\mathbb{R} \times \mathbb{R}$

This is the reason that the Cartesian plane $\mathbb{R}^2$ and the Cartesian product $\mathbb{R} \times \mathbb{R}$ are equal. *Each point in the Cartesian plane can be represented by a unique pair (x, y) of real numbers.*

We then have a picture that we can use to diagram a number of Cartesian products. Such an example is *the pairs of natural numbers* $\mathbb{N} \times \mathbb{N}$.



(2.5)

The Array of Natural Numbers $\mathbb{N} \times \mathbb{N}$

$$\boxed{\mathbb{N} \times \mathbb{N} \quad = \quad \{(n,m) \,|\, n,m \in \mathbb{N}\}.}$$

Here is how we picture $\mathbb{N} \times \mathbb{N}$ as a subset of the Cartesian plane. This array extends indefinitely upward and to the right. Then $\mathbb{N} \times \mathbb{N}$ contains $(1,1)$ and $(101,102)$, but it does not contain $(1/2,3)$ or $(\sqrt{2},0)$. We will use an array similar to this one when we count the set

$$\boxed{\mathbb{Q}^+ = \{x \in \mathbb{Q} \,|\, x > 0\}.}$$

The next result may strike you as a bit odd. It is another example of the care with which the empty set must be treated. The proof is one by contradiction. Briefly, $A \times \emptyset$ is empty because it does not contain any *pairs.*

Theorem 2.3.2

$$\boxed{A \times \emptyset \;=\; \emptyset.}$$

for any set A.

Proof: Assume for the sake of contradiction that there is an element $p \in A \times \emptyset$. Then $p = (x, y)$ for some $x \in A$ and $y \in \emptyset$. This contradicts the fact that $\emptyset$ has no elements. Thus, our initial assumption is False, and it follows that $A \times \emptyset = \emptyset$. This concludes the proof.

For instance, $\mathbb{N} \times \emptyset$, $\mathbb{R} \times \emptyset$, and, for sets A, $\emptyset \times A$ are the empty set. You can understand these mind stretchers by thinking of multiplication by 0. Given any number x, $x \cdot 0 = 0$ and in much the same way, $A \times \emptyset = \emptyset$ for sets A.

This is not the last time we will make an analogy between operations on sets and operations on numbers. In the later portion of this book, we make the analogy precise.

Exercise 2.3.3 1. Draw a rectangle of dots that represents $\{a, b, c, d\} \times \{u, v, w, x\}$.

2. Prove that $\emptyset \times \emptyset = \emptyset$.

3. Prove that $A = \{H, T\}$ and that $B = \{1, 2, 3, 4, 5, 6\}$. Find $(A \times \{6\}) \cap (\{H\} \times B)$.

2.4 Power Sets

Let A be a set. The reader is acquainted with the idea of a subset of A. We will take that idea to a new level by introducing *the set of all subsets of* A. This power set will be an important device in the later chapters of this book.

Definition 2.4.1 *Let A be a set. The* power set *of A is the set of all subsets of A. We write*

$$\boxed{\mathcal{P}(A) \quad = \quad \{\text{sets } X \,|\, X \subset A\}.}$$

Seeing a set as an element in a larger set can be done, but it takes practice. Also, it may be hard to accept at first that *the elements of $\mathcal{P}(A)$ are themselves sets.* Some examples will be useful. Since power sets grow quickly, our examples will of necessity be small.

If $A = \{w, x, y, z\}$ then $\mathcal{P}(A)$ contains *as elements* the sets $\emptyset$, $\{w, z\}$, $\{x, y, z\}$, and 13 more sets. See if you can find them. If we let

$$\boxed{\mathbb{R}^+ = \{x \in \mathbb{R} \,|\, x > 0\}}$$

denote the set of *positive real numbers*, then the power set $\mathcal{P}(\mathbb{R}^+)$ of $\mathbb{R}^+$ contains *as an element* the set $\mathbb{Q}^+$. $\mathcal{P}(\mathbb{R}^+)$ also contains the element $\{1, 2, 3\}$ and the element $\{\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{6}, \ldots\}$. (What is the next element on the last implied list, reader?)

Since $\emptyset$ and A are always subsets of A, $\mathcal{P}(A)$ contains $\emptyset$ and A.

$$\boxed{\emptyset, A \in \mathcal{P}(A) \quad \text{for any set } A.}$$

For example,

$$\mathcal{P}(\{\bullet\}) = \{\emptyset, \{\bullet\}\}$$

and

$$\mathcal{P}(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}.$$

Given $A = \{a, b, c\}$, the eight elements in $\mathcal{P}(A)$ will include $\emptyset$, $\{b\}$, $\{a, c\}$, and $\{a, b, c\}$.

Because $\emptyset$ is the only subset of $\emptyset$ we have the equation

$$\boxed{\mathcal{P}(\emptyset) \quad = \quad \{\emptyset\}.}$$

The importance of the power set is that it allows us to construct a much larger set from a smaller one. In general, if A is a finite set with exactly n elements then

$$\boxed{\mathcal{P}(A) \text{ contains exactly } 2^n \text{ elements.}}$$

Thus $\mathcal{P}(\{a,b\})$ has $2^2 = 4$ elements while $\mathcal{P}(\{a,b,c\})$ has exactly $2^3 = 8$ elements. Try writing down the 8 sets that comprise $\mathcal{P}(\{a, b, c\})$.

In the case where A is an infinite set, we can ask for a self-contained method for describing the elements of A. For example, $\mathcal{P}(\mathbb{N})$ contains *as elements* the set $\{1\}$, the set $\mathbb{E} = \{2, 4, 6, \ldots\}$ of even numbers, the set $\mathbb{P}$ of prime numbers, and $\mathbb{N}$. But when we try to make a list of the elements of $\mathcal{P}(\mathbb{N})$ we are met with the following deep result of mathematics.

$$\boxed{\text{We cannot make an implied list of the elements in } \mathcal{P}(\mathbb{N}).}$$

Any list that we make of $\mathcal{P}(\mathbb{N})$ will miss some element of $\mathcal{P}(\mathbb{N})$, and any correction we attempt to make will still result in an incomplete list of the elements of $\mathcal{P}(\mathbb{N})$. We will have much more to say about this mathematical mystery in later chapters.

Exercise 2.4.2 1. Find $\mathcal{P}(A)$ when $A = \{a, b, c\}$. How many sets did you find?

2. Find $\mathcal{P}(A)$ when $A = \{a, b, c, d\}$. How many sets did you find?

3. Find the elements of $\mathcal{P}(A)$ that have exactly 3 elements when $A = \{a, b, c, d, e\}$. How many sets did you find?

2.5 Something From Nothing

Here is an application of sets that appears infrequently in the popular press. Although humans have used the natural numbers since they could herd sheep, most of us are unaware of the mathematically precise construction of, say, 1. We will use sets and subsets to construct the first few natural numbers. The process with which we start is easily extended to a construction of $\mathbb{N}$ from nothing. For this we must initiate a discussion of *cardinality* that will be improved on throughout the chapters of this book.

We give a superficial and *mathematically imprecise* definition of counting that can be understood without a significant mathematical background. In the next chapter, we will present the mathematical background needed to make a precise accounting of the counting process. Let A be a finite set. The *cardinality of A* is defined to be

$$\boxed{\begin{array}{rcl}\text{card}(A) & = & \text{all the sets } X \text{ that can be matched} \\ & & \text{element by element with } A.\end{array}}$$

When treating card(A), we must be careful because card(A) *is not a set*. It is proper to call it something else like a collection or a class. But card(A) is just too big to be a set. We will avoid the issue of what card(A) is by carefully avoiding set theoretic issues surrounding collections and classes.

We will use card(A) as a precise way of counting the number of elements in A. That is why we need a precise understanding of how to *match element by element* the elements of X and A. On the one hand, we have a word *cardinality* that you should interpret as meaning *the number of elements in* a set. On the other hand, we have a notation card(A) that represents all sets that have the same number of elements as A. Some examples will help you visualize what we mean by this.

For the set

$$\{\bullet\}$$

the set card$(\{\bullet\})$ consists of all sets that can be matched element by element with $\{\bullet\}$. This is a simple thing to see, mostly because the set $\{\bullet\}$ is so small. The sets $\{a\}, \{*\}, \{1\}$ are in card$(\{\bullet\})$ because they can be matched element by element with $\{\bullet\}$. The matching between $\{a\}$ and $\{\bullet\}$ is obvious enough. This matching is

$$a \longmapsto \bullet.$$

Similarly, $\mathrm{card}(\{a,b\})$ consists of the sets that can be matched element by element with $\{a,b\}$. The sets $\{x,y\},\{0,1\},\{H,W\}$ are in $\mathrm{card}(\{a,b\})$. One matching between $\{0,1\}$ and $\{a,b\}$ is

$$\begin{cases} 0 & \longmapsto & a \\ 1 & \longmapsto & b. \end{cases}$$

There are others, and you are invited to write them down, but we only need one. At this point I hope that the pattern is clear to you. The set $\{a,b,c\}$ is in $\mathrm{card}(\{1,2,3\})$ because there is a matching

$$\begin{cases} a & \longmapsto & 3 \\ b & \longmapsto & 2 \\ c & \longmapsto & 1. \end{cases}$$

Notice that we didn't need to pick the one you were thinking of. Any matching of elements will do. We just chose a different one.

Let us continue the discussion by looking at the set without elements, $\emptyset$. We have

$$\boxed{\mathrm{card}(\emptyset) = \text{all sets that have no elements.}}$$

Inasmuch as the notation $\mathrm{card}(\emptyset)$ is a bit cumbersome, we will use the symbol 0 to denote $\mathrm{card}(\emptyset)$.

$$\boxed{0 = \mathrm{card}(\emptyset).}$$

This notation is perfect. The identification of $\mathrm{card}(\emptyset)$ with the traditional symbol 0 makes perfect sense. We want to choose a symbol for $\mathrm{card}(\emptyset)$ that will not confuse your mathematical sensibilities. The cardinality of $\emptyset$ is denoted by 0 because we say so and for no other reason. But the use of 0 is compelling. It makes you think about the right ideas, the right quantities, and the right magnitudes. We could just as easily have decided that $\mathrm{card}(\emptyset)$ should be denoted by Z. But since the world has been using 0 for a similar purpose for centuries, we will continue the tradition. We have thus *constructed the natural number 0 from nothing.*

Next we construct the number 1. Take the power set $\mathcal{P}(\emptyset)$ of $\emptyset$. $\mathcal{P}(\emptyset)$ is the set of all subsets of $\emptyset$. Thus, $A \in \mathcal{P}(\emptyset)$ exactly when $A \subset \emptyset$. But since $\emptyset$ has no elements, $A \subset \emptyset$ implies that $A = \emptyset$. We conclude that the only element of $\mathcal{P}(\emptyset)$ is $\emptyset$, or equivalently that

$$\boxed{\mathcal{P}(\emptyset) = \{\emptyset\}.}$$

Observe that $\mathcal{P}(\emptyset)$ is not the empty set since it contains an element, namely, $\emptyset$. Initially, you may not like thinking of $\emptyset$ as an element in a set, so keep trying. Since it has been used in a similar way for ages, we will use the symbol 1 to denote the cardinality of $\mathcal{P}(\emptyset)$.

$$\boxed{1 = \mathrm{card}(\{\emptyset\}).}$$

This is in perfect agreement with the ages old use for 1, isn't it? Before you read this book, if I asked you to count the number of elements in the set $\{\emptyset\}$, you would most certainly say *One*. So this notation agrees with your intuition about 1, with your experiences, and with your senses. It is the idea of *defining* 1 in this way that may unsettle your inner moral mathematical compass. But since this use is not in conflict with the rest of society, then why not make the identification. Nothing spiritual will be harmed. What we identify and what we believe are the same thing.

Okay, let's define the natural number 2 in such a way that it does not conflict with our education of what 2 means. I will respect your difficulty in seeing $\emptyset$ as an element in a set, and we will for the moment look at $\{\bullet\}$. We will see together that

$$\boxed{\mathcal{P}(\{\bullet\}) = \{\emptyset, \{\bullet\}\}.}$$

Write down a subset $A \subset \{\bullet\}$. Since A is a set either $\bullet \in A$ or $\bullet \notin A$, there can be no other cases to consider. Because $\bullet$ is the only element of $\{\bullet\}$, either $A = \{\bullet\}$ or A is empty, $A = \emptyset$. Thus, as we predicted, $\mathcal{P}(\{\bullet\}) = \{\emptyset, \{\bullet\}\}$.

Now force yourself to see $\emptyset$ as the lone element of $\{\emptyset\}$. Convince yourself that $\{\emptyset\}$ is the set whose only element is $\emptyset$. The substitution of $\bullet = \emptyset$ yields the power set of $\{\emptyset\}$.

$$\boxed{\mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}.}$$

We have constructed 2 once we have defined

$$\boxed{2 = \text{card}(\{\emptyset, \{\emptyset\}\}).}$$

This definition of 2 agrees with everything you learned as a child. There are just enough elements to make perfect sense for this use of the *symbol* 2. You can think of 2 as card($\{\emptyset\}$), or you can think of it as a number you learned long ago. It doesn't matter. They represent the same idea, don't they? That is the beauty of this discussion. We are constructing natural numbers in an entirely precise mathematical format, and without losing the traditional meaning of the symbols we are using.

Next, construct 4. We will be careful not to upset our mathematical traditions concerning the number or symbol 4. Use bullets for $\emptyset$ if you need to, but see the set $\{\emptyset, \{\emptyset\}\}$ as a set with two elements. We will examine the subset structure of $\{x, y\}$ and then restrict our attention to $\{\emptyset, \{\emptyset\}\}$. This is the last time we will do this, so get comfy.

Let $A \subset \{x, y\}$. If A has no elements, then $A = \emptyset$. If A has exactly one element, then $A = \{x\}$ or $A = \{y\}$. If A has other than one element, then $A = \{x, y\}$. There are no other choices since all of the elements of $\{x, y\}$ are exhausted at this point. We then write

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

If we replace x by $\emptyset$ and y by $\{\emptyset\}$ then we have written the power set of $\{\emptyset,\ \{\emptyset\}\}$.

$$\boxed{\mathcal{P}(\{\emptyset,\ \{\emptyset\}\}) = \{\emptyset,\ \{\emptyset\},\ \{\{\emptyset\}\},\ \{\emptyset, \{\emptyset\}\}\}.}$$

We define the number 4 to be the cardinality of $\mathcal{P}(\{\emptyset, \{\emptyset\}\})$. Thus,

$$\boxed{4 = \text{card}(\{\emptyset,\ \{\emptyset\},\ \{\{\emptyset\}\},\ \{\emptyset, \{\emptyset\}\}\}),}$$

which agrees with our previous uses of the symbol 4. *We have thus constructed the natural numbers* $0, 1, 2, 4$.

Wait, there is another way to see this. We observe that

$$\boxed{\begin{array}{rcl} \emptyset & = & \{\} \\ \mathcal{P}(\emptyset) & = & \{\emptyset\} \\ \mathcal{P}(\mathcal{P}(\emptyset)) & = & \mathcal{P}(\{\emptyset\}) \;=\; \{\emptyset, \{\emptyset\}\} \\ \mathcal{P}(\mathcal{P}(\mathcal{P}(\emptyset))) & = & \mathcal{P}(\mathcal{P}(\{\emptyset\})) \;=\; \mathcal{P}(\{\emptyset, \{\emptyset\}\}) \\ & = & \{\emptyset,\ \{\emptyset\},\ \{\{\emptyset\}\},\ \{\emptyset, \{\emptyset\}\}\} \end{array}}$$

From this chart, we can see that

$$\boxed{\begin{array}{rcl} 0 & = & \text{card}(\emptyset) \\ 1 & = & \text{card}(\mathcal{P}(\emptyset)) \\ 2 & = & \text{card}(\mathcal{P}(\mathcal{P}(\emptyset))) \\ 4 & = & \text{card}(\mathcal{P}(\mathcal{P}(\mathcal{P}(\emptyset)))) \end{array}}$$

This last list shows us that the numbers 0, 1, 2, 4 can be realized as an iterated application of the power set operation $\mathcal{P}(\cdot)$ by starting with $\emptyset$. So in a real sense we have constructed the numbers 0,1,2,4 from nothing.

I leave it to you to construct the natural number 3 from the number 4 in the spirit of the above construction. However, 3 will not be $\mathcal{P}(A)$ for any set A.

Exercise 2.5.1 1. Construct 3 in the manner of this section.

2. Construct 5 in the manner of this section.

3. Construct 8 in the manner of this section.

2.6 Indexed Families of Sets

To this point we have only considered $\cup$, $\cap$, and $()'$ for sets A and B. These were used to make us comfortable with the logical properties of the words *and*, *or*, and *not*. In this section, we will work with $\cup$, $\cap$, and $()'$ for *families* of sets. To make our discussion as accessible as possible, we will work only with *finite* or *countable families* of sets $\{A_1, A_2, A_3, \ldots\}$. These countable families will prove to be enough of a mind expander for this book.

The smallest countable families of sets are the finite families. Then $\{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$ is a finite family of sets as is $\{\emptyset, \{\bullet\}\}$. You know what a finite family of sets is. When you list the students in the classrooms at the local elementary school, you are making finitely many finite sets. You make one finite set for each classroom. There may be six such classes. The lists of enrolled students in a classroom form a family of six sets as in the list

$$\mathrm{List}_1, \mathrm{List}_2, \mathrm{List}_3, \mathrm{List}_4, \mathrm{List}_5, \mathrm{List}_6$$

of lists. Each list List_k is the list of names of students in classroom k. The symbol k will be a natural number between 1 and 6.

Our examples of countable families of sets begin with

$$\begin{gathered} \{1\}, \{2\}, \{3\}, \ldots \\ \{1\}, \{1,2\}, \{1,2,3\}, \ldots. \end{gathered}$$

Oft times, the implied list of sets will be listed vertically, as in the following family

$$\begin{aligned} A_1 &= \{0\} \\ A_2 &= \{0,1\} \\ A_3 &= \{0,1,2\} \\ A_4 &= \{0,1,2,3\} \\ &\vdots \end{aligned}$$

of sets. This list is infinite and it consists of sets. That is why we call such an implied list an *infinite family of sets*.

Here is another. For each $m \in \mathbb{N}$ let $\mathbb{N}_m$ be the set of natural numbers divisible by m. Then the list

$$\begin{aligned}
\mathbb{N}_2 &= \{2, 4, 6, \ldots\} \\
\mathbb{N}_3 &= \{3, 6, 9, \ldots\} \\
&\vdots \\
\mathbb{N}_m &= \{m, 2m, 3m, \ldots\} \\
&\vdots
\end{aligned}$$

is an infinite family of sets.

For each natural number m, let

> $\mathbb{Q}_m^+$ be the positive rational numbers whose numerator is m.

We can make an implied list of this set since the denominators of a positive fraction can be only $1, 2, 3, 4, \ldots$.

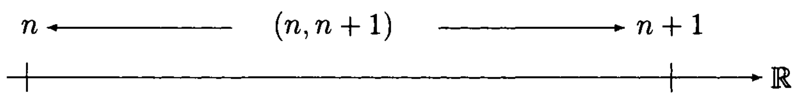
$$\mathbb{Q}_m^+ = \left\{\frac{m}{1}, \frac{m}{2}, \frac{m}{3}, \frac{m}{4}, \cdots\right\}.$$

Notice that the denominator increases while the numerator remains m.

$$\begin{aligned}
\mathbb{Q}_1^+ &= \left\{\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots\right\} \\
\mathbb{Q}_2^+ &= \left\{\frac{2}{1}, \frac{2}{2}, \frac{2}{3}, \frac{2}{4}, \cdots\right\} \\
\mathbb{Q}_3^+ &= \left\{\frac{3}{1}, \frac{3}{2}, \frac{3}{3}, \frac{3}{4}, \cdots\right\} \\
\mathbb{Q}_{24}^+ &= \left\{\frac{24}{1}, \frac{24}{2}, \frac{24}{3}, \frac{24}{4}, \cdots\right\} \\
&\vdots
\end{aligned}$$

Here is another example. Given an $n \in \mathbb{N}$, let $(n, n+1) = \{x \in \mathbb{R} \,|\, n < x < n+1\}$ be the set of real numbers strictly between

n and $n+1$. In picture form we would draw



Then
$$\{(n, n+1) \mid n \in \mathbb{N}\} = \{(0,1), (1,2), (2,3), \ldots\}$$
is an infinite family of sets.

At this point we will define $\cap$ and $\cup$ for countable families of sets. The point behind doing this is that the notation corresponds to the language of *quantifiers* that we encountered on page 20. In particular, we will have the opportunity to use phrases like *for some*, *at least*, and *for all.* Recall that the words *for some* mean at least one and at most all.

Definition 2.6.1 *Let $\{A_0, A_1, A_2, \ldots\}$ be an infinite family of sets.*

1. $\bigcap\{A_n \mid n \in \mathbb{N}\} = \bigcap_{n\in\mathbb{N}} A_n = \{x \mid x \in A_n \ \forall n \in \mathbb{N}\}$.
2. $\bigcup\{A_n \mid n \in \mathbb{N}\} = \bigcup_{n\in\mathbb{N}} A_n = \{x \mid x \in A_n$ *for some* $n \in \mathbb{N}\}$.

Equivalently, $x \in \cap_{n\in\mathbb{N}} A_n$ exactly when x is an element of A_n for all $n \in \mathbb{N}$. And $x \in \cup_{n\in\mathbb{N}} A_n$ exactly when $x \in A_n$ for some $n \in \mathbb{N}$.

We certainly can use some examples here. Let
$$A_0, A_1, A_2, A_3, \ldots$$
be an infinite family of sets. Then

$$\boxed{\bigcup_{n\in\mathbb{N}} A_n = A_0 \cup A_1 \cup A_2 \cup \ldots.}$$

Example 2.6.2 Let
$$\begin{aligned} A_1 &= \{0\} \\ A_2 &= \{0,1\} \\ A_3 &= \{0,1,2\} \\ A_4 &= \{0,1,2,3\} \\ &\vdots \end{aligned}$$

Then $1 \in A_2$, $2 \in A_3$, $3 \in A_4$, and continue. In general we see that

$$n \in A_{n+1} \ \forall n \in \mathbb{N}.$$

Hence, every $n \in \mathbb{N}$ satisfies the predicate $n \in A_{n+1}$. Put another way,

If $k \in \mathbb{N}$ then $k \in A_n$ for some $n \in \mathbb{N}$.

We conclude that $k \in \cup_{n\in\mathbb{N}} A_n$. In fact, since there was no extra condition on k we can say that *each* $k \in \mathbb{N}$ is in $\cup_{n\in\mathbb{N}} A_n$. Thus,

$$\boxed{\mathbb{N} \subset \bigcup_{n\in\mathbb{N}} A_n.}$$

Suppose that we look at $\cap_{n\in\mathbb{N}} A_n$ now. Since $0 \in A_n \ \forall n \in \mathbb{N}$, we see that

$$0 \in \bigcap_{n\in\mathbb{N}} A_n.$$

If $m \in \mathbb{N}$ is a natural number other than 0, then $m \notin A_m$ so that $m \notin A_n$ for some natural number n. In fact, $m \notin A_n$ for many natural numbers n, but we only need the one occurrence to conclude that

$$m \notin \bigcap_{n\in\mathbb{N}} A_n.$$

Therefore $\bigcap_{n\in\mathbb{N}} A_n = \{0\}$.

Example 2.6.3 Let us make another infinite family of sets. Examine the list

$$\begin{aligned}
B_0 &= \{0, 1\cdot 5\} \\
B_1 &= \{1\cdot 5, 2\cdot 5\} \\
B_2 &= \{2\cdot 5, 3\cdot 5\} \\
B_3 &= \{3\cdot 5, 4\cdot 5\} \\
&\vdots
\end{aligned}$$

of sets of two natural numbers. Let

$$\mathbb{F} = \{5n \mid n \in \mathbb{N}\}$$

denote the set of *multiples of 5.* We will prove that

$$\bigcup_{n\in\mathbb{N}} B_n = \mathbb{F}.$$

Proof: Let $x \in \cup_{n\in\mathbb{N}} B_n$. Then $x \in B_n$ for some $n \in \mathbb{N}$. By the definition of B_n, $x \in \{5n, 5(n+1)\}$, so that x is a multiple of 5. We conclude that $x \in \mathbb{F}$. As sets this means that

$$\bigcup_{n\in\mathbb{N}} B_n \subset \mathbb{F}.$$

The second inclusion begins with an element $x \in \mathbb{F}$. By the definition of $\mathbb{F}$, there is an $n \in \mathbb{N}$ such that $x = 5n$. Then $x \in \{5n, 5(n+1)\} = B_n$ for some $n \in \mathbb{N}$. We conclude that $x \in \bigcup_{n\in\mathbb{N}} B_n$, and so that

$$\mathbb{F} \subset \bigcup_{n\in\mathbb{N}} B_n.$$

Therefore $\bigcup_{n\in\mathbb{N}} B_n = \mathbb{F}$, which completes the proof.

Example 2.6.4 In this example, we let $p \geq 2$ be a prime and we let $\mathbb{N}_p$ denote the set of *nonnegative natural numbers divisible by p*. Then $\mathbb{N}_2$ is the set of positive even numbers $= \{0, 2, 4, 6, \ldots\}$, and $\mathbb{N}_3$ is the set of nonnegative numbers divisible by 3. We let $\mathbb{N}_1 = \{0, 1\}$ by default. This is just a notation we are using. It has nothing to do with the prime property. Then we have an infinite family

$$\begin{aligned} \mathbb{N}_1 &= \{0,1\} \\ \mathbb{N}_2 &= \{0,2,4,6,\ldots\} \\ \mathbb{N}_3 &= \{0,3,6,9,\ldots\} \\ \mathbb{N}_5 &= \{0,5,10,15,\ldots\} \\ &\vdots \end{aligned}$$

of sets. Furthermore, we have an infinite family

$$\mathbb{N}_1, \mathbb{N}_2, \mathbb{N}_3, \mathbb{N}_5, \ldots$$

indexed by 1 and the set of *prime numbers* $\mathbb{P}$. Thus $\mathbb{N}_7$ is a member of the family but $\mathbb{N}_9$ is not included in the family.

Example 2.6.5 Let

$$\bigcup_{n\in\mathbb{P}} \mathbb{N}_n \;=\; \mathbb{N}_2 \cup \mathbb{N}_3 \cup \mathbb{N}_5 \cup \ldots$$

be the union of sets, an let $\mathbb{D}$ = the set of natural numbers divisible by some prime number. We will show that

$$\mathbb{D} = \cup_{p\in\mathbb{P}}\mathbb{N}_p.$$

Proof: Given $x \in \mathbb{D}$, then x is divisible by some prime p. Then there is an $a \in \mathbb{N}$ such that $x = pa$. By definition of $\mathbb{N}_p$, $x = pa \in \mathbb{N}_p$, so that $x \in \cup_{p\in\mathbb{P}}\mathbb{N}_p$. Thus, $\mathbb{D} \subset \cup_{p\in\mathbb{P}}\mathbb{N}_p$.

Conversely, let $x \in \cup_{p\in\mathbb{P}}\mathbb{N}_p$. Then $x \in \mathbb{N}_p$ for some $p \in \mathbb{P}$. By definition of $\mathbb{N}_p$, $x = pa$ for some $a \in \mathbb{N}$, so that $x \in \mathbb{D}$. Hence, $\cup_{p\in\mathbb{P}}\mathbb{N}_p \subset \mathbb{D}$, and therefore,

$$\bigcup_{p\in\mathbb{P}} \mathbb{N}_p = \mathbb{D} = \text{those natural numbers divisible by some prime.}$$

This concludes our argument.

Since each natural number $n \neq 1$ is divisible by some prime number, I leave it to you to justify that $\mathbb{D}$ = the set of natural numbers $n \neq 1$. Summarizing, we have proved that

$$\bigcup_{p\in\mathbb{P}} \mathbb{N}_p = \mathbb{D} = \{0, 2, 3, 4, \ldots\}.$$

Because 1 is not divisible by any prime p, $1 \notin \mathbb{N}_p$ for any prime number p. Then $1 \notin \bigcup_{p\in\mathbb{P}} \mathbb{N}_p$, and therefore,

$$\begin{aligned} \mathbb{N}_1 \cup \bigcup_{p\in\mathbb{P}} \mathbb{N}_p &= \mathbb{N}_1 \cup \mathbb{N}_2 \cup \mathbb{N}_3 \cup \mathbb{N}_5 \cup \ldots \\ &= \{0, 1, 2, 3, \ldots\} \\ &= \mathbb{N}. \end{aligned}$$

Example 2.6.6 Using the notation in the previous example, let us examine the set

$$X = \mathbb{N}_2 \cap \mathbb{N}_3 \cap \mathbb{N}_5 \cap \ldots = \bigcap_{p\in\mathbb{P}} \mathbb{N}_p.$$

1. We observe that $0 \in X$ since $0 = p0 \in \mathbb{N}_p$ for all primes p. Suppose, for the purposes of a proof by contradiction, that we assume that there is an element $x \neq 0 \in X$. Then $x \in \mathbb{N}_p$ for

each prime $p \in \mathbb{P}$ so that x is divisible by each prime $p \in \mathbb{P}$. What are the nonzero natural numbers divisible by all prime numbers? You are passingly familiar with the answer each time you factor a nonzero natural number. Nonzero natural numbers are divisible by only finitely many prime numbers. This contradiction (that $x \neq 0$ is divisible by all primes and that x can be divisible by only finitely many primes) shows us that our assumption is incorrect. Hence, $X = \emptyset$, and so

$$\mathbb{N}_2 \cap \mathbb{N}_3 \cap \mathbb{N}_5 \cap \ldots = \emptyset.$$

2. We will show that the intersection of any two of these sets is nonempty. Let $p, q \in \mathbb{P}$ be prime numbers. Then

$$pq \in \mathbb{N}_p \cap \mathbb{N}_q \text{ since } pq \text{ is a multiple of } p \text{ and a multiple of } q$$

so that $\mathbb{N}_p \cap \mathbb{N}_q \neq \emptyset$.

Example 2.6.7 Let $A_0, A_1, A_2, \ldots$ be an infinite family of sets. The following negation of the definitions of $\cap$ and $\cup$ is a good exercise in the proper use of universal and existential quantifiers.

$$\boxed{x \notin \bigcap_{n\in\mathbb{N}} A_n \text{ exactly when } \exists n \in \mathbb{N},\ x \notin A_n.}$$

Notice the change in phrasing. We went from *for all* to its logical negation *there exists.*

The statement *all colors are impressive* is negated by stating *some colors are not impressive.* Try negating $x \in S$ *for all sets* S before reading on. You would be correct if your negation reads $x \notin S$ *for some set* S. It could also read $\exists x,\ x \notin S$. Your choice would depend on the English surrounding the mathematics.

The statement $n < n+1,\ \forall n \in \mathbb{N}$ is negated by writing $n \geq n+1$ *for some* $n \in \mathbb{N}$. Said another way, this negation reads as

$$\exists n \in \mathbb{N},\ n \geq n + 1.$$

This is obviously a False statement, which we should expect since we are negating the True statement $n < n + 1,\ \forall n \in \mathbb{N}$.

Recall that *for some* means *at least one and at most all.* Its meaning will be reinforced with several examples.

The statement *some children have blue eyes* means that there is one child or perhaps several children who have blue eyes. A quick check of European children will verify that *many* children have blues eyes, so logically *some* children have blue eyes. The statement $x \in A_n$ *for some* $n \in \mathbb{N}$ means that there is one, and possibly more than that, natural number n such that $x \in A_n$. The statement $3x = 0$ *for some* $x \in \mathbb{R}$ means that there is one, and perhaps more, real number x such that $3x = 0$. Algebra shows us immediately that in this case there is exactly 1 number x such that $3x = 0$. But still it is proper to say that *there are some real numbers* x *such that* $3x = 0$.

If we have a predicate P and we write P *is satisfied by at least one natural number*, then we mean that there is a natural number n that satisfies P, and that there may be more natural numbers m that satisfy P. We do not rule out the possibility that every natural number n might satisfy X. For instance, let

$$P = \text{there are } n \text{ musicians in a rock-and-roll band.}$$

Then P is satisfied by at least the natural number 4. In fact, P is satisfied by 3, 5, and 7 as well. Rock-and-roll bands do not run into the thousands of musicians, so there are some natural numbers that do not satisfy P. I leave it to you to name the bands that have 3, 4, or 7 people in them, keeping in mind that your author's narrow musical taste runs to sixties and seventies rock-and-roll.

Example 2.6.8 Let

$$\begin{array}{rcl} A_0 & = & \{0\} \\ A_1 & = & \{0,1\} \\ A_2 & = & \{0,1,2\} \\ & \vdots & \end{array}$$

be an infinite family of sets. Then $2 \notin A_1$, $3 \notin A_2$, and in general, $n \notin A_{n-1}$. Hence, given $m \in \mathbb{N}$,

$$2,3,\ldots,m \quad \notin \quad \bigcap_{n\in\mathbb{N}} A_n$$

because $m \notin A_m$ in this infinite family of sets. Since $0 \in A_n$ for all $n \in \mathbb{N}$,

$$\{0\} = \bigcap_{n\in\mathbb{N}} A_n.$$

If $x \in A_n$ for some $n \in \mathbb{N}$, then $x \in \cup_{n\in\mathbb{N}} A_n$. The statement defines the union of an infinite family of sets $\{A_n \mid n \in \mathbb{N}\}$. The negation of this statement can be written as

$$x \notin \bigcup_{n\in\mathbb{N}} A_n \text{ exactly when } x \notin A_n, \forall n \in \mathbb{N}.$$

The reader will show that

$$\{0, 1, 2, \ldots\} = \bigcup_{n\in\mathbb{N}} A_n.$$

Exercise 2.6.9 Let $\{A_n \mid n \in \mathbb{N}\}$ be an infinite family of sets contained in some universal set $\mathcal{U}$.

1. State in terms of quantifiers what $x \in \cup_{n\in\mathbb{N}} A_n$ means.

2. State in terms of quantifiers what $x \notin \cup_{n\in\mathbb{N}} A_n$ means.

3. State in terms of quantifiers what $x \in \cap_{n\in\mathbb{N}} A_n$ means.

4. State in terms of quantifiers what $x \notin \cap_{n\in\mathbb{N}} A_n$ means.

5. Show that $\left(\bigcap_{n\in\mathbb{N}} A_n\right)' = \bigcup_{n\in\mathbb{N}} A'_n$.

6. Show that $\left(\bigcup_{n\in\mathbb{N}} A_n\right)' = \bigcap_{n\in\mathbb{N}} A'_n$.

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

Chapter 3

Functions

Throughout this book we will want to compare sets. We do this by studying the *functions* that exist between them. A function from a set A into a set B is a rule that associates each element $x \in A$ with exactly one element $y \in B$. If we name the rule f, then we will write

$$f(x) = y$$

to denote the fact that the element x is associated with y. We will also say that y is *the image of x under f*. Some of you are familiar with functions from that third year of high school math. In that case a function was usually a mapping or a general rule that acted on real numbers. For example,

$$f(x) = x^2$$

would be the function from $\mathbb{R}$ into $\mathbb{R}$ that associates each number x with its square x^2. Others may have written this function as $x^2 = y$. In either case we are using some algebra to denote the image of x.

Our functions will only occasionally operate on real numbers. The reason is that our deliberations are on general sets and not exclusively on real numbers. For example, we might define the function f that takes each person on Earth to his or her age in years, or we might define a function that associates each woman with exactly one man. You might even associate the elements of

$\{a, b, c\}$ with the elements from the set $\{x, y, z\}$ by the rule

$$\begin{cases} a & \longmapsto & x \\ b & \longmapsto & z \\ c & \longmapsto & z \end{cases}$$

which sends both b and c to z, and nothing to y. This type of abstract association is more closely related to the functions that we will encounter as our discussion evolves.

3.1 Functional Preliminaries

Suppose we are given sets $\{x, y, z\}$ and $\{0, 1, 2\}$. Some of the more important questions we will pursue in this book are the following.

1. What do we really mean by an *element by element matching between* $\{x, y, z\}$ *and* $\{0, 1, 2\}$?

2. Can we make the notion of the *cardinality of* $\{x, y, z\}$ a mathematically precise idea?

3. It seems clear that $\{a, b\}$ has fewer elements than $\{x, y, z\}$. Can we make this precise?

We will approach these questions by considering a general comparison of sets called a *function* that requires us to investigate the idea of a function on *abstract* sets.

The following takes our definition of function and gives it some mathematical style. It may read as a dry definition to you, but it has the advantage of being mathematically precise. Perhaps you could try to define functions in such a way that it encompasses the examples we give in this section. There is something to be learned by trying to do that, so please do.

Definition 3.1.1 *Let A and B be sets. A* function *from A to B is a rule f with the following two properties.*

1. *f associates each $x \in A$ with an element $f(x) \in B$.*

2. *Given $x \in A$ there is exactly* one *value $f(x)$.*

A bit of very useful notation is the following. We will write

$$\boxed{f : A \longrightarrow B}$$

if f is a function from A to B. We call A *the domain of* f, and we call B *the codomain of* f. We call $f(x)$ *the image of* x *under* f just as we did in our algebra class in high school.

Another *description* of a function that you may find a bit easier to absorb is the following. A function f taking values in A to values in B assigns to each element $x \in A$ an image $f(x) \in B$. There must be no ambiguity in the value of the image $f(x)$. This definition gives you the mental picture of a function that you had in high school, but it contains so much more. You may have read that

$$f(x) = y$$

and that is not wrong. We often let y denote the image of x. It is common to do so especially when graphing the function f as we will do presently. You may also have read that

$$f(x) = 2x + 1,$$

which is an example of a function on real numbers. This is not the only type of function, which is why we had to define functions in such a broad abstract way. There is a function $f : \{0, 1, 2\} \longrightarrow \{x, y, z\}$ defined by the rule

$$\begin{cases} 0 & \longmapsto & z \\ 1 & \longmapsto & x \\ 2 & \longmapsto & z \end{cases}$$

which sends both 0 and 2 to z. Notice that under this function, y is not associated with any element in $\{0, 1, 2\}$.

Our next example of a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by setting

$$f(x) = 7x + 6 \quad \text{for each } x \in \mathbb{R}.$$

We should all be familiar with the fact that the graph of this function is a line that is inclined upward when we read it from left to right. Notice that given $x \in \mathbb{R}$, there is exactly one real number $y = 7x + 6$. Observe that the images of 0, 1, and -1 can be calculated using the formula $f(x) = 7x + 6$.

$$\begin{aligned} f(0) &= 7 \cdot 0 + 6 = 6 \\ f(1) &= 7 \cdot 1 + 6 = 13 \\ f(-1) &= 7 \cdot (-1) + 6 = -1. \end{aligned}$$

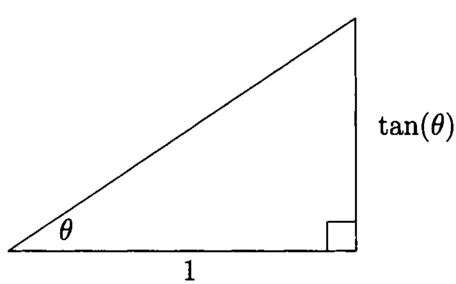
Another familiar function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by $f(x) = x^2$. The value x^2 is a real number that is unique to x so f is indeed a function. We can find the images of 0, 1, and -1 easily enough.

$$\begin{aligned} f(0) &= 0^2 = 0 \\ f(1) &= 1^2 = 1 \\ f(-1) &= (-1)^2 = 1. \end{aligned}$$

If you had any trigonometry in your education, then you may know that

$$f(\theta) = \tan(\theta)$$

defines a function that associates with each angle θ its tangent, $\tan(\theta)$. We can construct the value $\tan(\theta)$ as follows. Draw a triangle whose base length is 1 and whose base angles are θ and a right angle, as in the diagram below. Then the side opposite θ has length $\tan(\theta)$. If you possessed a ruler that was mathematically accurate to all decimal places, you could read $\tan(\theta)$ from picture (3.1).



(3.1)

If we use the definition of $\tan(\theta)$ as given in figure (3.1), then the domain of $f(\theta) = \tan(\theta)$ is the set

$$[0, \pi/2) = \{\theta \in \mathbb{R} \mid 0 \le \theta < \pi/2\}.$$

Note that this domain does not include negative values θ. Then we would write $f : [0, \pi/2) \to \mathbb{R}$.

The function $f(\theta) = y$ can be extended to

$$(-\pi/2, \pi/2) = \{\theta \in \mathbb{R} \mid -\pi/2 < \theta < \pi/2\}$$

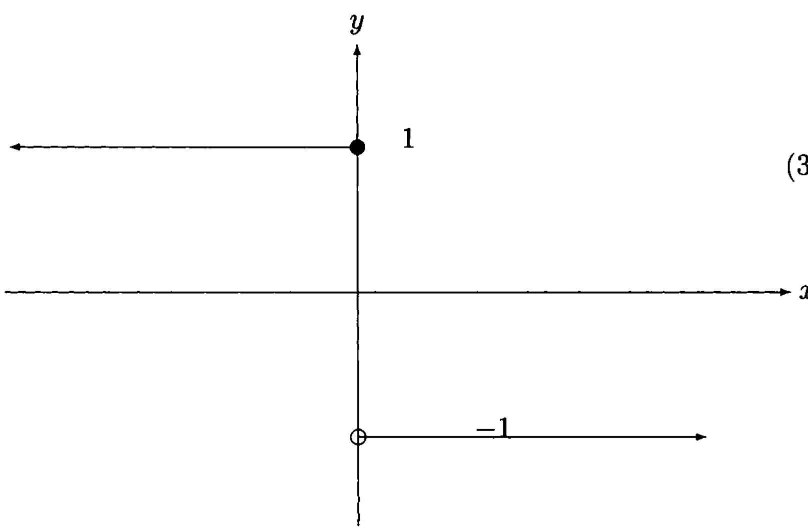
by defining $\tan(-\theta) = -\tan(\theta)$ for each $\theta \in [0, \pi/2)$. The resulting function

$$\tan(\theta) : (-\pi/2, \pi/2) \to \mathbb{R}$$

yields the tangent values you learned about in a high school trigonometry course.

The functions $f(x) = 7x + 6$, $f(x) = x^2$, and $f(x) = \tan(x)$ will take real numbers x and associate them with or *send them to* a real number $f(x)$. Most of the functions that you have experienced in your lifetime are like this. Now let's see something more abstract.

Consider the graph (3.2).



(3.2)

It has a naturally occurring feature that you may not have seen before. It has a gap and a hole in it. These are common features of graphs, even though they have been kept from you. They occur more often than the smooth connected curves with which you are familiar. We will present some of these functions with broken graphs.

Example 3.1.2 Define a function $f : \mathbb{R} \longrightarrow \{1, -1\}$ by setting

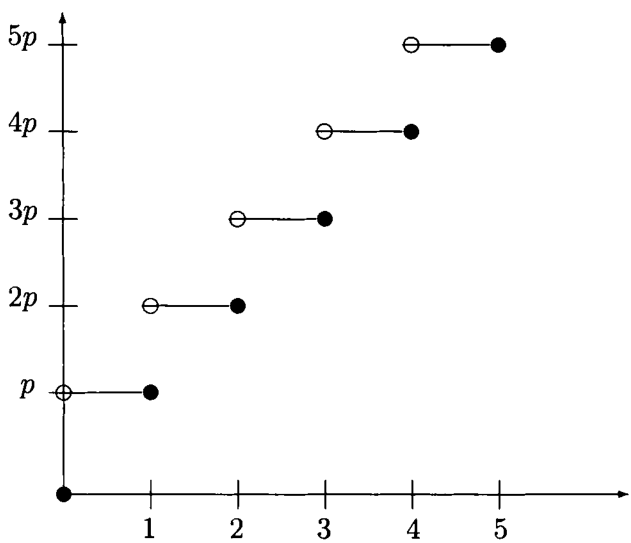
$$f(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ -1 & \text{if } x > 0 \end{cases}.$$

We observe that given $x \in \mathbb{R}$, then $f(x) = 1$ or $f(x) = -1$ but not both. Thus, f satisfies our definition of function. The graph of f is drawn in (3.2).

Example 3.1.3 A function whose graph has many more steps is the *postage rate function.*

$$f(x) \;=\; \text{cost in pennies to send a letter that weighs } x \text{ ounces by snail mail.}$$

Its graph is the graph (3.3) below.



(3.3)

If a letter weighs up to and including 1 ounce, the cost of postage is p. (At the time of this writing $p = 44$ cents.) If it weighs more than 1 ounce but at most (and including) 2 ounces, then the cost to mail the letter is $2p$. This continues on until you get tired of it. The graph of this function must reflect that jump in price at 1 ounce, at 2 ounces, at 3 ounces, and so on.

The fact that we have a hole $\circ$ on a graph means that the price is not evaluated at that line. The line corresponding to p has a hole at the left end. This is the cost of mailing a letter weighing 0 ounces. There is no letter to mail in this case so the cost is 0 cents. The line corresponding to p has a solid dot $\bullet$ at its right end, and the line corresponding to $2p$ has a whole at the left end. That means that the cost for a 1 ounce letter is to be evaluated from the lower line. For these reasons, we pay p cents when mailing a 1 ounce letter.

The graph tells us that it will cost $2p$ cents to mail a letter that weighs more than 1 ounce and up to 2 ounces. The black dot $\bullet$ on that line means that if the letter weighs more than 1 ounce, then you must pay $2p$ to mail it. Similarly if the letter weighs more than 2 ounces or up to 3 ounces, then the letter costs $3p$ to mail. You can guess what happens when the letter weighs more than 3 ounces but not more than 4 ounces.

Example 3.1.4 Define a function $f : \mathbb{R} \longrightarrow \{0, 1\}$ by setting

$$f(x) = \begin{cases} 1 & \text{if} \quad x \in \mathbb{Q} \quad \text{That is, } x \text{ is a fraction} \\ 0 & \text{if} \quad x \notin \mathbb{Q} \end{cases}.$$

This one cannot be graphed, but it is still a function. If we were to try to graph this function, we would have to graph a horizontal line that is full of holes and full of dark spots. For instance,

$$f(0) = 1, \ f\left(\frac{1}{2}\right) = 1, \text{ and } f\left(\frac{4}{7}\right) = 1,$$

while

$$f\left(\frac{1}{\sqrt{2}}\right) = 0, \ f\left(\frac{1}{\pi}\right) = 0, \text{ and } f\left(\frac{2}{\sqrt[3]{13}}\right) = 0.$$

Furthermore, between every two real numbers x and y there is a rational number q

$$x < q < y,$$

so between every two numbers x and y at which $f(x) = f(y) = 0$ there is a number q such that $f(q) = 1$.

$$f(x) = 0, f(q) = 1, f(y) = 0.$$

No technology will ever permit us to graph such a line. However, since a given real number x is either a rational number or an irrational number but not both, the rule f defines a function $f : \mathbb{R} \longrightarrow \{0, 1\}$.

Next let us consider functions that are not so algebraic.

Example 3.1.5 Let $\mathbf{P}$ be the set of people on the Earth, and define a function $a : \mathbf{P} \longrightarrow \mathbb{N}$ as the function that assigns to each person on Earth his/her age in years as of January 1, 2005. Then $a(\text{Wendy}) = 21$, $a(\text{new born}) = 0$, and $a(\text{author}) > 25$. You cannot at this time determine $a(\text{author})$, but there is a natural number that the author calls his age. Observe that at this time there is no person x such that $a(x) = 200$.

Example 3.1.6 Let $\mathbf{L}$ be the set of locations on the Earth, and define a function $t : \mathbf{L} \longrightarrow \mathbb{N}$ to be the function that assigns to each location on Earth the temperature in degrees Celsius at that location at exactly 12 noon on January 1, 2005. So you might find that $t(\text{South Pole}) = -80°$ C, $t(\text{Guam}) = 30°$C, and $t(\text{outside}) = 20°$C. You cannot at this time determine $t(300$ miles under New York City$)$, but there is a natural number that we would all agree is that temperature. There is no place x on Earth such that $t(x) =$ absolute zero.

A rule f must satisfy both conditions in our definition of function if it is to be a function.

Example 3.1.7 The rules in this example do not define functions.

1. (Failure of Condition 1.) Let f be the association from $\mathbb{R}$ to $\mathbb{R}$ defined by $f(x) = \sqrt{x}$. Then f is not a function since $f(-1) = \sqrt{-1} \notin \mathbb{R}$.

2. (Failure of Condition 2.) Let $\mathbb{R}^+$ be the set of positive real numbers, and let f be the association from $\mathbb{R}^+$ to $\mathbb{R}$ defined by $f(x) = \pm\sqrt{x}$. Then f is not a function since $f(1) = 1, -1$ is more than one value.

We think we know what an implied list is, but here is a mathematically precise way to think of it. Let A be a set. Any set will do. We define an implied list of elements in A to be a function

$$f : \mathbb{N} \longrightarrow A.$$

In this case we will write

$$f(0) = a_0,\ f(1) = a_1,\ f(2) = a_2,\ \ldots$$

and in general

$$f(n) = a_n \text{ for each element } n \in \mathbb{N}.$$

For instance, we can define $f : \mathbb{N} \longrightarrow \mathbb{R}$ as

$$f(n) = n.$$

Then f is the implied list

$$0, 1, 2, 3, \ldots.$$

Another function $f : \mathbb{N} \longrightarrow \mathbb{R}$ defined by

$$f(n) = \frac{1}{n+1} \text{ for all } n \in \mathbb{N}$$

is the implied list

$$\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots.$$

The following is a pretty little rule that is not a function unless we are very careful.

In this age of computers, it should be no surprise that we can find for every real number $0 < x < 1$, a *binary series*

$$\boxed{x = \frac{b_1}{2} + \frac{b_2}{2^2} + \frac{b_3}{2^3} + \ldots,}$$

where the digits $b_1, b_2, b_3 \ldots$ take values in the set $\{0, 1\}$. We would then write the *binary expansion* of x as the binary decimal number

$$x \;=\; .b_1 b_2 b_3 \ldots \tag{3.4}$$

in which the dot, ".", is a decimal point. For example,

$$\frac{1}{2} \;=\; \frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \ldots$$

and

$$\frac{1}{3} = \frac{0}{2} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{1}{2^4} + \ldots,$$

where the coefficients for $\frac{1}{3}$ are 1 for the even exponents of $\frac{1}{2}$ and 0 elsewhere. The associated binary expansions are

$$\begin{aligned} \frac{1}{2} &= .100\bar{0} \\ \frac{3}{4} &= .1100\bar{0} \\ \frac{1}{3} &= .0101\overline{01} \end{aligned}$$

where the expansion $.1100\bar{0}$ comes from the sum $\frac{3}{4} = \frac{1}{2} + \frac{1}{2^2}$.

A *binary sequence* is a sequence $b_1 b_2 b_3 \cdots$ in which $b_i \in \{0, 1\}$ for all positive $i \in \mathbb{N}$. In the general spirit of mathematics, we let

$$\boxed{\mathbf{S} = \text{ the set of all binary sequences } b_1 b_2 b_3 \ldots}$$

denote the set of all binary sequences. That is, $\mathbf{S}$ consists of all sequences of 0's and 1's. For instance, $\mathbf{S}$ contains elements like $10\bar{0}\ldots$ and $01\overline{01}\ldots$, but it does not contain the number .333 or the sequence $\overline{999}$.

We define the *unit interval* $(0, 1)$ to be the real numbers that lie properly between 0 and 1 on the real number line.

$$\boxed{(0, 1) = \{x \in \mathbb{R} \mid 0 < x < 1\}.}$$

Let us define a rather obvious function

$$f : (0, 1) \longrightarrow \mathbf{S}$$

as follows. Given an $x \in (0, 1)$, write x as a binary expansion as in (3.4),

$$x = .b_1 b_2 b_3 \ldots,$$

and then define

$$f(x) = b_1 b_2 b_3 \ldots .$$

Notice that $f(x)$ simply drops the decimal point of the binary expansion for x. Evidently, given $x \in (0,1)$, $f(x)$ is a binary sequence, so that $f(x) \in \mathbf{S}$ for each $x \in (0,1)$.

However, the value $f(x)$ has some ambiguity surrounding it. For example, let's show that $\frac{1}{2}$ has two binary expansions.

$$\begin{aligned}
\frac{1}{2} &= \frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \ldots \\
&= \frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots .
\end{aligned}$$

This will require some trickery and guile that we will use several times in the future. Given

$$\frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots,$$

let's multiply by $\frac{1}{2} = 1 - \frac{1}{2}$ to realize the following equations. Use the distributive law or your friend foil to do the multiplication.

$$\begin{aligned}
&\left(\frac{1}{2}\right)\left(\frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \ldots\right) \\
= &\left(1 - \frac{1}{2}\right)\left(\frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \ldots\right) \\
= &\left\{ \begin{array}{l} \left(\frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \ldots\right) \\ -\frac{1}{2}\left(\frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \ldots\right) \end{array} \right. \\
= &\left\{ \begin{array}{ccccccccc} \frac{1}{2^2} & + & \frac{1}{2^3} & + & \frac{1}{2^4} & + & \frac{1}{2^5} & + & \ldots \\ & - & \frac{1}{2^3} & - & \frac{1}{2^4} & - & \frac{1}{2^5} & - & \ldots \end{array} \right. \\
= &\frac{1}{2^2}.
\end{aligned}$$

Then

$$\left(\frac{1}{2}\right)\left(1\frac{1}{2^2}+1\frac{1}{2^3}+1\frac{1}{2^4}+\ldots\right) = \frac{1}{2^2},$$

so that multiplication by 2 yields the equation

$$1\frac{1}{2^2}+1\frac{1}{2^3}+1\frac{1}{2^4}+\ldots = \frac{1}{2}.$$

So, reader, which one is it? According to our definition of f, there are two choices for $f(x)$. Which choice do you make?

$$f\left(\frac{1}{2}\right)=\begin{cases} 100\bar{0} \text{ or} \\ 011\bar{1} \end{cases}.$$

The point here is that *there should be no choice in the value* $f(x)$. The value $f(x)$ should be unique, unambiguous, and singular to x. If two values present themselves, they should be equal. They might look different, but they should be the same thing. This is not the case here. The value for $f\left(\frac{1}{2}\right)$ has on the one hand only one entry of 1, while on the other hand it has infinitely many entries 1. As the rule f is defined, *it does not satisfy Condition 2* of the definition of function, so f is not a function.

The only way to correct this deficiency is to take advantage of the number of 1's in these binary sequences. This, it turns out, is all we need to change the rule f into a function. Define a rule

$$F:(0,1)\longrightarrow \mathbf{S}$$

related to f by requiring that

$$F(x) = b_1b_2b_3\ldots, \text{ where we choose the binary expansion for } x=.b_1b_2b_3\ldots \text{ with the smallest number of 1's.} \quad (3.5)$$

Under this new rule F, the sequence $01\bar{1}\ldots$ is ruled out as the image of $\frac{1}{2}$ since there is a "shorter" binary sequence available. Hence

$$F\left(\frac{1}{2}\right)=100\bar{0}\ldots.$$

This definition for the new rule F avoids the ambiguity that the rule f possessed.

Let us examine an operation on functions. Functions in abstract sets cannot be added or multiplied, so we will not be dealing with the addition or multiplication of functions as we might have done in algebra. There is, however, an important operation on functions called *composition*.

Definition 3.1.8 *Let $f : A \longrightarrow B$ and $g : B \longrightarrow C$ be functions. (Note that B serves as both the domain of g and the codomain of f.) There is a function*

$$g \circ f \ : \ A \longrightarrow C$$

whose rule is

$$\boxed{g \circ f(x) = g(f(x)) \text{ for each } x \in A.}$$

For example, let $f : \mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by $f(x) = x + 1$ and $g(x^2)$. Then

$$g \circ f(x) \quad = \quad g(f(x)) \quad = \quad g(x+1) \quad = \quad (x+1)^2$$

and

$$f \circ f(x) \quad = \quad f(f(x)) \quad = \quad f(x+1) \quad = \quad (x+1)+1 \quad = \quad x+2.$$

Example 3.1.9 Let $f(x) = \dfrac{1}{x}$. Then

$$f \circ f(x) = f(f(x)) = f\left(\frac{1}{x}\right) = \frac{1}{1/x} = x.$$

That is, $f \circ f(x) = x$.

Example 3.1.10 Let $f(x) = \sqrt{x}$ and $g(x) = x^2$. Then

$$f \circ g(x) = f(g(x)) = f(x^2) = \sqrt{x^2} = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}.$$

Let us experiment with $f(x) = \sqrt{x}$ and $g(x) = x^2$. Compare $f(-1)$ and $f(1)$. Then

$$f \circ g(-1) = f(g(-1)) = f\left((-1)^2\right) = f(1) = \sqrt{1} = 1,$$

while

$$f \circ g(1) = f(g(1)) = f\left(1^2\right) = f(1) = \sqrt{1} = 1.$$

We note that $f \circ g(-1) = 1 \neq -1$, so that in general $\sqrt{x^2} = f \circ g(x) \neq x$. As an exercise, show that for $x = -2$, $\sqrt{(-2)^2} \neq -2$. How do you reconcile this fact with what you know today?

Example 3.1.11 Let $F : (0,1) \longrightarrow \mathbf{S}$ be the function defined in (3.5). Then $F(x)$ is the binary sequence $b_1 b_2 b_3 \ldots$ with fewest 1's such that $x = .b_1 b_2 b_3 \cdots$. Define a function

$$G : \mathbf{S} \longrightarrow \{0,1\}$$

by requiring that

$$G(b_1 b_2 b_3 \ldots) = b_1 = \text{the first entry in the sequence } b_1 b_2 b_3 \ldots.$$

The composition $G \circ F$ when applied to $x = .b_1 b_2 b_3 \cdots$ has the rule

$$G \circ F(x) = G(F(x)) = G(b_1 b_2 b_3 \ldots) = b_1 \in \{0,1\}.$$

Then $G \circ F$ is a function that assigns to each $x \in (0,1)$ either a 0 or a 1, and it makes the assignment without ambiguity. We challenge the reader to find the set of numbers $x \in (0,1)$ such that $G \circ F(x) = 1$. Hint: Look at the definition of F. The answer to this challenge is given at the end of this section.

We will need a notation for the set of all functions from A into B.

Definition 3.1.12 *Let A and B be sets. We let*

$$\boxed{B^A = \text{the set of all functions } f : A \longrightarrow B.}$$

Let $A = \{a, b\}$ and $B = \{0, 1\}$. We will write down all of the functions in B^A by defining the values $f(x)$ for all f and all $x \in \{a, b\}$.

$$\begin{cases} f(a) = 0 \\ f(b) = 0 \end{cases} \qquad \begin{cases} g(a) = 0 \\ g(b) = 1 \end{cases}$$

$$\begin{cases} h(a) = 1 \\ h(b) = 0 \end{cases} \qquad \begin{cases} k(a) = 1 \\ k(b) = 1 \end{cases} .$$

Since we have completely exhausted the possible images for a and b, we have written a complete list of the elements of B^A. Another way of writing these functions presents itself.

Let us agree that each function f will be written as

$$f = [x, y] \text{ provided that } f(a) = x \text{ and } f(b) = y.$$

For instance,

$$f = [1, 0] \text{ means that } f(a) = 1 \text{ and } f(b) = 0.$$

We can then write down all of the functions in $\{0, 1\}^{\{a,b\}}$ by systematically writing down the possible pairs $[x, y]$. They are

$$\begin{matrix} [0,0] & [0,1] \\ [1,0] & [1,1]. \end{matrix}$$

We can then easily count the number of functions in $\{0, 1\}^{\{a,b\}}$. Recall that the *cardinality of* X is denoted by $\operatorname{card}(X)$. Then

$$\operatorname{card}(\{0,1\}^{\{a,b\}}) = 4 = 2^2 = \operatorname{card}(\{0,1\})^{\operatorname{card}(\{a,b\})},$$

which explains why we chose this notation.

Other examples are found by letting $A = \{a, b, c\}$ and $B = \{0, 1\}$. To count the number of functions in $\{0, 1\}^{\{a,b,c\}}$, let us agree to write the function

$$f : \{a, b, c\} \longrightarrow \{0, 1\}$$

in terms of its values at a, b, and c. Then f is denoted by

$$\boxed{f = [x, y, z], \quad \text{where } f(a) = x,\ f(b) = y, \text{ and } f(c) = z.}$$

For example, the function f such that $f(a) = 1$, $f(b) = 1$, $f(c) = 0$ is denoted by

$$f = [1, 1, 0].$$

Counting entries from left to right, the first entry in $[1, 1, 0]$ is 1 because $f(a) = 1$. The second entry is 1 because $f(b) = 1$, and the third entry is 0 because $f(c) = 0$. Every function $f : \{a, b, c\} \longrightarrow \{0, 1\}$ can be represented in this way. This provides us with a very simple way to count the elements in $\{0, 1\}^{\{a,b,c\}}$. There are two possible values for each of the three entries of $f = [x, y, z]$ so that there are $2 \cdot 2 \cdot 2 = 8$ possible entries for f and therefore there are 8 possible functions in $\{0, 1\}^{\{a,b,c\}}$. That is,

$$\text{card}\left(B^A\right) = \text{card}(\{0, 1\}^{\{a,b,c\}}) = 8 = \text{card}(\{0, 1\})^{\text{card}(\{a,b,c\})}.$$

We leave it to the reader to write down the eight functions in B^A.

As a further exercise, the functions in $\{0, 1, 2\}^{\{a,b\}}$ can be written as

$$[x, y] \text{ where } f(a) = x \text{ and } f(b) = y,$$

where x, y are numbers in $\{0, 1, 2\}$. Since there are three numbers that can fill the two entries $[x, y]$, there are exactly $3 \cdot 3 = 9$ functions in $\{0, 1, 2\}^{\{a,b\}}$.

$$\text{card}\left(A^B\right) = \text{card}(\{0, 1, 2\}^{\{a,b\}}) = 3^2 = \text{card}(\{0, 1, 2\})^{\text{card}(\{a,b\})}.$$

Four of these functions are

$$[0, 0], \quad [1, 0], \quad [0, 2], \quad [1, 2].$$

The first function takes each of a and b to 0. $[1, 2]$ is a function that takes a to 1 and that takes b to 2. The reader is invited to list all 9 functions and their rules using the $[x, y]$ notation that we described above.

This is the answer to the $G \circ F$ challenge. You can use the sum (3.4) to convince yourself that $G \circ F(x) = 1$ exactly when $\frac{1}{2} \leq x < 1$.

Exercise 3.1.13 1. If $\tan(\theta)$ is the function you learned about in high school trigonometry, then show that $\tan(-\theta) = -\tan(\theta)$ for $\theta \in [0, \pi/2)$.

2. Graph $f(x) = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$.

3. A candy store owner sells red licorice by the centimeter. Candy of length x such that $n \leq x < n + 10$ costs n cents. Draw the graph of cost as a function of length.

4. Determine which of the following is a function with indicated domain and codomain.

   (a) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$.

   (b) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sqrt[4]{x}$.

   (c) $f : (-\pi/2, \pi/2) \to \mathbb{R}$, $f(x) = \dfrac{1}{\tan(\theta)}$.

   (d) $f : [0, 1] \to \{0, 1\}$, $f(x) = 0$ if $0 \leq x \leq 1/2$, and $f(x) = 1$ if $1/2 \leq x \leq 1$.

5. Define a function $H : \mathbf{S} \to \{00, 01, 10, 11\}$ by letting $H(x) = b_1b_2$ where $x = .b_1b_2b_3\ldots$. That is, b_1b_2 are the first two places in the binary expansion for x. Show that H is a function, or define H so that it is a function.

6. Define H as in the previous exercise. What is $H^{-1}(00)$?

7. Define H as in the previous exercise. What is $H^{-1}(\{00, 01\})$?

8. Write out the array that is the set of functions B^A where $A = \{0, 1\}$ and $B = \{a, b, c, d\}$.

9. Write out the array that is the set of functions B^A where $A = \{0, 1, 2, 3\}$ and $B = \{a, b\}$.

3.2 Images and Preimages

Consider the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$f(x) = x^2.$$

A moment's thought reveals that

$$f(1) = f(-1) = 1.$$

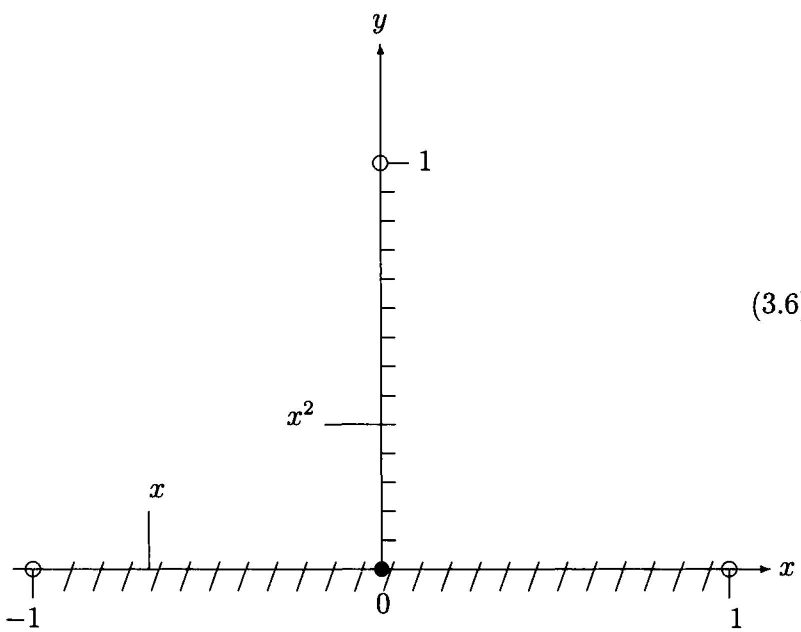
Let's write that in a more useful notation.

$$f(\{1,-1\}) = \{1\}.$$

Furthermore, if we recall that $(-1,1) = \{x \in \mathbb{R} \mid -1 < x < 1\}$ then

$$f(\,(-1,1)\,) \;=\; \{f(x) \mid -1 < x < 1\} \;=\; \{y \mid 0 \leq y < 1\}.$$

The following picture may help you see this. The interval $(-1,1)$ defined by the slashes "/" is mapped into the interval $[0,1)$ defined by the dashes "–" . The open circles $\circ$ mean that $x = -1$, $x = 1$, and $y = 1$ are not included.



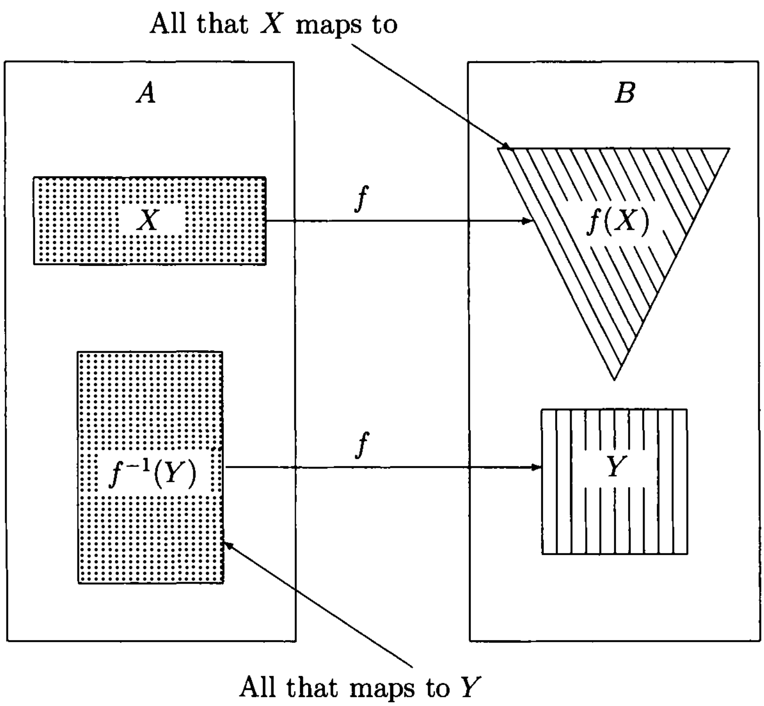
(3.6)

We will say that $[0,1)$ is *the image of* $(-1,1)$ under the function $f(x) = x^2$. Since $(-1,1)$ is everything that $f(x)$ sends into $[0,1)$, we say that $(-1,1)$ is *the preimage of* $[0,1)$. Another example would be that $\{1\}$ is the image of $\{-1\}$, and since $\{-1,1\}$ is all that f maps to $\{1\}$, $\{-1,1\}$ is the preimage of $\{1\}$. In more general terms, we have the following definition.

Definition 3.2.1 *Let $f : A \longrightarrow B$ be a function, and let $X \subset A$ and $Y \subset B$ be subsets.*

1. $f(X) = \{f(x) \mid x \in X\}$. $f(X)$ *is called* the image of X.
2. $f^{-1}(Y) = \{x \in X \mid f(x) \in Y\}$. $f^{-1}(Y)$ *is called* the preimage of Y.

If $x \in A$ and $y \in B$ then we will write $f(x)$ for $f(\{x\})$ and $f^{-1}(y)$ for $f^{-1}(\{y\})$. In a simple language, $f(X)$ is the set of elements that X is mapped to, and $f^{-1}(Y)$ is the set of elements that map into Y. Picture (3.7) is there to help you visualize images and preimages. Suppose we think of A and B as boxes, and think of f is an arrow between them. Inside A is a box we call X, and inside B is a box we call Y.



(3.7)

The arrows in picture (3.7) indicate that the box X maps to the triangle $f(X)$ and that the square Y comes from the rectangle $f^{-1}(Y)$. In fact, *everything* that maps into Y is in the rectangle $f^{-1}(Y)$.

Here are more examples to illustrate these ideas.

Example 3.2.2 Let $a :$ People of Earth $\longrightarrow \mathbb{N}$ be defined as

$$a(x) \quad = \quad \text{age in years of the person } x \text{ on January 1, 2005.}$$

Then $a(\text{your author}) > 50$,

$$\begin{aligned} a^{-1}(15) &= \{\text{people } x \mid a(x) = 15\} \\ &= \{\text{people } x \mid x \text{ is 15 years old}\}. \end{aligned}$$

Suppose that x is some individual on Earth whose age is 50. Then $a(x) = 50$, and hence

$$\begin{aligned} a^{-1}(a(x)) &= a^{-1}(50) \\ &= \{\text{people } z \mid z \text{ is 50 years old}\}. \end{aligned}$$

Notice that $x \in a^{-1}(a(x))$ but that x is hardly everyone who is 50. We leave it to the reader to justify that $a^{-1}(200) = \emptyset$.

Example 3.2.3 In this example, let $t :$ places on Earth $\longrightarrow \mathbb{R}$ be the function that takes a place on Earth x and returns the temperature $t(x)$ in degrees Celsius of that place at 12 noon on January 1, 2005. Then $t(\text{outside}) = 20°\text{C}$,

$$\begin{aligned} t^{-1}(30°\text{C}) &= \{\text{places } x \mid t(x) = 30°\text{C}\} \\ &= \{\text{places } x \mid \text{ temperature is } 30°\text{C at } x\} \end{aligned}$$

and

$$t^{-1}(100°\text{C}) \quad = \quad \{\text{places } x \mid t(x) = 100°\text{C}\}.$$

You might also describe $t^{-1}(100°\text{C})$ as those places at which the temperature is the boiling point of water at sea level. Notice that $t^{-1}(100°\text{C})$ contains the places where people at sea level are boiling water. It also contains those places, like geysers and toaster ovens, at which the temperature is 100°C.

Example 3.2.4 An example from algebra is the following. Let $f(x) = x(x-1)(x-2)(x-3)(x-4)$. Then the rule f defines a function $f : \mathbb{R} \to \mathbb{R}$. The preimage $f^{-1}(0)$ is the set of all numbers x such that $f(x) = 0$. That is, $f^{-1}(0)$ is the set of solutions to

$$x(x-1)(x-2)(x-3)(x-4) = 0,$$

so that $f^{-1}(0) = \{0, 1, 2, 3, 4\}$.

Example 3.2.5 An example from trigonometry is given by the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(\theta) = \tan(\theta)$, where tan is the *tangent function.* We have already defined the tangent function for angles θ. (See page 70.) Then $f(0) = \tan(0) = 0$, while

$$f^{-1}(0) \quad = \quad \{\theta \in (-\pi/2, \pi/2) \mid \tan(\theta) = 0\} = \{0\}.$$

Here are some results on images and preimages that we are using as educational tools to prepare the reader for the more advanced arguments presented in the next chapter.

Theorem 3.2.6 *Let $f : A \longrightarrow B$ be a function. If $Y \subset Y' \subset B$ then $f^{-1}(Y) \subset f^{-1}(Y')$.*

Proof: We must show that every element of $f^{-1}(Y)$ is in $f^{-1}(Y')$. Let $x \in f^{-1}(Y)$. By definition of preimage, $f(x) \in Y$, and since $Y \subset Y'$, $f(x) \in Y'$. Then $x \in f^{-1}(Y')$ by definition of preimage. Hence $f^{-1}(Y) \subset f^{-1}(Y')$, which completes the proof.

Look at what we did in the above proof. We started with an element $x \in f^{-1}(Y)$ and we evaluated what that means to us. The definition of the preimage must be used. It tells us that x maps into Y. That is, $f(x) \in Y$. The subset hypothesis is now used. $f(x) \in Y \subset Y'$ implies that $f(x) \in Y'$. Since x maps into Y', x is in the set $f^{-1}(Y')$ of *all* elements that map into Y'. We write $x \in f^{-1}(Y')$ in this case. Hence $f^{-1}(Y) \subset f^{-1}(Y')$.

Notice the detail with which I am arguing. This detail is necessary for two reasons. Firstly, we cannot know what the other is thinking so I must show you all of the thoughts that I want you to think. Secondly, before we can skip steps without fear of error setting in, we must first pay our dues with this type of detail. I hope you are paying dues. The later material will more than make up for the present effort that I am asking you to exert.

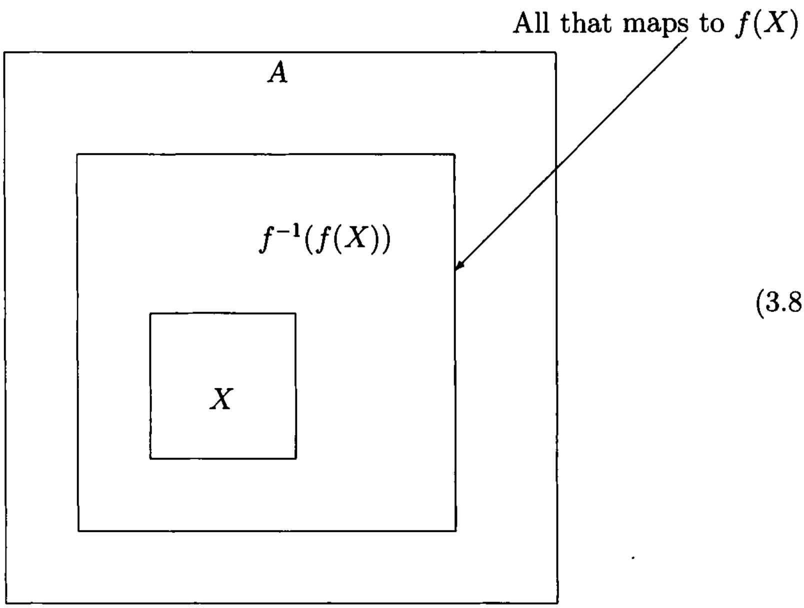
Theorem 3.2.7 *Let $f : A \longrightarrow B$ be a function and let $X \subset A$. Then $X \subset f^{-1}(f(X))$.*

Picture (3.8) is an illustration of what we will show in the proof of this theorem.

Proof: We must show that each element of X is an element of $f^{-1}(f(X))$. Let $x \in X$. By the definition of image $f(x) \in f(X)$. For the sake of clarity, let $W = f(X)$. Then $f(x) \in W$. The definition of preimage and the definition of W show us that

$$x \in f^{-1}(W) = f^{-1}(f(X)).$$

Thus, $X \subset f^{-1}(f(X))$, which completes the proof.



(3.8)

Once again we will detail and motivate the given proof. To prove that $X \subset f^{-1}(f(X))$, we must show that each given $x \in X$ is in $f^{-1}(f(X))$. Let $x \in X$. It must be clear to you by now that $f(x) \in f(X)$. Then x is a part of all that maps into $f(X)$. Put symbolically, $x \in f^{-1}(f(X))$. We conclude that $X \subset f^{-1}(f(X))$.

Example 3.2.8 Here are two examples of how $f^{-1}(f(X))$ can be much larger than X.

1. Consider the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = x^2$, and let $X = \{2\}$. Since $f(2) = 4$, $2 \in f^{-1}(4)$. The set $f^{-1}(4)$ is the set

of *all* x such that $x^2 = 4$. Inasmuch as $f(2) = f(-2) = 4$, we see that $\{-2, 2\} = f^{-1}(4)$. Thus, $X = \{2\} \neq \{-2, 2\} = f^{-1}(f(X))$.

2. Let a : People on Earth $\longrightarrow \mathbb{N}$ be the function that assigns to each person x on Earth his or her age as of January 1, 2005. Then $a(\text{your author}) = 50$, while $a^{-1}(50)$ is the set of all 50 year olds on the planet. Consequently,

$$a^{-1}(a(\text{your author})) = a^{-1}(50)$$

is quite a bit larger set than the set $\{\text{your author}\}$.

This is why the boxes in diagram (3.8) properly contain each other.

There is a relationship between subsets and the composition of functions. Notice the reversal of order in the functions f and g in the next result. This is also a test. The only question on this exam is: *Have you learned enough of the basics to understand a more terse mathematical proof?* Before you read on, remember that sets A and B are equal, $A = B$, exactly when $A \subset B$ and $B \subset A$. Nothing less than these two conditions may be proved if we wish to conclude that $A = B$. You will start with an $x \in A$ and prove that it is in B. Then you will start with an $x \in B$ and prove that it is in A. Notice that in the next several proofs, this is exactly what we are practicing.

Theorem 3.2.9 *Let $f : A \longrightarrow B$ and $g : B \longrightarrow C$ be functions, and let $Z \subset C$. Then*

$$(g \circ f)^{-1}(Z) \quad = \quad f^{-1}\left(g^{-1}(Z)\right).$$

Proof: To prove the equality, we must show that

$$(g \circ f)^{-1}(Z) \quad \subset \quad f^{-1}\left(g^{-1}(Z)\right)$$

and that

$$f^{-1}\left(g^{-1}(Z)\right) \quad \subset \quad (g \circ f)^{-1}(Z).$$

We will prove the inclusion $(g \circ f)^{-1}(Z) \subset f^{-1}\left(g^{-1}(Z)\right)$.

Let $x \in (g \circ f)^{-1}(Z)$. By definitions of composition and preimage,

$$g(f(x)) = (g \circ f)(x) \in Z.$$

Since $g(f(x)) \in Z$, $f(x) \in g^{-1}(Z)$, and then $x \in f^{-1}(g^{-1}(Z))$. Hence

$$(g \circ f)^{-1}(Z) \quad \subset \quad f^{-1}\left(g^{-1}(Z)\right).$$

Conversely, we will prove that $f^{-1}(g^{-1}(Z)) \subset (g \circ f)^{-1}(Z)$. Suppose that $x \in f^{-1}(g^{-1}(Z))$. We must show that $x \in (g \circ f)^{-1}(Z)$. Now $x \in f^{-1}(g^{-1}(Z))$ implies that $f(x) \in g^{-1}(Z)$ and hence that $g(f(x)) \in Z$. By definition of $\circ$ we have

$$(g \circ f)(x) = g(f(x)) \in Z,$$

so that

$$x \in (g \circ f)^{-1}(Z).$$

Consequently,

$$f^{-1}\left(g^{-1}(Z)\right) \subset (g \circ f)^{-1}(Z),$$

and therefore $f^{-1}(g^{-1}(Z)) = (g \circ f)^{-1}(Z)$. This completes the proof.

Finally, as an exercise in the use of the quantifiers *for all* and *there exists*, we will prove the following theorems. Please notice how the logic, the set theory, and the functions neatly interact.

Theorem 3.2.10 *Let $f : A \longrightarrow B$ be a function. If $\{Y_n \mid n \in \mathbb{N}\}$ is a family of subsets of B then*

$$\bigcup_{n \in \mathbb{N}} f^{-1}(Y_n) = f^{-1}\left(\bigcup_{n \in \mathbb{N}} Y_n\right).$$

Proof: In shorthand notation, we will first prove the inclusion

$$\bigcup_{n \in \mathbb{N}} f^{-1}(Y_n) \quad \subset \quad f^{-1}\left(\bigcup_{n \in \mathbb{N}} Y_n\right).$$

Let

$$x \in \bigcup_{n \in \mathbb{N}} f^{-1}(Y_n).$$

By the definition of $\cup$, $x \in f^{-1}(Y_m)$ *for some* $m \in \mathbb{N}$, so that $x \in f^{-1}(Y_m)$. By definition of preimage, $f(x) \in Y_m$, so that

$$f(x) \in \bigcup_{n \in \mathbb{N}} Y_n$$

by definition of $\cup$. The definition of preimage shows us that

$$x \in f^{-1}\left(\bigcup_{n\in\mathbb{N}} Y_n\right)$$

and hence that

$$\bigcup_{n\in\mathbb{N}} f^{-1}(Y_n) \quad \subset \quad f^{-1}\left(\bigcup_{n\in\mathbb{N}} Y_n\right).$$

Conversely, we will prove that

$$f^{-1}\left(\bigcup_{n\in\mathbb{N}} Y_n\right) \subset \bigcup_{n\in\mathbb{N}} f^{-1}(Y_n).$$

Let

$$x \in f^{-1}\left(\bigcup_{n\in\mathbb{N}} Y_n\right).$$

The definition of preimage implies that

$$f(x) \in \bigcup_{n\in\mathbb{N}} Y_n,$$

and the definition of $\cup$ implies that $f(x) \in Y_m$ *for some* $m \in \mathbb{N}$. Then $x \in f^{-1}(Y_m)$ for that $m \in \mathbb{N}$, and so

$$x \in \bigcup_{n\in\mathbb{N}} f^{-1}(Y_n).$$

Consequently,

$$f^{-1}\left(\bigcup_{n\in\mathbb{N}} Y_n\right) \quad \subset \quad \bigcup_{n\in\mathbb{N}} f^{-1}(Y_n)$$

and therefore

$$f^{-1}\left(\bigcup_{n\in\mathbb{N}} Y_n\right) \quad = \quad \bigcup_{n\in\mathbb{N}} f^{-1}(Y_n).$$

This completes the proof.

Exercise 3.2.11 Here are a few exercises that will build your mental mathematical muscles. The reader can sample the thrill of the creative and discovery processes by providing proofs of all these results. Rather than repeat the hypotheses over and over again, let us agree that, for each of these exercises, $f : A \longrightarrow B$ is a function, that $X, X' \subset A$, and that $Y, Y' \subset B$.

1. If $X \subset X' \subset A$ then $f(X) \subset f(X')$.
2. $f(f^{-1}(Y)) \subset Y$.
3. If $X, X' \subset A$ then $f(X \cap X') \subset f(X) \cap f(X')$.
4. If $X, X' \subset A$ then $f(X \cup X') = f(X) \cup f(X')$.
5. If $Y, Y' \subset B$ then $f^{-1}(Y \cap Y') = f^{-1}(Y) \cap f^{-1}(Y')$.
6. If $Y, Y' \subset B$ then $f^{-1}(Y \cup Y') = f^{-1}(Y) \cup f^{-1}(Y')$.
7. If $\{X_n \,|\, n \in \mathbb{N}\}$ is a family of subsets of A then $f(\bigcup_{n\in\mathbb{N}} X_n) = \bigcup_{n\in\mathbb{N}} f(X_n)$.
8. If $\{Y_n \,|\, n \in \mathbb{N}\}$ is a family of subsets of B then $\bigcap_{n\in\mathbb{N}} f^{-1}(Y_n) = f^{-1}\left(\bigcap_{n\in\mathbb{N}} Y_n\right)$.

3.3 One-to-One and Onto Functions

Functions come in a variety of different colors. For instance, $f(x) = x^2$ defines a function on $\mathbb{R}$ such that $f(-1) = f(1)$. That is, f takes two different numbers to the same number. The function $f(x) = x + 2$, on the other hand, takes different numbers $x \neq x'$ to different numbers $x + 2 \neq x' + 2$. When you stand in a particularly long line to see a movie and you count the number of people in front of you, you are forming a function f that associates a natural number n with a person in that line. Then $f(1)$ is the first person in line, $f(2)$ is the second person in line, and so on. In this function, we see that if $n \neq m$, then the people $f(n)$ and $f(m)$ are different. One person will not hold two different places in line.

This type of *counting* function goes back to the first humans to herd animals. As the goats would enter the pasture land, the ancient goatherd would drop a rock or stone in a pile. In this way, he sets up a function from the goats to the pile of stones. As the goats leave the pasture, he removes one stone for each goat that passes. If he has stones left, he might conclude that the wolf has found one of his goats. If there are more goats than stones, he might conclude that he has picked up another goat from somewhere. Different stones are associated with different goats. Thus, when the goats and the stones match up element by element, he concludes that he has just as many goats that evening as he had that morning.

The next definition identifies this property of functions.

Definition 3.3.1 *Let $f : A \longrightarrow B$ be a function.*

1. *We say that f is* a one-to-one function *if*

$$f(x) = f(x') \text{ implies that } x = x'$$

or equivalently if

$$x \neq x' \text{ implies that } f(x) \neq f(x').$$

Said in a colloquial manner, f is one-to-one if different elements of A map to different elements of B.

2. *We say that f is* an onto function *if*

$$\text{given } y \in B \text{ then we can find an } x \in A \text{ such that } f(x) = y.$$

In an informal way, f is onto if every element in B is the image of an element from A.

Consider what this definition represents. We have uncovered a pair of properties about functions that the reader probably has not encountered before. This discovery is the tip of an iceberg, or a general principle, that is called *abstraction*. Abstraction is one of the processes through which mathematics grows. For example, you can study lines on a sheet of paper and you will understand lines and linear processes. But once you abstract lines to linear approximation, you have discovered the calculus. Now you can leave that sheet of paper and study nature and the universe around us. If you allow your definition of line to be abstracted, you grow from points and lines on a plane to studying the geometry of points and lines on a sphere, or a globe. Lines here would be different because *they must curve along the sphere.* Quite a change, isn't it? Abstraction usually has this effect on mathematics.

Some examples will help you visualize one-to-one and onto functions.

Example 3.3.2 Start with a function $f : \{a, b\} \longrightarrow \{x, y, z\}$ defined by

$$f(a) = x \quad f(b) = y.$$

A glance at the definition shows us that $f(a) \neq f(b)$ so that f is a one-to-one function. f is not an onto function since $f(a), f(b) \neq z$. That is, z is not an image of an element from $\{a, b\}$.

Define a function $f : \{a, b, c\} \longrightarrow \{x, y\}$ by

$$f(a) = x \quad f(b) = x \quad f(c) = y.$$

Since $f(a) = x$ and $f(c) = y$, each element of the codomain $\{x, y\}$ is an image of an element from $\{a, b, c\}$. Then f is an onto function. Since the different elements $a \neq b$ map to the same element $f(a) = f(b) = x$, f is not a one-to-one function.

You may recall that we examined sets that could be *matched element by element.* The matching is a limited way of describing a function that is both one-to-one and onto. These one-to-one and onto functions will thus replace the previous notion of an element by element matching of sets.

Example 3.3.3 Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by $f(x) = 7x + 6$. We claim that f is both one-to-one and onto.

1. f is one-to-one: Suppose that $x, x' \in \mathbb{R}$ are such that $f(x) = f(x')$. Then $7x + 6 = 7x' + 6$ implies that $7x = 7x'$. We can divide by 7 to show that $x = x'$. As claimed, f is a one-to-one function.

2. f is onto: Suppose that $y \in \mathbb{R}$. We must find a number $x \in \mathbb{R}$ such that $f(x) = y$. Let $x = \dfrac{y-6}{7} \in \mathbb{R}$. A little algebra shows us that

$$f(x) = f\left(\frac{y-6}{7}\right) = 7\left(\frac{y-6}{7}\right) + 6 = y.$$

Then $f(x) = y$, so that f is an onto function, as claimed.

In the above example, we made use of the arithmetic of 7 and 6. We added and subtracted at will, and we divided by 7 like there was no problem with it. (There isn't.) But now let us abstract this problem. We will need to acknowledge the fact that we could divide by 7 because $7 \neq 0$. We also need to see that $6 - 6 = 0$. Let's see how that is used in a more abstract setting.

Example 3.3.4 Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by $f(x) = mx + b$ for some $m \neq 0$. We claim that f is both one-to-one and onto.

1. f is one-to-one: Suppose that $x, x' \in \mathbb{R}$ are such that $f(x) = f(x')$. Then

$$\begin{aligned} mx + b &= mx' + b \\ (mx + b) - b &= (mx' + b) - b \\ mx &= mx' \\ \frac{mx}{m} &= \frac{mx'}{m} \\ x &= x'. \end{aligned}$$

We could divide by m in the second to last step only because we chose $m \neq 0$. As claimed, f is a one-to-one function.

2. f is onto: Suppose that $y \in \mathbb{R}$. We must find a number $x \in \mathbb{R}$ such that $f(x) = y$. Let $x = \dfrac{y - b}{m}$. Since $m \neq 0$, we can divide by m to form the real number x. A little algebra shows that

$$f(x) = f\left(\frac{y - b}{m}\right) = m\left(\frac{y - b}{m}\right) + b = y.$$

Then as claimed, $f(x) = y$, so that f is an onto function.

Our choice of x in the above example is typical of mathematical proof. Before setting down to write this little mathematical essay, several drafts of the essay were made. We found the value $x = \dfrac{y - b}{m}$ before we set pen to paper to write out our proof. Our choice of x may at first seem magical, coming from nowhere, without motivation as it did, but that is because you were not there while I worked out all of the details in the drafts of this proof. That is the sign of a well-written well-constructed argument. You have done all of your homework surrounding the proof so that when you attempt to type that final draft, you can appeal to values as if they come naturally from the ether as opposed to the preparation. I will use this style of mathematics as little as possible, but I will use it.

Example 3.3.5 Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by $f(x) = x^2$. We will show that f is neither one-to-one nor onto.

1. f is not one-to-one: We need only produce two different numbers that map to the same number. $f(-2) = 4 = f(2)$ shows that f is not one-to-one. (Where did -2 and 2 come from, reader? If you write out a draft of this proof yourself, you might find their origins.)

2. f is not onto: We need only produce one element in $\mathbb{R}$ that cannot be written as $f(x)$. Since $x^2 \geq 0$ for all $x \in \mathbb{R}$, $f(x) = x^2 \neq -1$ for any $x \in \mathbb{R}$. (A little preparation gives us the value -1 as a good example of a number that is not an image under $f(x) = x^2$.)

A simple change of the domain or the codomain can significantly change the function. That is, by changing the domain and the codomain of a function that is neither one-to-one nor onto, we change the function into one that is both one-to-one and onto. As we said previously, these one-to-one and onto functions are the element by element matchings that we encountered earlier as functions for counting sets. Such a special function deserves a name. We use the French term.

Definition 3.3.6 *The function $f : A \longrightarrow B$ is a* bijection *if f is a* one-to-one *and* onto *function.*

Exercise 3.3.7 Which of the functions in exercises 1 through 5 is one-to-one and which is onto. Provide proof or justification for your answer.

1. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2 + 1$.

2. $f : \mathbb{R} \to \mathbb{R}^+$, $f(x) = x^4$.

3. $f : \mathbb{R} \to \mathbb{R}^+ \cup \{0\}$, $f(x) = x^4$.

4. $f : \mathbb{R}^+ \to \mathbb{R}$, $f(x) = \ln(x)$.

5. $f : \mathbb{R} \rightarrow \mathbb{R}^+$, $f(x) = e^x$.

6. Show that if $f : A \rightarrow B$ and $g : B \rightarrow C$ are one-to-one functions, then $g \circ f : A \rightarrow C$ is one-to-one.

7. Show that if $f : A \rightarrow B$ and $g : B \rightarrow C$ are onto functions, then $g \circ f : A \rightarrow C$ is onto.

8. Show that if $g \circ f : A \rightarrow C$ is a one-to-one function, then $f : A \rightarrow B$ is one-to-one.

9. Show that if $g \circ f : A \rightarrow C$ is an onto function, then $g : B \rightarrow C$ is onto.

3.4 Bijections

Some functions on finite sets will further develop your intuition about bijections, or one-to-one and onto functions.

The function $f : \{a, b, c\} \longrightarrow \{x, y, z\}$ defined by

$$f(a) = x, \quad f(b) = y, \quad f(c) = z$$

or by a list of assignments

$$\begin{cases} a & \longmapsto & x \\ b & \longmapsto & y \\ c & \longmapsto & z \end{cases}$$

is a bijection since a glance at the list shows us that different elements in the domain $\{a, b, c\}$ are mapped to different elements in $\{x, y, z\}$, and each element in the codomain $\{x, y, z\}$ is the image of an element from $\{a, b, c\}$. We might have said as in Chapter 1 that there is an element by element matching of $\{a, b, c\}$ with $\{x, y, z\}$. Our use of a bijection makes this intuitive matching of elements mathematically concrete.

Example 3.4.1 Example 3.3.3 shows that if $f : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by $f(x) = 7x+6$ then f is one-to-one and onto. Hence, $f(x) = 7x+6$ defines a bijection $f : \mathbb{R} \longrightarrow \mathbb{R}$.

Example 3.4.2 Example 3.3.4 shows that if $m \neq 0$ is a real number, and if $f : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by $f(x) = mx + b$, then f is one-to-one and onto. Thus, $f(x) = mx + b$ defines a bijection $f : \mathbb{R} \longrightarrow \mathbb{R}$.

Example 3.4.3 Let $\mathbb{R}^+$ be the set of positive real numbers, and let $f : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ be defined by $f(x) = \dfrac{1}{x}$. We will show that f is a bijection.

To show that f is one-to-one suppose we choose x, $x' \in \mathbb{R}^+$ such that $f(x) = f(x')$. The definition of f shows us that $\dfrac{1}{x} = \dfrac{1}{x'}$, and so a little algebra shows us that $x = x'$. Then f is one-to-one.

To show that f is onto begin with a $y \in \mathbb{R}^+$. We must choose an $x \in \mathbb{R}^+$ such that $f(x) = y$. Since $y \in \mathbb{R}^+$, $y \neq 0$, so we can form the real number $x = \dfrac{1}{y} \in \mathbb{R}^+$. Observe that

$$f(x) = f\left(\frac{1}{y}\right) = \frac{1}{1/y} = y$$

to conclude that f is onto. Therefore, f is a bijection.

Example 3.4.4 Recall that $\mathbb{R}^+$ is the set of positive real numbers. Let $g : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ be defined by $g(x) = x^2$. We will show that g is a bijection. That is, it is both one-to-one and onto.

g is one-to-one: Suppose that $x, x' > 0$ are such that $g(x) = g(x')$. Then $x^2 = (x')^2$. Since x, $x' > 0$, we can take square roots to see that

$$x = \sqrt{x^2} = \sqrt{(x')^2} = x'.$$

Thus, g is one-to-one.

The above chain of equations holds only because $x, x' > 0$. To see this, notice that $1^2 = (-1)^2$, while $1 \neq -1$. Only the *positive* numbers x satisfy $\sqrt{x^2} = x$.

g is onto: Let $y \in \mathbb{R}^+$. Since $y > 0$, we can let $x = \sqrt{y} \in \mathbb{R}^+$. Then

$$g(x) = (\sqrt{y})^2 = y,$$

which shows us that g is onto. (Where did that value x come from? Obviously, it is from the preparations we made before writing down the final draft of this proof.)

These examples concerning the rule x^2 should be compared. Even though the functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ have the same rule, namely, $f(x) = g(x) = x^2$, their domains and their codomains are different. This difference makes g both one-to-one and onto, while f is neither one-to-one nor onto. These must be different functions because they behave differently. They enjoy different properties. We conclude that the sets used to define a function will have a significant effect on the properties possessed by the function.

Example 3.4.5 Example 3.3.2 shows us that the function $f : \{a, b\} \longrightarrow \{x, y, z\}$ defined by

$$f(a) = x, \quad f(b) = y$$

is one-to-one but not onto. Then f is not a bijection.

However, the function $g : \{a, b\} \longrightarrow \{x, y\}$ defined by

$$g(a) = x, \quad g(b) = y$$

is a bijection. Notice that f and g have the same rule. The difference in their codomains makes them different functions.

Example 3.4.6 By Example 3.3.2 the function $f : \{a, b, c\} \longrightarrow \{x, y\}$ defined by

$$f(a) = x, \quad f(b) = x, \quad f(c) = y$$

is onto and not one-to-one. Then f is not a bijection.

However, the function $g : \{b, c\} \longrightarrow \{x, y\}$ defined by

$$g(b) = x, \quad g(c) = y$$

is a bijection. Again we see that a change in domain changes a function that is not one-to-one into a function that is one-to-one.

3.5 Inverse Functions

The operation of composition $g \circ f$ looks very much like a multiplication of functions. It is natural to ask if we can divide functions as well. Unfortunately, *the division of functions is not defined in general*, so we must examine an analogous operation, called *the inverse function.* We consider each function f as a rule that manipulates or that changes an input value x. The value $f(x)$ is seen then as some kind of variation on x. Our inverse functions will be this manipulation in reverse. If f squares, then its inverse g takes square roots. If f adds 3 to x, then g will subtract 3 from x. Do you see it? The inverse of f is the function that performs the opposite rule of computation that f performs.

The best solution in our search for an inverse function is to find the largest class of functions that have inverse functions. Specifically, we will discuss functions for which the composition division exists. Within this collection a specific function f will be associated with a function g such that

$$\boxed{\begin{aligned} g \circ f(x) &= g(f(x)) = x \text{ and} \\ f \circ g(x) &= f(g(x)) = x. \end{aligned}}$$

In this case we will say that g *is the inverse of* f. This inverse is not found by dividing by f but by *undoing* whatever it is that f does. It happens that this kind of functional inverse exists only for bijections.

Definition 3.5.1 *Let $f : A \longrightarrow B$ and $g : B \longrightarrow A$ be functions. Notice the change in the domain and codomain of f and g. Then g is called* the inverse of f *if*

1. *$g \circ f(x) = x$ for each $x \in A$, and*

2. *$f \circ g(y) = y$ for each $y \in B$.*

In this case, f is also the inverse of g. If f has an inverse then we say that f is invertible.

Observe that the compositions $g \circ f$ and $f \circ g$ are defined since their domains and codomains match up.

Example 3.5.2 Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by

$$f(x) = mx + b$$

for some numbers $m \neq 0$ and b. Define a function $g : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$g(x) = \frac{x - b}{m}.$$

Since $m \neq 0$, $g : \mathbb{R} \longrightarrow \mathbb{R}$ is a function. We will show that g is the inverse of f.

Let $x \in \mathbb{R}$. We must show that $g \circ f(x) = f \circ g(x) = x$. A little algebra shows us that

$$\begin{aligned} g \circ f(x) &= g(f(x)) \\ &= g(mx + b) \\ &= \frac{(mx + b) - b}{m} \\ &= \frac{mx}{m} \\ &= x \end{aligned}$$

and that

$$\begin{aligned} f \circ g(x) &= f(g(x)) \\ &= f\left(\frac{x - b}{m}\right) \\ &= m\left(\frac{x - b}{m}\right) - b \\ &= (x - b) - b \\ &= x. \end{aligned}$$

Thus, g is the inverse of f.

Now where do you suppose the inverse g of f came from in the above example? In preparing for this example, I assumed that $f(x) = mx + b$ had an inverse $g(x)$. Then I calculated as follows. Since $f(g(x)) = x$, we write

$$m \cdot g(x) + b = x.$$

Now solve for $g(x)$.

$$\begin{aligned} m \cdot g(x) + b &= x \\ m \cdot g(x) &= x - b \\ g(x) &= \frac{x - b}{m}. \end{aligned}$$

The only reason that $g(x)$ can be found in this case is because $m \neq 0$. If $m = 0$ then f is not a bijection and its inverse g cannot be found.

The next theorem shows that bijections and *invertible functions* (i.e., functions possessing an inverse) are the same functions. Try thinking of its proof as another examination of how far your mathematical thought processes have progressed over the last 10 pages or so. This approach is drier than the examples you have met in this chapter, and the presentation will cause you to think about my line of reasoning. The idea of choosing values x or y in some previous draft is used throughout this argument. However, all of the details are there, and if you can read it without too much difficulty, then you are ready for the next chapter. There we will begin our investigation of infinite sets.

Theorem 3.5.3 *Let $f : A \longrightarrow B$ be a function.*

1. *If f is invertible then f is a bijection.*

2. *If f is a bijection then f is invertible.*

Proof: 1. Suppose that f is invertible. We must show that f is one-to-one and onto. Since f is invertible, it has an inverse g.

Suppose that $f(x) = f(x')$ for some $x, x' \in A$. Then

$$g(f(x)) = g(f(x')),$$

and since g is the inverse of f we have

$$x = g(f(x)) = g(f(x')) = x'.$$

Then f is one-to-one.

Let $y \in B$, and let $x = g(y)$. Since g is the inverse of f, we have

$$f(x) = f(g(y)) = y,$$

so that f is onto. Therefore, f is a bijection.

2. Suppose that f is a bijection. We must *construct* a function g that will serve as the inverse of f. Define a rule g from B into A by

$$\boxed{g(x) = y \text{ exactly when } x = f(y).}$$

This is a reasonable definition for an inverse, but there are several properties that g must possess. To begin with, we must show that g is a function.

The definition of function (page 68) asks us to define g on all of B. So let $x \in B$. Since f is onto, there is a $y \in A$ such that $x = f(y)$. Then by definition of g, $g(x) = y$.

Also the definition of function askes us to show that $g(x)$ is exactly one value. To prove that there is exactly one of something you assume that you have two perhaps equal things, and then you prove that they are the same. So suppose there are two elements $y, y' \in A$ such that

$$g(x) = y \text{ and } g(x) = y'.$$

By the definition of g,

$$x = f(y) \text{ and } x = f(y')$$

and because f is one-to-one, $y = y'$. Thus $g(x)$ is exactly one value.

We have proved our first goal: $g : B \longrightarrow A$ is a function.

To complete the proof, we must show that $g(f(x)) = x$ and that $f(g(y)) = y$ for each $x \in A$ and $y \in B$.

Let $x \in A$, and write

$$f(x) = y.$$

Then by the definition of g,

$$x = g(y)$$

so that

$$g(f(x)) = g(y) = x.$$

On the other hand, if $y \in B$ is given, then write

$$g(y) = x.$$

By the definition of g, $y = f(x)$, so that

$$f(g(y)) = f(x) = y.$$

Hence, g is the inverse of f, which is what we had to prove.

Theorem 3.5.3 can be used to anticipate the existence of inverses without telling us what that inverse might be. For instance,

$$f(x) = mx + b \text{ where } m \neq 0$$

is known to be a bijection. Then Theorem 3.5.3 shows us that f in invertible, but it does not hint at the rule for the inverse of $f(x)$. We found (see Example 3.3.4) that

$$g(x) = \frac{x-b}{m}$$

is the inverse of f.

The function

$$h(x) : \mathbb{R}^+ \longrightarrow \mathbb{R}^+ \text{ such that } h(x) = x^2$$

is known to be a bijection. (See Example 3.4.4.) Then h has an inverse. Its inverse can be found by emulating the proof of the above theorem. Solve the equation

$$x = y^2$$

for y. Of course, we just take a square root, so that

$$\sqrt{x} = y.$$

Hence,

$$s(x) = \sqrt{x}$$

is the inverse of h.

Now that we have plowed through the preliminaries, let me present a story.

When we learn to play a game like checkers or chess, we play with a person learned in the ways of the game. As we play, we learn the rules over several initial games. It might take three or four games to learn the rules. There is a deeper understanding that comes with playing more games. The only way to learn the game well is to lose a lot of games to a better player. If we keep coming back to challenge this master of the game, we eventually will learn enough to challenge him or her. However, maybe you don't play games like checkers or chess. Maybe you like video games.

How do we learn to play video games? We start with a two dimensional character who moves across the screen prodded by our control box. Initially we play for 15 seconds and then we die. Do we give up? Do we go onto something else? No, we do not. We hit *reset* and begin again. This time we get a little further. Eventually, after resetting the game for what feels like an infinite number of times, we reach the end of the game. We defeat the Monster who was holding your Sister with the Golden Hair hostage, and we rejoice. We won the game. But it took a while. It took a lot of time.

I ask you to devote a fraction of that energy into the topics that we will cover in the next few chapters. The derived rewards will be far more than those enjoyed by checkers or video games, I assure you.

Exercise 3.5.4

1. Show that $f : \mathbb{R} \to \mathbb{R}^+$, $f(x) = e^x$ is a bijection.
2. Show that $f : \mathbb{R}^+ \to \mathbb{R}$, $f(x) = \ln(x)$ is a bijection.
3. Show that $e^x : \mathbb{R} \to \mathbb{R}^+$ and $\ln(x) : \mathbb{R}^+ \to \mathbb{R}$ are inverse functions.
4. If $f : A \to B$ is a bijection and if $g \circ f : A \to C$ is a bijection then $g : B \to C$ is a bijection.
5. Show that if $f : A \to B$ and $g : B \to C$ are one-to-one then $g \circ f : A \to C$ is a bijection.
6. Refer to the previous problem. Show that the inverse of $g \circ f$, usually written as $(g \circ f)^{-1}$, is equal to the function. $f^{-1} \circ g^{-1}$.

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

Chapter 4

Counting Infinite Sets

4.1 Finite Sets

The *natural numbers* are the numbers introduced by the Hindu and Arabic cultures. Today they are written as

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\}.$$

This is an implied list. The three dots, ..., indicate that the pattern continues indefinitely, or without stopping. Although this set appears to be small, it contains numbers that we do not encounter in our daily lives. Here are some examples of natural numbers that you have probably not met before.

1. The number of people on the Earth is about 7,000,000,000 or 7×10^9.

2. The bailouts funded by the federal government in 2010 amounted to over 700 billion, or 7×10^{11}, dollars.

3. The distance across the observable universe is about 15×10^9 light years. This is 15 followed by 9 zeros. It takes light 15×10^9 years to travel this distance. This number doesn't have a name, only a scientific significance.

4. The number of fish in the ocean on January 1, 2005 is quite large. Even though we can't write down this number, it is still a number and we can refer to it by the symbol, ϕ, pronounced *fee*.

5. The number of grains of sand on the Earth was first computed by Archimedes in about 100 BC. We will figure out how large this number is, but I think we can all agree that even though the number exists, it is impossible to write the exact number down.

Let us see how these numbers compare. To do this we will have to expand your imagination a bit. Once enlarged we will use the natural numbers $\{0, 1, 2, \ldots\}$ to write down some really large numbers. Suppose we write down the largest number that you can imagine, say,

$$X.$$

We can get a larger number by adding 1:

$$X + 1.$$

How big was X? Was it the number of people on Earth, 7 billion? Was it the speed of light, 186,000 miles/second? Maybe it was a *light year*, the distance light travels in a year, or about 1.6×10^9 miles. Now those were large numbers. But this is hardly big enough. We can produce a larger number by adding 1 to 7 billion.

$$7,000,000,000,001.$$

Not really much larger, though, is it? Maybe we can form an exponential number:

$$\left(7 \times 10^{12}\right)^2 = 49 \times 10^{24}.$$

This is larger than a light year since it has 24 zeros following the 36. We can make a larger number:

$$49 \times 10^{24} + 1.$$

Ones are not what we need here. I want numbers that grow so large that we cannot print them with meaning. Let me explain that with examples.

Now that you have seen how big these numbers can be, let us write down implied lists of these numbers called *sequences*. One such sequence starts with 100.

$$100, 101, 102, 103, 104, \ldots.$$

These numbers should not seem as large to you as they once might have been, but they are getting larger. They just grow slowly. Adding 1 moves us only a little bit when compared with our first number.

So let us take multiples of 100.

$$100, 200, 300, 400, \ldots.$$

The second number is twice as large as the first number, and the third number is three times as large as the first, etc. Once you've done this 100 times you have reached a number that is 100 times larger than 100. This is the number

$$100 \times 100 = 10,000.$$

These numbers get larger more quickly than the numbers do in the sequence 100, 101, 102, 103, ..., but they are still quite small compared with the numbers we can form.

The sequence

$$100, 100^2, 100^3, 100^4, \ldots \tag{4.1}$$

gets large relatively quickly. Each time we produce another entry in this sequence, we add 2 zeros to the previous entry. The first number has 2 zeros, the second number has 4 zeros, the third has 6, the fourth number

$$100,000,000$$

has 8 zeros, and so on. The number 100^2 is already as large as the hundreth entry in the sequence $100, 200, 300, 400, \ldots$. The hundreth entry in the sequence 100, 100^2, 100^3, 100^4, ... is 100^{100}, which is a 1 followed by 200 zeros. These numbers are getting larger *in magnitude* with each entry in the sequence. So things are quickly getting large, but maybe the values are not large enough.

For the next demonstration, we will write

$$100 = 10^2.$$

The reasoning here is that the exponential form is more easily manipulated. Consider the sequence

$$10^2, 10^{10^2}, 10^{10^{10^2}}, 10^{10^{10^{10^2}}}, \ldots$$

of exponents of 10. These numbers get large so fast that the second one, the number

$$10^{100} = 1 \quad \text{followed by a hundred zeros}$$

is too large to print comfortably in this space. The usual way of writing it down as $1,000,\ldots$ is utter folly since when written as a sequence of zeros, you cannot tell the different between 10^{100} and 10^{101} without a good deal of effort. No one can read that many zeros with meaning. Did we read 99 zeros or did we read 100 zeros? This number 10^{100} is larger than the number of grains of sand you would need to fill a bag as big as the Earth. (More on that sand number later.) It is hard to understand just how big the other numbers are in this sequence. And yet each of them is finite.

Now let us compare the sizes of the numbers in the different sequences. We will all agree that

$$10^{10} + 1 < 2 \times 10^{10} < 10^{20} < 10^{100}.$$

One way to see that a number x is large is to see how small its reciprocal $\frac{1}{x}$ is. The number 10^{100} is so large that if you used your calculator to find a *decimal equivalent* for the fraction

$$\frac{1}{10^{100}}$$

your display would probably read 0 or error. And yet this number is not 0, no matter what the display shows.

Now that we have some large numbers, we can examine how large the relatively small numbers 10^{100} and $10^{10^{100}}$ are. We will use the sand that makes up the Earth to give us all a mental picture of these numbers.

To write down the number of grains of sand on the Earth we will have to use *scientific notation.* Scientific notation is the use of powers of 10 to abbreviate numbers that have a daunting length. Certainly we can write down

$$10.$$

This number is large if you happen to be 9 years old. We can also write

$$10^2 = 100.$$

One hundred is a large number if you earned a 75 on your last mathematics examination, but it may not require the scientific notation 10^2. Larger numbers look like this:

$$10^{15} = 1{,}000{,}000{,}000{,}000{,}000.$$

You see that 10^{15} is a more compact way of writing 1 followed by 15 zeros. As indicated previously, $X = 10^{100}$ is a more practical method for writing down the number X. It is an example of a number that can only be written in a meaningful way as a power of 10.

Now that is one large number. And that is where we draw the line. One hundred zeros is beyond the patience of most people. Scientific notation prevails.

Small numbers can also be written in scientific notation. For instance,

$$\begin{aligned}
10^{-1} &= .1 \\
10^{-2} &= .01 \text{ notice the 1 zero.} \\
10^{-10} &= .0000000001 \text{ notice the 9 zeros.} \\
10^{-100} &= \text{a decimal point ``.'' followed by 99 zeros and then a 1.}
\end{aligned}$$

The only useful way to write down 10^{-100} is as a power of 10. In other words, there exist numbers that are only realized with scientific notation.

The reader may ask: *Why examine this? Surely* 10^{100} *and* 10^{-100} *are unique enough that they do not apply to anything in the real world.* This is not the case. For example, when banks transfer sums of money between themselves, they need to encode their transaction. They need to hide this little bit of business from the rest of the world. The numbers used to encode these transactions have more than 600 digits. That is, in order to transfer information between banks, a 600 digit number, a number that is much larger than 10^{100}, is required. Just how this encoding takes place is too far afield for our elementary intuitive discussion.

Another large number is found in chemistry. There are about 6×10^{23} atoms per *mole of matter.* All we need to know about *a mole of matter* is that it is a fundamental unit of measurement in chemistry. There are at least 10^{28} atoms in your body, and at least

$10^{80} = 10^{8\times 10}$ protons, neutrons, and electrons exist in the space of the observable universe. This is the fourtieth entry in the implied list $100, 100^2, 100^3, \ldots$ given by (4.1), or a 1 followed by 80 zeros. The point is, we have discovered something new. Numbers with up to 100 digits have a very practical purpose.

The story goes that two mathematicians and a child were sitting around the kitchen table one afternoon when one adult asks the other just how many grains of sand there are on the Earth. To make this problem something we can see in our mind's eye, suppose we have a typical grain of sand. The typical grain of sand has the following dimensions.

1. weight of one grain of sand $= 10^{-2}$ grams.

2. diameter of one grain of sand $= 10^{-4}$ meters.

We duplicate this grain many times over until we have a bag of sand that, in our minds eye, is the size of the Earth. By size I mean that the bag of sand and the Earth have the same size and shape and mass. The question we are asking now becomes *How many grains of sand do we need to fill a bag that large?*

The grain of sand is light, but not too light. It seems that 100 copies of this grain of sand weigh

$$100 \cdot 10^{-2} = 10^2 \cdot 10^{-2} = 1 \text{ gram.}$$

So 100 grains of sand would not make a large or heavy bag. Suppose we have 1 trillion or 10^{12} copies of this grain of sand. Then the mass of that bag of sand is

$$1 \text{ trillion } \cdot 10^{-2} = 10^{12} \cdot 10^{-2} = 10^{12-2} = 10^{10}$$

or 10 billion grams. Now that is a large number of grains of sand, and yet modern computer storage devices have many more than 10 billion bytes (10 gigabytes) of available storage. So this number is still something you might have in front of you right now. Suppose that we have 10^{102} copies of that grain of sand. Then the mass of that bag of sand is

$$10^{102} \cdot 10^{-2} = 10^{102-2} = 10^{100} \text{ grams,}$$

which it turns out is pretty close to the mass of the Earth, so we can say that there are about 10^{102} grains of sand on the Earth.

The mathematicians I mentioned were so impressed by this number, that they asked a nine month old child to name the number. The child said *googol*, and that name has stuck.

The number 10^{100} is called a *googol*.

Even though we have answered our question of just how many grains of sand make up the Earth, that is not the end of the story. As mathematicians will do, they asked for some kind of concrete realization of the number *googolplex*.

The number $10^{10^{100}}$ is called a *googolplex*.

Let us demonstrate just how large a googolplex is.

For this demonstration we will need a common object from around the house, toilet paper. But this roll of toilet paper has a special printing on it. On the first square of paper is a 1, and on each following square there is a 0. The roll contains enough squares so that the number 1 followed by all of those zeros is a googolplex. Just how large is that roll? It is so large that you would never run out of this household necessity. This roll could not be kept in your house; it is too large. You might try keeping it somewhere on your block. No. The roll is much larger than that. Perhaps the mayor of your city would allow you to use the city as a storage for this roll. This also is too small. The roll expands beyond the boundaries of any city. How about the country you live in? The roll is larger than that. Perhaps you would keep it on the North American continent. Nope. Too small for this roll. Let us jump and ask if the Earth will hold it. No. The Earth is too small to contain such a roll.

Please be patient. We have a long journey to make here. Perhaps the roll would fit into the solar system as encompassed by the orbit of the planets? This roll is too big to be kept in that space. Perhaps the Galaxy is large enough to encompass this roll? No. The roll will spill out into intergalactic space. How much larger should we go?

Is this roll large enough to fit into the observable universe? Now that would be one huge roll of toilet paper. And yet this roll is too large to fit into the observable universe.

We are faced with the following thought: *This roll is larger than our physical universe can hold.* That googolplex is one large number. It would seem that a googolplex is a number for which there is no physical significance. However, it is still a small value when compared with numbers like the exponential

$$\boxed{10^{\text{googolplex}}.}$$

I can't even begin to hint at how large that monster might be. These numbers are large but they are still finite. The numbers that we will encounter in the next section, called *infinite cardinals*, outstrip these natural numbers for size.

The numbers googol and googolplex seem to have struck a nerve with modern American culture. The numbers turn up in cartoons, movies, and a number of novels published in the United States. For example, there are duplexes, (two family homes), there is a cineplex, (a movie house), and there is a 10-plex, (a cineplex showing ten movies). There is a yellow colored cartoon character whose family goes to a googolplex for cinematic release. The marque for this movie house goes on into the clouds. Another example is a cowboy/scientist who has just lost his girl. He sits in a saloon moaning that she was one in a million, one in 10 billion, and one in a goolgolplex. Now that was one unique woman. These are the first examples I know of an obscure mathematical idea that appears in popular American culture. I believe there is also an internet tool named for it, but I cannot be sure.

Exercise 4.1.1 1. This old problem can be a good deal of fun. Take a 3×5 card, and try to write the largest number possible on it. You can write 1 or 1,000, but these are not large enough. You can write a 1 followed by a lot of zeros, (let us say 100 zeros), but that is not large enough, as 10^{600} is larger. Can you write a number on the card that is larger than any other number you can write on the card? Remember, ∞ is not a number.

4.2 Hilbert's Infinite Hotel

In Chapter 2, we started a process that would eventually define each natural number n from nothing. We will assume that we have constructed the set

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\}$$

of natural numbers from the empty set. To decide what an infinite set is, we first decide what a finite set is.

Definition 4.2.1 *Let $n \in \mathbb{N}$ and let A be a set. We say that A* has cardinality *n if there is a bijection $f : A \longrightarrow \{1, \ldots, n\}$. In this case, we write*

$$\boxed{\operatorname{card}(A) \quad = \quad n.}$$

The set A is finite *if $\operatorname{card}(A) = n$ for some $n \in \mathbb{N}$.*

Example 4.2.2 $\operatorname{card}(\{\bullet\}) = 1$ since every function $f : \{\bullet\} \longrightarrow \{1\}$ is a bijection.

The set $\{\bullet, :\}$ has cardinality 2 since the map

$$f : \{\bullet, :\} \longrightarrow \{1, 2\}$$

defined by

$$f(\bullet) = 2 \quad \text{and} \quad f(:) = 1$$

is a bijection.

Example 4.2.3 We claim that $\mathbb{N}$ is not a finite set. Even though this is an intuitively obvious fact, mathematicians require some kind of justification or proof. We will show that there are no bijections from $f : \mathbb{N} \longrightarrow \{1, \ldots, n\}$. To do this, we will show that no onto function $f : \mathbb{N} \longrightarrow \{1, \ldots, n\}$ is a one-to-one function. (In this way, f cannot be both one-to-one and onto.) To prove that f is not one-to-one, we must produce $x \neq x' \in \mathbb{N}$ such that $f(x) = f(x')$. That is, we find two *distinct* or *unequal* natural numbers x and x' that map to the same value.

Given an integer $n > 0$, suppose we are given an onto function $f : \mathbb{N} \longrightarrow \{1, \ldots, n\}$. By the definition of onto, there are $x_1, \ldots, x_n \in \mathbb{N}$

such that $f(x_1) = 1, \ldots, f(x_n) = n$. Since $\{x_1, \ldots, x_n\}$ is a finite set of numbers, it has a largest number. We will assume without loss of generality that this number is x_1. Since x_1 is the largest element and since $x_1 + 1 > x_1$, $x_1 + 1 \notin \{x_1, \ldots, x_n\}$. To simplify the notation, we will let $X = x_1 + 1$. Because $X \in \mathbb{N}$, $f(X) \in \{1, \ldots, n\}$. Hence, $f(X) = k = f(x_k)$ for some $k \in \{1, \ldots, n\}$. Since $X \notin \{x_1, \ldots, x_n\}$, $X \neq x_k$, so we have found number $X \neq x_k$ such that $f(X) = f(x_k)$. Therefore, f is not one-to-one.

Subsequently, there can be no bijections $f : \mathbb{N} \longrightarrow \{1, \ldots, n\}$. Since n was arbitrarily chosen (i.e., there is no special property about n other than $n \in \mathbb{N}$), $\mathbb{N}$ is not a finite set.

Let us look at that proof again. A specific n is called for. Suppose that $f : \mathbb{N} \longrightarrow \{1, \ldots, 101\}$ is a bijection. Then there are natural numbers $\{x_1, \ldots, x_{101}\} \subset \mathbb{N}$ such that $f(x_k) = k$. There is a largest number, say, $x_1 \in \{x_1, \ldots, x_{101}\}$. "What made you do that?" you ask. "Writing and rewriting this example enables me to do that," I respond. The number x_1 is just assumed to be the largest number in $\{x_1, \ldots, x_{101}\}$. There is at least one largest number, and we will denote it by x_1. *Why look there, why look for x_1? Tell me, please, the thoughts that inspired you to do that*, you ask. *My preparation for this proof is what inspired that choice*, I say. That and a number of years of experience. Because $x_1 < x_1 + 1 = X$, we conclude that $X \notin \{x_1, \ldots, x_{101}\}$. (This is because x_1 is the largest element in $\{x_1, \ldots, x_n\}$.) Thus, $X \neq x_k$ for any of the $x_k \in \{x_1, \ldots, x_{101}\}$. Meanwhile, however,

$$f(X) \in \{1, \ldots, 101\} = \{f(x_1), \ldots, f(x_{101})\}$$

so that $f(X) = f(x_k)$ for some $x_k \in \{x_1, \ldots, x_{101}\}$. We have shown that there is an x_k such that $X \neq x_k$ but such that $f(X) = f(x_k)$. These X and x_k are two different numbers in $\mathbb{N}$ that map under f to the same number, and so f is not one-to-one.

Congratulations! You have just learned *how* mathematicians count and *what* mathematicians count. We count elements in sets via bijections. If the object is not in a set, then we cannot count it. Moreover, you have seen how a mathematician does his work. He counts on his experiences, he counts on his mathematical instincts, and he counts on his mathematical inspirations. Do not discount the wonderful consequences of mathematical inspiration.

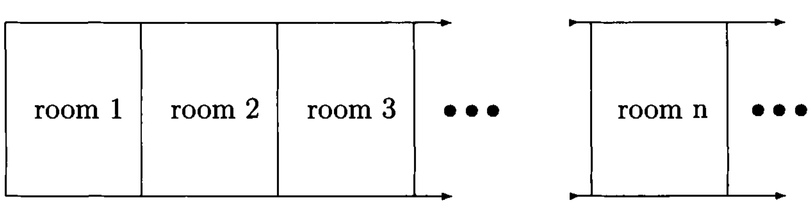
With this, we can define what mathematics means by infinity or by an infinite set.

Definition 4.2.4 *A set A is said to be* infinite *if A is not finite.*

Of course, we proved in Example 4.2.3 that $\mathbb{N}$ is not a finite set, so by our definition, $\mathbb{N}$ is an infinite set. Since $\mathbb{N} \subset \mathbb{Q}$, and since $\mathbb{N}$ is infinite, then $\mathbb{Q}$ is infinite. Since $\mathbb{Q} \subset \mathbb{R}$, $\mathbb{R}$ is infinite.

We will attempt to expand your horizons in preparation for a more mathematical treatment of infinite sets. The thought experiment that we will use is called *Hilbert's Infinite Hotel.* Do not try to think of this Hotel as a *hammer and nail* construction. It exists only in the fertile, creative, and imaginative minds of human beings. If you require your mathematics to be practical, you are about to get your mind significantly stretched.

Hilbert owns a unique hotel. It has just one floor, but it has *infinitely many rooms* in it.



The three dots • • • indicate that the room numbers continue in the same manner. Obviously, Hilbert's Hotel does not exist in our universe, but in some other universe of imagined mathematical objects.

Initially the Hotel is empty. A set $\{P_1, P_2\}$ of people, (people travel in sets in this universe), visits the Hotel. They want different rooms. So Hilbert assigns them rooms as follows.

room 1	room 2	room 3	room 4	•••
P_1	P_2			

(4.2)

Hilbert observes that infinitely many rooms $3, 4, 5, \ldots$ are vacant.

Next, a set $\{N_1, N_2, N_3, \ldots\}$ of Naturalists visits the Hotel. They are in town for a convention protesting the overuse of paradoxes involving infinity in modern mathematics. They all want private rooms. Hilbert hits upon the following scheme for assigning them rooms. Since rooms 1 and 2 are occupied Hilbert looks at picture (4.2) and assigns rooms as follows.

room 1	room 2	room 3	room 4	room 5	•••
P_1	P_2	N_1	N_2	N_3	

(4.3)

Notice that in general, the nth Naturalist N_n is put up in room $n + 2$ in picture (4.3). In this way, everyone has a single room and every room is occupied.

Next to arrive at Hilbert's Hotel is a lone stranger L who wants a private room. Hilbert studies picture (4.3) for a while and comes up with the following scheme. Every occupant will move down one room.

$$\begin{array}{ccc} \text{room } 1 & \longmapsto & \text{room } 2 \\ \text{room } 2 & \longmapsto & \text{room } 3 \\ \text{room } 3 & \longmapsto & \text{room } 4 \\ & \vdots & \end{array}$$

The result is summarized by the picture (4.4).

room 1	room 2	room 3	room 4	room 5	•••
	P_1	P_2	N_1	N_2	

(4.4)

In this way, everyone has a private room. As he anticipated, Hilbert sees that room 1 in picture (4.4) is unoccupied. (There is no room 0 to send to room 1.) Thus, Hilbert assigns the lone stranger L to room 1 as in picture (4.5).

room 1	room 2	room 3	room 4	room 5	•••
L	P_1	P_2	N_1	N_2	

(4.5)

Everyone has a single room, and every room is occupied as in picture (4.5). There are no vacancies at Hilbert's Infinite Hotel.

A troop of 1023 Boy Scouts arrives at the Hotel. After extracting a promise from them that they will not practice the art of making fire from two sticks, Hilbert assigns them rooms in his filled Hotel as follows. First he moves his guests.

$$
\begin{array}{lcl}
\text{room } 1 & \longmapsto & \text{room } 1023+1=1024 \\
\text{room } 2 & \longmapsto & \text{room } 1023+2=1025 \\
\text{room } 3 & \longmapsto & \text{room } 1023+3=1026 \\
 & \vdots &
\end{array}
$$

Draw this picture for yourself, reader. This empties rooms 1 through 1023 so he assigns scout n Room n. Each of the 1023 Scouts then has a single room. The Hotel is filled to capacity once again. I leave it to the reader to draw the pictures that relate to this assignment of rooms.

The next day everyone checks out. The Hotel is not empty for long, though, as an infinite set $\{S_1, S_2, S_3, \ldots\}$ of college students arrive for spring break. Hilbert assigns student S_n to room n. The Hotel is filled as in picture (4.6).

room 1	room 2	room 3	room 4	room 5		
S_1	S_2	S_3	S_4	S_5	•••	(4.6)

Just as Hilbert thinks that the worst of this continual stream of guests is over, an infinite set of Psychologists $\{X_1, X_2, X_3, \ldots\}$ arrives at Hilbert's Hotel. They are here for The Conference on the Theory of Group Therapy for Infinite Groups being held in the Hotel's Infinite Conference Room. Hilbert tells his staff to set up infinitely many chairs and a podium for the lectures, so now he has to assign these new guests rooms. Of course, everyone wants a single room. But the Hotel is filled. How can he do this? He decides to make room by making every other room vacant. Try to figure this one out yourself before reading on.

Here's how he does it. The student S_n in room n is sent to room $2n$ as follows.

$$\begin{array}{ccl} S_1 & \longmapsto & \text{room } 2 \\ S_2 & \longmapsto & \text{room } 4 \\ S_3 & \longmapsto & \text{room } 6 \\ & \vdots & \\ S_n & \longmapsto & \text{room } 2n \\ & \vdots & \end{array}$$

The room assignments in the Hotel look like picture (4.7).

room 1	room 2	room 3	room 4	room 5		
	S_1		S_2		•••	(4.7)

Each of the odd numbered rooms is empty, so Hilbert assigns the Psychologists rooms as follows.

$$\begin{array}{rcll}
X_1 & \longmapsto & \text{the first odd numbered room} & 1 \\
X_2 & \longmapsto & \text{the second odd numbered room} & 3 \\
X_3 & \longmapsto & \text{the third odd numbered room} & 5 \\
 & \vdots & & \\
X_n & \longmapsto & \text{the } n\text{th odd numbered room} & 2n\text{-}1 \\
 & \vdots & &
\end{array}$$

This gives each student a single room, and each odd numbered room is occupied as in the following picture.

room 1	room 2	room 3	room 4	room 5		
X_1	S_1	X_2	S_2	X_3	•••	(4.8)

In general, we are assigning rooms as follows. The rooms $2n - 1$ and $2n$ receive people X_n and S_n as follows.

	room $2n$-1	room $2n$	
•••	X_n	S_n	•••

Thus, each room is occupied by exactly one person. The Hotel is filled once again.

There is an interesting thing going on in these room assignments. Hilbert assigns rooms so that when he is done, his Hotel is filled to capacity. That is right. He fills the hotel so that each person has a room and each room is filled. So far he has been able to do that for any group that arrives, but shortly we will see that there is a large collection of Realists that cannot be housed in Hilbert's Hotel. I will not give details, only the punch line. Hilbert has filled his Hotel with everyone who has arrived so far, but he will be unable to accommodate all of the Realists.

The next day, all the guests at Hilbert's Infinite Hotel check out. Suddenly, infinitely many groups of Rationalists show up and require single rooms.

$$\begin{aligned} \mathbf{L}_1 &= \{A_{11}, A_{12}, A_{13}, \ldots\} \\ \mathbf{L}_2 &= \{A_{21}, A_{22}, A_{23}, \ldots\} \\ \mathbf{L}_3 &= \{A_{31}, A_{32}, A_{33}, \ldots\} \\ &\vdots \end{aligned}$$

Notice that

$$\boxed{A_{nm}}$$

is Rationalist m in implied list $\mathbf{L}_n$. Thus, A_{12} is Rationalist 2 in the implied list $\mathbf{L}_1$, A_{47} is Rationalist 7 in the implied list $\mathbf{L}_7$, and $A_{101\text{googol}}$ is Rationalist number googol in the implied list $\mathbf{L}_{101}$. We will demonstrate how to accommodate these guests so that each room is occupied and each Rationalist has a single room.

Write the names in each implied list $\mathbf{L}_n$ as row n in the infinite rectangular array (4.9).

$$\begin{array}{ccccc} A_{11} & A_{12} & A_{13} & A_{14} & \cdots \\ A_{21} & A_{22} & A_{23} & A_{24} & \cdots \\ A_{31} & A_{32} & A_{33} & A_{34} & \cdots \\ A_{41} & A_{42} & A_{43} & A_{44} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{array} \tag{4.9}$$

Draw parallel arrows starting at the elements A_{k1} and ending at the elements A_{1k} as in array (4.10).

$$
\begin{array}{ll}
\#1 & A_{11} \quad A_{12} \quad A_{13} \quad A_{14} \quad A_{15} \\
\#2 & A_{21} \quad A_{22} \quad A_{23} \quad A_{24} \quad A_{25} \\
\#3 & A_{31} \quad A_{32} \quad A_{33} \quad A_{34} \quad . \\
\#4 & A_{41} \quad A_{42} \quad A_{43} \quad \ddots \\
\#5 & A_{51} \quad A_{52}
\end{array}
\tag{4.10}
$$

These arrows can be used to define a bijection

$$f : \mathbb{N} \longrightarrow A$$

as follows. These arrows define a means of reading the array and of assigning rooms to the Rationalists. Hold the page at an angle if you must, but start reading the array along the arrows as though they were words to be read. The first thing you read is A_{11}. The next two things you read are A_{21} and A_{12}. The next things you would read would be A_{31}, A_{22}, A_{13}. The process continues indefinitely. In the more precise language of functions we would write

$$
\begin{array}{llll}
f(0) = A_{11} & & & \\
f(1) = A_{21} & f(2) = A_{12} & & \\
f(3) = A_{31} & f(4) = A_{22} & f(5) = A_{13} & \\
f(6) = A_{41} & f(7) = A_{32} & f(8) = A_{23} & f(9) = A_{14} \\
\quad \ldots & & &
\end{array}
$$

Trace these values for f through the first five arrows in the above array. We will show you that this rule f is the promised bijection. Try to plow through this argument. However, if you feel overwhelmed, then try to understand that in drawing these arrows we have intersected each of the entries of the array in such a way that each arrow intersects finitely many entries in the array, and that each entry is met by some arrow.

Since no Rationalist A_{nm} occupies two places in the array, f is one-to-one. Let $A_{nm} \in \mathbf{L}_n$. By the above scheme, we see that the subscripts lying along the first arrow add up to $1+1=2$, the subscripts lying along the second arrow add up to 3, the subscripts lying along arrow #3 add up to 4, and in general the subscripts lying along arrow $\#k$ add up to $k+1$. Thus A_{nm} lies along arrow $\#n+m-1$. Now consider the number that we will assign A_{nm} with f.

On arrow #1, there is 1 entry. On arrow #2, , there are 2 entries, and in general on arrow $\#n$, there are n entries. Thus, when we have finished counting the entries from A_{11} to A_{1n}, we will have counted a total of

$$\begin{array}{rccccccc} & 1 & + & 2 & + \ldots + & \ell & = & \dfrac{\ell(\ell+1)}{2} \\ \text{from arrow} & \#1 & & \#2 & \ldots & \#\ell & & \end{array}$$

entries. We will see why this sum is $\dfrac{\ell(\ell+1)}{2}$ in a little bit. For now, just accept it knowing that there is the promise of a proof. Meanwhile, back at the argument, we see that A_{nm} is on arrow $\#n+m-1$ so that

$$\text{there is some } k \leq \frac{(n+m-1)(n+m)}{2} \text{ such that } f(k) = A_{nm}.$$

This proves that $f(k) = A_{mn}$ for some number k, and this in turn proves that f is onto. Since we have already stated that f is one-to-one, $f : \mathbb{N} \longrightarrow A$ is a bijection. We would then send Rationalist A_{nm} to room k. The assignment of rooms is complete.

Let us take a break at this time and examine the sum $1+2+\ldots+\ell$. A German school teacher was feeling a bit hungry so he prepared to go out for coffee and a bagel. To keep his pupils busy, he instructed them to add up the numbers from 1 to 100. We will do this problem for an arbitray positive integer ℓ. A child, Carl F. Gauss, decided that there must be a faster way. In a flash of genius, the 11 year old Gauss discovered the following identity. Let us call the sum S.

$$S = 1+2+\ldots+\ell.$$

Find the sum in two different ways. Start at 1 and end at ℓ, and in the other start at ℓ and end at 1.

$$\begin{array}{rcccccccccc} S & = & 1 & + & 2 & + & \dots & + & \ell-1 & + & \ell \\ S & = & \ell & + & \ell-1 & + & \dots & + & 2 & + & 1 \end{array}$$

Notice that in each column, we have a sum

$$\ell+1.$$

Add these two sums together as follows.

$$\begin{array}{rcccccccccc} S & = & 1 & + & 2 & + & \dots & + & \ell-1 & + & \ell \\ +S & = & \ell & + & \ell-1 & + & \dots & + & 2 & + & 1 \\ \hline 2S & = & (\ell+1) & + & (\ell+1) & + & \dots & + & (\ell+1) & + & (\ell+1) \end{array}$$

There are ℓ terms $\ell+1$, so we have $2S = \ell(\ell+1)$ or equivalently

$$S = 1+2+\ldots+\ell = \frac{\ell(\ell+1)}{2}.$$

Quite a pretty little argument for a child, eh? Before the school teacher could wrap himself in his muffler, Carl turned his slate over and folded his hands. The teacher, thinking something was amiss, examined the slate and saw the correct sum written there. History does not record whether the school teacher got his coffee that day, or if he suffered for a lack of it.

The purpose behind the above stories is to develop an intuition for the infinite. The story about Hilbert's Infinite Hotel shows us that in an infinite set there is always *plenty of room*. It would seem that we can make room for any number of guests. Or can we?

The conventions and conferences are over so that the Hotel is empty. Hilbert breathes a sigh of relief and counts his revenue. Accommodating all of those guests has taxed him. There is a knock at the door, and Hilbert sees that an infinite collection of Realists

$$\{R_x \mid x \in \mathbb{R} \text{ and } 0 \leq x \leq 1\}$$

is in town for a continuing conference on complex issues. Hilbert sets out to assign each of them a private room, but his first attempt at assigning rooms leaves infinitely many Realists without a room. He tries again, and still he cannot assign every Realist a single room.

For some reason he does not yet fathom, no matter how he assigns rooms to the Realists, some Realists do not get a single room.

The mathematics behind the roomless Realists conundrum is a jolt to our *common sense* and will have to wait until Theorem 5.3.1. The problem of assigning rooms to Realists shows us that our *intuition* about finite sets is useless when we consider matters about infinite sets. We will have to develop entirely new insights if we are to understand the infinite. Hilbert does not yet have the mathematical experience to address this problem. He sends most of these potential guests packing.

Let us discover what Hilbert did not fathom. Let us see why there is no way to accommodate these Realists with single rooms in Hilbert's Infinite Hotel.

Suppose, for the sake of contradiction, that we can make an implied list of these guests and their single rooms. That is, we assign Realists to single rooms and then we make an implied list of the rooms n and the associated occupants R_{x_n}.

$$\begin{array}{ccrlllll}
\text{room} & \text{occupant} & \text{decimal} & & & & & \\
1 & R_{x_1} & x_1 = & .\boxed{d_{11}} & d_{12} & d_{13} & d_{14} & \cdots \\
2 & R_{x_2} & x_2 = & .d_{21} & \boxed{d_{22}} & d_{23} & d_{24} & \cdots \\
3 & R_{x_3} & x_3 = & .d_{31} & d_{32} & \boxed{d_{33}} & d_{34} & \cdots \\
4 & R_{x_4} & x_4 = & .d_{41} & d_{42} & d_{43} & \boxed{d_{44}} & \cdots \\
\vdots & \vdots & & & & & &
\end{array}$$

We have also given the decimal expansion of x_n. In this implied list, each d_{ij} is a digit, a number in the set $\{0, 1, \ldots, 9\}$, and we choose the shortest decimal expansion for x_n. That is, if the decimal expansion for x_n ends with $\overline{999}$, then we replace it with the number that ends in 0's. For instance, the following numbers are equal.

$$\begin{aligned}
1.00\bar{0} &= .\overline{999} \\
.100\bar{0} &= .0\overline{999} \\
.123\bar{0} &= .122\overline{999}.
\end{aligned}$$

In these cases, we choose the shorter finite decimal expansion, *except for* 1. The number 1 is written as a decimal $.\overline{999}$.

Furthermore, d_{11} represents the first digit in the decimal form of x_1, d_{23} represents digit number 3 in the decimal form of x_2, and in general, d_{nm} represents digit number m in the decimal form of x_n. We can now produce a Realist R_x who does not get a room.

Let us define a digit d_1 as follows.

$$d_1 = \begin{cases} d_{11}+1 & \text{if } d_{11} \neq 9 \\ 0 & \text{if } d_{11} = 9 \end{cases}.$$

Notice that we have defined d_1, so that $d_1 \neq d_{11}$. For example, if $d_{11} = 5$, then $d_1 = 6$, and if $d_{11} = 9$, then $d_1 = 0$.

We do the same thing for d_{22}. Let

$$d_2 = \begin{cases} d_{22}+1 & \text{if } d_{22} \neq 9 \\ 0 & \text{if } d_{22} = 9 \end{cases}.$$

Notice that we have defined d_2 so that $d_2 \neq d_{22}$. For example, if $d_{22} = 0$, then $d_2 = 1$, and if $d_{22} = 7$, then $d_2 = 8$.

In general, let us define

$$d_n = \begin{cases} d_{nn}+1 & \text{if } d_{nn} \neq 9 \\ 0 & \text{if } d_{nn} = 9 \end{cases}.$$

Then $d_n \neq d_{nn}$.

The roomless Realist R_x is given by writing down the decimal expansion

$$\boxed{x = .d_1d_2d_3d_4\ldots.}$$

If x ends in $\overline{999}$, then we replace it with the shorter decimal representation that ends in $\overline{000}$. Then $x \neq x_1$ because by our choice of d_1, x and x_1 differ in the first decimal place, $d_1 \neq d_{11}$. Also, $x \neq x_2$ because by our choice of d_2, x and x_2 differ in decimal place number 2, $d_2 \neq d_{22}$. In general, $x \neq x_n$ because by our choice of d_n, x and x_n differ in decimal place number n, $d_n \neq d_{nn}$. Thus, R_x is not on the implied list of Realists who got a room.

This is the person we have sought. This is why Hilbert could not give each Realist a single room. There are simply too many Realists! No matter what rooms he assigned, there would always be a Realist R_x who does not get a room. There are more Realists than there are rooms.

How can that be, reader? There are infinitely many rooms and infinitely many Realists. How can there be more Realists than rooms? Read on.

Let's examine the above choices for d_n and the construction of x on a concrete example. Suppose that we are given specific real numbers x_1, x_2, x_3, ... that form the following implied list.

$$\begin{array}{lcllllll}
x_1 & = & .\boxed{1} & 2 & 3 & 4 & 5 & \ldots \\
x_2 & = & .0 & \boxed{9} & 0 & 1 & 0 & \ldots \\
x_3 & = & .1 & 4 & \boxed{9} & 5 & 7 & \ldots \\
x_4 & = & .4 & 1 & 4 & \boxed{2} & 4 & \ldots \\
x_5 & = & .0 & 1 & 0 & 1 & \boxed{3} & \ldots \\
 & \vdots & & & & & &
\end{array}$$

Then

$$x = .2\ 0\ 0\ 3\ 4\ \cdots.$$

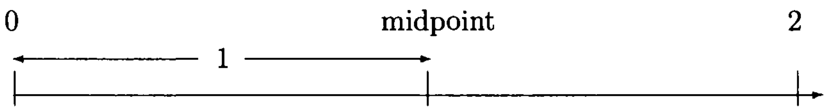
The first decimal of x is 2 because the first decimal of x_1 is 1.
The second decimal of x is 0 because the second decimal of x_2 is 9.
The second decimal of x is 0 because the third decimal of x_3 is 9.
The fourth decimal of x is 3 because the fourth decimal of x_4 is 2.
The fifth decimal of x is 4 because the fifth decimal of x_5 is 3.

The comparisons of x with the numbers on our implied list are obvious.

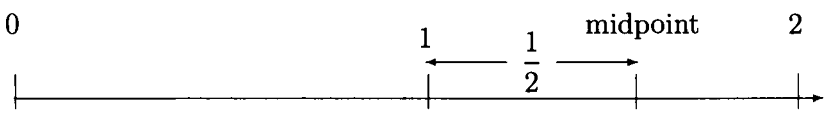
$x \neq x_1$ because they differ in decimal place 1,
$x \neq x_2$ because they differ in decimal place 2,
$x \neq x_3$ because they differ in decimal place 3.

These comparisons continue indefinitely. In this way, we show that x is not on the implied list $x_1, x_2, x_3, \ldots$.

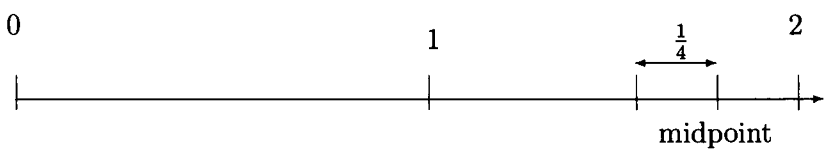
There is a person at the Hotel who haunts the Lobby. Each day he tries to walk a 2 meter long straight line across the Lobby floor. He starts at one end of the line and wants to get to the other end, but he has a peculiar way of walking. His next step is always half the length of his previous step. He starts with a step of one meter. One meter remains to be crossed.



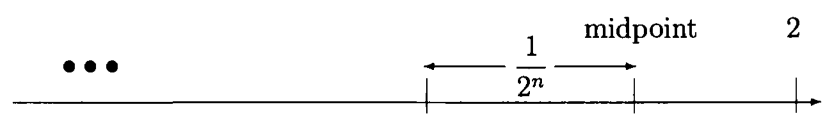
He follows this with a $\frac{1}{2}$ meter step. Half a meter remains.



Then he steps off a $\frac{1}{4} = \frac{1}{2^2}$ meter step. A $\frac{1}{4}$ meter remains.



Continuing on, on his nth step, he moves $\frac{1}{2^n}$ meters, and that leaves $\frac{1}{2^n}$ meters left to be covered. The scale is magnified.



He continues on indefinitely. Upon completion, there are 0 meters left to cover and so he has crossed

$$2 = 1 + \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^n} + \ldots$$

meters. He has walked the entire 2 meter length.

Each of the rooms at Hilbert's Infinite Hotel has to be cleaned and there is only one cleaning lady to do the job. Her name is Maria. (Hilbert's sister needed a job.) Her routine is to clean the rooms in order. She begins in the morning with room 1, followed by room 2, and then room 3, and on inductively. She is an industrious woman who picks up momentum as her day progresses, so she cleans the first room in 1 hour, the second room in $\frac{1}{2}$ hour, the next in $\frac{1}{2^2}$ hour, and so on, cleaning room n in $\frac{1}{2^n}$ hours. When she has cleaned each room, Maria is done. Thus she works

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} + \ldots = 2$$

hours a day. She certainly has a good job. She only has a 2 hour work day. But something else is wrong here. Where is she at the end of her day? There is no last room for her to clean because the Hotel has *infinitely many rooms.* However, we can be assured that each of the rooms is spotless. We will learn the fate of the cleaning lady when we finish our discussion of ordinals on page 225.

Exercise 4.2.5 1. Suppose that the Hotel is filled with infinitely many students $S_1, S_2, S_3, \ldots$. Two infinite groups $\{Y_1, Y_2, Y_3, \ldots\}$ and $\{Z_1, Z_2, Z_3, \ldots\}$ show up looking for rooms for the night. Show how to accommodate these new guests with single rooms.

2. The Hotel is empty when a group $\{U_1, U_4, U_9, \ldots, U_{n^2}, \ldots\}$ shows up. Assign them rooms so that the Hotel is filled.

3. Make a specific list of 12 real numbers $0 < x < 1$, and then use the argument concerning Realists to construct a number x that is not on your list.

4.3 Equivalent Sets and Cardinality

In this section, we will define the equivalence of cardinalities.

Definition 4.3.1 *Sets A and B are said to be* equivalent *if there is a bijection $f : A \longrightarrow B$.*

Theorem 4.3.2 *A and B are equivalent iff* $\text{card}(A) = \text{card}(B)$.

If we look at the different steps in Hilbert's Infinite Hotel, we see that in making space for the lone stranger Hilbert used the fact that the function

$$f : \{1,2,3,\ldots\} \longrightarrow \{2,3,4,\ldots\}$$

defined by

$$f(n) = n + 1$$

defines a bijection. Then $\{1,2,3,\ldots\}$ is equivalent to the proper subset $\{2,3,4,\ldots\}$.

In moving the students $\{S_1, S_2, S_3, \ldots\}$ into new rooms to accommodate all of the Psychologists $\{X_1, X_2, X_3, \ldots\}$, we are making use of the bijection

$$f : \{1,2,3,\ldots\} \longrightarrow \{2,4,6,\ldots\}$$

given by

$$f(n) = 2n.$$

Then $\{1,2,3,\ldots\}$ is equivalent to the proper subset $\{2,4,6,\ldots\}$.

One of the exercises asks you to prove a result due to Galileo Galilei. Galileo observed that *there are no more and no fewer perfect squares than there are natural numbers.* In our language, Galileo was observing that there is a bijection

$$f : \{1,2,3,\ldots\} \longrightarrow \{1,4,9,\ldots\}$$

given by

$$f(n) = n^2.$$

Then $\{1,2,3,\ldots\}$ is equivalent to its proper subset $\{1,4,9,\ldots\}$ of perfect squares. This property turns out to be a characteristic of infinite sets that can be used to define infinite sets without referring to finite sets.

Theorem 4.3.3 *The set A is* infinite *iff A is equivalent* to a *proper subset* of *itself.*

For example, by the functions f that we discussed above, $\mathbb{N}$ is an infinite set.

Example 4.3.4 Let us show that $\mathbb{Z}$ is equivalent to $\mathbb{N}$. This may seem strange. $\mathbb{N}$ is unbounded to the right, while $\mathbb{Z}$ is unbounded in both directions. A bijection between them exists nonetheless.

Define a rule $f : \mathbb{Z} \longrightarrow \mathbb{N}$ by writing

$$f(z) = \begin{cases} 2z & \text{if } z \geq 0 \\ -1 - 2z & \text{if } z < 0 \end{cases}$$

for all $z \in \mathbb{Z}$. *Now where did that come from?* you ask. *From a great deal of preparation*, I respond. Let us get a feel for the rule given here. For $z \geq 0$ we have

$$f(0) = 0, f(1) = 2, f(2) = 4,$$

and in general, each nonnegative integer n is sent to the even number

$$f(n) = 2n.$$

The negative $n \in \mathbb{Z}$ are a different matter.

$$\begin{aligned} f(-1) &= -1 - 2(-1) = 1 \\ f(-2) &= -1 - 2(-2) = 3 \\ f(-3) &= -1 - 2(-3) = 5 \end{aligned}$$

In general, each negative integer $-n$ is sent to the odd number

$$f(-n) = 2n - 1.$$

We casually observe that f is onto since an integer is either even or odd. Furthermore, f is one-to-one. This will require us to look at a few cases. Let $z \neq z' \in \mathbb{Z}$.

1. If $z \neq z' \geq 0$ then $f(z) = 2z \neq 2z' = f(z')$.

2. If $z \neq z' < 0$ then $f(z) = -1 - 2z \neq -1 - 2z' = f(z')$.

3. In the last case, we can assume without loss of generality that $z < 0 < z'$. Then $f(z)$ is odd and $f(z')$ is even so that $f(z) \neq f(z')$.

In any case, $f(z) \neq f(z')$ so that f is one-to-one. Hence f is a bijection, and consequently $\mathbb{Z}$ is equivalent to $\mathbb{N}$. Subsequently, since $\mathbb{Z}$ is equivalent to its proper subset $\mathbb{N}$, $\mathbb{Z}$ is infinite.

Example 4.3.5 We will show that $\mathbb{R}$ is equivalent to the proper subset $\mathbb{R}^+$ of positive real numbers, thus proving that $\mathbb{R}$ is infinite. This may seem obvious, but remember that every statement in mathematics must be accompanied by a justification or proof. Besides, the exercise prepares us for more complex arguments later.

The bijection $f : \mathbb{R} \longrightarrow \mathbb{R}^+$ is given by

$$f(x) = 2^x.$$

The argument requires a little high school algebra involving the logarithm $\log_2(x)$. All you have to know about $\log_2(x)$ is that

$$\log_2(2^x) = x = 2^{\log_2(x)}.$$

To see that f is one-to-one, suppose that $f(x) = f(x')$. Then

$$2^x = 2^{x'}.$$

Apply $\log_2$ to each side of this equation to find that

$$\log_2(2^x) = \log_2(2^{x'}),$$

so that $x = x'$. Then $f : \mathbb{R} \longrightarrow \mathbb{R}^+$ is one-to-one.

To see that f is onto let $y \in \mathbb{R}^+$. Then $x = \log_2(y)$ is a number such that

$$f(x) = 2^x = 2^{\log_2(y)} = y.$$

Then $f : \mathbb{R} \longrightarrow \mathbb{R}^+$ is onto, thus f is a bijection, and hence $\mathbb{R}$ is equivalent to $\mathbb{R}^+$.

Example 4.3.6 We will show in Theorem 5.3.1 that the function

$$\tan(\theta) : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \longrightarrow \mathbb{R}$$

is a bijection. Then $\mathbb{R}$ is equivalent to its proper subset $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, hence $\mathbb{R}$ is infinite, and therefore $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is infinite. The algebraically educated reader should try to recognize and then fill in all of the details in this example.

We present two classic examples due to Georg Cantor, which will demonstrate that our intuition about infinity is inaccurate. These examples will present ideas that are contrary to what we learned as children, if indeed we learned anything about infinity as children. We begin with an alternative approach to cardinality.

Let A be an infinite set. The cardinality of A can be defined as follows.

$$\boxed{\text{card}(A) \;=\; \text{the sets } B \text{ that are equivalent to } A.}$$

In this case, we call card(A) a *cardinal number* or a *cardinal.* It is traditional to use the first letter of the Hebrew alphabet $\aleph$ to designate a general cardinal. Thus, $\aleph$ is a cardinal exactly when there is a set A such that $\aleph = \text{card}(A)$. Although this appears to be different from the original definition of cardinal number, it defines the same thing.

For example, card$(\emptyset)$ is the collection of sets that have no elements. Naturally $\emptyset$ is the only such set, so we will use *the symbol* 0 to denote card$(\emptyset)$.

$$\boxed{0 \;=\; \text{card}(\emptyset).}$$

Since any bijection

$$f : \{\bullet\} \longrightarrow X$$

is onto, we can write

$$X = \{f(\bullet)\}$$

or equivalently

$$X = \{x\} \text{ for some element } x.$$

Then it is natural to use *the symbol* 1 to denote card$\{\bullet\}$.

$$\boxed{1 \;=\; \text{card}\{\bullet\}.}$$

A tradition started by Georg Cantor [1] is to write

$$\boxed{\aleph_0 \quad = \quad \mathrm{card}(\mathbb{N}).}$$

Thus, given a set A, then $\mathrm{card}(A) = \aleph_0$ iff A is equivalent to $\mathbb{N}$, iff there is a bijection $f : A \to \mathbb{N}$. We say that A is a *countably infinite set*, or simply that A is *countable*, if $\mathrm{card}(A) = \aleph_0$.

Exercise 4.3.7 1. Show that $\{f \mid f : \mathbb{N} \to \{0,1\} \text{ is a function}\}$ is an infinite set.

2. Show that $\mathbb{R}^2$, the Euclidean plane, is an infinite set.

3. Show that $\{1, 10, 100, \ldots\}$ is an infinite set.

4. Show that $\mathrm{card}(\{1, 10, 100, \ldots\}) = \aleph_0$.

5. Let $A = \{(0, n) \mid n \in \mathbb{N}\}$. Show that $\mathrm{card}(A) = \aleph_0$.

6. Show that $\mathrm{card}(\mathbb{N} \times \mathbb{N}) = \aleph_0$. Recall the Rationalists in Hilbert's Infinite Hotel.

7. Show that $\mathrm{card}(A) = \mathrm{card}(A)$ by producing a bijection $f : A \to A$.

8. If $\mathrm{card}(A) = \mathrm{card}(B)$ then $\mathrm{card}(B) = \mathrm{card}(A)$. Use bijections to do this exercise.

9. If $\mathrm{card}(A) = \mathrm{card}(B)$, and if $\mathrm{card}(B) = \mathrm{card}(C)$, then $\mathrm{card}(A) = \mathrm{card}(C)$. Use bijections to do this exercise.

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

Chapter 5

Infinite Cardinals

As we defined in the previous chapter, the *cardinality of a set* A, card(A), is the class of all sets B that are *equivalent to* A. For a finite set $A = \{1, \ldots, n\}$, card(A) is the collection of sets that contain exactly n elements. Then card(A) is a good way to describe the size of A, *even when* A *is infinite.* We might naively say that if A and B are infinite sets, then $\text{card}(A) = \text{card}(B)$. However, we showed in Chapter 3 that Hilbert's Infinite Hotel, whose rooms are numbered $1, 2, 3, \ldots$ cannot fit the Realists whose group is indexed by real numbers $(0, 1) = \{x \in \mathbb{R} \,|\, 0 < x < 1\}$. In another light, there is no bijection from the set $(0, 1)$ onto the set of natural numbers $\mathbb{N}$. If such a bijection does not exist, then we must conclude intuitively and accurately that

card($\mathbb{N}$) and card$(0, 1)$ are different cardinalities.

This should come as a shock to your common sense. Both cardinals are infinite, so how can they be different values? The inequality $\text{card}(\mathbb{N}) \neq \text{card}(\ (0, 1)\ )$ is True enough, so we have learned that our common sense is wrong when we deal with infinity. This is an important point. We do not have any common sense when it comes to infinite sets. We have to develop a new sense for size by relying on the mathematical properties of infinite sets. Once this is done, we can find some new intuitive feelings about infinity.

We will find in this chapter that there are two kinds of infinite sets: the *countable* ones and the *uncountable* ones. Any set equivalent to $\mathbb{N}$ is said to be a *countable set*, so we say that $\text{card}(\mathbb{N})$ is a *countable cardinal.* If S is an infinite set, and if $\text{card}(\mathbb{N}) \neq \text{card}(S)$, then S is said to be *uncountable.* We will show that $\text{card}(\mathbb{N}) \neq \text{card}(\ (0,1)\ )$, so the set $(0,1)$ is uncountable. Other uncountable sets include $\mathcal{P}(\mathbb{N})$ and $\{0,1\}^{\mathbb{N}}$ uncovering the mathematical fact there are several uncountable sets. We end this chapter with a proof of the somewhat amazing fact that there is *an infinite chain of infinite cardinals.* We might colloquially rephrase this by saying that *there are infinitely many infinites.*

5.1 Countable Sets

A set A is said to be *countable* if A is finite or if $\text{card}(A) = \text{card}(\mathbb{N})$. Equivalently, A is countable if A is equivalent to some subset of $\mathbb{N}$. Countable sets are appropriately named since these are the sets A that we can write as an implied list

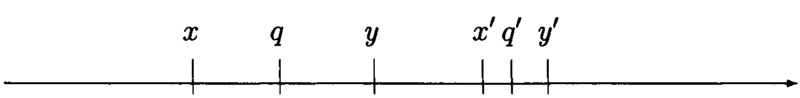
$$A = \{a_1, a_2, a_3, \ldots\}$$

of possibly infinite length. In any countable set, we can use elements the of $\mathbb{N}$ to write the elements of A as an implied list. Indeed, this can be taken as the definition of a countable set. The sets $\emptyset$, $\{1,2,3\}$, and $\{2,4,6,\ldots\}$ are countable. There are many countable sets, and we will give plenty of examples of them in this section.

Let

$$\boxed{\mathbb{Q}^+ \quad = \quad \left\{ \frac{n}{m} \,\middle|\, n,m > 0 \text{ and } n,m \in \mathbb{Z} \right\}}$$

be the set of *positive rational numbers.* It is a fact (not proved here) that between any two positive real numbers $x < y$ there is a positive rational number q. If you pick two different real numbers $x' < y'$, then there is another rational number q' between them. Let's draw that picture.



Thus, between 0 and $\frac{1}{\pi^2}$ there is a positive rational number. You can check that one such number is $\frac{1}{4^2}$. Compare this with the fact that the distance between two different *natural* numbers is at least 1. We say that $\mathbb{Q}^+$ is a *dense subset* of $\mathbb{R}^+$, while $\mathbb{N}$ is a *discrete subset* of $\mathbb{R}^+$. Another way to think of this property of $\mathbb{Q}^+$ is that each real number can be approximated by a rational number to any degree of accuracy. We are familiar with the notion of the density of the rationals since each calculator readout demonstrates that each real number x can be approximated by a number with a finite decimal expansion. Finite decimal numbers are rational numbers. For example, the readout on your calculator for $\sqrt{2}$ is not completely accurate even though it may give 64 decimal places of $\sqrt{2}$. The calculator value for $\sqrt{2}$ is accurate only to those few decimal places. The reason is that the number $\sqrt{2}$ is neither a finite decimal nor a repeating decimal. An infinite number of decimal places is necessary to write down $\sqrt{2}$ with complete accuracy, but we can approximate $\sqrt{2}$ with rational numbers.

We might be led to believe that $\mathbb{N}$ and $\mathbb{Q}^+$ are not equivalent since $\mathbb{N}$ is more thinly distributed in $\mathbb{R}^+$ than is $\mathbb{Q}^+$. That is, we might suspect that $\text{card}(\mathbb{N}) \neq \text{card}(\mathbb{Q}^+)$. However, the next theorem, due to Georg Cantor, shows us that $\mathbb{Q}^+$ is a *countable* set. Thus the distribution or density of numbers on the real line does not necessarily reflect cardinality.

Theorem 5.1.1 *$\mathbb{Q}^+$ is a countable set.*

Proof: We must produce a bijection $f : \mathbb{N} \longrightarrow \mathbb{Q}^+$. Begin by enumerating $\mathbb{Q}^+$ as the infinite rectangular array (5.1).

$$\begin{array}{ccccc} \frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\ \frac{2}{1} & \frac{2}{2} & \frac{2}{3} & \frac{2}{4} & \cdots \\ \frac{3}{1} & \frac{3}{2} & \frac{3}{3} & \frac{3}{4} & \cdots \\ \frac{4}{1} & \frac{4}{2} & \frac{4}{3} & \frac{4}{4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{array} \tag{5.1}$$

Notice that the numbers in Row 1 have numerator 1, the numbers in Row 2 have numerator 2, and in general the numbers in Row n have numerator n. Similarly, the denominator of the fractions in Column 1 is 1, the denominator of the fractions in Column 2 is 2, and in general the fractions in Column m have denominator m. Since each number in $\mathbb{Q}^+$ is a fraction $\frac{n}{m}$, the array (5.1) contains all of the numbers in $\mathbb{Q}^+$. (Which Row and Column contain that last fraction?)

Delete from this array any fractions that are not in reduced form. For example, we will delete the fractions

$$\frac{3}{3}, \frac{4}{2}, \frac{9}{27}, \frac{212}{424}$$

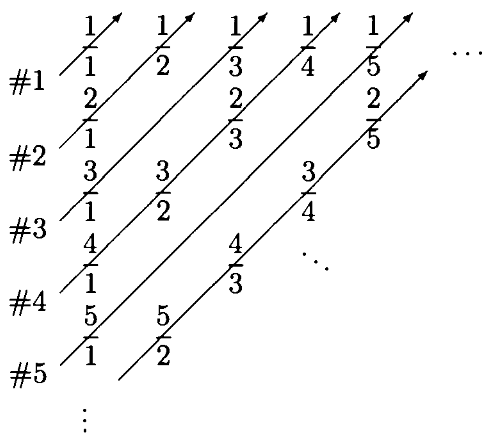
as well as any other fraction that can be reduced. The remaining fractions cannot be reduced. Some of those fractions include

$$\frac{1}{1}, \frac{1}{2}, \frac{19}{27}, \frac{101}{53}.$$

A portion of the resulting array is

$$\begin{array}{ccccc}
\frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\
\frac{2}{1} & & \frac{2}{3} & & \cdots \\
\frac{3}{1} & \frac{3}{2} & & \frac{3}{4} & \cdots \\
\frac{4}{1} & & \frac{4}{3} & & \cdots \\
\vdots & \vdots & \vdots & \vdots &
\end{array} \tag{5.2}$$

At this point we introduce arrows into array (5.2). The arrow #1 passes through $\frac{1}{1}$ only. The arrow #2 starts at $\frac{2}{1}$ and extends to $\frac{1}{2}$. In general, arrow #n starts at $\frac{n}{1}$ on the left edge of the array and extends to $\frac{1}{n}$ at the top edge of the array. The array (5.3) will help you envision this process.



(5.3)

These arrows give us a means of writing down a bijection $f : \mathbb{N} \longrightarrow \mathbb{Q}^+$ as follows. We will define f by reading the fractions as they appear along the arrows.

$$
\begin{aligned}
\textbf{LIST}: \quad & f(0) = \frac{1}{1}, \\
& f(1) = \frac{2}{1}, \quad f(2) = \frac{1}{2}, \\
& f(3) = \frac{3}{1}, \quad f(4) = \frac{1}{3}, \\
& f(5) = \frac{4}{1}, \quad f(6) = \frac{3}{2}, \qquad f(7) = \frac{2}{3}, \quad f(8) = \frac{1}{4}, \\
& \quad \vdots
\end{aligned}
$$

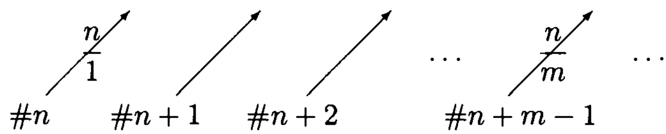
Before reading on, try to use arrows #5 and #6 to fill in the next two rows of this rule.

You can peek now. The next three arrows will produce the values

$$
\begin{aligned}
& f(9) = \frac{5}{1}, \quad f(10) = \frac{1}{5}, \\
& f(11) = \frac{6}{1}, \quad f(12) = \frac{5}{2}, \quad f(13) = \frac{4}{3}, \quad f(14) = \frac{3}{4}, \quad f(15) = \frac{2}{5}, \quad f(16) = \frac{1}{6}, \\
& f(17) = \frac{7}{1}, \quad f(18) = \frac{5}{3}, \quad f(19) = \frac{3}{5}, \quad f(20) = \frac{1}{7}
\end{aligned}
$$

for the function f. Now that we have some intuition about the rule f, our next task is to show that f is a bijection.

To see that f is onto let $\frac{n}{m}$ be on the array. Then $\frac{n}{m}$ is in Row n and Column m and the number $\frac{n}{1}$ begins the nth arrow in our array as in the picture (5.4).



(5.4)

By shifting to the right one place, we shift to the next arrow. It is clear from the diagram that $\frac{n}{m}$ is on arrow $\#n+m$. How many numbers have we counted when we finish arrow $\#n+m$? On arrow #1 we counted 1 number, on arrow #2 we counted 2 more, and in general on arrow $\#\ell$ we counted *at most* ℓ fractions. Subsequently, when we have finished counting along arrow $\#n+m$, we have counted *at most* a total of

$$1+2+\ldots+(n+m-1) \quad = \quad \frac{(n+m-1)(n+m)}{2}$$

fractions. You may remember this sum as the identity found by an eighteenth century elementary school student Carl F. Gauss. See page 124. Since sometime before we reach the end of arrow $\#n+m-1$ we will reach $\frac{n}{m}$, there will be some number

$$k \leq \frac{(n+m-1)(n+m)}{2}$$

such that

$$f(k) = \frac{n}{m}.$$

For example, before we reach the end of arrow #7, we will count at most

$$1+2+\ldots+7 = \frac{7(7+1)}{2} = 28$$

fractions. Then there is some number $k \leq 28$ such that $f(k) = \frac{3}{5}$. The reader will find by reading Row 7 of the above **LIST** that $f(19) = \frac{3}{5}$. That is, $\frac{3}{5}$ is the nineteenth fraction on **LIST**. Hence, f is onto.

It is clear that different natural numbers k and k' are sent to different fractions $f(k)$ and $f(k')$ on the **LIST**, so f is one-to-one. Thus, f is a bijection, and therefore $\mathbb{N}$ is equivalent to $\mathbb{Q}^+$. This completes the proof.

The inclusions

$$\{2, 4, 6, \ldots\} \subset \mathbb{N} \subset \mathbb{Q}^+$$

can be used to intuitively deduce that

$$\text{card}(\{2, 4, 6, \ldots\}) \leq \text{card}(\mathbb{N}) \leq \text{card}(\mathbb{Q}^+).$$

But the compelling work that we have done to date allows us to deduce the stronger thought

$$\boxed{\text{card}(\{2, 4, 6, \ldots\}) = \text{card}(\mathbb{N}) = \text{card}(\mathbb{Q}^+).}$$

Some might try to say that because $\mathbb{N}$ and $\mathbb{Q}^+$ are infinite they must be equivalent. This represents old common sense so it has to be rejected as an argument. Some of our subsequent work will show why that kind of reasoning is not a part of mathematics. Remember that card(A) is the set of all sets B that are equivalent to A, and that the cardinalities card(A) and card(B) are equal only when we can produce a bijection between their elements A and B.

The next result shows us that *the countable union of countable sets is countable.* We will start with a family of sets

$$A_1, A_2, A_3, \ldots$$

such that each A_k is itself a countable set. For instance, we can take set A_1 and list it as

$$A_1 = \{a_{11}, a_{12}, a_{13}, \ldots\},$$

where a_{12} represents element 2 in set A_1, and where $a_{1,11}$ is the 11th element in set A_1. We will also list A_2 as

$$A_2 = \{a_{21}, a_{22}, a_{23}, \ldots\}.$$

Then $a_{2,15}$ is element 15 in the set A_2. Notice that in general, in A_2, the first subscript 2 gives the set A_2 that the element is in, and the second subscript 15 gives the place that the element occupies on the implied list A_2.

Given an arbitrary $n \in \mathbb{N}$, we will follow this pattern of subscripts, and make a systematic implied list

$$\boxed{A_n = \{a_{n1}, a_{n2}, a_{n3}, \ldots\}}$$

of the elements of A_n. For instance, in the implied list A_{121}, the 200th element on the implied list is $a_{121,200}$. The next result shows that if we have a countable implied list $A_1, A_2, A_3, \ldots$ of countable sets, then when we put them together using the union operation, we will have a countable set.

Theorem 5.1.2 *Let $A_1, A_2, A_3, \ldots$ be an implied list of countable sets. Then $\bigcup_{n\in\mathbb{N}} A_n$ is a countable set.*

Proof: We have used this argument several times before.

Since each set A_n is a countable set, we can make an implied list of their elements as we did in the discussion preceding this theorem.

$$\begin{array}{rcl} A_1 & = & \{a_{11}, a_{12}, a_{13}, a_{14}, \ldots\}, \\ A_2 & = & \{a_{21}, a_{22}, a_{23}, a_{24}, \ldots\}, \\ A_3 & = & \{a_{31}, a_{32}, a_{33}, a_{34}, \ldots\}, \\ A_4 & = & \{a_{41}, a_{42}, a_{43}, a_{44}, \ldots\}, \\ & \vdots & \end{array}$$

To make the notation easier, let $A = \bigcup_{n\in\mathbb{N}} A_n$ and list the elements of A as in array (5.5).

$$\begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{14} & \cdots \\ \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots \\ \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots \\ \\ a_{41} & a_{42} & a_{43} & a_{44} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{array} \tag{5.5}$$

We are going to use arrows to count the elements in the array (5.5). Your instincts might say that we should count left to right, the way we read. But that approach leads us to exhaust the counting numbers $\mathbb{N}$ before we have left the first row. No, the best way is to use these diagonal arrows because *each of them is finite* and each a_{ij} lies on some arrow. We are thus assured that each element will be counted.

Draw parallel line segments starting at the elements a_{k1} and ending at the elements a_{1k} as in the array (5.6).

$$\begin{array}{llllll} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \quad \cdots \\ \#1 & & & & & \\ & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \#2 & & & & & \\ & a_{31} & a_{32} & a_{33} & a_{34} & \\ \#3 & & & & & \\ & a_{41} & a_{42} & a_{43} & \ddots & \\ \#4 & & & & & \\ & a_{51} & a_{52} & & & \\ \#5 & & & & & \\ & \vdots & & & & \end{array} \tag{5.6}$$

Use these arrows to define a bijection

$$f : \mathbb{N} \longrightarrow A$$

whose images $f(n)$ satisfy the following rule:

$$\begin{array}{llll} f(0) = a_{11}, & & & \\ f(1) = a_{21}, & f(2) = a_{12}, & & \\ f(3) = a_{31}, & f(4) = a_{22}, & f(5) = a_{13}, & \\ f(6) = a_{41}, & f(7) = a_{32}, & f(8) = a_{23}, & f(9) = a_{14}, \\ f(10) = a_{51}, & \quad \cdots & & \end{array}$$

It is clear that f is one-to-one. Let $a_{mn} \in A$. By the above scheme we see that the subscripts lying along arrow #1 add up to $1+1 = 2$, the subscripts lying along arrow #2 add up to 3, the subscripts lying along arrow #3 add up to 4, and in general the subscripts lying along arrow $\#\ell$ add up to $\ell + 1$. Then a_{mn} lies along arrow $\#m + n - 1$, which proves that $f(k) = a_{mn}$ for some number k. In fact, using the argument used to prove Theorem 5.1.1, we could show that

$$k \leq 1 + 2 + \ldots + (n + m - 1) = \frac{(n + m - 1)(n + m)}{2}.$$

Try to do that, reader. This proves that f is onto and hence that $f : \mathbb{N} \longrightarrow A$ is a bijection. Consequently, A is equivalent to $\mathbb{N}$ and therefore A is a countable set. This completes the proof.

The next result shows that if we increase $\mathbb{N}$ by finitely many elements, then we have not increased the cardinalty of $\mathbb{N}$.

Theorem 5.1.3 *Let $n \in \mathbb{N}$, and suppose that $\{a_1, \ldots, a_n\}$ is a set. Then $\mathbb{N} \cup \{a_1, \ldots, a_n\}$ is a countable set.*

Proof: The sets $\mathbb{N}$ and $\{a_1, \ldots, a_n\}$ are countable, so by Theorem 5.1.2, their union $\mathbb{N} \cup \{a_1, \ldots, a_n\}$ is also countable. This completes the proof.

Theorem 5.1.2 can be used to give a quick proof that $\text{card}(\mathbb{N}) = \text{card}(\mathbb{Q})$.

Theorem 5.1.4 *$\mathbb{Q}$ is countable.*

Proof: Observe that $\mathbb{Q} = \mathbb{Q}^+ \cup \{0\} \cup \mathbb{Q}^-$, where $\mathbb{Q}^-$ is the set of negative rational numbers. We will prove that $\mathbb{Q}^-$ is countable.

Define a function $f : \mathbb{Q}^+ \longrightarrow \mathbb{Q}^-$ as

$$f\left(\frac{n}{m}\right) = -\frac{n}{m}.$$

That is, the function $f(x)$ sends a rational number x to its additive inverse $-x$. Thus $f(1) = -1$ and $f\left(\frac{1}{9}\right) = -\frac{1}{9}$. We claim now, and prove below, that f is a bijection.

To see that f is one-to-one let $x \neq x'$ be different numbers in $\mathbb{Q}^+$. It is then clear that $-x \neq -x'$, so that $f(x) \neq f(x')$. Then f is one-to-one.

Prove that f is onto as follows. Let $y \in \mathbb{Q}^-$ be a negative fraction. Then $x = -y$ is a positive fraction, now isn't it? By forming $-y$, you are simply removing the negative sign from y. Try it with $-\frac{1}{2}$ and $-\frac{5}{3}$. Then $x \in \mathbb{Q}^+$, and

$$f(x) = -x = -(-y) = y,$$

so that f is an onto function.

This completes the proof of our claim that $f : \mathbb{Q}^+ \longrightarrow \mathbb{Q}^-$ is a bijection, and so $\mathbb{Q}^+$ and $\mathbb{Q}^-$ are equivalent sets.

By Theorem 5.1.1, $\mathbb{Q}^+$ is countable, so the bijection f given above proves that $\mathbb{Q}^-$ is countable. Then $\mathbb{Q}^+$, $\{0\}$, and $\mathbb{Q}^-$ are countable sets, so clearly, $\mathbb{Q} = \mathbb{Q}^+ \cup \{0\} \cup \mathbb{Q}^-$ is a union of three countable sets. Theorem 5.1.2 then implies that $\mathbb{Q}$ is a countable set. This completes the proof.

To summarize the above proof, we showed that $\mathbb{Q}^+$ and $\mathbb{Q}^-$ have the same cardinality because there is a bijection $f : \mathbb{Q}^+ \longrightarrow \mathbb{Q}^-$. The bijection f is defined by $f(x) = -x$ for each $x \in \mathbb{Q}^+$. Then $\mathbb{Q}^+$ and $\mathbb{Q}^-$ are countable sets, and so Theorem 5.1.2 implies that the union $\mathbb{Q} = \mathbb{Q}^+ \cup \{0\} \cup \mathbb{Q}^-$ of three countable sets is a countable set.

The next result provides us with at least one application of Theorem 5.1.2 that shows that a combination of countable sets is countable. For example, $\mathbb{N}$ is a countable set, and one can prove that the square root function

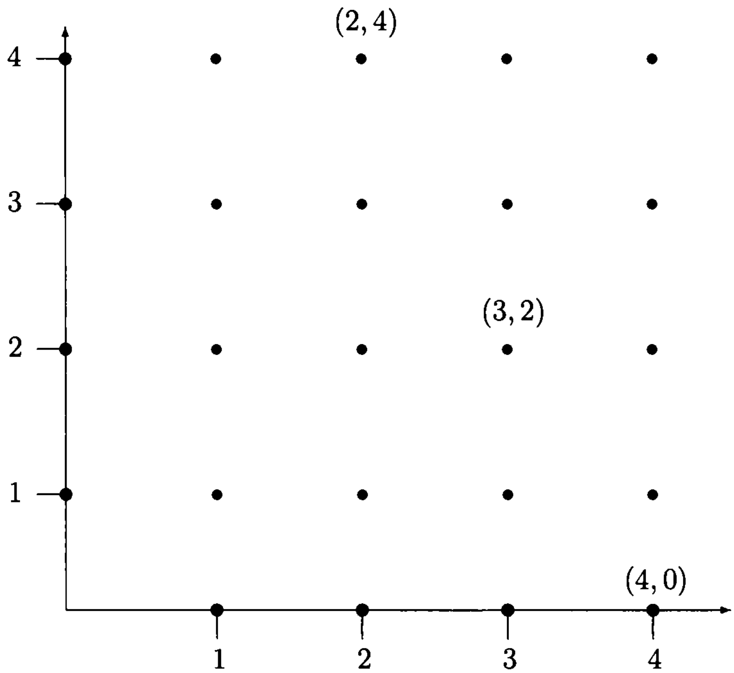
$$\sqrt{\cdot} : \mathbb{N} \to \{\sqrt{n} \mid n \in \mathbb{N}\}$$

is a bijection. Then Theorem 5.1.2 shows us that their union

$$\mathbb{N} \cup \{\sqrt{n} \mid n \in \mathbb{N}\}$$

is also a countable set.

Consider the set of pairs $\mathbb{N} \times \mathbb{N} = \{(n, m) \mid n, m \in \mathbb{N}\}$. Since $\mathbb{N} \subset \mathbb{R}$, we can graph $\mathbb{N} \times \mathbb{N}$ as points in the plane $\mathbb{R}^2$ as in figure (5.7). The pairs $(2, 4)$ and $(3, 2)$ label a couple of randomly selected points. This pleasant geometric object of evenly spaced points in the plane is called a *lattice*. The lattice extends indefinitely in all directions. Given the theme of this chapter, it is natural to ask for the cardinality of the lattice.



(5.7)

The Lattice of Positive Integer Points in $\mathbb{R}^2$

Theorem 5.1.5 $\mathbb{N} \times \mathbb{N}$ *is a countable set.*

Proof: Arrange $\mathbb{N} \times \mathbb{N}$ as an infinite rectangular array (5.8).

$$\begin{array}{ccccc}
(0,0) & (0,1) & (0,2) & (0,3) & \ldots \\
(1,0) & (1,1) & (1,2) & (1,3) & \ldots \\
(2,0) & (2,1) & (2,2) & (2,3) & \ldots \\
(3,0) & (3,1) & (3,2) & (3,3) & \ldots \\
\vdots & \vdots & \vdots & \vdots &
\end{array} \tag{5.8}$$

We will avoid the arrows used in Theorems 5.1.1 and 5.1.2. The rows of the array (5.8) are actually countable sets $A_0, A_1, A_2, \ldots$, and there are countably many of them.

$$\begin{array}{llllll} A_0 = & \{(0,0), & (0,1), & (0,2), & (0,3), & \ldots\} \\ A_1 = & \{(1,0), & (1,1), & (1,2), & (1,3), & \ldots\} \\ A_2 = & \{(2,0), & (2,1), & (2,2), & (2,3), & \ldots\} \\ A_3 = & \{(3,0), & (3,1), & (3,2), & (3,3), & \ldots\} \\ \vdots & & & & & \end{array} \tag{5.9}$$

In general, we make an implied list of A_n by writing

$$A_n = \{(n,0), (n,1), (n,2), (n,3), \ldots\},$$

and we note that (n,m) is the m-th element in A_n.

We claim that A_n is a countable set. Let $f : \mathbb{N} \to A_n$ be defined by the rule $f(m) = (n,m)$ for all $m \in \mathbb{N}$. The reader will prove as an exercise that f is a bijection. Hence A_n is a countable set for each $n \in \mathbb{N}$.

By definition, $(n,m) \in A_n$ for each $n, m \in \mathbb{N}$, so that

$$\mathbb{N} \times \mathbb{N} = A_0 \cup A_1 \cup A_2 \cup \ldots.$$

This proves that $\mathbb{N} \times \mathbb{N}$ is a countable union of countable sets, whence $\mathbb{N} \times \mathbb{N}$ is countable by Theorem 5.1.2. This completes the proof.

Theorem 5.1.5 was a warm up for the next more general result. While Theorem 5.1.5 shows us that pairs of natural numbers form a countable set, its proof is the real mathematical issue. We will apply its proof to Theorem 5.1.6, thus showing that we can increase the dimension of a set and still preserve its countability.

Theorem 5.1.6 *Let A and B be countable sets. Then $A \times B$ is a countable set.*

Proof: Since A and B are countable, we can make an implied list them.

$$\begin{aligned} A &= \{a_1, a_2, a_3, \ldots\}, \\ B &= \{b_1, b_2, b_3, \ldots\}. \end{aligned}$$

Then we can list the elements of $A \times B$ as an infinite array or implied list (5.10).

$$\begin{array}{llll}
(a_1, b_1) & (a_1, b_2) & (a_1, b_3) & (a_1, b_4) \dots \\
(a_2, b_1) & (a_2, b_2) & (a_2, b_3) & (a_2, b_4) \dots \\
(a_3, b_1) & (a_3, b_2) & (a_3, b_3) & (a_3, b_4) \dots \\
(a_4, b_1) & (a_4, b_2) & (a_4, b_3) & (a_4, b_4) \\
\vdots & \vdots & \vdots & \vdots
\end{array} \tag{5.10}$$

Name the rows of array (5.10) with symbols X_1, X_2, $\dots$ as in array (5.11).

$$\begin{aligned}
X_1 &= \{(a_1, b_1), (a_1, b_2), (a_1, b_3), (a_1, b_4), ..\} \\
X_2 &= \{(a_2, b_1), (a_2, b_2), (a_2, b_3), (a_2, b_4), ..\} \\
X_3 &= \{(a_3, b_1), (a_3, b_2), (a_3, b_3), (a_3, b_4), ..\} \\
X_4 &= \{(a_4, b_1), (a_4, b_2), (a_4, b_3), (a_4, b_4), ..\} \\
&\quad \vdots \qquad \vdots \qquad \vdots \qquad \vdots
\end{aligned} \tag{5.11}$$

Each X_n is a countable set (this is an exercise for you), and there are countably many sets X_1, X_2, X_3, $\dots$, so by Theorem 5.1.2, the union

$$A \times B = X_1 \cup X_2 \cup X_3 \cup \dots$$

is a countable set. This completes the proof.

Since $\mathbb{Q}$ is countable, we have proved the following theorem.

Theorem 5.1.7 $\mathbb{Q} \times \mathbb{Q}$ *is a countable set.*

Given a set A, let

$$\boxed{A^n = \{(x_1, \dots, x_n) \mid x_1, \dots, x_n \in A\}}$$

be the set of *t-tuples of elements in A*. For example, if $A = \mathbb{N}$, then

$$\boxed{\mathbb{N}^n = \{(x_1, \ldots, x_n) \,|\, x_1, \ldots, x_n \in \mathbb{N}\}}$$

be the set of *t-tuples of natural numbers*. For instance, , $\mathbb{N}^2$ is the set of pairs of natural numbers, and $\mathbb{N}^3$ is the set of triples (3-tuples) of natural numbers. We note that $(0,0,0) \in \mathbb{N}^3$. Using the same kind of argument, the reader should try to prove the next result.

Theorem 5.1.8 *$\mathbb{N}^n$ is a countable set for each positive natural number $\mathbb{N}$.*

Exercise 5.1.9 1. Let $A = \{\frac{2n}{m}$ written in lowest terms $\big| n, m \in \mathbb{N}$, $m \neq 0\}$. These are the positive fractions with even numerator. Prove that A is countable.

2. Let $n \in \mathbb{N}$ and let $A_n = \{(n,0),(n,1),(n,2),\ldots\}$. Show that A_n is countable.

3. Suppose that $b_1, b_2, b_3, \ldots$ is a sequence of distinct elements. Let $n \in \mathbb{N}$, let a_n be some element, and let $X_4 = \{(a_n, b_1), (a_n, b_2), (a_n, b_3), \ldots\}$. Show that X_n is countable.

4. Given $n \in \mathbb{N}$, let $A_n =$ the set of sequences of natural numbers that have exactly n terms. For example, $A_1 =$ the set of sequences x_1 with 1 term $x_1 \in \mathbb{N}$, and $A_2 =$ the set of sequences x_1, x_2 with $\{x_1, x_2\} \subset \mathbb{N}$. Prove that $\cup_{n\in\mathbb{N}} A_n$ is a countable set.

5. Prove that $\mathbb{N} \times \mathbb{Q}$ is countable.

6. Prove that $\mathbb{Q} \times \mathbb{Q}$ is countable.

7. Prove that $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ is countable.

5.2 Uncountable Sets

The next stage of our work requires us to compare the size of cardinals.

Definition 5.2.1 *Let A and B be sets. We write*

$$\mathrm{card}(A) \leq \mathrm{card}(B)$$

if there is a one-to-one function $f : A \longrightarrow B$.

This is quite a reasonable definition of inequality for cardinals. For example, $3 < 4$ because $\{1, 2, 3\} \subset \{1, 2, 3, 4\}$, and in general

$$m < n \text{ exactly when } \{1, \ldots, m\} \subset \{1, \ldots, n\}.$$

For larger sets, we see that

$$\operatorname{card}(\mathbb{N}) \leq \operatorname{card}(\mathbb{Z}) \leq \operatorname{card}(\mathbb{Q}) \leq \operatorname{card}(\mathbb{R})$$

because $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. One property that real numbers share with cardinals is the following property of natural numbers.

$$\text{If } n \leq m, \text{ and if } m \leq n, \text{ then } n = m.$$

What this says is that we can deduce the equality of numbers if we can show that they are mutually related by $\leq$. It is a remarkable fact that this property also holds for *infinite cardinals.* Its proof is too far afield for us to write here, so we will just accept its truth in this book.

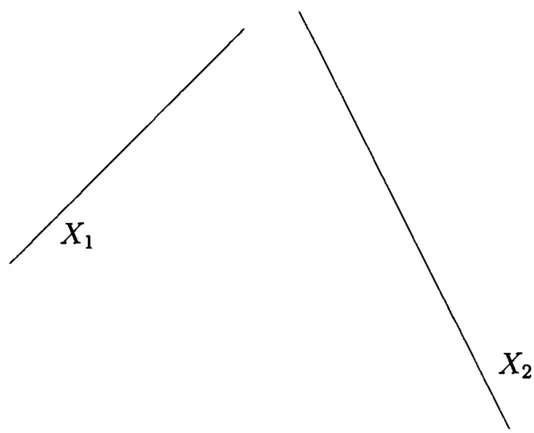
$$\text{If } \operatorname{card}(A) \leq \operatorname{card}(B), \text{ and if, } \operatorname{card}(B) \leq \operatorname{card}(A),$$
$$\text{then } \operatorname{card}(A) = \operatorname{card}(B).$$

What we need now are some good examples. The set A is said to be *uncountable* if $\operatorname{card}(\mathbb{N}) \neq \operatorname{card}(A)$. The question is, are there any uncountable sets at all, or is every set countable? We will show the remarkable fact that there are indeed uncountable sets. This must seem strange if you think about it for a second. Why should there be two types of infinite set? And yet that is where we are headed.

Let $(0, 1)$ denote the set of real numbers properly between 0 and 1. We write

$$(0,1) \quad = \quad \{x \in \mathbb{R} \mid 0 < x < 1\}.$$

The next series of results provides us with a collection of uncountable sets by showing that many of the geometric objects you are familiar with are themselves uncountable. For example, every line segment is uncountable. Thus, the cardinality of a set is independent of its length. The first result shows us that all line segments of *finite length* have the same cardinality.



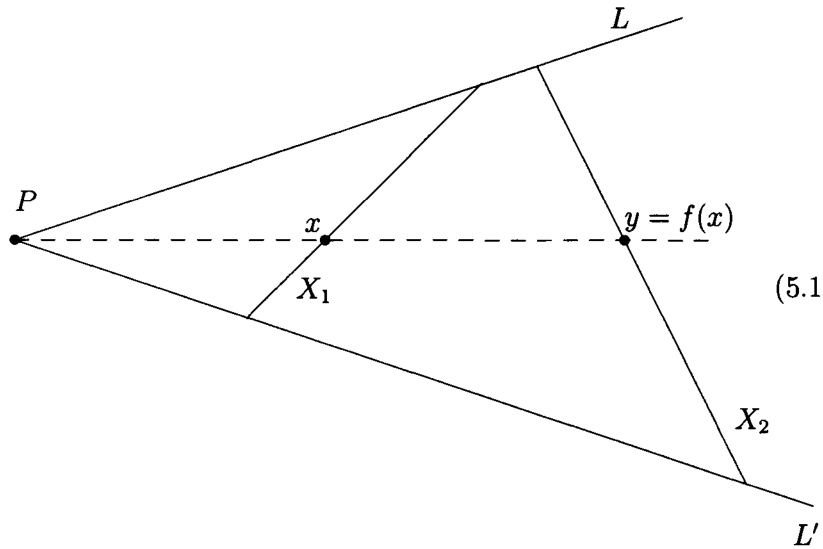
(5.12)

Theorem 5.2.2 *If* X_1 *and* X_2 *are finite line segments then* $\mathrm{card}(X_1) = \mathrm{card}(X_2)$. *Equivalently,* any two line segments are equivalent.

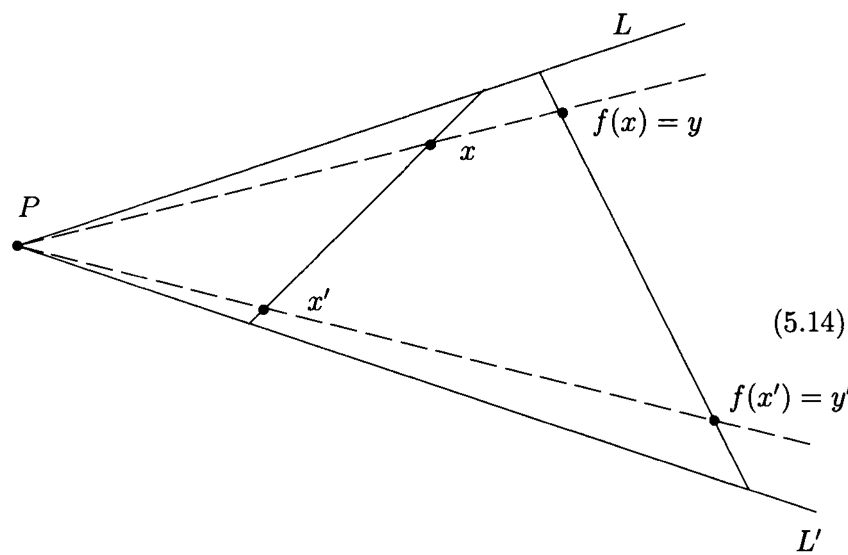
Proof: Begin with two line segements as in picture (5.12). The segments X_1 and X_2 in picture (5.12) are representative of all segments. We will use only their most general properties in this discussion. In picture (5.13), the end points of these segments are contained in lines L and L'. Lines L and L' intersect at a point P.

We define a function $f : X_1 \longrightarrow X_2$ as follows. Given any point $x \in X_1$, draw a line, (the dotted one in picture (5.13)), from P through x such that it intersects the segment X_2 at the point y. We define

$$\boxed{f(x) = y.}$$



(5.13)



(5.14)

To see that f is onto, refer to picture (5.13). Let $y \in X_2$ be a point. Extend a dotted line from P to y. This line intersects X_1 at some point x. Then by the definition of f, $f(x) = y$, and hence f is onto.

To see that f is one-to-one, suppose that $x \neq x'$. Then as in picture (5.14), x and x' determine different (dotted) lines through P. Let y and y' be the points where the dotted lines intersect X_2. The intersections of these dotted lines with X_2 are different points y and y'. (If otherwise, the intersection is one point y, and then y and the point P yield *exactly one* dotted line, not two.) Then $f(x) \neq f(x')$, which implies that f is one-to-one.

Thus, f is a bijection, and therefore X_1 is equivalent to X_2. This ends the proof.

A colloquial way of stating the above result is that

Any two line segments have the same cardinality

or the same number of elements. However, our later work will show that to say "the same number of elements" is a terribly inadequate way to describe the cardinality of a set.

For instance,

$$\text{card}(0,1) = \text{card}\left(0, 10^{10^{100}}\right) = \text{card}\,(-2,-1) = \text{card}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

Consequently,

$\text{card}(0,1) = \text{card}(a,b)$ for any numbers $a < b$.

Therefore, any two line segements have cardinality $\text{card}(0,1)$.

The next example shows that all line egments have the cardinality of $\mathbb{R}$.

Theorem 5.2.3 $\text{card}(0,1) = \text{card}(\mathbb{R})$.

Proof: By the comment preceding this theorem,

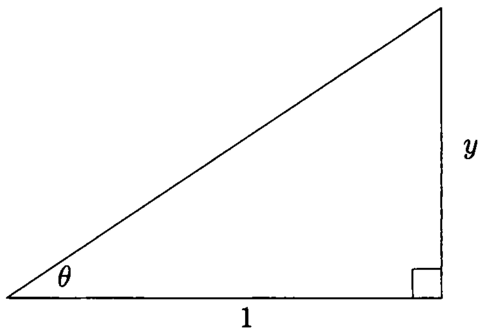
$$\text{card}(0,1) = \text{card}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

so, without losing generality, we can prove that $\text{card}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right) = \text{card}(\mathbb{R})$.

Consider the function

$$\tan(\theta) : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \longrightarrow \mathbb{R}.$$

Given a real number y, draw a right triangle labeled as in the diagram below.



Elementary trigonometry shows that

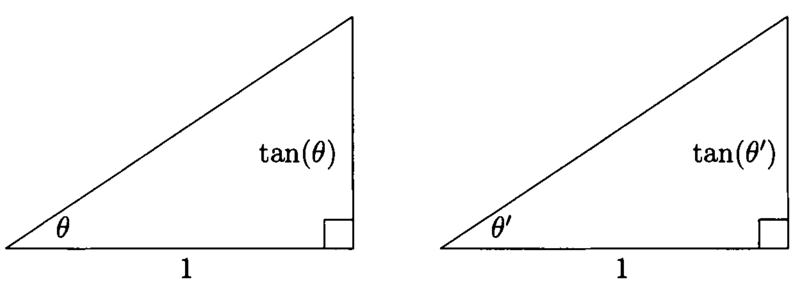
$$\tan(\theta) = \frac{y}{1} = y,$$

so that $\tan(\theta)$ is onto.

To see that $\tan(\theta)$ is one-to-one, suppose that θ and θ' are given such that

$$\tan(\theta) = \tan(\theta').$$

Draw two triangles as follows.



Certainly, the bases of these triangles are equal, as are the indicated right angles, and by hypothesis, the right hand legs are equal in length. Then by a theorem you learned in high school Geometry, the two triangles are congruent. Then

$$\theta = \theta',$$

or equivalently, the measures θ and θ' of the corresponding base angles are equal. Then $\tan(\theta)$ is one-to-one.

Hence, $\tan(\theta) : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \longrightarrow \mathbb{R}$ is a bijection, and therefore,

$$\text{card}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right) = \text{card}(\mathbb{R}).$$

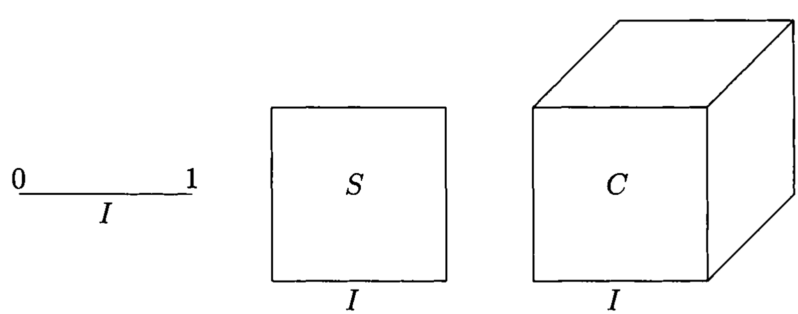
This completes the proof of the theorem.

Theorem 5.2.3 may come as a surprise to you. The cardinality of a finite segment and the cardinality of the unbounded real number line are the same. This implies that greater length does not produce a greater cardinality. That is, more length does not produce more points. In simple but inaccurate terms, the number of points on a finite segment is equal to the number of points in all of the real number line.

Let

$$\boxed{I = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}.}$$

This is the set whose elements are 0 and 1 and *all* of the numbers between 0 and 1. Let S be the *unit square and its interior*, and let C be the *unit cube and its interior*. In pictures we have



Let us be clear about which figures we mean. I includes the end points and the interior of the open unit interval $(0, 1)$, S is the interior and the boundary of the square whose sides have length 1, and C is the solid cube whose sides have length 1. The following example will show us that the cardinality of a set is independent of its geometry and dimension. If you draw a line segment on a piece of paper, that segment has the same number of points as the piece of paper, which has the same number of points as the room in which you are reading this. Thus, even though these sets have different geometric dimension, they possess the same cardinality.

Theorem 5.2.4 $\operatorname{card}(I) = \operatorname{card}(S) = \operatorname{card}(C)$.

Proof: It seems clear enough that

$$I \subset S \subset C,$$

so that

$$\operatorname{card}(I) \leq \operatorname{card}(S) \leq \operatorname{card}(C). \tag{5.15}$$

To complete the proof, we will produce a one-to-one function $f : C \longrightarrow I$. This will allow us to conclude that

$$\operatorname{card}(C) \leq \operatorname{card}(I)$$

and hence that $\operatorname{card}(I) = \operatorname{card}(C)$. Play with that one for a while. Suppose that a, b are numbers such that $a \leq b$ and $b \leq a$. The only way for this to happen is for $a = b$. We are simply using the fact that this same property holds for cardinals. If you find the following argument difficult, just skip forward. There is plenty of time and no shame in coming back to these deliberations.

From analytic geometry, we recall that the unit cube C is a set of ordered triples,

$$C = \{(x, y, z) \,|\, x, y, z \in I\}.$$

The one-to-one function $f : C \to I$ that we need is defined as follows. Given an ordered triple $(x, y, z) \in C$, write x, y, z as decimals

$$\begin{array}{rcllll} x & = & x_0. & x_1 & x_2 & x_3 \ \ldots \\ y & = & y_0. & y_1 & y_2 & y_3 \ \ldots \\ z & = & z_0. & z_1 & z_2 & z_3 \ \ldots \end{array}$$

where $x_0 = 0$ or 1, and where $x_1, x_2, x_3, \ldots$ are digits in the set $\{0, 1, \ldots, 9\}$. Similar statements apply to the decimal expansions of y and z. Furthermore, to avoid that nasty case where numbers end in $\overline{999}$, we assume that the shorter decimal expansion that ends in $\overline{000}$ is used for x, y, and z. Define

$$\boxed{f(x, y, z) = .x_0y_0z_0 \; x_1y_1z_1 \; x_2y_2z_2 \; x_3y_3z_3 \; \ldots.}$$

Since the numbers x_k, y_k, z_k are in the set $\{0, 1, \ldots, 9\}$, $f(x, y, z)$ is a decimal between 0 and 1. That is, $f(x, y, z) \in I$ and hence

$$f : C \to I$$

is a function. Since the numbers x, y, and z do not have an infinite string of 9's in their decimal expansion, the number $f(x, y, z)$ does not end in $\overline{999}$.

For example, to find

$$f(1.00\bar{0}, 0.41415\ldots, 0.1\overline{222}),$$

we write the three numbers as

$$\begin{array}{rcllll} x & = & 1. & 0 & 0 & 0 & \cdots \\ y & = & 0. & 4 & 1 & 4 & \cdots \\ z & = & 0. & 1 & 2 & 2 & \cdots \end{array}$$

and then shuffle them as follows.

$$f(1.0\bar{0}, 0.41415\ldots, 0.1\overline{222}) = .100 \; 041 \; 012 \; 042 \; \cdots.$$

We must show that f is one-to-one. Suppose that

$$w = f(x', y', z') = f(x, y, z) \;\; = \;\; .x_0y_0z_0 \; x_1y_1z_1 \; x_2y_2z_2 \; x_3y_3z_3 \; \cdots$$

for some ordered triples $(x, y, z), (x', y', z') \in C$. This is the unique short way of writing w since by design w does not end in $\overline{999}$. By applying the definition of f in reverse, we have

$$\begin{array}{rclllllcl} x & = & x_0. & x_1 & x_2 & x_3 & \ldots & = & x' \\ y & = & y_0. & y_1 & y_2 & y_3 & \ldots & = & y' \\ z & = & z_0. & z_1 & z_2 & z_3 & \ldots & = & z'. \end{array}$$

Then $(x, y, z) = (x', y', z')$, which implies that f is a one-to-one function. By our definition of $\leq$ for cardinals,

$$\operatorname{card}(C) \leq \operatorname{card}(I).$$

Combining this with the inequalities (5.15), we see that

$$\operatorname{card}(I) = \operatorname{card}(S) = \operatorname{card}(C)$$

as required to complete the proof.

The above example should strike you as contrary to the conservation of mass or some such concept from physics. How could a line and a cube have the same anything? The issue here is that we are counting the points in a mathematical object, and points do not have any mass and they do not have any other physical properties. Physics and physical properties of lines and cubes never enter the picture.

The surprises continue in the next result where we state, but do not prove, that infinite space, a child's toy block, and a 1 inch long segment have the same cardinality.

Theorem 5.2.5 *Let $\mathbb{R}$ denote the one dimensional real line, let $\mathbb{R}^2$ denote the two dimensional Euclidean plane, and let $\mathbb{R}^3$ denote three dimensional Euclidean space. Then*

$$\operatorname{card}(0, 1) = \operatorname{card}(\mathbb{R}) = \operatorname{card}(\mathbb{R}^2) = \operatorname{card}(\mathbb{R}^3).$$

That is, cardinality is independent of spacial dimension.

This example should also upset your common sense. If we identify $(0, 1)$ with an inch long line segment, then you might believe that certainly an inch has fewer points than the entire real line, which has fewer points than the plane or all of three dimensional space. Once again we are shown that our intuition regarding infinity does not agree with the surrounding mathematical facts. The theorems in the next section will show you why we are being so careful in our discussion.

Exercise 5.2.6 1. Show that $J = \{x^2 \mid 0 \leq x \leq 1\}$ is uncountable.

2. The set of complex numbers is $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$ where $i = \sqrt{-1}$. Show that $\mathbb{C}$ is equivalent to $\mathbb{R}$.

3. Let $A = \left\{(x, \dfrac{1}{x}) \mid x \in \mathbb{R}^+\right\}$. Show that A is uncountable.

4. Let $\mathbf{C}$ be the unit circle in the Euclidean plane. Show that $\mathbf{C}$ is equivalent to $\mathbb{R}$.

5. Let $I = (0,1)$, and let S be the unit square. Show that $\mathrm{card}(S) = \mathrm{card}(I)$.

6. Let $A_1, A_2, A_3, \ldots$ be a countable implied list of *uncountable* sets. Show that $A = \cup_{n\in\mathbb{N}} A_n$ is an uncountable set.

7. Let $\{A_r \mid r \in \mathbb{R}\}$ be a uncountable *family* of distinct *countable* sets. Show that $A = \cup_{n\in\mathbb{N}} A_n$ is an uncountable set.

8. Let $S_n \subset \mathbb{R}^3$ be the cube whose edges measure n units in length and whose center (obvious definition) is the origin. Show that

   (a) S_n is uncountable.

   (b) $\cup_{n\in\mathbb{N}} S_n$ is uncountable.

   (c) $\cup_{n\in\mathbb{N}} S_n = \mathbb{R}^3$, so that $\mathbb{R}^3$ is uncountable.

5.3 Two Infinities

Let us summarize what we have learned to this point. The *cardinality* of a set X is a way of measuring in precise mathematical terms the size of X. We saw that

$$\mathrm{card}(\mathbb{N}) = \mathrm{card}\{1^2, 2^2, 3^2, \ldots\} = \mathrm{card}(\mathbb{Q})$$

and that

$$\mathrm{card}(\mathbb{N}) \neq \mathrm{card}(\mathbb{R}).$$

Sets A for which $\mathrm{card}(A) = \mathrm{card}(\mathbb{N})$ are called *countable sets.* Then A is countable if we can make an implied list of the elements in A. For instance, $\mathbb{N}$ is a countable set. We saw that

$$\mathrm{card}(0,1) = \mathrm{card}(\text{unit cube}) = \mathrm{card}(\mathbb{R}) = \mathrm{card}(\text{space}).$$

A set A for which $\text{card}(A) \geq \text{card}(\mathbb{R})$ is called *uncountable* . For instance, $(0,1)$ and a square, and a cube are each uncountable sets.

We ask the natural question, *How are* $\text{card}(\mathbb{N})$ *and* $\text{card}(\mathbb{R})$ *otherwise related?* Are the uncountable sets and the countable sets two different types of sets? The uninitiated reader might argue that since both sets are infinite, $\mathbb{N}$ and $\mathbb{R}$ have the same cardinality. If you argued that way, don't feel too badly. Its just your inexperience showing. The next result answers our question. A colloquial and accurate interpretation of this theorem is the dark philosophical shocker that *there is more than one infinity.*

Theorem 5.3.1 $\text{card}(\mathbb{N}) < \text{card}(\mathbb{R})$.

Proof: Evidently $\mathbb{N} \subset \mathbb{R}$, so by the definition of inequality of cardinals,

$$\text{card}(\mathbb{N}) \leq \text{card}(\mathbb{R}).$$

What we have to prove now is that $\text{card}(\mathbb{N}) \neq \text{card}(\mathbb{R})$. To do this, we borrow from the Realists who visited Hilbert's Infinite Hotel on page 126.

Suppose, for the sake of contradiction, that we have a one-to-one function

$$f : \mathbb{N} \longrightarrow (0,1).$$

Make a list of f and its values in $(0,1)$.

$$\begin{array}{lllllll}
f(1) & = x_1 = & .\boxed{d_{11}} & d_{12} & d_{13} & d_{14} & \cdots \\
f(2) & = x_2 = & .d_{21} & \boxed{d_{22}} & d_{23} & d_{24} & \cdots \\
f(3) & = x_3 = & .d_{31} & d_{32} & \boxed{d_{33}} & d_{34} & \cdots \\
f(4) & = x_4 = & .d_{41} & d_{42} & d_{43} & \boxed{d_{44}} & \cdots \\
\vdots & & \vdots & & & &
\end{array}$$

In the above implied list we have given the shorter decimal expansion of $f(n) = x_n$ for each $n \in \mathbb{N}$, if the shorter decimal expansion exists. In this implied list, each d_{ij} is a digit, a number in the set $\{0,1,\ldots,9\}$, and $f(n)$ does not end in $\overline{999}$. Furthermore, d_{11} represents the first digit in the decimal form of x_1, d_{23} represents digit number 3 in the decimal form of x_2, and in general d_{nm} represents digit number m in the decimal form of x_n. We can now produce a real number that is not on this implied list.

Define the digit d_1 as

$$d_1 = \begin{cases} d_{11}+1 & \text{if } d_{11} \neq 9 \\ 0 & \text{if } d_{11} = 9 \end{cases}.$$

Notice that we have defined d_1 so that $d_1 \neq d_{11}$.

Let

$$d_2 = \begin{cases} d_{22}+1 & \text{if } d_{22} \neq 9 \\ 0 & \text{if } d_{22} = 9 \end{cases}.$$

Notice that we have defined d_2 so that $d_2 \neq d_{22}$.

In general, define

$$d_n = \begin{cases} d_{nn}+1 & \text{if } d_{nn} \neq 9 \\ 0 & \text{if } d_{nn} = 9 \end{cases}.$$

Then $d_n \neq d_{nn}$.

The extra number x is given by the decimal expansion

$$\boxed{x = .d_1d_2d_3d_4\ldots}$$

thus making it a real number in $(0,1)$. If the number x ends in $99\bar{9}$, replace it with the shorter decimal expansion that ends in $00\bar{0}$.

Then $x \neq x_1$ because by our choice of d_1, x and x_1 differ in the first decimal place, $d_1 \neq d_{11}$. Also, $x \neq x_2$ because by our choice of d_2, x and x_2 differ in decimal place number 2, $d_2 \neq d_{22}$. In general, $x \neq x_n$ because by our choice of d_n, x and x_n differ in decimal place number n, $d_n \neq d_{nn}$. Thus x is not on the implied list $x_1, x_2, x_3, \ldots$ of real numbers in $(0,1)$. This is what we wanted. Hence f is not a bijection.

We have thus shown that there are no bijections $f : \mathbb{N} \longrightarrow (0,1)$, which proves that

$$\text{card}(\mathbb{N}) \neq \text{card}(0,1).$$

Since Theorem 5.2.3 states that $\text{card}(0,1) = \text{card}(\mathbb{R})$, we have shown that

$$\text{card}(\mathbb{N}) \neq \text{card}(\mathbb{R})$$

and therefore that

$$\text{card}(\mathbb{N}) < \text{card}(\mathbb{R}).$$

This completes the proof.

Theorem 5.3.2 card($\mathbb{N}$) < card($[0, 1]$).

Proof: By the above theorem, we know that card($\mathbb{N}$) < card($\mathbb{R}$) = card$(0, 1)$, and because $(0, 1) \subset [0, 1]$, card($\mathbb{N}$) < card($[0, 1]$). This is what we had to prove.

Theorem 5.3.3 card$(0, 1)$ = card($[0, 1]$).

Proof: Since $(0, 1) \subset [0, 1]$, card$(0, 1) \leq$ card($[0, 1]$). Next, choose $a < b \in (0, 1)$. Then $[a, b] \subset (0, 1)$, so that card($[a, b]$) $\leq$ card($[0, 1]$). By Theorem 5.2.2, card($[0, 1]$) = card($[a, b]$), so that card($[0, 1]$) $\leq$ card$(0, 1)$. Therefore, card$(0, 1)$ = card($[a, b]$), and the proof is complete.

Subsequently, there cannot be a list of the real numbers $\mathbb{R}$. Furthermore, we conclude that the terms *countable* and *uncountable* describe two different types of infinite sets. Those sets equivalent to $\mathbb{N}$ are not equivalent to $\mathbb{R}$. Since $\mathbb{Q}$ is a dense subset of $\mathbb{R}$, the following result shows us that cardinalty does not depend on the density of the set in the real line.

Theorem 5.3.4 card($\mathbb{Q}$) < card($\mathbb{R}$).

Proof: Theorem 5.1.4 shows us that card($\mathbb{N}$) = card($\mathbb{Q}$), and Theorem 5.3.1 shows us that card($\mathbb{N}$) < card($\mathbb{R}$). Then card($\mathbb{Q}$) < card($\mathbb{R}$). This is what we wanted to prove.

Summarizing some of our results on cardinality of sets, we have

$$\boxed{\text{card}(\mathbb{N}) = \text{card}(\mathbb{Q}) < \text{card}(\mathbb{R}) = \text{card}(\text{space}).}$$

Are you surprised? When you first read this result, do you rebel at it? But there it is, in a plain mathematically justifiable language. There are at least two infinities, or infinite cardinals. We will see later that there are more than just these two. That should jolt you. Infinity is infinity and that is that, you might have said. The answer to that argument is that yes, all infinite sets are not finite, but infinity is not a number. Infinity is not a cardinal. Infinity is not the last word on the size of a set. It is only the beginning, like saying that something is *short*. There will always be larger sets.

Saying that a set is infinite is like saying that a car has color. Yes, it has color, but which one? Is it blue or red? So when we encounter cardinals we will decide if it is finite or infinite, and if it is infinite we will decide if it is countable or uncountable. This sort of discussion will describe somewhat which cardinal we have, but it will not be the last word on how large that cardinal is. Keep an open mind as you read this section.

Because $\mathrm{card}(\mathbb{Q}) < \mathrm{card}(\mathbb{R})$, there must be real numbers x that are not fractions, or equivalently, the real number x does not possess a repeating decimal. We will say that A number x is *irrational* if it is not rational. In other words, irrational numbers are in $\mathbb{Q}'$, which is the set theoretic complement of $\mathbb{Q}$ in $\mathbb{R}$. Then the real number x is irrational iff $x \notin \mathbb{Q}$. Equivalently, $x \in \mathbb{R}$ is irrational iff its decimal expansion is infinite and not repeating. We let

$$\boxed{\mathbb{Q}' = \text{ the set of irrational numbers.}}$$

Unlike $\mathbb{Q}$, there can be no implied list of irrational numbers $\mathbb{Q}'$. The proof of this fact is your old friend the indirect argument.

Theorem 5.3.5 $\mathrm{card}(\mathbb{Q}')$ *is uncountable.*

Proof: Suppose to the contrary that $\mathbb{Q}'$ is not uncountable. Then it is countable, as is $\mathbb{Q}$, (Theorem 5.1.1). Since $\mathbb{Q}'$ is the complement of $\mathbb{Q}$ in $\mathbb{R}$, $\mathbb{R}$ is a finite union of countable sets $\mathbb{R} = \mathbb{Q} \cup \mathbb{Q}'$. Then Theorem 5.1.2 states that $\mathbb{R}$ is a countable set, which contradicts Theorem 5.3.1 in which we proved that $\mathbb{R}$ is uncountable. This contradiction shows that $\mathbb{Q}'$ is uncountable, which proves the theorem.

It is interesting to observe that this simple counting of $\mathbb{Q}'$ shows us that *most numbers are not rational.* This is contrary to your experience since every number you have ever seen has been a rational number. That changes nothing. There are many more irrational numbers than there are rational ones. However, professional mathematicians consider it to be very hard to prove that any given number is irrational. While it is True that π is irrational, the proof is quite hard. We will show shortly that $\sqrt{2}$ is an irrational number, but we cannot show that $\pi^{\sqrt{2}}$ is irrational. Currently, no one knows its rationality. If you have some facility with logarithms, you might find it challenging to try to prove that

$$\log_2(3) \text{ is irrational.}$$

(Hint: Read the proof of Theorem 5.3.6 first.) A simple question that currently remains unanswered is this. *If x and y are irrational numbers, under what conditions is $x + y$, xy, or x^y irrational?* For example, it is known that $\sqrt{3} + \sqrt{2}$ is irrational. But what can be said of more mysterious numbers like $2^{\sqrt{2}}$?

Let's settle back and catch our breath now. The earliest example of an irrational number occurs around 600 BC in ancient Greece, where the Pythagoreans showed that $\sqrt{2}$ is an irrational number. They found this fact to be so disturbing that they threatened people with a removal of their hands if they betrayed the great secret that $\sqrt{2}$ is not a fraction of integers. We laugh at this today, or we gape in awe that rational people would react so extremely to a mathematical fact. Today this fact is considered to be one of the highlights of early mathematics. The proof is just as compelling today as it was 2600 years ago. The proof that $\sqrt{2}$ is irrational is a pivotal moment in mathematics. Prior to this, people who thought about the universe thought that all of it could be described by fractions, rational numbers. Thus, the following result was a moment of transformation in mathematics, physics, and science in general.

Theorem 5.3.6 *$\sqrt{2}$ is an irrational number.*

Proof: Before we begin, we need to reestablish the fact

$$\text{If } n^2 \text{ is divisible by 2 then } n \text{ is divible by 2.} \tag{5.16}$$

Equivalently, we will prove its contrapositive

$$\text{If } n \text{ is not divisible by 2 then } n^2 \text{ is not divisible by 2.}$$

The proof is a calculation. Suppose that n is not divisible by 2. Then $n = 2k + 1$ for some $k \in \mathbb{N}$, so that

$$n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1.$$

Then $n^2 = 2K + 1$, where K is the integer $2k^2 + 2k$, and hence n^2 is an odd number. This proves statement (5.16). Armed with this fact, we can attack the irrationality of $\sqrt{2}$. Once again, we will use the indirect proof.

Assume for the sake of contradiction that $\sqrt{2}$ is rational. Then

$$\sqrt{2} = \frac{n}{m}$$

for some integers $n, m > 0$. We assume without losing our generality that the fraction $\frac{n}{m}$ is in reduced form. Specifically,

$$\begin{array}{c}\text{at least one of the numbers } n \text{ and } m \\ \text{is not divisible by 2.}\end{array} \tag{5.17}$$

Squaring both sides yields

$$2 = \frac{n^2}{m^2}.$$

Then $2m^2 = n^2$, so that n^2 is divisible by 2. Statement (5.16) implies that n is divisible by 2, so we can write

$$n = 2k$$

for some integer k. Squaring both sides we find that

$$2m^2 = n^2 = 4k^2,$$

so division by 2 yields

$$m^2 = 2k^2.$$

Then m^2 is divisible by 2, and so m is divisible by 2 by statement (5.16). Both n and m are divisible by 2, which is contrary to the statement (5.17). Hence, our original assumption is in error, and we conclude that $\sqrt{2}$ is irrational. This completes the proof.

The reader can use the same argument to show that $\sqrt{3}$ is irrational and that in general $\sqrt{p}$ is irrational for each prime p. This reinforces the above work that there are infinitely (actually uncountably) many irrational numbers. The present work has the advantage that we have implied the existence of a list of *some* of the irrational numbers.

$$\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}, \ldots \text{ are some of the irrational numbers.}$$

Compare this with Theorem 5.3.5, which states that there are infinitely many irrational numbers but which does not list them.

There are at least two types of cardinals, the countable and the uncountable. We will see in the next section that uncountable cardinals also come in many different sizes.

Exercise 5.3.7 1. Write down a list of 12 different numbers $0 < x < 1$. Use the argument in Theorem 5.3.1 to produce a number $0 < y < 1$ that is not on your list.

2. Show that $\sqrt{3}$ is irrational.

3. Let $p \in \mathbb{N}$ be a prime number. Show that $\sqrt{p}$ is irrational.

4. Show that $\log_2(3)$ is irrational.

5. Let $p \in \mathbb{N}$ be a prime. Show that $\log_2(p)$ is irrational.

6. Let $p, q \in \mathbb{N}$ be primes. Show that $\log_q(p)$ is irrational.

7. A number $a \in \mathbb{R}$ is said to be *algebraic* if there is a polynomial $f(x) = (x - a)g(x)$ whose coefficients are rational numbers. It is known that there are countably many such polynomials. Show that the set A of algebraic numbers in $\mathbb{R}$ is countable.

8. The complement A' in $\mathbb{R}$ of the set of algebraic numbers in $\mathbb{R}$ is called the set of *transcendental* numbers. Show that A' is uncountable.

9. Let Ω be the set of all binary sequences $b_1 b_2 b_3 \cdots$, where $b_j \in \{0, 1\}$. Using Cantor's argument that $\text{card}(\mathbb{N}) \leq \text{card}(0, 1)$, show that $\text{card}(\mathbb{N}) < \text{card}(\Omega)$.

5.4 Power Sets

Let A be a set. Recall from Definition 2.4.1 that *the power set of* A, $\mathcal{P}(A)$, is the set of all subsets of A. We will use the power set to show that there are many infinite cardinals.

From Theorem 5.3.1 we know that $\text{card}(\mathbb{N})$ and $\text{card}(\mathbb{R})$ are different infinite cardinals. The next result shows us that $\mathbb{R}$ and $\mathcal{P}(\mathbb{N})$ are equivalent sets.

Theorem 5.4.1 $\text{card}(\mathbb{R}) = \text{card}[0,1] = \text{card}(\mathcal{P}(\mathbb{N}))$.

Proof: If you find this proof is stretching you a bit too far, skip to the end. There is no examination at the end of this discussion.

We will produce a one-to-one function

$$f : (0,1) \longrightarrow \mathcal{P}(\mathbb{N}),$$

thus proving that $\text{card}(0,1) \leq \text{card}(\mathcal{P}(\mathbb{N}))$. Then by Theorem 5.2.3, we will have shown that $\text{card}(\mathbb{R}) = \text{card}(0,1) \leq \text{card}(\mathcal{P}(\mathbb{N}))$.

In this age of computers it should come as no surprise that each real number can be written as a *binary decimal*. In general, we can write down a binary decimal expansion

$$x = .b_1 b_2 b_3 \ldots$$

of x for some bits $b_1, b_2, b_3, \ldots \in \{0,1\}$ iff x is the specific sum

$$x = b_1 \frac{1}{2} + b_2 \frac{1}{2^2} + b_3 \frac{1}{2^3} + \ldots$$

of powers of $\frac{1}{2}$ whose coefficients b_k are $b_0, b_1, b_2, \ldots$.

Example 5.4.2 1. This one is simple. Evidently,

$$\frac{1}{2} = 1\frac{1}{2} + 0\frac{1}{2^2} + 0\frac{1}{2^3} + \ldots$$

is a sum of powers of $\frac{1}{2}$ whose sum is $\frac{1}{2}$. The coefficients of the remaining fractions $\frac{1}{2^n}$ are 0. In this case, we write

$$\frac{1}{2} = .10\bar{0}$$

which is a binary expansion for $\frac{1}{2}$.

2. A sum of powers of $\frac{1}{2}$ whose sum is $\frac{5}{8}$ is

$$\frac{5}{8} = 1\frac{1}{2} + 0\frac{1}{2^2} + 1\frac{1}{2^3} + 0\frac{1}{2^4} + 0\frac{1}{2^5} + \ldots,$$

where the coefficients of the remaining fractions $\frac{1}{2^n}$ are 0. The associated binary expansion of $\frac{5}{8}$ is

$$\frac{5}{8} = .1010\overline{0}.$$

3. Try to use the Geometric Series found in freshmen Calculus courses to show that

$$\frac{1}{3} = 0\frac{1}{2} + 1\frac{1}{2^2} + 0\frac{1}{2^3} + 1\frac{1}{2^4} + 0\frac{1}{2^5} + \dots$$

where the pattern of coefficients $0, 1, 0, 1, 0, 1, \dots$ continues indefinitely. Then

$$\frac{1}{3} = .0101\overline{01}$$

is the binary expansion of $\frac{1}{3}$.

Now, write the elements of $(0, 1)$ in the usual way. Given an $x \in (0, 1)$, write x as a binary expansion,

$$x = .b_1 b_2 b_3 \dots,$$

and pair the decimal bits $b_1, b_2, b_3, \dots$ with $\mathbb{N}$ as follows.

$$\begin{array}{ccccc} 1 & 2 & 3 & 4 & \dots \\ b_1 & b_2 & b_3 & b_4 & \dots \end{array}$$

Construct a subset $f(x) = U_x \subset \mathbb{N}$ as follows.

$$\begin{cases} n \in U_x \text{ exactly when} & b_n = 1 \\ n \notin U_x \text{ exactly when} & b_n = 0 \end{cases}.$$

That this defines a subset of $\mathbb{N}$ is clear since the bits b_n can only be 0 or 1. For example, let's see what set U_x corresponds to $\frac{1}{3}$. Because $\frac{1}{3} = .0101\overline{01}$, the pairing for $\frac{1}{3}$ looks like this.

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ 0 & 1 & 0 & 1 & 0 & 1 & \dots \end{array}$$

Then the elements of U_x are those paired with a 1. That is,

$$f\left(\frac{1}{3}\right) = \{2, 4, 6, \ldots\}.$$

The reader can show that the pairing for

$$x = .101001000100001000001\ldots$$

results in the set

$$U_x = \{1, 3, 6, 10, 15, 21, \ldots\}.$$

Search for the pattern in the decimal expansion for x.

There is a possible point of confusion that we should discuss. There are two very different ways to write $\frac{1}{2}$ as a sum of powers of $\frac{1}{2}$. They are

$$\begin{aligned} \frac{1}{2} &= 1\frac{1}{2} + 0\frac{1}{2^2} + 0\frac{1}{2^3} + 0\frac{1}{2^4} + \ldots \\ \frac{1}{2} &= 0\frac{1}{2} + 1\frac{1}{2^2} + 1\frac{1}{2^3} + 1\frac{1}{2^4} + \ldots, \end{aligned}$$

or in other words

$$\frac{1}{2} = .1\overline{0} = .0\overline{1}.$$

Which of these sums should we use in defining $f\left(\frac{1}{2}\right)$? Let us agree that, in case there are two or more ways to write a fraction x as a binary decimal expansion, we will take the one with fewer 1's. Then we take $\frac{1}{2} = .1\overline{0}$, and so

$$f\left(\frac{1}{2}\right) = \{1\}.$$

Our goal now is to show that f is one-to-one. This is done by showing that

$$x \neq y \text{ implies that } f(x) \neq f(y).$$

So suppose that $x \neq y$, and write each as a binary decimal.

$$\begin{array}{rcl} x & = & .b_1 \;\; b_2 \;\; b_3 \;\; b_4 \;\; \ldots \\ y & = & .c_1 \;\; c_2 \;\; c_3 \;\; c_4 \;\; \ldots \end{array}.$$

Since $x \neq y$, they must differ in some position, say, the nth position. That is,

$$b_n \neq c_n.$$

We may assume without loss of generality that $b_n = 1$ and that $c_n = 0$. Then by the definition of f

$$\begin{array}{rcccl} n & \in & f(x) & = & U \text{ and} \\ n & \notin & f(y) & = & V, \end{array}$$

so that U and V do not contain the same elements. It follows that

$$f(x) \;=\; U \;\neq\; V \;=\; f(y),$$

and so f is a one-to-one function.

Then $f : (0,1) \longrightarrow \mathcal{P}(\mathbb{N})$ is a one-to-one function, and therefore

$$\operatorname{card}(0,1) \leq \operatorname{card}(\mathcal{P}(\mathbb{N})). \tag{5.18}$$

Conversely, we can compare $\mathcal{P}(\mathbb{N})$ with the closed unit interval $[0,1]$, thus proving that $\operatorname{card}(0,1) \leq \operatorname{card}(\mathcal{P}(\mathbb{N}))$. Let me give a brief analysis of the proof.

Our careful inspection of $\mathcal{P}(\mathbb{N})$ will reveal that

$$\operatorname{card}(\mathcal{P}(\mathbb{N})) \leq \operatorname{card}[0,1]. \tag{5.19}$$

Combining (5.19) with $\operatorname{card}(0,1) = \operatorname{card}[0,1]$ in Theorem 5.3.3 shows that $\operatorname{card}(\mathcal{P}(\mathbb{N})) \leq \operatorname{card}[0,1] = \operatorname{card}(0,1)$. When combined with the inequality (5.18), we will have shown that $\operatorname{card}(0,1) = \operatorname{card}(\mathcal{P}(\mathbb{N}))$. Since $\operatorname{card}(\mathbb{R}) = \operatorname{card}(0,1)$, we will have proved that

$$\operatorname{card}(\mathbb{R}) = \operatorname{card}(0,1) = \operatorname{card}(\mathcal{P}(\mathbb{N})).$$

Now to begin that proof. Partition $\mathcal{P}(\mathbb{N})$ as $\mathcal{P}(\mathbb{N}) = \mathbf{F} \cup \mathbf{I}$, where

$$\mathbf{F} = \{U \in \mathcal{P}(\mathbb{N}) \mid U \text{ does not contain a sequence of numbers of the form } n, n+1, n+2, \ldots \text{ for some } n \in \mathbb{N}\}$$

and

$$\mathbf{I} = \{U \in \mathcal{P}(\mathbb{N}) \mid U \text{ contains a sequence of numbers of the form } n, n+1, n+2, \ldots \text{ for any } n \in \mathbb{N}\}.$$

For instance, $U = \{0, 2, 4, 6, \ldots\}$ is an element of $\mathbf{F}$, while $U = \{5, 6, 7, 8, \ldots\}$ is an element of $\mathbf{I}$.

Define a one-to-one function

$$f : \mathbf{F} \longrightarrow [0, 1]$$

as follows. First, write each element $U \in \mathbf{F}$ as a binary sequence. Given a set $U \in \mathbf{F}$, define a binary bit b_n as

$$b_n = \begin{cases} 1 & \text{if } n \in U \\ 0 & \text{if } n \notin U \end{cases},$$

and then

$$f(U) = .0b_0b_1b_2 \ldots \in [0, 1].$$

Note that $f(U)$ begins with 0 if $U \in \mathbf{F}$. This correspondence is so detailed that we can identify U with $b_0b_1b_2\ldots$. They both contain the same information, namely, which numbers are in U and which are not.

For instance, to construct the binary sequence that corresponds to the set $U = \{0, 2, 4, \ldots\}$ of even numbers, we find each binary bit, beginning with 0.

$$\begin{aligned} b_0 &= 1 \text{ since } 0 \in U \\ b_1 &= 0 \text{ since } 1 \notin U \\ b_2 &= 1 \text{ since } 2 \in U \\ b_3 &= 0 \text{ since } 3 \notin U \\ &\vdots \end{aligned}$$

Then

$$f(U) = f(\{0, 2, 4, 6, \ldots\}) = .010\overline{10}.$$

Given the set $U = \{5\}$, then $f(U) = .0000001\overline{0}$. If U is the set of odd numbers, then $f(U) = .0010\overline{01}$. The set U of prime numbers corresponds to the binary sequence that begins with

$$.00011010100010100010 1.$$

You can find the next five terms in that sequence if you wish. The set $U = \{4, 5, 6, \ldots\}$ is not in $\mathbf{F}$. In this manner, we can treat each $U \in \mathbf{F}$ as a binary sequence $.0b_0b_1b_2\ldots$.

The images $f(U)$ for $U \in \mathbf{F}$ are binary numbers that begin with 0. Then $f(U)$ is a number such that $0 \leq f(U) < \frac{1}{2}$. Then $f(\mathbf{F}) \subset [0, 1)$.

In a similar manner (that means that the details here will be an exercise for the student), we can define

$$f(V) = .1b_1b_2b_3\cdots,$$

where $b_1b_2b_3\cdots$ is the binary sequence determining V. Note that $f(V)$ begins with 1, so that the number $f(V)$ satisfies $\frac{1}{2} \leq f(V) \leq 1$.

In this way we have pieced together a function

$$f : \mathbf{F} \cup \mathbf{I} \longrightarrow [0, 1]$$

such that $f(\mathbf{F}) \subset [0, 1)$ and $f(\mathbf{I}) \subset \left[\frac{1}{2}, 1\right]$. Inasmuch as $\mathcal{P}(\mathbb{N}) = \{\mathbf{F}\} \cup \mathbf{I}$, we have defined a function $f : \mathcal{P}(\mathbb{N}) \rightarrow [0, 1]$.

We will show that f is one-to-one by considering three cases for U and V.

Case 1: Suppose that

$$U \neq V \text{ for some sets } U, V \in \mathbf{F}.$$

We must show that $f(U) \neq f(V)$. As we did above, write the sets

$$U = b_0b_1b_2\ldots \text{ and } V = c_0c_1c_2\ldots$$

as *binary sequences.*. Since $U \neq V$, there is an $n \in \mathbb{N}$ such that $b_n \neq c_n$. Then $.0b_1b_2b_3\cdots \neq .0c_1c_2c_3\cdots$ are the different numbers because they differ in the n-th place. Thus,

$$f(U) = .0b_1b_2b_3\cdots \neq .1c_1c_2c_3\cdots = f(V).$$

Case 2: Suppose that

$$U \neq V \text{ for some sets } U, V \in \mathbf{I}.$$

We must show that $f(U) \neq f(V)$. As we did above, write the sets

$$U = b_0 b_1 b_2 \dots \text{ and } V = c_0 c_1 c_2 \dots$$

as *binary sequences*.. Since $U \neq V$, there is an $n \in \mathbb{N}$ such that $b_n \neq c_n$. Then $.1b_1b_2b_3 \cdots \neq .1c_1c_2c_3 \cdots$ are the different numbers because they differ in the n-th place. Thus,

$$f(U) = .1b_1b_2b_3 \cdots \neq .1c_1c_2c_3 \cdots = f(V).$$

CASE 3: The reader will show that if $U \neq V$, if $U \in \mathbf{F}$, and if $V \in \mathbf{I}$, then $f(U) \neq f(V)$. Since this exhausts all of the possible cases for U and V, we conclude that f is a one-to-one function.

Consequently, $\text{card}(\mathcal{P}(\mathbb{N})) \leq \text{card}[0,1]$. Since we have already proved that $\text{card}(\mathbb{R}) = \text{card}[0,1] \leq \text{card}(\mathcal{P}(\mathbb{N}))$, we conclude that

$$\text{card}(\mathbb{R}) = \text{card}[0,1] = \text{card}(\mathcal{P}(\mathbb{N})).$$

This concludes the proof of Theorem 5.4.1.

In Theorem 5.3.1, we showed that

$$\text{card}(\mathbb{N}) < \text{card}[0,1] = \text{card}(\mathbb{R}),$$

thus showing that there are at least two different infinite cardinals. This presents the problem of determining how many infinite cardinals there are. It seems unlikely that there are just two, so we look for more.

The following beautiful result shows us that no matter which set A we choose, $\mathcal{P}(A)$ is always a larger set than A. That is, the cardinality of $\mathcal{P}(A)$ is always more than the cardinality of A. This fact is certainly True of *finite* cardinals as we have proved that

$$\text{if } \text{card}(A) = n \text{ is finite then } \text{card}(\mathcal{P}(A)) = 2^n.$$

We see immediately that

$$\text{card}(A) = n < 2^n = \text{card}(\mathcal{P}(A))$$

for any $n \in \mathbb{N}$, thus proving that $\mathcal{P}(A)$ has more elements than the finite set A. If $A = \{a_1, \dots, a_n\}$ has exactly n elements and if we let

$$\begin{aligned} \mathcal{P}^2(A) &= \mathcal{P}(\mathcal{P}(A)) & & \\ \mathcal{P}^3(A) &= \mathcal{P}(\mathcal{P}(\mathcal{P}(A))) &&= \mathcal{P}(\mathcal{P}^2(A)) \\ \mathcal{P}^4(A) &= \mathcal{P}(\mathcal{P}(\mathcal{P}(\mathcal{P}(A)))) &&= \mathcal{P}(\mathcal{P}^3(A)) \\ &\vdots \end{aligned}$$

(note the use of 4 $\mathcal{P}$'s in the third line), then we see that

$$\begin{array}{lclcll} \text{card}(\mathcal{P}(A)) &=& 2^n & & & \\ \text{card}(\mathcal{P}^2(A)) &=& \text{card}(\mathcal{P}(\mathcal{P}(A))) &=& 2^{2^n} & \text{since card}(\mathcal{P}(A)) = 2^n \\ \text{card}(\mathcal{P}^3(A)) &=& \text{card}(\mathcal{P}(\mathcal{P}^2(A))) &=& 2^{2^{2^n}} & \text{since card}(\mathcal{P}^2(A)) = 2^n \\ \text{card}(\mathcal{P}^4(A)) &=& \text{card}(\mathcal{P}(\mathcal{P}^3(A))) &=& 2^{2^{2^n}} & \text{since card}(\mathcal{P}^3(A)) = 2^{2^n} \\ \vdots & & & & & \end{array}$$

This leads us to the infinite sequence of *finite cardinals*

$$\boxed{n < 2^n < 2^{2^n} < 2^{2^{2^n}} < \ldots.}$$

Our next result shows that there is a similar sequence of *infinite* cardinals.

We challenge the reader to find the similarities between the argument used below, and the argument used to discuss the Epimenides Paradox in Theorem 1.7.1.

Theorem 5.4.3 *If A is a set then* $\text{card}(A) < \text{card}(\mathcal{P}(A))$.

Proof: First we establish that $\text{card}(A) \leq \text{card}(\mathcal{P}(A))$. The symbol λ is the Greek letter called *lambda*. The one-to-one function we need is

$$\lambda : A \longrightarrow \mathcal{P}(A)$$

defined by

$$\lambda(x) = \{x\}.$$

Certainly $\lambda(x) = \lambda(x')$ implies that $\{x\} = \{x'\}$ or equivalently that $x = x'$. Then λ is a one-to-one function, and hence $\text{card}(A) \leq \text{card}(\mathcal{P}(A))$.

To prove that $\text{card}(A) < \text{card}(\mathcal{P}(A))$, we need to show that no function

$$F : A \longrightarrow \mathcal{P}(A)$$

is onto, thus showing that there are no bijections between A and $\mathcal{P}(A)$. It will then follow that $\text{card}(A) \neq \text{card}(\mathcal{P}(A))$ and hence that $\text{card}(A) < \text{card}(\mathcal{P}(A))$.

Given a function $F : A \longrightarrow \mathcal{P}(A)$ and an $x \in A$, observe that $F(x) \in \mathcal{P}(A)$ so that $F(x)$ is a subset of A, $F(x) \subset A$. We can ask, *Is $x \in F(x)$ or is $x \notin F(x)$?* Exactly one of these inclusions is True. Define a special set $X \subset A$ by

$$X = \{x \in A \mid x \notin F(x)\}.$$

The set X is a remarkable observation made by Bertrand Russell in 1920 and has no motivation. Let's accept the genius that uncovered this idea, that it works well, and then see how it is used in our argument. We claim that there does not exist an $x \in A$ such that $F(x) = X$.

Suppose to the contrary that there is an $x \in A$ such that $F(x) = X$. We seek a contradiction. Since X is a set we ought to be able to decide whether or not $x \in F(x) = X$ or $x \notin F(x) = X$. We consider cases.

CASE 1: Suppose that $x \in F(x) = X$. Then x satisfies the predicate defining X, which is $x \notin F(x)$. This is not possible since $F(x) = X$ is a set, so Case 1 does not apply.

CASE 2: So it must be that $x \notin F(x) = X$. In this case, x satisfies the predicate for X, so that $x \in X$. But this is also impossible, since X is a set. Then Case 2 does not apply.

The failure of Case 1 and Case 2 shows us that we cannot decide if $x \in F(x) = X$ or if $x \notin F(x) = X$. This contradiction to the definition of set shows us that our initial assumption that $F(x) = X$ for some x is a Falsehood. Thus, there is no $x \in A$ such that $F(x) = X$. This proves the claim that $F : A \rightarrow \mathcal{P}(A)$ is not an onto function, and therefore $\text{card}(A) \neq \text{card}(\mathcal{P}(A))$. Hence $\text{card}(A) < \text{card}(\mathcal{P}(\mathbb{N}))$, which completes our proof.

Theorem 5.4.4 $\text{card}(\mathbb{N}) < \text{card}(\mathcal{P}(\mathbb{N}))$.

Theorem 5.4.3 allows us to prove the following counterintuitive result. In a language devoid of mathematics it implies that *there are infinitely many infinities.*

Theorem 5.4.5 *There is an infinite sequence $\alpha_0 < \alpha_1 < \alpha_2 < \ldots$ of infinite cardinals.*

Proof: If we let $\alpha_0 = \text{card}(\mathbb{N})$ and let $\alpha_1 = \text{card}(\mathcal{P}(\mathbb{N}))$, then Theorem 5.4.3 shows us that $\alpha_0 < \alpha_1$. Next, consider the sequence

$$\mathcal{P}(\mathbb{N}), \mathcal{P}^2(\mathbb{N}) = \mathcal{P}(\mathcal{P}(\mathbb{N})), \mathcal{P}^3(\mathbb{N}) = \mathcal{P}(\mathcal{P}^2(\mathbb{N})), \ldots.$$

and let

$$\alpha_1 = \text{card}(\mathcal{P}(\mathbb{N})), \alpha_2 = \text{card}(\mathcal{P}^2(\mathbb{N})), \alpha_3 = \text{card}(\mathcal{P}^3(\mathbb{N})), \ldots.$$

In other words, if we let $\mathcal{P}(\mathbb{N}) = A$ then $\mathcal{P}^2(\mathbb{N}) = \mathcal{P}(A)$ is the power set of $A = \mathcal{P}(\mathbb{N})$. Theorem 5.4.3 shows us that

$$\alpha_1 = \text{card}(\mathcal{P}(\mathbb{N})) = \text{card}(A) < \text{card}(\mathcal{P}(A)) = \text{card}(\mathcal{P}^2(\mathbb{N})) = \alpha_2.$$

That is, $\alpha_o < \alpha_1 < \alpha_2$.

In general, let $A = \mathcal{P}^n(\mathbb{N})$. Because

$$\mathcal{P}(A) = \mathcal{P}(\mathcal{P}^n(\mathbb{N})) = \mathcal{P}^{n+1}(\mathbb{N}),$$

Theorem 5.4.3 shows us that

$$\text{card}(\mathcal{P}^n(\mathbb{N})) = \text{card}(A) < \text{card}(\mathcal{P}(A)) = \text{card}(\mathcal{P}(\mathcal{P}^n(\mathbb{N}))) = \text{card}(\mathcal{P}^{n+1}(\mathbb{N})).$$

By setting

$$\boxed{\alpha_n = \text{card}\left(\mathcal{P}^n(\mathbb{N})\right) \qquad \text{for each } n \in \mathbb{N},}$$

we have constructed *an infinite chain of infinite cardinals*

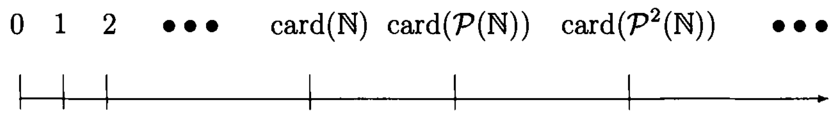
$$\boxed{\alpha_o < \alpha_1 < \ldots < \alpha_n < \alpha_{n+1} < \ldots.}$$

Specifically,

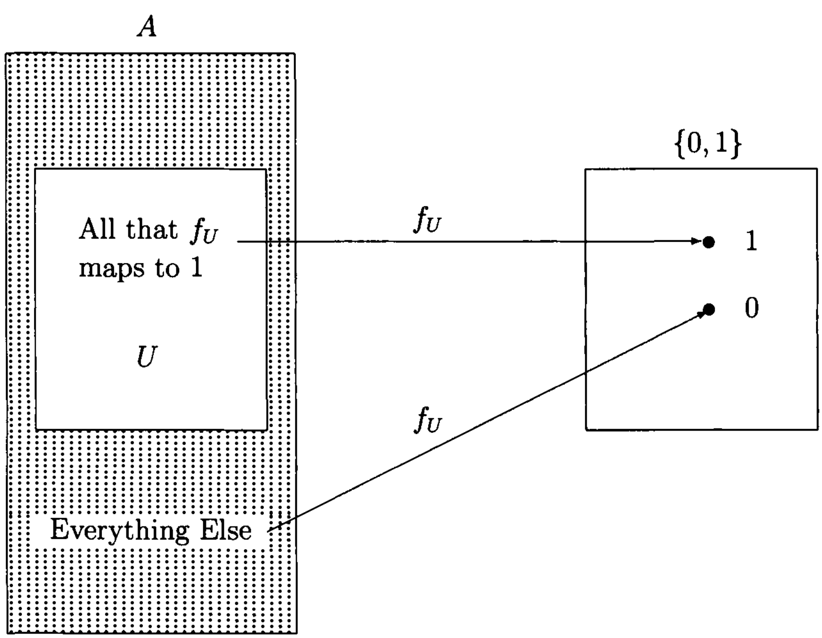
$$\boxed{\text{card}(\mathbb{N}) < \text{card}\left(\mathcal{P}(\mathbb{N})\right) < \text{card}\left(\mathcal{P}^2(\mathbb{N})\right) < \text{card}\left(\mathcal{P}^3(\mathbb{N})\right) < \ldots.}$$

This ends the proof.

In pictures, the above chain can be realized as in the following line. This is not the real line. It is just a visual means of describing a chain of cardinals.



Let us examine another method for constructing cardinals and infinities. Recall that if A is a set, then $\{0,1\}^A$ is the set of functions $f : A \longrightarrow \{0,1\}$.



Theorem 5.4.6 *Let A be a set. Then*

$$\boxed{\operatorname{card}(\mathcal{P}(A)) = \operatorname{card}\left(\{0,1\}^A\right).}$$

Proof: We will produce a bijection

$$F \;:\; \mathcal{P}(A) \longrightarrow \{0,1\}^A$$

from which we will conclude that

$$\mathrm{card}(\mathcal{P}(A)) \;=\; \mathrm{card}(\{0,1\}^A).$$

Given a subset $U \subset A$, define a function

$$f_U \;:\; A \longrightarrow \{0,1\}$$

as in the above diagram.

$$f_U(x) \;=\; \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{if } x \notin U \end{cases}.$$

Then $f_U(x) = 1$ precisely when $x \in U$. Otherwise it is 0. Clearly

$$f_U \in \{0,1\}^A,$$

so that the rule

$$F(U) = f_U$$

defines a function $F : \mathcal{P}(A) \longrightarrow \{0,1\}^A$. We will show that F is a bijection.

To see that F is one-to-one, let

$$U \neq V$$

be subsets of A. By the definition of equal sets, some element x is in one set and not the other. We assume without loss of generality that there is an element $x \in U$ that is not in V. Let f_V be the function that takes V to 1 and all else to 0. Then

$$x \in U \quad \text{and} \quad x \notin V,$$

so that

$$f_U(x) = 1 \quad \text{while} \quad f_V(x) = 0$$

by the definitions of the functions f_U and f_V. Then f_U and f_V have different rules, so that

$$F(U) = f_U \neq f_V = F(V).$$

We conclude that $F : \mathcal{P}(A) \longrightarrow \{0,1\}^A$ is a one-to-one function.

To see that $F : \mathcal{P}(A) \longrightarrow \{0,1\}^A$ is an onto function, let $f \in \{0,1\}^A$. Define a subset of A,

$$U = \{x \in A \,|\, f(x) = 1\}.$$

That is, $U = f^{-1}(1)$, or in other words, U is the set of all $x \in A$ that are sent by f to 1. (See the above diagram.) Then

$$\begin{aligned} f(x) &= 1 = f_U(x) \quad \text{for all } x \in U \text{ and} \\ f(x) &= 0 = f_U(x) \quad \text{for all } x \notin U. \end{aligned}$$

Since f and f_U have the same rules, $f = f_U$, and since $F(U) = f_U$, we have

$$F(U) = f_U = f.$$

It follows that $F : \mathcal{P}(A) \longrightarrow \{0,1\}^A$ is onto.

Hence F is a bijection, and therefore

$$\operatorname{card}(\mathcal{P}(A)) = \operatorname{card}\left(\{0,1\}^A\right),$$

which completes the proof.

Theorem 5.4.7 *1.* $\operatorname{card}(\mathbb{N}) < \operatorname{card}\left(\{0,1\}^{\mathbb{N}}\right)$.

2. $\operatorname{card}(\mathbb{R}) < \operatorname{card}\left(\{0,1\}^{\mathbb{R}}\right)$.

Proof: 1. By Theorem 5.4.3, $\operatorname{card}(\mathbb{N}) < \operatorname{card}(\mathcal{P}(\mathbb{N}))$, and by Theorem 5.4.6, $\operatorname{card}(\mathcal{P}(\mathbb{N})) = \operatorname{card}(\{0,1\}^{\mathbb{N}})$. Then $\operatorname{card}(\mathbb{N})$ is strictly smaller than $\operatorname{card}(\{0,1\}^{\mathbb{N}})$. Part 2 is handled in a similar fashion. This completes the proof.

The notation we will use for the exponents is also an exponent. Given a set A, let

$$\boxed{\operatorname{card}(A) = \aleph}$$

(read as *aleph*) and then let

$$\boxed{\operatorname{card}\left(\{0,1\}^A\right) = 2^{\aleph}.}$$

This does not imply that we are taking a power of 2 to some infinite cardinal. It is just a convenient method for writing some cardinalities. There is something more to see here though.

Theorem 5.4.8 *Let $\aleph$ be a cardinal. That is, $\aleph = \text{card}(A)$ for some set A. Then*

$$\boxed{\aleph < 2^{\aleph}.}$$

Proof: From Theorems 5.4.3 and 5.4.6 we see that

$$\aleph = \text{card}(A) < \text{card}(\mathcal{P}(A)) = \text{card}(\{0,1\}^A) = 2^{\text{card}(A)} = 2^{\aleph}.$$

This completes the proof.

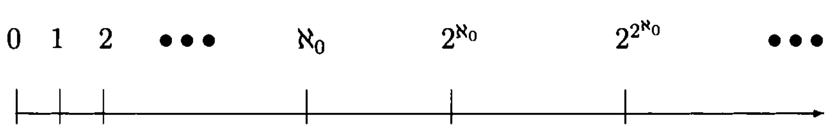
It is a tradition dating back to Georg Cantor [1] in the late nineteenth century to use the first letter of the Hebrew alphabet, $\aleph$, read *aleph*, to symbolize infinite cardinals. Following in this tradition, we let

$$\boxed{\text{card}(\mathbb{N}) = \aleph_0,}$$

where $\aleph_0$ is read as *aleph nought.* With Theorem 5.4.6 and our new notation, we can write

$$\begin{aligned} \aleph_0 &= \text{card}(\mathbb{N}), \\ 2^{\aleph_0} &= \text{card}\left(\{0,1\}^{\aleph_0}\right) = \text{card}(\mathcal{P}(\mathbb{N})), \\ 2^{2^{\aleph_0}} &= \text{card}\left(\{0,1\}^{\{0,1\}^{\aleph_0}}\right) = \text{card}(\mathcal{P}(\mathcal{P}(\mathbb{N}))), \\ &\vdots \end{aligned}$$

which by Theorem 5.4.6, leads us to an infinite sequence of infinite cardinals



The next question we will explore is, can we find some inequality between card $(\{0,1\}^{\mathbb{N}})$ and card$(\mathbb{R})$?

Exercise 5.4.9 1. Write $\frac{3}{4}$ and $\frac{1}{5}$ as binary decimals.

2. Fill in the details of CASE 2 in Theorem 5.4.1.

3. Fill in the details of CASE 3 in Theorem 5.4.1.

4. Go over the details in Theorem 5.4.3 when $A = \mathbb{N}$.

5. Discuss the similarities between the proofs in Example 1.7.1 and Theorem 5.4.3.

5.5 The Arithmetic of Cardinals

We will use the Greek letters α (alpha) and β (beta), and the Hebrew letter $\aleph$ (aleph) to denote general cardinals. We will also use the traditional notation of card$(\mathbb{N}) = \aleph_0$.

What you are about to read should be interesting. We are about to define the addition and multiplication of natural numbers. This is the arithmetic that you learned in elementary school. We will define $1+1$ and show that it is 2. We will define $2 \cdot 3$ and show that it is 6. We will eventually use this arithmetic to define operations like addition and multiplication on infinite cardinals. And surprise! $\aleph_0 + \aleph_0 = \aleph_0$. Now isn't that what you expected of infinity? But what value would you expect of the following sum?

$$\aleph_0 + 2^{\aleph_0} = \ ?$$

Let us define *the addition of natural numbers.* Sets E and F are said to be *disjoint* if $E \cap F = \emptyset$.

Definition 5.5.1 *Let $n, m \in \mathbb{N}$. Choose disjoint finite sets E and F such that* card$(E) = n$ *and* card$(F) = m$. *Then we define*

$$\boxed{n + m = \text{card}(E \cup F).}$$

As our first example of how to use this definition, we will prove an elementary arithmetic fact. Recall from Section 2.5 that 1 is the set of all sets that are equivalent to the set $\{\bullet\}$, and similarly that 2 denotes the set of all sets that are equivalent to $\{\bullet, \square\}$. Then

$$\operatorname{card}(\{\bullet\}) = \operatorname{card}(\{\square\}) = 1$$

so that

$$\begin{aligned} 1+1 &= \operatorname{card}(\{\bullet\} \cup \{\square\}) \\ &= \operatorname{card}(\{\bullet, \square\}) \\ &= 2. \end{aligned}$$

We have just proved that $1+1=2$. We did not state $1+1=2$ as a numerical identity, or as we used to call them, *a number fact.* We did not ask for measurements, and we did not accept the teacher's word for it. We proved it with the same certainty that a proof about congruent triangles would possess in a Geometry course. If we were to expand this investigation to another level, we would discover that Set Theory is a point of view that allows us to discuss all of the mathematical fundamentals with logical certainty. That is the power that these axioms of mathematics possess.

This idea of considering the addition of natural numbers as the union of finite sets is what we will use to define *the addition of infinite cardinals.*

Definition 5.5.2 *Suppose that at least one of the cardinals α and β is infinite. Choose any sets A and B such that* $\operatorname{card}(A) = \alpha$ *and* $\operatorname{card}(B) = \beta$. *Then we define*

$$\boxed{\alpha + \beta = \operatorname{card}(A \cup B).}$$

Note the difference between the addition of numbers, and the addition of infinite cardinals.

For the addition of the natural numbers n and m we required *disjoint* sets E and F such that $\operatorname{card}(E) = n$ and $\operatorname{card}(F) = m$, and then we define $n + m = \operatorname{card}(E \cup F)$. The power possessed by the addition of infinite cardinals α and β is that to add α and β, we choose *any* of set A and B such that $\text{card}(A) = \alpha$ and $\text{card}(B) = \beta$. In other words, unlike the addition of finite cardinals, the chosen sets A and B used in the addition of infinite cardinals *do not have to be disjoint.*

We are taking advantage of a theorem in Set Theory when we define cardinal addition in this way. Professionals use a weaker statement as a definition. But for our purposes, the definition I have presented is a good working definition. Thus, $\alpha + \beta$ is independent of the intersection $A \cap B$ of our chosen sets A and B. The professional mathematician would say that our presented definition of the addition of infinite cardinals is *well defined.*

The addition of infinite cardinals has some of the properties associated with the addition of real numbers, and as we will see, the addition of infinite cardinals can be quite unexpected. One similarity is the *commutative law of addition.* Given cardinals α and β such that at least one them is infinite, then

$$\boxed{\alpha + \beta = \beta + \alpha.}$$

To see the truth of this matter, choose sets A and B such that $\alpha = \text{card}(A)$ and $\beta = \text{card}(B)$. Then

$$\begin{aligned} \alpha + \beta &= \text{card}(A \cup B) \\ &= \text{card}(B \cup A) \\ &= \beta + \alpha, \end{aligned}$$

which is what we wanted to prove.

We have just *proved* that the addition of infinite cardinals is commutative. We proved that the commutative law of addition holds for *all* infinite cardinals, not just a chosen few used for demonstration purposes. In the case of the arithmetic of numbers, the commutative law has been handed down to you through educational tradition. But no one (I'm betting) in your past proved that this property was True *of all* cardinals or numbers.

Example 5.5.3 Traditionally one uses the German letter

$$\boxed{\mathfrak{c} = \text{card}(\mathbb{R}),}$$

to denote the cardinality of the real line. The symbol $\mathfrak{c}$ stands for *the continuum*, an old term used to describe the real line. Then

$$\begin{aligned} \aleph_0 + \mathfrak{c} &= \mathfrak{c} + \aleph_0 \text{ and} \\ 2^{\aleph_0} + \mathfrak{c} &= \mathfrak{c} + 2^{\aleph_0}. \end{aligned}$$

This shows us that the addition of these specific cardinals is commutative, but it does not show us that addition is in general commutative, nor does it indicate what the indicated sums are.

As children, we might have heard an adage that *infinity plus 1 is infinity* or that *infinity plus infinity is infinity.* This is mildly True, since any number of elements added to an infinite set still yields an infinite set. However, as we saw in the previous sections, the term *infinite* is just too coarse a measurement for the size of a set. It is like saying, *the house has color.* Sure it has color, but which ones? Saying that something is infinite, simply says that it is not finite. It does not tell us a thing about the cardinality of the set. Is its cardinality $\aleph_0$, or $\mathfrak{c}$, or some other value?

As an example of how to add infinite cardinals, consider the following result. Our work to date now brings mathematical truth to the old adage *Infinity plus one is infinity.*

Theorem 5.5.4 *Let* $\aleph$ *be any infinite cardinal.*

1. $0 + \aleph = \aleph$.
2. $n + \aleph = \aleph$ *for any* $n \in \mathbb{N}$.

Proof: 1. By definition, there is a set A such that $\aleph = \text{card}(A)$, and by its definition in Section 2.5, $0 = \text{card}(\emptyset)$. Then

$$0 + \aleph = \text{card}(\emptyset \cup A) = \text{card}(A) = \aleph.$$

This proves part 1.

2. Since A is infinite and n is finite, there is a set $\{a_1, \ldots, a_n\} \subset A$ of n elements. Specifically, $n = \text{card}\{a_1, \ldots, a_n\}$ and $\{a_1, \ldots, a_n\} \cup A = A$. Then

$$n + \aleph = \text{card}(\{a_1, \ldots, a_n\} \cup A) = \text{card}(A) = \aleph,$$

which proves part 2. This completes the proof.

For example, part 2 above shows us that

$$\aleph_0 + 1 = \aleph_0$$

and that

$$\aleph_0 + 10^{10^{10^{10^{10}}}} = \aleph_0$$

even though $10^{10^{10^{10^{10}}}}$ is a huge finite natural number.

The next result might be read as *infinity plus infinity is infinity.* This vague statement may sound True, but it is void of the mathematical language necessary to bring it truth. If we use infinite cardinals, however, then we can say with certainty which infinite cardinal the sum is.

Theorem 5.5.5 *Let* $\aleph$ *be* any *infinite cardinal. Then*

$$\boxed{\aleph + \aleph = \aleph.}$$

Proof: We know that $\aleph = \text{card}(A)$ for some set A. Hence

$$\aleph + \aleph = \text{card}(A \cup A) = \text{card}(A) = \aleph,$$

which ends the proof.

Example 5.5.6 Using Theorem 5.5.5, we readily see that

$$\aleph_0 + \aleph_0 = \aleph_0,$$

that

$$\mathfrak{c} + \mathfrak{c} = \mathfrak{c},$$

and that

$$2^{\aleph} + 2^{\aleph} = 2^{\aleph}$$

for any infinite cardinal $\aleph$.

You see? An infinite cardinal plus the same infinite cardinal is the same infinite cardinal. Or in other words, an infinite cardinal added to itself is itself. For that reason, we will not subtract cardinals because *subtraction makes no sense on infinite cardinals.* To see the truth of this matter, consider the following example.

Example 5.5.7 Let $\aleph$ be an infinite cardinal. We will show that

There is no way to define $\aleph - \aleph$ as a cardinal.

Proof: Assume to the contrary, that we could define the subtraction $\aleph - \aleph$ as the opposite of addition. The implied value of the subtraction $\aleph - \aleph$ is

$$\aleph - \aleph = 0.$$

By Theorem 5.5.4(1),

$$1 + \aleph \;=\; \aleph.$$

Using the assumed subtraction, we would then subtract $\aleph$ from both sides of the equation, thus revealing that

$$\begin{aligned} 1 + (\aleph - \aleph) &= \aleph - \aleph \\ 1 + 0 &= 0 \\ 1 &= 0, \end{aligned}$$

which is ridiculous. This mathematical mistake shows us that our initial assumption is wrong, and thus $\aleph - \aleph$ cannot be defined. This concludes the proof.

So given *any* infinite cardinal $\aleph$, then we cannot define $\aleph - \aleph$ in a satisfying arithmetical manner. If we cannot define $\aleph - \aleph$, then *a more general subtraction of infinite cardinals is not possible.* In general, there is the more daunting rule of *a lack of cancellation* of infinite cardinals.

You cannot cancel infinite cardinals from an equation.

For instance, if we are given an infinite cardinal α, and the equation $\alpha+1 = \alpha+\aleph_0$, then we cannot cancel the α's and deduce the obvious Falsehood, $1 = \aleph_0$.

So let us decide what the addition of two cardinals is. That is, what is $\alpha + \beta$? The next result gives further mathematical certainty to the old chestnut *infinity plus infinity is infinity.*

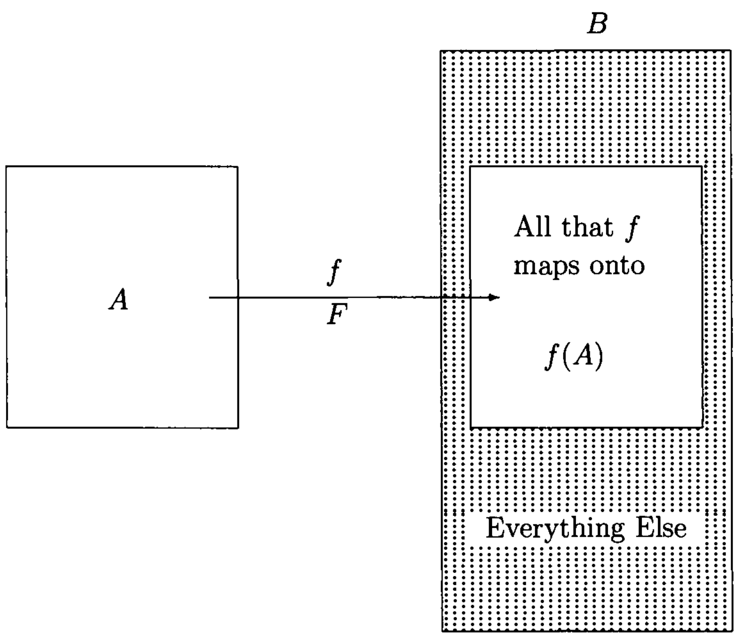
Theorem 5.5.8 *Let $\alpha < \beta$ be infinite cardinals. Then*

$$\boxed{\alpha + \beta = \beta = \text{ the larger of two cardinals.}}$$

Proof: To prove this result, we will need sets A and B such that $\alpha = \text{card}(A)$ and $\beta = \text{card}(B)$. By the definition of inequality of cardinals (see page 151), there is a one-to-one function

$$f : A \longrightarrow B.$$

We will show that A and $f(A)$ are equivalent sets. The function $f : A \longrightarrow B$ restricts to one $F : A \longrightarrow f(A)$. Then f and F have the same rule, but they map into different sets. The following picture will help you see what we are doing with f, F, and $f(A)$.



Specifically, f may not be onto, but we will show that F is a bijection. The function F is one-to-one because that is how we chose f. To see that F is onto, let $y \in f(A)$. By definition of $f(A)$, there is an $x \in A$ such that

$$F(x) = f(x) = y.$$

Then F is onto, hence F is a bijection, whence A and $f(A)$ are equivalent.

Since $f(A) \subset B$, $f(A) \cup B = B$, so that

$$\begin{aligned}
\alpha + \beta &= \operatorname{card}(A) + \operatorname{card}(B) \\
&= \operatorname{card}(f(A)) + \operatorname{card}(B) \\
&= \operatorname{card}(f(A) \cup B) \\
&= \operatorname{card}(B) \\
&= \beta.
\end{aligned}$$

This completes the proof.

Example 5.5.9 For more familiar cardinals, there are the following examples.

1. Because by Theorem 5.3.1,

$$\operatorname{card}(\mathbb{N}) = \aleph_0 < \mathfrak{c} = \operatorname{card}(\mathbb{R}),$$

Theorem 5.5.8 states that

$$\aleph_0 + \mathfrak{c} = \mathfrak{c},$$

2. By Theorem 5.4.8, $\aleph < 2^{\aleph}$, so

$$\aleph + 2^{\aleph} = 2^{\aleph},$$

by Theorem 5.5.8. Specifically,

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

A curious consequence of this addition is that $\aleph_0$ behaves like 0 on the infinite cardinals. You see,

$$\aleph_0 \leq \aleph$$

for each infinite cardinal $\aleph$, so by Theorem 5.5.8,

$$\boxed{\aleph_0 + \aleph = \aleph \text{ for infinite cardinals } \aleph.}$$

Don't try to cancel that $\aleph$ from the equation. That cancellation would employ the subtraction of cardinals $\aleph - \aleph$ which we just showed was not defined. Again, our finite intuition is worthless in the infinite setting.

Since we can add cardinals to themselves we can define *multiples of cardinals*. Let $\aleph$ be an infinite cardinal and let $n \in \mathbb{N}$. Then

$$n \cdot \aleph = \underbrace{\aleph + \ldots + \aleph}_{n \text{ summands}} = \aleph.$$

Specifically,

$$\begin{aligned} 2 \cdot \aleph_0 &= \aleph_0 \text{ and} \\ n \cdot \mathfrak{c} &= \mathfrak{c} \text{ for each } n \in \mathbb{N}. \end{aligned}$$

Example 5.5.10 We could not define the subtraction of infinite cardinals, and we cannot define the division of infinite cardinals. Thus, the multiplication of infinite cardinals is not associated with a *division* operation. Specifically, for an infinite cardinal $\aleph$,

$$\boxed{\text{There is no way to define } \aleph \cdot \frac{1}{\aleph} \text{ as a cardinal.}}$$

Proof: Here's why. Assume to the contrary that we can divide by the infinite cardinal $\aleph$. One of the fundamental identities implied by the existence of a division is that

$$\aleph \cdot \frac{1}{\aleph} = 1.$$

By applying the assumed division we would then have the following sequence of equations.

$$\begin{aligned} 2 \cdot \aleph &= \aleph \\ 2 \cdot \aleph \cdot \frac{1}{\aleph} &= \aleph \cdot \frac{1}{\aleph} \\ 2 \cdot 1 &= 1 \\ 2 &= 1, \end{aligned}$$

another clear contradiction. Therefore our initial assumption must be tossed out, and we conclude that *we cannot divide by* $\aleph$. This concludes the proof.

This may be your first experience with a multiplication that is not associated with a division. It might seem strange at first but the laws of mathematics demand it.

Moving on to the *multiplication of infinite cardinals*, observe that the Cartesian Product

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

consists of the 6 pairs

$$\begin{array}{ccc} (1,1) & (1,2) & (1,3) \\ (2,1) & (2,2) & (2,3) \end{array}$$

of numbers. We have just proved the arithmetic fact

$$2 \cdot 3 = 6.$$

You should be sitting straight up in your chair when I tell you something like this. Here is a fundamental concept of human intellect, and I am showing you how to prove it from more basic facts.

The reader can write down the 15 pairs that make up $\{1,2,3\} \times \{1,2,3,4,5\}$. The point behind these examples is that we have found a way to *define the multiplication* of natural numbers.

Theorem 5.5.11 if E and F are finite sets, *then*

$$\boxed{\operatorname{card}(E \times F) \quad = \quad \operatorname{card}(E)\operatorname{card}(F).}$$

Proof: To see the truth of the matter, suppose that $E = \{1, \ldots, m\}$ and that $F = \{1, \ldots, n\}$. Arrange $E \times F$ as a rectangular array

$$\begin{array}{cccc} (1,1) & (1,2) & \ldots & (1,n) \\ (2,1) & (2,2) & \ldots & (2,n) \\ \vdots & & & \vdots \\ (m,1) & (m,2) & \ldots & (m,n) \end{array}$$

of pairs (x, y) of natural numbers. There are $m = \text{card}(E)$ rows and $n = \text{card}(F)$ columns in this array, so there must be $m \cdot n$ pairs in the rectangular array. Then $\text{card}(E \times F) = mn = \text{card}(E)\text{card}(F)$, which is what we had to prove.

Thus motivated, it is reasonable to define the product of infinite cardinals as follows.

Definition 5.5.12 *Let α and β be infinite cardinals. Choose sets A and B such that $\text{card}(A) = \alpha$ and $\text{card}(B) = \beta$. We define*

$$\boxed{\alpha \cdot \beta \;=\; \text{card}(A \times B).}$$

As before, we want to verify that *the multiplication of cardinals is commutative.* That is,

$$\boxed{\alpha \cdot \beta = \beta \cdot \alpha.}$$

To prove this, choose sets A and B such that $\alpha = \text{card}(A)$ and $\beta = \text{card}(B)$. The reader is invited to show that the function

$$f : A \times B \longrightarrow B \times A,$$

which is defined by switching components

$$f(a, b) = (b, a),$$

is a bijection. Then $\text{card}(A \times B) = \text{card}(B \times A)$, and hence

$$\begin{aligned} \alpha \cdot \beta &= \text{card}(A \times B) \\ &= \text{card}(B \times A) \\ &= \beta \cdot \alpha. \end{aligned}$$

Thus, the multiplication of infinite cardinals satisfies the commutative law.

The commutative law may be the totality of the similarity between the multiplication of natural numbers, and the multiplication of infinite cardinals.

Example 5.5.13 Let us examine the product $\aleph_0 \cdot \aleph_0$ under this multiplication. Since $\aleph_0 = \text{card}(\mathbb{N})$ we can write

$$\begin{aligned} \aleph_0 \cdot \aleph_0 &= \text{card}(\mathbb{N} \times \mathbb{N}) \\ &= \text{card}(\mathbb{N}) \qquad \text{by Theorem 5.1.5} \\ &= \aleph_0. \end{aligned}$$

Hence,

$$\boxed{\aleph_0 \cdot \aleph_0 = \aleph_0.}$$

This multiplication must strike you as strange. The only real numbers x for which $x \cdot x = x$ are 0 and 1. And $\aleph_0$ is not special in this regard. In general,

$$\boxed{\aleph \cdot \aleph = \aleph \quad \text{for any infinite cardinal } \aleph.}$$

Furthermore, since $\aleph \cdot \aleph = \aleph^2$ we are led to consider square roots of cardinals. Things are not what our old common sense might have said to us.

$$\boxed{\sqrt{\aleph} = \aleph \quad \text{for any infinite cardinal } \aleph.}$$

This is another difference between natural numbers and infinite cardinals. The square root of most natural numbers is not a natural number. However, the square root of *any* infinite cardinal is an infinite cardinal. Here is another peculiarity. Question: How many natural numbers do you know of that are their own square roots? Answer: Only two, 0 and 1. And yet *every* infinite cardinal is its own square root. The land of infinite cardinals is indeed strange.

You might skip this discussion if you have never had calculus. Here is a property of multiplication that may not agree with your Calculus lectures.

Limits of the form $\frac{0}{0}$ require an application of L'Hopital's Rule. In using L'Hopital's Rule, the student immediately sees that if $f(x)$ and $g(x)$ are functions such that $f(a) = g(a) = 0$, then the limit

$$\lim_{x \to a} \frac{f(x)}{g(x)}$$

can be almost any real number. The value of the limit depends on the functions $f(x)$ and $g(x)$. For example,

$$\lim_{x \to 0} \frac{\sin(x)}{x} = 1$$

and

$$\lim_{x \to 1} \frac{x^2 - 1}{x^2 - 3x + 2} = -2.$$

Similarly, the limits of the form $0 \cdot \infty$ do not conform to the general rules for multiplication by 0, as they too can be any real number. In Set Theory, no such rule is necessary.

Theorem 5.5.14 *Let $\aleph$ be an infinite cardinal. Then*

$$\boxed{0 \cdot \aleph = 0.}$$

Proof: This one is tricky. There is a set A such that $\mathrm{card}(A) = \aleph$. See Example 2.3.2 to see that $\emptyset \times A = \emptyset$ for any set A. Hence $0 \cdot \aleph = \mathrm{card}(\emptyset \times A) = 0$. This completes the proof.

The following theorem shows us that like addition, the multiplication of cardinals is dominated by the larger cardinal. So of course, numbers and infinite cardinals do not share in a similar multiplication.

Theorem 5.5.15 *Let $\alpha \leq \beta$ be infinite cardinals. Then*

$$\boxed{\alpha \cdot \beta = \beta = \text{ the larger of the two cardinals.}}$$

We do not prove this one, but we do offer the following examples.

Example 5.5.16

$$\begin{aligned}
\aleph_0 \cdot \aleph_0 &= \aleph_0, \\
\mathfrak{c} \cdot \mathfrak{c} &= \mathfrak{c}, \\
\aleph_0 \cdot \mathfrak{c} &= \mathfrak{c}, \quad \text{by Theorem 5.3.1,} \\
\aleph_0 \cdot \aleph &= \aleph \quad \text{for any infinite cardinal } \aleph, \\
\aleph \cdot \aleph \cdot \aleph &= \aleph \quad \text{for any infinite cardinal } \aleph.
\end{aligned}$$

The results in this section show us that the arithmetic of infinite cardinals will require some thought before we can apply it or manipulate with it. Read the above arithmetic facts while reminding yourself that these equations do not hold for real numbers other than 0 and 1.

Inasmuch as we have defined

$$\aleph^2 = \aleph \cdot \aleph,$$

it is important that we define powers of cardinals. Recall that given sets A and B

$$\boxed{B^A \text{ is the set of all functions } f : A \longrightarrow B.}$$

Let us write down the rules of all of the functions in $\{1,2\}^{\{1,2,3\}}$. A function $f : \{1,2,3\} \longrightarrow \{1,2\}$ is given by its images of 1, 2, and 3. Then we can write down the rule for f by just listing how it behaves on 1, 2, and 3. A very effective method is to list

$$f = [a, b, c],$$

where $a, b, c \in \{1,2\}$ and where

$$f(1) = a, f(2) = b, f(3) = c.$$

For instance, suppose that we have a function

$$f : \{1,2,3\} \to \{1,2\}$$

whose rule is given by

$$f(1) = 1, f(2) = 1, f(3) = 1.$$

Then the very effective method for writing down f is

$$f = [1, 1, 1].$$

The function

$$g : \{1, 2, 3\} \to \{1, 2\}$$

whose rule is given by

$$g(1) = 2, g(2) = 1, g(3) = 2$$

is written down as

$$g = [2, 1, 2].$$

Furthermore, the given triple

$$[1, 2, 2],$$

defines a function

$$h : \{1, 2, 3\} \to \{1, 2\}$$

whose rule is given by

$$h(1) = 1, h(2) = 2, h(3) = 2.$$

Let us count the number of functions in $\{1, 2\}^{\{1,2,3\}}$. Since $[a, b, c]$ is a random function, and since there are 2 choices for a, 2 choices for b, and 2 choices for c, we see that there are 2^3 functions f in $\{1, 2\}^{\{1,2,3\}}$. Then

$$\text{card}(\{1, 2\}^{\{1,2,3\}}) \quad = \quad 2^3 \quad = \quad \text{card}(\{1, 2\})^{\text{card}(\{1,2,3\})}.$$

In general, if $E = \{1, \ldots, n\}$ and $F = \{1, \ldots, m\}$ are finite sets, then each function in F^E can be written as a finite sequence

$$[y_1, y_2, \ldots, y_n],$$

where $y_1, \ldots, y_n \in F$, and where

$$f(1) = y_1, \ldots, f(n) = y_n.$$

For each k, there are exactly m choices y_k for the image $f(k)$ under f. Since there are m choices made for each of the n values $y_1, \ldots, y_n$, there are exactly m^n functions in F^E. In other words, if E and F are finite sets, then

$$\boxed{\operatorname{card}(E^F) = \operatorname{card}(E)^{\operatorname{card}(F)}.}$$

This motivates the following natural definition of exponential values of infinite cardinals.

Definition 5.5.17 *Let α and β be cardinals. Choose any sets A and B such that $\alpha = \operatorname{card}(A)$ and $\beta = \operatorname{card}(B)$. We define β^α to be*

$$\boxed{\beta^\alpha \quad = \quad \operatorname{card}(B^A).}$$

In particular,

$$\boxed{2^{\aleph_0} = \operatorname{card}(\{0,1\})^{\operatorname{card}(\mathbb{N})} = \operatorname{card}\left(\{0,1\}^{\mathbb{N}}\right).}$$

Theorems 5.4.3 and 5.4.6 show us that

$$\operatorname{card}(\mathbb{N}) \;<\; \operatorname{card}(\mathcal{P}(\mathbb{N})) \;=\; \operatorname{card}(\{0,1\}^{\mathbb{N}})$$

for the *infinite set* $\mathbb{N}$, so that

$$\boxed{\aleph_0 < 2^{\aleph_0}}$$

by Theorem 5.3.1. Consider this inequality in a more general context. Let $\aleph$ be an infinite cardinal and let A be a set such that $\aleph = \operatorname{card}(A)$. Then

$$\boxed{2^{\aleph} = \operatorname{card}\left(\{0,1\}^A\right),}$$

and hence

$$\aleph < 2^{\aleph}$$

by Theorem 5.4.6. Moreover, since $2^{\aleph}$ is also a cardinal, we can form the cardinal

$$2^{2^{\aleph}},$$

which is properly larger than $2^{\aleph}$. We have constructed

$$2^{\aleph} < 2^{2^{\aleph}} < 2^{2^{2^{\aleph}}}.$$

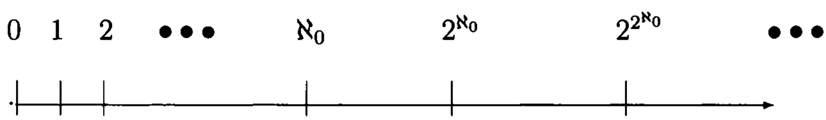
Continuing this inductive process indefinitely will produce *an infinite chain of infinite cardinals*

$$\boxed{\aleph < 2^{\aleph} < 2^{2^{\aleph}} < 2^{2^{2^{\aleph}}} < \cdots.}$$

Beginning this process with $\aleph_0 = \aleph$ produces the infinite chain

$$\boxed{\aleph_0 < 2^{\aleph_0} < 2^{2^{\aleph_0}} < 2^{2^{2^{\aleph_0}}} < \cdots.}$$

The linear diagram below illustrates this inequality. *This is not the real line.* It is just a visual means of describing a chain of cardinals.



We leave this section with a question and a list of small infinite cardinals that summarizes our research to date.

$$\boxed{\begin{gathered}\aleph_0 = \mathrm{card}(\mathbb{N}) = \mathrm{card}(\mathbb{Q}) = \mathrm{card}(\mathbb{N}\times\mathbb{N}) \\ \mathrm{card}(0,1) = \mathrm{card}(\mathbb{R}) = \mathrm{card}(\mathbb{R}^3) = \mathrm{card}(\mathcal{P}(\mathbb{N})) = 2^{\aleph_0} \\ \aleph_0 < 2^{\aleph_0} < 2^{2^{\aleph_0}} < 2^{2^{2^{\aleph_0}}} < \ldots\end{gathered}}$$

Question: Is there a cardinal between card($\mathbb{N}$) and card($\mathbb{R}$)? Specifically, is there a cardinal $\aleph_1$ such that

$$\text{card}(\mathbb{N}) < \aleph_1 < \text{card}(\mathbb{R})?$$

Exercise 5.5.18 1. Let E, F, and G be sets containing exactly m, n, and p elements. Prove that $m + n + p = \text{card}(E \cup F \cup G)$.

2. Let E, F, and G be sets containing exactly m, n, and p elements. Prove that $(m \cdot n) \cdot p = \text{card}((E \times F) \times G)$.

3. Let E be a set containing exactly m elements. Find $(m^m)^2$ as a cardinal number.

4. Let E, F, and G be sets containing exactly m, n, and p elements. Prove that $p(m + n) = \text{card}(G \times (E \cup F))$.

5. Fill in the details of the discussion on pages 177 to 200.

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

Chapter 6

Well-Ordered Sets

The set of real numbers $\mathbb{R}$ is *linearly ordered.* That is, given $x, y \in \mathbb{R}$, then $x \leq y$ or $y \leq x$. Such an ordering of a set X allows us to think of X as some part of a line. For instance, $\mathbb{Q}$ is linearly ordered since it is a subset of $\mathbb{R}$. It is the rational part of the real line. The integers, $\mathbb{Z}$, form a linearly ordered set, but it has the very interesting property that to each $x \in \mathbb{Z}$ there is a *next integer* or a *successor* to x, usually denoted by x^+. One easily sees that $x + 1$ is the successor to x in $\mathbb{Z}$. The existence of a successor for each element in $\mathbb{Z}$ is called the *Well-Ordering Principle.* We say that $\mathbb{Z}$ is *well-ordered.* Because $\mathbb{N} \subset \mathbb{Z}$, $\mathbb{N}$ is well-ordered, but $\mathbb{Q}$ is not well-ordered. In this chapter, we introduce the well-ordered *chain of ordinals*, and we examine the well-ordered set of cardinals.

6.1 Successors of Elements

The set of natural numbers $\mathbb{N}$ has an interesting property.

The Well Ordered Property: Given any element $n \in \mathbb{N}$, there is a *unique* next element or *successor* element n^+.

Given an integer $n \in \mathbb{N}$, there is exactly one successor element n^+. You know n^+ as $n+1$, since for instance $0^+ = 1$, $12^+ = 13$, and $121^+ = 122$. There is no element $n \in \mathbb{N}$ such that $n^+ = 0$. Other common sets have this property, and we tend to think that *every* collection has the Well-Ordered Property.

For example, suppose that $\mathbf{P} = \{P_1, P_2, P_3\}$ is a set of three people. Our culture automatically endows $\mathbf{P}$ with several well orderings. We give $\mathbf{P}$ a well ordering by saying that the smallest element will be the shortest person, say, P_1, the next element will be the next tallest person, say, P_2, and then the tallest person, P_3. We would write

$$P_1 < P_2 < P_3.$$

Good. This is a well ordering of $\mathbf{P}$ by saying that the symbol $<$ points to the shorter person. Furthermore,

$$P_1^+ = P_2, \quad P_2^+ = P_3, \quad P_3^+ = P_3.$$

There is no element x in $\mathbf{P}$ such that $x^+ = P_1$, and P_3 has no successor.

Height is not the only way to well order $\mathbf{P}$. Suppose that P_1 is older than P_3 and that P_3 is older than P_2. Then we might well order $\mathbf{P}$ by age by writing

$$P_2 < P_3 < P_1.$$

In this case, $<$ points to the younger person. Then

$$P_2^+ = P_3, \quad P_3^+ = P_1, \quad P_1^+ = P_1$$

under the age well ordering. There is no element x in $\mathbf{P}$ such that $x^+ = P_2$, and P_1 has no successor.

You might also well order $\mathbf{P}$ alphabetically on the last names of the people in $\mathbf{P}$. Then, if P_1 is named Himmel, if P_2 is named Frugal, and if P_3 is named Briar, then under the alphabet ordering

$$P_3 < P_2 < P_1.$$

Then $P_3^+ = P_2$, $P_2^+ = P_1$, and P_1 has no successor.

We have linearly ordered $\mathbf{P}$ in different ways by considering different numbers associated with people. We are constantly well ordering people and their lives when we say "I love you more" or "You're so much better than him." These sentences imply that, given any two people we can assign an inequality to them.

$$P_1 < P_2 \text{ if } P_1 \text{ is "better than" } P_2.$$

In this example $<$ points to the better person. Most of us will well order people on a continual basis even when an ordering is not called for. After all, can a parent declare which of their children they most love? I don't think so.

The *power set* $\mathcal{P}(A)$ of a set A is a naturally occurring example of a set that is not linearly ordered in the natural way. We define an order $<$ on $\mathcal{P}(A)$ that is not a linear order as follows.

$$\boxed{\text{Given} \quad X, Y \in \mathcal{P}(A) \quad \text{then} \quad X < Y \quad \text{if} \quad X \subset Y.}$$

Take, for example, $\mathcal{P}(\{1,2,3\})$. The elements $\{2\}$, $\{1,2\}$, and $\{2,3\}$ $\in \mathcal{P}(\{1,2,3\})$ satisfy

$$\{2\} \subset \{1,2\} \quad \text{and} \quad \{2\} \subset \{2,3\}.$$

We might consider the two elements to be successors of $\{2\}$. Since $\{2\}$ does not have a *unique* successor element, $\mathcal{P}(\{1,2,3\})$ is not a well-ordered set. A similar argument shows that $\emptyset^+$ does not exist because it cannot be defined *uniquely* under the subset ordering $<$.

$$\emptyset \subset \{1\}, \quad \emptyset \subset \{2\}, \quad \emptyset \subset \{3\}.$$

However, we can define an *abstract* well ordering $\leq$ on $\mathcal{P}(\{1,2,3\})$ by picking a linear ordering at random on the 8 elements of $\mathcal{P}(\{1,2,3\})$ as follows.

$$\{1\} \leq \emptyset \leq \{1,2,3\} \leq \{2\} \leq \{2,3\} \leq \{1,3\} \leq \{1,2\} \leq \{3\}.$$

Notice that this linear ordering does not have much to do with the subset inclusion $\subset$. It is simply a linear ordering that your author chose to make a point. Abstract well orderings are still well orderings, and they will be treated as legitimate well orderings on sets.

The set $\mathbb{R}$ of *real numbers* is an example of a set that possesses a linear order $<$ but that is not well-ordered. For example, $\sqrt{2}$ is a real number, but there is no *next real number*, $\sqrt{2}^+$. That is, there is no real number that is next on the real number line after $\sqrt{2}^+$. $\sqrt{2}+1$ will not do since

$$\sqrt{2} < \sqrt{2} + \frac{1}{2} < \sqrt{2} + 1.$$

There should be nothing between a number and its successor.

It is also apparent that $\mathbb{Q}$ is not well-ordered. 0^+ does not exist *as a rational number.* We cannot use 1 as 0^+ in $\mathbb{Q}$ since 1 is not the next *rational* number larger than 0. This is clear once we write

$$0 < \frac{1}{2} < 1.$$

So well-ordered sets are going to be thin sets. Given an element x, there is a next value or successor value x^+.

There is another interesting example of a set that is not well-ordered but that is still quite thin. Let

$$W = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right\},$$

and let $<$ be the usual linear ordering of real numbers. Then in the set W, $\dfrac{1}{2}^+ = 1$, $\dfrac{1}{4}^+ = \dfrac{1}{3}$, but 0^+ is not to be found in W. Let's argue about that. Suppose 0^+ exists *in* W. Then

$$0^+ = \frac{1}{n} \text{ for some } n \in \mathbb{N}.$$

But then $\dfrac{1}{n}$ should be the next value in the set that is larger than 0. This is not the case since

$$0 < \frac{1}{n+1} < \frac{1}{n}.$$

Hence, in W, 0^+ does not exist, and therefore W is not well-ordered.

I leave it to the reader as a challenge problem to justify to yourself that the set

$$U = \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right\}$$

is well-ordered. Just try to find x^+ for each $x \in U$.

The well-ordered set of natural numbers $\mathbb{N}$ satisfies some interesting properties that we will use later in a more general setting. $\mathbb{N}$ satisfies the following property.

The Trichotomy Property: Given $n, m \in \mathbb{N}$, then *exactly one* of the following relationships holds.

$$n < m \quad \text{or} \quad m < n \quad \text{or} \quad n = m.$$

For example, the solution to $2 = x + 1$ satisfies exactly one of the comparisons $x < 0$, $0 < x$, or $x = 0$.

Another example is the age n_o of the youngest Mathematics professor in Canada. This number n_o satisfies exactly one of the following options. Either

$n_o < 22$ and we say that the Mathematics professor is younger than 22,
or $n_o > 22$ and we say that the Mathematics professor is older than 22,
or $n_o = 22$ and we say that the Mathematics professor is 22 years old.

Even though we do not know the age n_o, we can still state that it satisfies the Trichotomy Property.

A property closely tied to the Trichotomy Property is the *Minimum Property* in a well-ordered set.

The Minimum Property: If A is a well-ordered set, then each nonempty subset of A has a unique minimum element.

Each nonempty set in $\mathbb{N}$ contains a unique least element. That is, given $\emptyset \neq X \subset \mathbb{N}$ there is an element $x_0 \in X$ such that $x_0 \leq x$ for each $x \in X$. For example, $\{7, 8, 9, 10, 11\} \subset \mathbb{N}$ has least element 7, while

$$\{n \in \mathbb{N} \mid n \text{ is the age in years of a Canadian Mathematician}\}$$

has a unique least age n_o that we cannot specifically identify. We would all agree, though, that this age n_o is somewhat more than 11 and somewhat less than 121.

A hard to verify minimum natural number is the minimum naturally occurring temperature in degrees Celsius on the surface of the Earth. For our purposes here, this temperature will not have a decimal value but only a whole number value. While we can safely say that this minimum temperature is smaller than 200°C and more than -200°C, we cannot say what that minimum temperature is without first measuring temperatures everywhere on the Earth's surface. However, the Minimum Property tells us that such a minimum temperature exists, whether or not we have made the measurements.

If you are given a set of people measured only as closely as inches (i.e., their heights are in inches but not in fractions of inches), then these heights have a unique minimum height. It might be that the heights are smaller than 120 inches and that they are greater than 1 inch, but there is a unique minimum height h_o and there is a person whose height is h_o.

The Well-Ordered Property and the Minimum Property combine as follows. Given $n \in \mathbb{N}$ then

n^+ is the smallest of all the elements that are larger than n.

In other words,

$$\text{Given } n < m \in \mathbb{N} \text{ then } n^+ \leq m.$$

For example, $1 < 5$ so that $1^+ = 2 \leq 5$; $12 < 14$ so that $12^+ \leq 14$; and $121 < 122$ so that $121^+ \leq 122$.

Another example is as follows. We will argue in this chapter that *the set of cardinals is a well-ordered set.* A complete justification is well beyond the scope of this book, but the framed inequality on page 198 shows us that

$$\aleph_0 < 2^{\aleph_0}.$$

So where is the successor $\aleph_0^+$ to $\aleph_0$? The best we can do at this time is to say that

$$\aleph_0 < \aleph_0^+ \leq 2^{\aleph_0}.$$

Furthermore, $\aleph_0^+ \neq \aleph_0 + 1$ since, as must be apparent, $\aleph_0^+ \neq \aleph_0 = \aleph_0 + 1$.

With these properties in mind, let us abstract the notion of a well-ordered set.

Definition 6.1.1 *Let A be any nonempty set. We say that A is a* well-ordered set *if it satisfies the following two properties.*

1. *A satisfies* the Trichotomy Property. *That is, given $x, y \in A$ then x and y satisfy exactly one of the following options.*

$$x < y,\ y < x,\ \text{or } x = y.$$

2. *A satisfies* the Minimum Property. *That is, each nonempty subset of A contains a unique least element. Equivalently, to each element $x \in A$, there is a unique element $x^+ \in A$ such that*

$$\textit{Given } y \in A \textit{ such that } x < y \textit{ then } x^+ \leq y.$$

We call x^+ the successor of x.

Notice that we cannot claim that the sum $x + 1$ is the successor x^+ of x since in general there is no addition on the elements of A. Thus, $x + 1$ is without Mathematical meaning for well-ordered sets.

Under this definition, $\mathbb{N}$ is a well-ordered set. We have already discussed that $\mathbb{N}$ satisfies the Trichotomy Property, and that unlike general well-ordered sets, each element $n \in \mathbb{N}$ possesses the successor element $n^+ = n + 1$. The fact that $n^+ = n + 1$ for each $n \in \mathbb{N}$ follows from the structure of $\mathbb{N}$. Addition in $\mathbb{N}$ is defined, and there are no natural numbers between n and $n+1$. This makes $n + 1 = n^+$.

We can also show that $\mathbb{N}$ satisfies the Minimum Property. Each subset of $\mathbb{N}$ has a least element found as follows. Let $X \neq \emptyset$ be a subset of $\mathbb{N}$. If $0 \in X$ then we are done. 0 is the least element of $\mathbb{N}$, so 0 is the least element in X. Otherwise, test 1. If $1 \in X$, then it is the least element of X. If $1 \notin X$, then test 2. If $2 \in X$, then 2 is the least element in X. If $2 \notin X$, then continue on in this inductive manner, taking 3, and on until you reach a first $n \in X$. Since $X \neq \emptyset$, you are guaranteed that under the process described here, you will find that first element n in X. This first number n is the least element of X. Anything smaller has been ruled out by our examination of $0, 1, 2, \ldots$. Thus, $\mathbb{N}$ satisfies the Minimum Property.

Given a set $A = \{x_0, x_1, \ldots, x_n\}$, we can write down a well ordering on the elements of A by requiring that

$$x_0 < x_1 < \ldots < x_n.$$

Notice that every element $x \neq x_n \in A$ has a successor $x^+ \in A$. For example,

$$x_0^+ = x_1, \quad x_1^+ = x_2.$$

The successor x_n^+ to x_n does not exist since

$$\{x \in A \,|\, x_n < x\}$$

is empty. Then this linear ordering on A is not *well* ordering on A. Moreover, it is improper to say that $x_n^+ = x_{n+1}$ in A since x_{n+1} *is not an element of* A.

We can make $\{x_1, x_2, x_3, \ldots\}$ a well-ordered set if we define

$$x_{11} < x_1 < x_{22} < x_2 < x_{33} < x_3 < x_{44} < x_4 < \cdots.$$

Notice the peculiar order of the subscripts when compared with the linear ordering of the x's. The order of the subscripts does not follow the linear ordering of the elements. It is the linear order of the elements that matters, not how we label them or name them. Observe that in this case, $x_1^+ = x_{22}$ and $x_{55} = x^5$.

Linearly order the pages of this book by the rule

$$x \leq y \text{ if page } y \text{ of the book follows page } x.$$

If x is the last page in the book, then x^+ does not exist since no page follows the last page of the book. You should try to convince yourself that this defines a linear ordering on the pages, but not a well-ordering.

Convince yourself that the subset ordering $\subset$ makes the set

$$A = \{\emptyset, \{a\}, \{a, b\}, \{a, b, c\}\}$$

a linearly ordered set, but not a well-ordered set.

When you try to linearly order the set

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

using the subset order $\subset$, you are met with defeat because the elements of A are *incomparable.* That is, $\{1,2\} \not\subset \{2,3\}$ and $\{2,3\} \not\subset \{1,2\}$. (Read $\not\subset$ as *not a subset of.*) Because the set A does not satisfy the Trichotomy Property, it is not a well-ordered set. Write this one down for yourself.

If we use the linear ordering $<$ of the real numbers, then

$$W = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$$

is *not* a well-ordered set since the set

$$U = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$$

does not contain a minimum element. The following diagram will help show us why. The line below is the line $\mathbb{R}^+$ of positive real numbers. It is not to scale.

This picture shows us that in W we have $\frac{1}{2}^+ = 1, \frac{1}{3}^+ = \frac{1}{2}$, and in general,

$$\frac{1}{n+1}^+ = \frac{1}{n}.$$

If we try to choose an element $\frac{1}{n}$ in U, then we have a smaller one $\frac{1}{n+1}$. Then U cannot contain a minimum element, since a minimum element is an element $x \in U$ that is smaller than any other element of U. The point is, a minimum element of U is an element of U. Thus W does not have the Minimum Property.

Moreover, 0 is without a successor 0^+ in W. To see this, suppose that 0^+ exists in W. Then 0^+ must satisfy

$$0^+ = \frac{1}{n}$$

for some $n \in \mathbb{N}$. In that case, $0 < \dfrac{1}{n+1}$ so that $0^+ \leq \dfrac{1}{n+1}$ by the definition of the successor element. But then

$$\frac{1}{n+1} < \frac{1}{n} = 0^+ \leq \frac{1}{n+1},$$

a clear contradiction. Then 0^+ does not exist in W.

Exercise 6.1.2 1. Show that $U = \left\{ \dfrac{1}{n} \mid n \in \mathbb{N}, n \neq 0 \right\}$ is a well-ordered set by defining a linear ordering on U and then showing that $\dfrac{1}{n}^+$ has a successor in U.

2. Show that under the subset ordering $\subset$, $A = \{\{1,2\},\{2,3\}\}$ does not satisfy the Trichotomy Property.

3. If a successor ω_0^+, then guess at what ω_0^+ should be.

6.2 Constructing Well-Ordered Sets

We can construct new well-ordered sets from old ones in the following way. For reasons that will become clear later, let us assume that

$$\boxed{\omega_0}$$

is a *symbol* not in $\mathbb{N}$. Any element not in $\mathbb{N}$ will do. It is traditional to choose ω_0. We reserve the symbol ∞ for another purpose.

We define a *linear order* on the set

$$\mathbb{N} \cup \{\omega_0\},$$

that will eventually become a well-ordering on a larger set, by requiring that

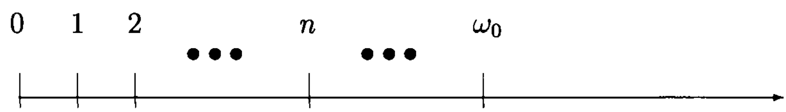
$$n < \omega_0 \text{ for each } n \in \mathbb{N}.$$

We will define ω_0^+ a little later in the discussion.

We list the elements in this linearly ordered set $\mathbb{N} \cup \{\omega_0\}$ as

$$\boxed{\mathbb{N} \cup \{\omega_0\} \quad = \quad 0 < 1 < 2 < 3 < \ldots < \omega_0.}$$

A picture of $\mathbb{N} \cup \{\omega_0\}$ will help.



This is not a picture of the real line. The placement of ω_0 indicates that ω_0 is larger than each $n \in \mathbb{N}$. For example,

$$10^{10^{10^{10^{10}}}} < \omega_0.$$

Even the very large natural numbers are smaller than ω_0. For this reason, we refer to ω_0 as a *limit ordinal.*

Reader Be Warned: We are dealing with an abstract Mathematical construction. We treat $\mathbb{N}$ as a featureless set. Recall how we constructed the natural numbers. 2 is the collection of all sets that are equivalent to $\{x, y\}$. 2 is not a distance under this construction. $1 < 2$ does not mean that 1 and 2 are a distance 1 unit apart. We have not introduced a means of defining distance on $\mathbb{N}$. You should not see ω_0 as being infinitely far away. ω_0 is not a number. It is simply another *element* in a larger set.

We have added ω_0 in such a way that ω_0 is the *unique largest element in* $\mathbb{N} \cup \{\omega_0\}$. Furthermore,

$$\boxed{\text{there is no element } x \in \mathbb{N} \cup \{\omega_0\} \text{ such that } x^+ = \omega_0.}$$

This is not hard to show. Let $n \in \mathbb{N}$ be any element. Then by elementary arithmetic, $n^+ = n + 1 \in \mathbb{N}$. Then $n^+ \neq \omega_0$.

For this reason, we say that

$$\boxed{\omega_0 \text{ is the first infinite ordinal}}$$

or *the first countably infinite ordinal.* It is natural to think of ω_0 as a countable infinity since it is larger than each $n \in \mathbb{N}$. We also refer to ω_0 as a *limit ordinal.*

Some of you may be having trouble thinking of this element ω_0. You are so accustomed to seeing $\mathbb{N}$ as a set whose elements are a distance 1 apart that the idea of $\mathbb{N}$ as a set without distance makes you uncomfortable. You think of $<$ as a measure of the distance between elements. This is not the case.

Think of $\mathbb{N}$ this way. Envision a Mathematical bag that contains each of the elements $n \in \mathbb{N}$. There is no distance and there is no unit of measurement. There is just a bag that contains all of the natural numbers n. You can do this, you know. It has been done on a smaller scale. Pick up a large dictionary and look at the back of the book. There is a page of measurements in various units of distance. Then we have on one page, included numbers like 12 inches to a 1 foot. Here we squeezed 12 numbers into one place. We also have 1 mile is 5280 feet. We put several thousands of natural numbers into one page. You might also find 186,000 miles per second if you looked up the speed of light. These are large numbers that are not far away. They are within our grasp. Think of our bag of natural numbers in the same way. They exist as numbers in some kind of set. Nothing more.

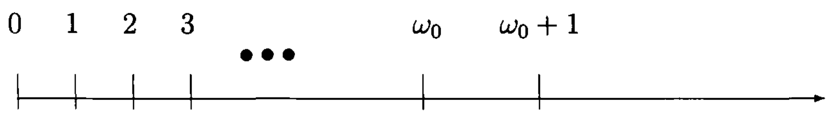
We now have a linearly ordered set $\mathbb{N} \cup \{\omega_0\}$ that contains a symbol ω_0 that can be identified with an infinite value. All we had to do was to introduce a new symbol ω_0 that *we defined* to be larger than each of the elements in $\mathbb{N}$. No distance was implied since infinite distance is undefined. The symbol ω_0 is just that, a symbol. Let us continue using this idea of adding a symbol and eventually extend the linearly ordered set $\mathbb{N} \cup \{\omega_0\}$ to a larger *well-ordered* set.

We will choose a symbol to serve as ω_0^+ in a linearly ordered set.

We choose for convenience

$$\boxed{\omega_0 + 1}$$

as the successor ω_0^+ of ω_0. We place $\omega_0 + 1$ as the maximal element in a new linearly ordered set as in the following line diagram.



This diagram helps us to define a linear order on $\mathbb{N} \cup \{\omega_0, \omega_0 + 1\}$. The linear order $<$ is defined on $\mathbb{N} \cup \{\omega_0\}$ as we defined on page 211, and we extend this linear order to a linear order

$$x < \omega_0 + 1 \quad \forall x \in \mathbb{N} \cup \{\omega_0\}$$

for each $x \in \mathbb{N} \cup \{\omega_0, \omega_0 + 1\}$. By definition, we have

$$\omega_0^+ = \omega_0 + 1.$$

The symbol does not really mean the sum of ω_0 and 1. We have simply chosen this symbol that we will recognize as the successor ω_0^+ of ω_0. However, we will abuse the notation and read $\omega_0 + 1$ as *omega nought plus one.* In this way, we have begun an induction process with which we will construct a new set of ordinals.

We have already introduced two different infinite ordinals, $\omega_0, \omega_0 + 1$, and there is the promise of many more. Notice that this construction is in conflict with the adage that we were taught as children: *infinity plus one is infinity.* While this adage applies to cardinals, it does not anticipate the existence of ordinals. We might try bending the adage somewhat and say that *infinity plus one is infinite.*

Extend the above construction of $\omega_0 + 1$ by choosing different symbols

$$\omega_0 + 2, \omega_0 + 3, \ldots$$

that we will call *ordinals*. The construction yields a tower of linearly ordered sets as follows.

Define a linear order $<$ on $\mathbb{N} \cup \{\omega_0, \omega_0+1, \omega_0+2\}$ by defining $x < y$ as above on $x, y \in \mathbb{N} \cup \{\omega_0, \omega_0+1\}$, and

$$x < \omega_0 + 2 \quad \forall x \in \mathbb{N} \cup \{\omega_0, \omega_0+1\}.$$

We define $(\omega_0+1)^+ = \omega_0+2$. Then every element $x \in \mathbb{N} \cup \{\omega_0, \omega_0+1, \omega_0+2\}$ has a successor x^+ in $\mathbb{N} \cup \{\omega_0, \omega_0+1, \omega_0+2\}$, except for ω_0+2. Then $<$ is linear order on $\mathbb{N} \cup \{\omega_0, \omega_0+1, \omega_0+2\}$.

In general, let $n \in \mathbb{N}$ and construct a larger linearly ordered set $\mathbb{N} \cup \{\omega_0, \dots, \omega_0+n+1\}$ as follows. Assume that we have defined a linear order set $\mathbb{N} \cup \{\omega_0, \dots, \omega_0+n\}$ as we did above. (This is not such an extraordinary assumption because we have defined this set for $n = 2$.)

The linear order is defined by extending the linear order $<$ on $\mathbb{N} \cup \{\omega_0, \dots, \omega_0+n\}$ by defining

$$(\omega_0+n)^+ = \omega_0 + (n+1).$$

Then the linear order $<$ on $\mathbb{N} \cup \{\omega_0, \dots, \omega_0+(n+1)\}$ is defined by $x < y$ for all $x, y \in \mathbb{N} \cup \{\omega_0, \dots, \omega_0+n\}$, and

$$x < \omega_0 + (n+1) \quad \forall x \in \mathbb{N} \cup \{\omega_0, \dots, \omega_0+n\}.$$

Then every element $x \in \mathbb{N} \cup \{\omega_0, \dots, \omega_0+(n+1)\}$ but $\omega_0+(n+1)$ has a successor x^+ in $\mathbb{N} \cup \{\omega_0, \dots, \omega_0+(n+1)\}$. Thus $<$ is a linear order on $\mathbb{N} \cup \{\omega_0, \dots, \omega_0+(n+1)\}$.

To continue our construction to infinitely many ordinals, let *two omega nought* be the set

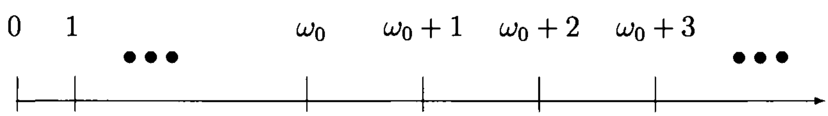
$$\begin{aligned} 2\omega_0 &= \bigcup_{n \in \mathbb{N}} \mathbb{N} \cup \{\omega_0, \dots, \omega_0+n\} \\ &= \mathbb{N} \cup \{\omega_0, \omega_0+1, \omega_0+2, \omega_0+3, \dots\}. \end{aligned}$$

We make $2\omega_0$ a *well-ordered* set by extending the linear order $<$ to $\mathbb{N} \cup \{\omega_0, \dots, \omega_0+1, \cdots\}$ as follows. Given $x, y \in 2\omega_0$, then by the definition of union, there is an $n \in \mathbb{N}$ such that $x, y \in \mathbb{N} \cup \{\omega_0, \dots, \omega_0+n\}$. In this set, $x < y$ is defined.

Quite naturally, the linear orders $<$ are then extended to $2\omega_0$ by defining

$$0 < 1 < 2 \ldots < \omega_0 < \omega_0 + 1 < \omega_0 + 2 < \omega_0 + 3 < \ldots.$$

A picture will help the reader visualize the ordering.



We have claimed that $2\omega_0$ is a well-ordered set. Let us prove it.

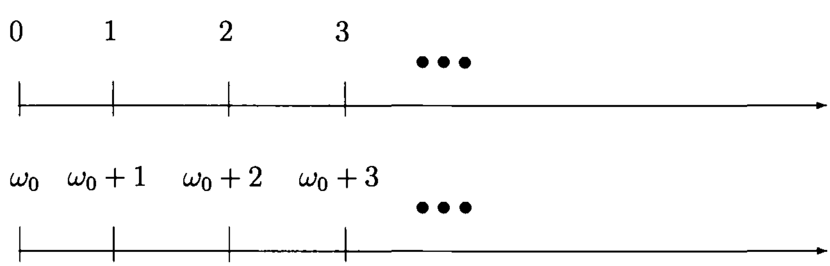
Theorem 6.2.1 *$2\omega_0$ is a well-ordered set.*

Proof: We have defined a linear order on $2\omega_0$, so we must show that each element in $2\omega_0$ possesses a successor in $2\omega_0$. Let $x \in 2\omega_0$. Then either $x \in \mathbb{N}$ or $x = \omega_0 + n$ for some $n \in \mathbb{N}$. If $n \in \mathbb{N}$ then $n^+ = n + 1$, and if $x = \omega_0 + n$ then $x^+ = (\omega_0 + n)^+ = \omega_0 + (n+1)$. Hence each $x \in 2\omega_0$ possesses a successor in $2\omega_0$, so that $2\omega_0$ is a well-ordered set. This completes the proof.

In effect, we have constructed a sequence of *infinite ordinals.* If we ignore the addition and multiplication of $\mathbb{N}$, then the set

$$\{\omega_0, \omega_0 + 1, \omega_0 + 2, \omega_0 + 3, \ldots\}$$

appears to be no different from $\mathbb{N}$. Another picture will help.



For all intents and purposes, the well-ordered set $2\omega_0$ is formed by attaching a copy of $\mathbb{N}$ to the right of $\mathbb{N}$. Thus it is natural to call $2\omega_0$

a *limit ordinal.* The way in which we associate $\mathbb{N}$ and ω_0 suggests the notation

$$\boxed{2\omega_0 = \omega_0 + \omega_0.}$$

This is indeed the case.

We notice the important fact that $2\omega_0$ is not an element of itself. Then as we search for a symbol with which we can extend our well-ordered set $2\omega_0$ to a linearly ordered set, we can use $2\omega_0$.

There is an interesting omission here. Where do we find the symbol $1 + \omega_0$ in this picture? The answer is hard to explain but easy to read.

$$\boxed{1 + \omega_0 = \omega_0 \neq \omega_0 + 1.}$$

This contradicts all of our intuition on addition of numbers. Why isn't $\omega_0 + 1 = 1 + \omega_0$? Common sense tells us that addition is commutative. The reality of it shows us that

> Our common sense about numbers cannot be applied meaningfully to infinite ordinals.

An explanation that might satisfy our curiosity and our common sense lies deep inside the study of well-ordered sets. We will not go that far. Suffice it to say that this startling unexpected equation is a property of ordinals and not of numbers.

Getting back to our construction, we have constructed a well-ordered set

$$2\omega_0 = \mathbb{N} \cup \{\omega_0, \omega_0 + 1, \omega_0 + 2, \ldots\}.$$

Note that $2\omega_0$ does not possess a largest ordinal. *The well-ordered set* $2\omega_0$ *is different from the previous well-ordered set* $\mathbb{N}$ in that $2\omega_0$ has a limit ordinal ω_0, while $\mathbb{N}$ has no limit ordinal. Let us extend $2\omega_0$ to a larger well-ordered set. In the interest of not losing your attention, some of the rigor used to this point will be abandoned.

There is no element $x \in 2\omega_0$ such that $x^+ = \omega_0$. Furthermore, to each element $x \in 2\omega_0$ there is an element $x^+ = x + 1$. Given the way we used this notation, it is natural and Mathematically proper to write

$$\boxed{(2\omega_0)^+ = 2\omega_0 + 1}$$

but given our constructions to date, we need a linearly ordered set X such that $X = 2\omega_0 + 1$. That linearly ordered set is

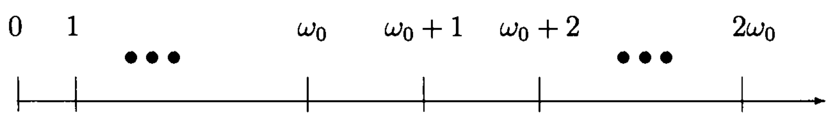
$$\begin{aligned} 2\omega_0 + 1 &= 2\omega_0 \cup \{2\omega_0\} \\ &= \mathbb{N} \cup \{\omega_0, \omega_0 + 1, \omega_0 + 2, \ldots\} \cup \{2\omega_0\} \end{aligned}$$

where we naturally require that

$$\omega_0 + n < 2\omega_0 \text{ for all } n \in \mathbb{N}.$$

We note that under this definition, $2\omega_0 \neq x^+$ for any $x \in 2\omega_0$.

In our new linearly ordered set, the largest element is $2\omega_0$, as in the accompanying line diagram.



Let us stretch our intuition a little. We have not defined a subtraction between ordinals because

$$\boxed{x = y + 1 \text{ does not have a solution in the well-ordered set } 2\omega_0.}$$

For example, recall that $\omega_0 \neq x^+ = x + 1$ for any $x \in 2\omega_0$. Then $\omega_0 - 1$ does not exist, or more plainly, $\omega_0 - 1 \neq x$ for any $x \in 2\omega_0$. Thus, if $y = \omega_0$, then $x + 1 = y$ does not have a solution in $2\omega_0$. Thus, the subtraction of ordinals cannot be defined.

Furthermore, we have defined the multiplication of at least two ordinals when we formed $2\omega_0$. We will show that

$$\boxed{\text{The fraction } \frac{1}{2}\omega_0 \text{ of ordinals is not an ordinal.}}$$

Proof: We assume for the sake of contradiction that

$$\frac{1}{2}\omega_0 = x$$

is a fraction of ordinals. The usual rules of arithmetic state that

$$\omega_0 = 2x > x.$$

By our construction of ordinals o date, the ordinals less than ω_0 are in $\mathbb{N}$, so that $x \in \mathbb{N}$. This implies that

$$\omega_0 = 2x \in \mathbb{N},$$

while we chose $\omega_0 \notin \mathbb{N}$. This contradiction shows us that $\frac{1}{2}\omega_0$ cannot be defined as an ordinal. Thus, the division of ordinals is not defined. This completes the proof.

Returning to our construction, to construct $3\omega_0$, we let

$$2\omega_0 \quad = \quad \mathbb{N} \cup \{\omega_0, \omega_0 + 1, \omega_0 + 2, \ldots\},$$

and then place a copy of $\mathbb{N}$ that follows $2\omega_0$. We have thus constructed the new well-ordered set

$$\begin{aligned} 3\omega_0 &= 2\omega_0 \cup \{2\omega_0, 2\omega_0 + 1, 2\omega_0 + 2, \ldots\} \qquad (6.1) \\ &= \mathbb{N} \cup \{\omega_0, \omega_0 + 1, \omega_0 + 2, \ldots\} \cup \{2\omega_0, 2\omega_0 + 1, 2\omega_0 + 2, \ldots\}. \end{aligned}$$

The line drawing

gives us a crude picture of $3\omega_0$.

Since we have adjoined three copies of $\mathbb{N}$, it is symbolically natural and Mathematically correct to write

$$\boxed{3\omega_0 = \omega_0 + 2\omega_0 = \omega_0 + \omega_0 + \omega_0.}$$

But by inserting $3\omega_0$ at the right, we construct a linearly ordered set

$$3\omega_0 \cup \{3\omega_0\}$$

that is not well-ordered. Evidently, $(3\omega_0)^+$ does not exist in $3\omega_0 \cup \{3\omega_0\}$. The set $3\omega_0$ is a linearly ordered set that is different from $2\omega_0$, since $2\omega_0$ has exactly one limit ordinal, while $3\omega_0$ contains the two limit ordinals, ω_0 and $2\omega_0$. Then the inclusion of the element $3\omega_0$, one that is larger than all elements in (6.1), creates a new linearly ordered set whose picture looks like this.

Now, let us abstract. Can you see it? Can you see how we will proceed next?

The above constructions suggest a chain of limit ordinals that are integer multiples of ω_0.

$$\omega_0 < 2\omega_0 < 3\omega_0 < \ldots$$

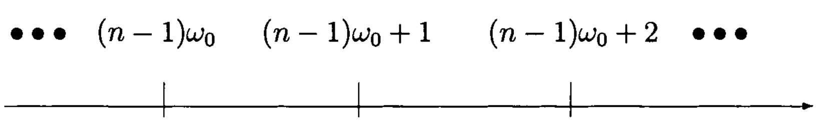
Given an ordinal $(n-1)\omega_0$ for some $n \in \mathbb{N}$, we would define an ordinal

$$n\omega_0$$

as follows. For each $k \in \mathbb{N}$, define the ordinals

$$\begin{aligned} (n-1)\omega_0^+ &= (n-1)\omega_0 + 1 \\ ((n-1)\omega_0 + k)^+ &= (n-1)\omega_0 + (k+1) \end{aligned}$$

as we did above with ω_0. A picture is given as follows.



Then we choose a symbol

$$\boxed{n\omega_0}$$

and define it to be the unique element that is larger than all of the elements

$$(n-1)\omega_0, (n-1)\omega_0+1, (n-1)\omega_0+2, \ldots.$$

The set that we have just defined is a well-ordered set since each element possesses a successor.

By adding in the symbol $n\omega_0$ we have described the well-ordered set that is described with the following picture.

$$\bullet\bullet\bullet \quad (n-1)\omega_0 \qquad (n-1)\omega_0+1 \qquad (n-1)\omega_0+2 \quad \bullet\bullet\bullet \quad n\omega_0$$

Before going on, we should examine the phrase *and so on*. We might be tempted to say *and so on* when we read ..., but what does that mean now? Indefinitely? Infinitely? What do these words mean? Do we mean that we should repeat the process once for each $n \in \mathbb{N}$, or once for each ordinal $< 3\omega_0$, or perhaps once for each ordinal less than $10^{100}\omega_0$? We need a more precise way of saying *and so on*, and we will see it in the next section.

We have indeed defined an unending well-ordered chain of limit ordinals

$$\omega_0 < 2\omega_0 < 3\omega_0 < \ldots < n\omega_0 < \ldots$$

that includes very ordinal $n\omega_0 + k$ between the successive ordinals $n\omega_0$ and $(n+1)\omega_0$. Furthermore, each ordinal $n\omega_0$ is a limit ordinal because we have assumed that this chain contains all of the ordinals that are $< n\omega_0$. For example, $100\omega_0$ and $186,000\omega_0 + 3$ are in this well-ordered set. Why, there is even a limit ordinal

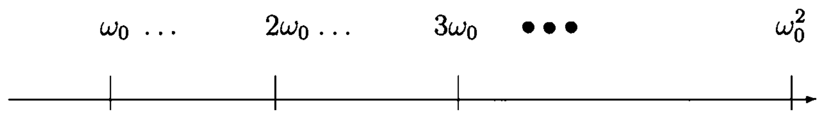
$$10^{10^{10^{10^{10}}}}\omega_0,$$

but this is so far out there that we would have trouble imagining just how large that ordinal is. Since this set contains infinitely many limit ordinals $\omega_0, 2\omega_0, 3\omega_0, \ldots$, the set is not like any of the well-ordered sets that we have defined to this point.

As long as we have a chain of limit ordinals $\omega_0, 2\omega_0, 3\omega_0, \ldots$ we can define a symbol

$$\boxed{\omega_0^2 = \omega_0 \cdot \omega_0}$$

that is larger than each of the limit ordinals $n\omega_0$ with $n \in \mathbb{N}$. Pictorially we have defined the new linearly ordered set whose limit ordinals look like this.



We have defined ω_0^2 by defining it to be the least symbol that is larger than all of existing symbols of the form $n\omega_0$ for $n \in \mathbb{N}$. Then the *ordinal* ω_0^2 is larger than all ordinals α such that $\alpha \leq n\omega_0$ for some $n \in \mathbb{N}$. Even the incredibly large limit ordinal

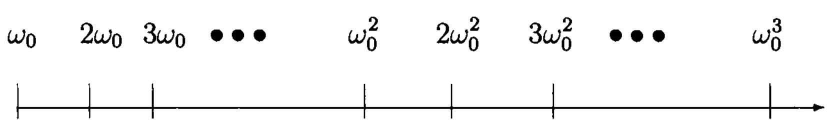
$$10^{10^{10^{10^{10}}}}\omega_0$$

is smaller than ω_0^2.

Just for the fun of it let's try to define ω_0^3. Let us define

$$\boxed{\omega_0^3 = \omega_0 \cdot \omega_0^2.}$$

Beginning with ω_0^2 we will define the ordinals $n\omega_0^2$ for each $n \in \mathbb{N}$, by placing n copies of ω_0^2 in a line that we call $n\omega_0^2$. The diagram shows us what we have done. It lacks scale, and only possesses direction.



If we start with $\omega_0 = \omega_0^1$, then we can proceed inductively and define

$$\boxed{\omega_0^n = \omega_0 \cdot \omega_0^{n-1} \text{ for each } n \in \mathbb{N}.}$$

This definition repeats the process used to define the small ordinal powers

$$\omega_0^2 = \omega_0 \cdot \omega_0, \quad \omega_0^3 = \omega_0 \cdot \omega_0^2, \quad \omega_0^4 = \omega_0 \cdot \omega_0^3.$$

In this way, the process continues, defining all of the powers ω_0^n.

We have then expanded our collection of limit ordinals by taking *powers of* ω_0 and all of their combinations. We define a new well-ordered set that contains

$$\boxed{\omega_0 < \omega_0^2 < \omega_0^3 < \dots}$$

and all of the ordinals α such that $\alpha \leq \omega_0^n$ for some $n \in \mathbb{N}$. For instance,

$$10\omega_0 < 121\omega_0^2 < 10^{10^2}\omega_0^{33} + 101\omega^{12} < 2\omega_0^{81} + \omega_0 < \omega_0^{100}$$

are elements in this set.

If we continue the chain $\omega_0, \omega_0^2, \omega_0^3, \dots$ of limit ordinals, we will arrive at the much larger limit ordinal

$$\boxed{\omega_0^{\omega_0} = \text{the least ordinal that is larger than all of the ordinals } \omega_0^n.}$$

This limit ordinal $\omega_0^{\omega_0}$ can be pictured as follows.

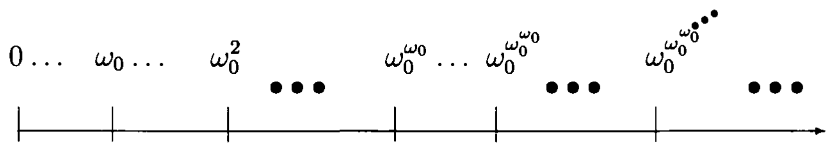
In fact, there is no end to the exponents we can form. We can even define an ordinal that has infinitely many exponents. Let us agree that

$$\boxed{\omega_0^{\omega_0^{\omega_0^{\cdot^{\cdot^{\cdot}}}}}}$$

is *the least ordinal that is larger than each of the exponentials*

$$\omega_0 < \omega_0^{\omega_0} < \omega_0^{\omega_0^{\omega_0}} < \omega_0^{\omega_0^{\omega_0^{\omega_0}}} < \dots.$$

Hence, $\mathbb{N}$ is part of a larger well-ordered set, called the *chain of ordinals*, a fraction of which is given below.



This picture is not complete since we can extend this well-ordered set by defining

$$\omega_0^{\omega_0^{\omega_0^{\cdot^{\cdot^{\cdot}}}}} < \omega_0^{\omega_0^{\omega_0^{\cdot^{\cdot^{\cdot}}}}} + 1 < \omega_0^{\omega_0^{\omega_0^{\cdot^{\cdot^{\cdot}}}}} + 2 < \dots.$$

It is difficult to explain which well-ordered set these ordinals represent. They are so large that they exceed every ordinal that we have discussed so far. Let us just agree that these are large ordinals whose existence we had not expected. Keep in mind, though, that like the other line graphs, there are many unlabeled ordinals between the labeled points in this picture. That is what $\dots$ and $\bullet\bullet\bullet$ should say to you.

And yet, as large as these ordinals are, there is a larger ordinal. To find it, we will have to construct the well-ordered set of ordinals in a different way.

Let ω_1 be the set of real numbers. It is best if we use a different symbol like ω_1, as it will enhance the discussion. In Theorem 5.3.1 we showed that

$$\text{card}(\mathbb{N}) < \text{card}(\mathbb{R}) = \text{card}(\omega_1).$$

One of the accepted axioms of Mathematics is that ω_1 can be turned into a well-ordered set. Recall that $\mathbb{R}$ is a linearly ordered set, but not well-ordered. The well ordering for ω_1 is not related to the usual linear order on $\mathbb{R}$ at all. The well ordering that we will choose for ω_1 cannot be constructed. We simply know it exists.

So let us assume that ω_1 is well ordered. This well-ordered set ω_1 is so large, that it contains a copy of every one of the ordinals we have discussed so far. For instance,

$$1, \quad \omega_0, \quad \omega_0^{\omega_0^{\omega_0^{\cdot^{\cdot^{\cdot}}}}}$$

are (essentially) in ω_1. But none of these ordinals is the largest element of ω_1.

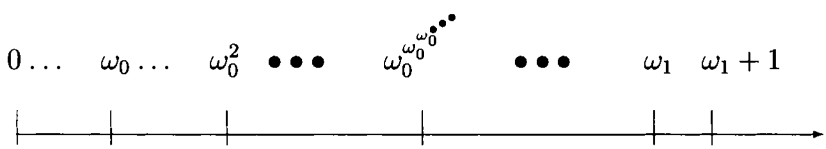
Define a new linear order on the set $\omega_1 \cup \{\omega_1\}$ by requiring that

$$x < \omega_1 \text{ for each } x \in \omega_1.$$

Then ω_1 becomes the largest element in $\omega_1 \cup \{\omega_1\}$. We say that

$$\boxed{\omega_1 \quad = \quad \text{the first uncountable ordinal.}}$$

And yet ω_1 is not the largest ordinal.



Indeed, if we seek the largest ordinal, then our search is fruitless. There is no such object. Ordinals just go on without limit. If we think that we have exhausted our chain of ordinals at some value α, we can always define a new ordinal $\alpha < \alpha + 1$, and kick-start our chain of ordinals.

You see, if Ω is the chain of ordinals, all of them, then the only thing that prevents us from creating new ordinals is the fact that Ω *is not a set.* Our construction of ordinals requires us to use a set in order to create another ordinal. See for yourself in the pages of construction that we just worked through. Arguments on chains that are not sets are best left to philosophers.

Let us end this section by reintroducing an old friend of ours. Do you remember the cleaning lady, Maria, in Hilbert's Infinite

Hotel from page 130? She's the one who cleaned all of the rooms immaculately, and was not heard from again. Well, as it happens, she was found resting quietly in a room outside the Infinite Hotel numbered ω_0. Naturally, this is the room for Hotel Staff.

Exercise 6.2.2 1. Construct $\omega_0 + 1$ and show that it is a linearly ordered set, but not a well-ordered set.

2. Construct $\omega_0 + 2$ and show that it is a linearly ordered set, but not a well-ordered set.

3. Construct $3\omega_0$ and show that it is a well-ordered set.

6.3 Cardinals as Ordinals

We have shown that it is possible to order a set *linearly* but not well-order the set. *Linear order* for $\mathbb{R}$ implies that the Trichotomy Property is True of $\mathbb{R}$, but we showed by example that $\mathbb{R}$ is not well-ordered.

Recall that $\text{card}(A) \leq \text{card}(B)$ if there is a one-to-one function $f : A \longrightarrow B$. This ordering $\leq$ of cardinals is linear because the given cardinals satisfy the Trichotomy Property. That is,

given any two cardinals α and β either $\alpha < \beta, \quad \beta < \alpha, \quad \text{or} \quad \alpha = \beta.$

This allows us to make the following conclusion.

Given two cardinals α and β then $\alpha \leq \beta$ and $\beta \leq \alpha \Longrightarrow \alpha = \beta.$

It is natural to ask whether this *linear* ordering of cardinals is a *well*-ordering of cardinals. For instance, if we want to know the comparative sizes of $\text{card}(\mathcal{P}^3(\mathbb{N}))$ and $2^{\text{card}(\mathcal{P}(\mathbb{N}))}$, then we can be assured that exactly one of the three conditions

$$\text{card}(\mathcal{P}^3(\mathbb{N})) < 2^{\text{card}(\mathcal{P}(\mathbb{N}))}, \quad 2^{\text{card}(\mathcal{P}(\mathbb{N}))} < \text{card}(\mathcal{P}^3(\mathbb{N})),$$

$$\text{or} \quad \text{card}(\mathcal{P}^3(\mathbb{N})) = 2^{\text{card}(\mathcal{P}(\mathbb{N}))}$$

applies to the two cardinals. Now that we know that the cardinals are linearly ordered, we make the next leap of the Mathematical imagination. The reason that we have taken such care with the ordinals and their arithmetic is the following deep result from Mathematics that links ordinals with cardinals.

Theorem 6.3.1 *Given an infinite cardinal $\aleph$, there is a successor cardinal $\aleph^+$. Thus, the set of cardinals is well-ordered.*

The theorem is deep, and its proof is far too complex for us to present in this book, so let us just accept the truth of it.

By Theorem 6.3.1, then there are cardinals $\aleph_0^+$ and $(2^{\aleph_0})^+$. Since

$$\aleph_0 < 2^{\aleph_0}$$

and since the successor element x^+ is the smallest element that is larger than x, we can say that

$$\aleph_0^+ \leq 2^{\aleph_0},$$

but we do not know if this is a proper inequality, or an equation. The answer to our observation is phrased in a language that takes us to edge of what can be proved from the Axioms of Twentieth Century Mathematics.

A more familiar inequality comes from Theorem 5.3.1:

$$\aleph_0 < \operatorname{card}(\mathbb{R}).$$

Since the successor element $\aleph_0^+$ is smaller than every element that is larger than $\aleph_0$, we can write that

$$\aleph_0^+ \leq \operatorname{card}(\mathbb{R}).$$

Questions about this inequality turn out to be some of the deepest questions in Mathematics. We will address this predicament presently.

The following notation will be used for the most often referred to cardinals:

$$\boxed{\begin{aligned} \aleph_0 &= \operatorname{card}(\mathbb{N}), \\ \aleph_1 &= \aleph_0^+ \text{ is the successor of } \aleph_0, \\ \aleph_2 &= \aleph_1^+ \text{ is the successor of } \aleph_1, \end{aligned}}$$

and so on. Inasmuch as

$$\aleph_0 = \text{card}(\mathbb{N}) < \text{card}(\mathbb{R}),$$

we can write that

$$\aleph_1 \leq \text{card}(\mathbb{R}).$$

The reasoning for this inequality is that the successor $\aleph_1 = \aleph_0^+$ is *the least element that is larger than* $\aleph_0$. We have thus arrived at our first mysterious property of cardinals. Given that $\aleph_1 \leq \text{card}(\mathbb{R})$ it is natural to ask if $\aleph_1 = \text{card}(\mathbb{R})$. This question is so subtle a matter that the Mathematician who answered it, Paul J. Cohen, was honored with the highest award in Mathematics: the Field's Medal.

Theorem 6.3.2 [Paul J. Cohen] *Using the Axioms of Twentieth Century Mathematics, it cannot be proved and it cannot be disproved that* $\aleph_1 = \text{card}(\mathbb{R})$.

We say that the condition $\aleph_1 = \text{card}(\mathbb{R})$ is *independent of the ZFC Axioms of Set Theory.* That is, it is unattainable as a theorem in Twentieth Century Mathematics. That statement needs some explanation. I suppose you have had experience in Geometry, so we start there.

The Geometry we studied in high school was based on ten elementary truths, or axioms and propositions, given by a Greek Mathematician Euclid (circa 300 BC). Euclid started with these ten statements and then proceeded to develop all of Geometry from them. The truth of most of those statements is quite transparent. For example, Euclid starts by assuming that *the whole is the sum of its parts.* Obviously, this is True. He also assumes that *all right angles have equal measure.* Again, obviously True. There are eight others that we will not state because they do not impact on our discussion.

From these ten obvious truths, Euclid could and did derive all of the Geometry that you studied in high school. By starting with True statements about points, lines, and triangles, and then extending his thoughts using *Aristotelian Logic*, (the logic of True and False statements), Euclid was able to develop a Mathematics that was free of any mistakes or misconceptions. Euclid's method, now called the

Axiomatic Method, is so compelling, that Mathematicians in every century over the last 2300 years have made it the basis for their work.

In using the Axiomatic Method, Mathematicians begin our work with a few basic statements, and proceed from there to demonstrate universal truths. These truths are not personal opinion. They are statements that would be True no matter who was reading them. They are truths taken from what is called *the model for the Axioms.* The model for Euclid's Geometry is the familiar plane, its points, and its lines. The Axioms from this plane are then truths that can be shared with advanced civilizations anywhere on Earth and, should the need arise, with extraplanetary civilizations. Assuming that a group of people is sufficiently advanced, assuming they know enough Mathematics, we would share with them the same theorems of Euclid's Geometry. That is the power and the nature of the axiomatic method. Try to make a similar statement concerning the literature, art, or religion of two different civilizations on Earth. That would not be possible.

Mathematics took on the axiomatic method with vigor in the early part of the twentieth century. In this time, every area of Mathematics from algebra to calculus tried to find those few elementary truths that could be used as a springboard with Aristotelian Logic to produce universal truths about that area of Mathematics. Sometime in the 1920's, a group of Mathematician came up with a few statements that all Mathematicians could agree on as obviously True, and that most agreed were simple statements. These axioms today are called the *the ZFC Axioms*, or *Zermelo-Franklin-Choice Axioms.* These are axioms about sets, and they are the basis for most modern Mathematics. For example, one of the statements is that there do not exist sets A and B that are elements of each other.

There do not exist sets A and B such that $A \in B$ and $B \in A$.

A moment's thought might convince you that this statement is True, and that its negation is

There are two sets A and B such that $A \in B$ and $B \in A$.

The point behind this is that by starting with elementary statements about sets, we are able to *prove* advanced ideas in such areas as Linear Algebra, Calculus, and Trigonometry. All of Mathematics assumes Set Theory so all of Mathematics depends upon Axioms.

What P. J. Cohen proved was that there is at least one statement about Mathematics that cannot be proved or disproved with the ZFC Axioms. No amount of Logic and Mathematics will allow us to start with the Axioms of Set Theory and then end up with $\aleph_1 = \text{card}(\mathbb{R})$. The Mathematics of the ZFC Axioms, as large as it might seem right now, is too small to include $\aleph_1 = \text{card}(\mathbb{R})$ as a theorem. Now there is a statement. Why would you ever think that Mathematics is small? And yet here we have an example of its limitations.

This is probably the first time you have seen such a statement. Think of this as a Catch-22 in your algebra course. Your teacher asks you to show that $x = y$ for some numbers x and y, and you tell your professor, *I cannot prove that $x = y$ but I cannot prove that $x \neq y$.* Your professor would certainly give you a high mark if $x = \aleph_1$ and $y = \text{card}(\mathbb{R})$.

Let's examine what it means to be independent of the ZFC Axioms of Set Theory. There are five axioms that define what set theory is. From these axioms, almost all of modern Mathematics can be proved as theorems. This occurs in exactly the same way you learned it in your high school Geometry course. You begin with five axioms or postulates (these from Euclid) from which you can prove all of the theorems in Elementary Geometry. Mathematics is handled in the same way. Most of the Mathematics we meet in elementary courses can be proved from the Axioms of Set Theory. Thus the ZFC Axioms of Set Theory become the foundation of modern Mathematics. With such a beginning, Mathematicians are sure that modern Mathematics is on the same firm foundation that Euclid's Geometry enjoys.

Before Cohen's Theorem 6.3.2, Mathematicians did not know of a *simply stated idea* that was independent of the ZFC Axioms of Set Theory. Indeed, these Axioms were relatively new discoveries when Cohen proved his theorem. It struck the Mathematicians of the day as quite strange. The fact that some statement about Mathematics is beyond Mathematical proof or disproof is thought to be curious in the early Twenty-first Century.

Before Cohen's Theorem, the powerful Mathematician David Hilbert, (remember him of Infinite Hotel fame), had suggested that Mathematicians should start looking for a program or computer of sorts that would produce all of Mathematics if it was allowed to run long enough. Cohen's Theorem 6.3.2 proved that Hilbert's suggestion could not be realized. There can be no program or computer that produces all of Mathematics, even if it was allowed to run in some thought experiment. Some statements of Mathematics are simply beyond the reach of proof form the ZFC Axioms of Set Theory. They can be stated, or they can be *assumed*, but they cannot be proved.

The cold fact is that the ZFC Axioms of Set Theory are simply not strong enough to prove every statement about Mathematics. Furthermore, this is not a weakness of the Axioms as it was proved by Curt Gödel in the 1930's that no collection of Mathematical Axioms will allow us to prove every statement about Mathematics.

The only way to logically include a statement like $\aleph_1 = \text{card}(\mathbb{R})$ in Mathematics is to *include it as an axiom.* However, once that inclusion is made *we create a larger and different Mathematics that properly contains the one we grew up with.* This new Mathematics will contain all of the work you did in your high school Mathematics class, including Algebra and Geometry, and it will contain a new and perhaps strange Mathematical statement. Before the end of this discussion we will encounter a simple statement that is True in the Mathematics that assumes that $\aleph_1 = \text{card}(\mathbb{R})$, but that is not True in a smaller Mathematics.

Georg Cantor did not know of Cohen's Theorem (since Cantor lived 80 years before Cohen's Theorem was published), so he accepted $\aleph_1 = \text{card}(\mathbb{R})$ as an axiom or hypothesis.

The Continuum Hypothesis: Assume that $\aleph_1 = \text{card}(\mathbb{R})$.

Because the Continuum Hypothesis is independent of our usual Mathematics, it can be used as an axiom with which we can construct an entirely new Mathematics:

CH.

CH is *the usual Mathematics together with the Continuum Hypothesis* used as an axiom. Many beautiful theorems are proved in **CH**, or by assuming the Continuum Hypothesis. The Mathematics you learned in high school is not strong enough to prove all of the theorems in **CH**.

On the other hand, because we cannot disprove the Continuum Hypothesis, we can include its *logical negation*, $\aleph_1 < \mathrm{card}(\mathbb{R})$, as an Axiom in Mathematics, thus constructing a new Mathematics:

notCH.

While many theorems are proved in **notCH**, these theorems cannot be compared to the theorems in **CH**. Some theorems in **CH** may seem to contradict some theorems in **notCH**, but this is an inaccurate and incorrect comparison because

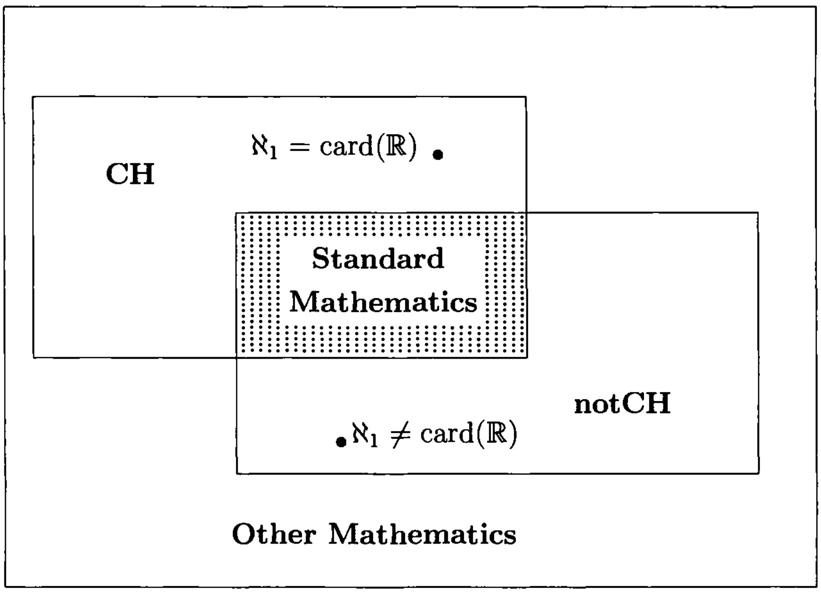
CH and **notCH** exist as two different Mathematical Universes.

We cannot compare the theorems from **CH** and **notCH** anymore than we could compare recipes for bread and cake, and claim that one is right and one is wrong. Each has its own uses and each has its own role on the dinner table. In the same way, the theorems of **CH** have their place in Mathematics, the theorems in **notCH** have their place in Mathematics, and the two places cannot be compared.

This must seem odd. We were all led to believe that there was only one Mathematics. This reaction is not unusual. The Mathematics we learned in high school is contained in both **CH** and **notCH**. No one except the experts ever has the opportunity to distinguish between the Mathematics in **CH** and the Mathematics in **notCH**. In picture (6.2), **CH** and **notCH** are the regions that contain the shaded region called Standard Mathematics. From this picture, we can imply certain facts about Mathematics.

Moreover, diagram (6.2) suggests that there are *other* regions of Mathematics to be discovered out there. This is indeed the truth of the matter. Here is what this means to you.

All of Mathematics (6.2)



1. **CH** and **notCH** overlap in the Standard Mathematics.

2. **CH** and **notCH** are different.

3. Neither **CH** nor **notCH** equals the Standard Mathematics.

4. Together **CH** and **notCH** do not encompass All of Mathematics.

In the following discussion, I will attempt to show you a Mathematical fact that changes when viewed first as a theorem in **CH**, and then as a theorem in **notCH**.

Recall that $\aleph_1 = \aleph_0^+$ is the successor to $\aleph_0$, and that $\aleph_2$ denotes the successor of $\aleph_1$.

$$\boxed{\aleph_2 = \aleph_1^+.}$$

The question we will consider is: *How do* $\aleph_0$, $\aleph_1$, card($\mathbb{R}$), *and* $\aleph_2$ *compare?* It turns out that these relationships depend on the Mathematics we choose to work in, **CH** or **notCH**. We have established that

$$\aleph_1 \leq \text{card}(\mathbb{R}),$$

so Theorems 5.3.1 and 5.4.3 can be used to construct a short chain of cardinals

$$\boxed{\aleph_0 < \aleph_1 \leq \text{card}(\mathbb{R}).}$$

To fine tune these inequalities we would use the Continuum Hypothesis or its logical negation. The next two boxes show us how these inequalities change in **CH** and in **notCH**. Note the different placement of $\aleph_1$ and $\aleph_2$.

In **CH**, we assume that $\aleph_1 = \text{card}(\mathbb{R})$, so there is a chain

$$\boxed{\aleph_0 < \boxed{\aleph_1 = \text{card}(\mathbb{R})} < \aleph_2,}$$

of cardinals. However, in **notCH**, $\aleph_1 \neq \text{card}(\mathbb{R})$, so there is a chain

$$\boxed{\aleph_0 < \aleph_1 < \boxed{\aleph_2 \leq \text{card}(\mathbb{R})}.}$$

of cardinals. This is real evidence that **CH** and **notCH** are different Mathematical systems.

Let us examine that thought. Based on its placement on the cardinal line, the cardinality of $\mathbb{R}$, card($\mathbb{R}$), which is essentially the number of elements in $\mathbb{R}$, seems to have different values when written in the different Mathematics **CH** and **notCH**. Thus, depending on which Mathematics you choose to work in, the infinite cardinal card($\mathbb{R}$) will compare differently to other cardinal values. The difference between **CH** and **notCH** could not be more plain.

$$\aleph_1 = \text{card}(\mathbb{R}) \text{ in } \mathbf{CH},$$
$$\aleph_2 \leq \text{card}(\mathbb{R}) \text{ in } \mathbf{notCH}.$$

How is that possible? Does the number of elements in $\mathbb{R}$ change as it passes between **CH** and **notCH**? Are there fewer real numbers in **CH** than there are in **notCH**? Of course not.

The set $\mathbb{R}$ does not change when we work in **CH** or in **notCH**. If x is a real number in the Mathematics **CH** then x is a real number in the Mathematics **notCH**. The reason for this is that $\mathbb{R}$ exists in the Mathematics you learned in high school. The real numbers will be the same no matter which Mathematical structure you work in.

So then how does card($\mathbb{R}$) change relative sizes between **CH** and **notCH**? The answer is that in **notCH** we add a cardinal between $\aleph_0$ and card($\mathbb{R}$). In our discussions and our line diagrams, we would treat card($\mathbb{R}$) as though it was a variable cardinal. We moved it like it was on wheels. This is where we erred.

Leave $\aleph_0$ and card($\mathbb{R}$) in one place on the cardinal line, and then write in the cardinals that **CH** provides for us. By reading the Continuum Hypothesis, we see that $\aleph_1$ should be placed where we placed card($\mathbb{R}$).

But when we form **notCH**, we must allow for the successor of $\aleph_0$, namely $\aleph_1$, to take a place less than card($\mathbb{R}$). In this way, **notCH** contains a cardinal for which the ZFC Axioms of Set Theory, and our knowledge of subsets of $\mathbb{R}$, do not prepare us. Consequently, there is the possibility that in **notCH**, $\aleph_1$ is not the cardinality of any set in Mathematics.

Now it is clear why card($\mathbb{R}$) has a different placement. We did not add to card($\mathbb{R}$). That cardinal remained unchanged. What we saw was that **notCH** introduced some new cardinal numbers into the line of infinite cardinals by allowing for the existence of other cardinals between $\aleph_0$ and card($\mathbb{R}$).

In other words, on the cardinal number line, there are two cardinals

$$\aleph_0 < \mathrm{card}(\mathbb{R}).$$

These cardinals exist no matter which of the two Mathematics you work in. But **CH** and **notCH** introduce different cardinals between $\aleph_0$ and $\mathrm{card}(\mathbb{R})$.

In **CH** there will not be cardinals properly between $\aleph_0$ and $\mathrm{card}(\mathbb{R})$. Therefore, when we place $(\aleph_0)^+ = \aleph_1$ in this line, then we have

$$\aleph_0 < \aleph_1 = \mathrm{card}(\mathbb{R}) < \aleph_2$$

Moreover, in **notCH** there is at least one cardinal properly between $\aleph_0$ and $\mathrm{card}(\mathbb{R})$. Then when we place $(\aleph_0)^+ = \aleph_1$ in this new line, we have

$$\aleph_0 < \aleph_1 < \mathrm{card}(\mathbb{R}),$$

so that

$$\aleph_0 < \aleph_1 < \aleph_2 \leq \mathrm{card}(\mathbb{R}).$$

It follows that the ways we *count* in **CH** and **notCH** are different because the number and density of cardinals in the Mathematics **CH** and **notCH** is quite different. The cardinal $\mathrm{card}(\mathbb{R})$ does not change values when we pass from the Mathematics **CH** to **notCH**. It is the existence of smaller cardinals that seems to change the relative size of $\mathrm{card}(\mathbb{R})$.

Now consider the cardinal $2^{\aleph_0}$. Under the ZFC Axioms of Set Theory we can write down that

$$\aleph_0 < \aleph_1 \leq 2^{\aleph_0}.$$

Thus in **CH** we have

$$\boxed{\aleph_0 < \aleph_1 = 2^{\aleph_0} < \aleph_2}$$

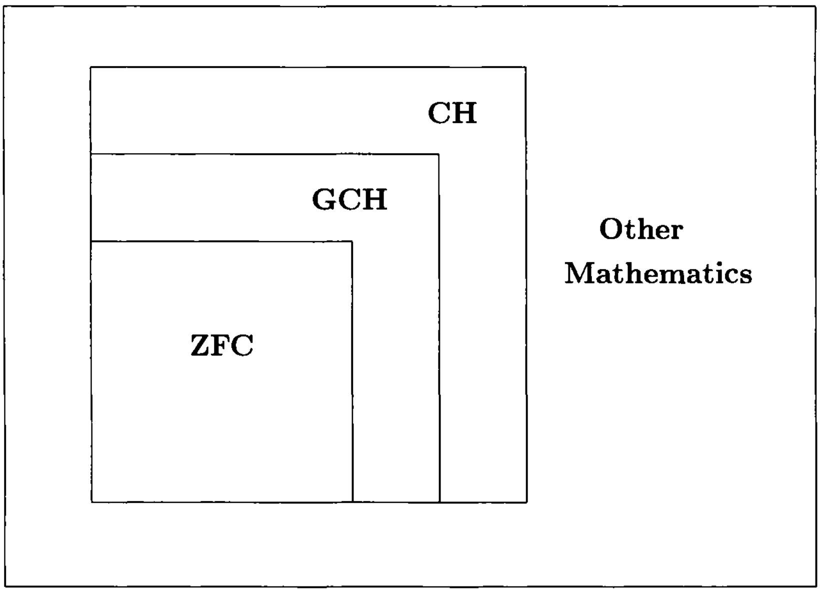
while in **notCH** we have

$$\boxed{\aleph_0 < \aleph_1 < \aleph_2 \leq 2^{\aleph_0}.}$$

Notice that the number $2^{\aleph_0}$ did not change size. When passing to **notCH**, a new cardinal, that we denote by $\aleph_1$, was placed properly between $\aleph_0$ and $2^{\aleph_0}$.

Once again, the size of $2^{\aleph_0}$ did not change. It just appeared to change because of the introduction of a cardinal that is not present in **CH**.

All of Mathematics (6.3)



Let me impart one final item. A more general version of the Continuum Hypothesis is

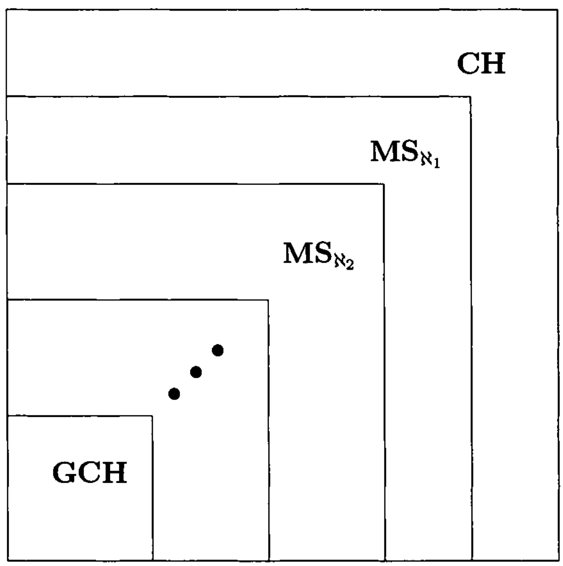
> **The Generalized Continuum Hypothesis** (GCH):
> Assume that $\aleph^+ = 2^{\aleph}$ for each cardinal $\aleph$.

That is, given a cardinal $\aleph$ assume that $2^{\aleph}$ is the next cardinal. Thus GCH assures us that all successor cardinals $\aleph^+$ come in essentially one form, that of a power of 2. Consequently, **GCH** shows us that $\aleph_n = 2^{\aleph_{n-1}}$ for each $n \in \mathbb{N}$. In other words,

$$\aleph_n = 2^{2^{\cdot^{\cdot^{\cdot^{\aleph_0}}}}}$$

where the tower contains exactly $n-1$ twos.

Picture (6.3) illustrates the containing relationship between the Mathematical system **GCH** that results from the addition of the Generalized Continuum Hypothesis GCH into the Standard Mathematics **ZFC**. Picture (6.3) suggests that **ZFC**, the Standard Mathematics, is *properly* contained in the Mathematical system **GCH**, (which is formed by adding GCH to **ZFC**), and that **CH** properly contains **GCH**. Thus we have found a Mathematics properly between **CH** and **ZFC**.



(6.4)

There is every reason to expect that there is an *infinite chain* of Mathematical systems containing **ZFC**. Indeed, given an infinite cardinal $\aleph$ we can define a Mathematics $\mathbf{MS}_\aleph$ which is formed by adding the equations

$$\alpha^+ = 2^\alpha \qquad \text{for all cardinals } \aleph_0 \leq \alpha \leq \aleph$$

to **ZFC**. For example, $\mathbf{MS}_{\aleph_0} = \mathbf{CH}$. Thus there results an infinite chain of Mathematical systems

$$\mathbf{ZFC} \subset \cdots \subset \mathbf{MS}_{\aleph_2} \subset \mathbf{MS}_{\aleph_1} \subset \mathbf{MS}_{\aleph_0} = \mathbf{CH}$$

between **ZFC** and **CH**.

As we stated earlier, if we assume the Generalized Continuum Hypothesis then each successor cardinal is a power of 2. In this way, cardinals are not like numbers at all. After all, not every number is a power of 2. Thus under the Generalized Continuum Hypothesis each of the cardinals $\aleph_0$, $\aleph_1$, $\aleph_2$, ... is a power of 2.

$$\boxed{\aleph_0, \quad \aleph_1 = 2^{\aleph_0}, \quad \aleph_2 = 2^{2^{\aleph_0}}, \quad \aleph_3 = 2^{2^{2^{\aleph_0}}}, \quad \ldots .}$$

There is one easy question that I will leave the reader with in this chapter. We know that the Continuum Hypothesis cannot be proved using the ZFC Axioms of Set Theory. There are also Mathematical statements that cannot be proved within **CH**. The Generalized Continuum Hypothesis is one such statement. Moreover, there is a True statement that cannot be proved within **GCH**. It exists but is unknown at this time.

6.4 Magnitude versus Cardinality

As with every discussion about the infinite, the ideas take place in a Mathematical thought experiment or a Platonic Universe. So if these ideas do not agree with what you see around you, good! There is nothing infinite about this physical world around us. Let us expand on that for a little bit.

When you were introduced to the infinite as a child, you thought that infinity was an infinite distance from us. Somehow, infinity was a place that we could reach if we could move infinitely far. This is not possible within our physical universe. In our universe, we have only a finite distance that we can travel before we come upon our starting point again. There are two infinities in that idea that have to be addressed, namely, infinite distance and infinite time.

The universe, as far as is known at the time of this printing, is a lumpy sphere having at least four dimensions. Its beginning can be pinned down to within such a small fraction 10^{-33} of *a second* that for our purposes we can assume that scientists have glimpsed the origin of the universe. Of course, this is not really the universe's origin, and perhaps that origin is beyond the scope of scientific observation, but it gives us a beginning, a starting point to talk about age *in years.*

The beginning of our universe seems to have occurred at least 15 billion years ago. It might be older or younger than that but we can assume with scientific certainty that the universe is no more than 30 billion years old. Moreover, outside this framework of time, it is relatively certain that time as we know it does not exist. Time seems to be peculiar to our universe. Thus, there can be no infinite time prior to us. Time has not existed forever.

The expansion of the universe probably occurred in fits and starts. By this I mean that the expansion of the universe was not smooth and not of uniform speed. But it has a maximum value, the speed of light. Given the age of the universe and a finite rate of expansion, it is clear that the universe is a bounded place whose size is dependent upon the speed of light. This gives us a bounded universe, so that distances in the universe do not appear to be infinite.

Before anyone starts to suggest that we can reach the edge of the universe, let me bring you back to reality. There is no edge to the universe. If you start out in a direction and hope to travel in a straight line, you are thwarted by gravitational considerations since gravity warps space and time. Furthermore, you are on a four dimensional lumpy spheroid, not a three dimensional object. Then your travels are necessarily distorted by the fourth dimension. For instance, suppose you are a point living on a sphere. You live in three dimensions, but you only experience two. As you travel around the bounded spherical universe of yours, you are impressed with the fact that you can never reach the edge of the universe. And yet you know the universe is bounded. Try it on an orange and see wether that helps you visualize the point moving about and seeking the edge of its universe.

Thus, we do not find our universal edge. It bends away from us in the fourth dimension. In fact, cosmologists, people who study the universe, suspect that our universe exists as an 11 dimensional object. It would then be impossible for us to know today where that bending or curvature in the universe might be.

But just because the point's universe has no edge does that mean that the distances in the universe extend to the infinite? Are there two points in our universe that are infinitely far apart? Probably not. The distances might be large, perhaps even larger than our imagined point ever imagined, but the distances on the sphere are quite finite.

In the same way, but in a higher number of dimensions, our universe does not possess two points that have an infinite distance between them. Infinite distances, it seems, must somehow extend beyond the universe. What then do we mean by infinite distances? This is the difference between cardinals and the infinite distances you may have become familiar with. Infinite distances do not exist while our notion of a cardinal does exist.

A cardinal is in a very loose sense a count of the number of elements in a set. As we saw there are many cardinals including the first infinite cardinal $\text{card}(\mathbb{N}) = \aleph_0$ and the cardinality of the reals $\text{card}(\mathbb{R})$. We can continually add elements to these sets to form sets with larger cardinality. For example, if we let $\mathcal{P}(X)$ denote the *power set* of X, or equivalently $\mathcal{P}(X)$ is the set of subsets of the set X, then

$$\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathcal{P}(\mathcal{P}(\mathbb{N})), \ldots$$

is a list of sets with *different cardinalities.* Since we can always take the power set of a set, this list is infinite. Thus, if we compare sets carefully, we find that there are *infinitely many infinite cardinals.* These cardinals do not measure distance. They do not measure how far it is across a set. These cardinalities simply measure *how many elements are in the set* using one-to-one and onto functions.

The infinite ideas that you have seen in life are more closely related to distance because their size is determined relative to the natural numbers, $\mathbb{N}$. Mathematics does not recognize distance as an infinite number. That symbol you may have seen, the lazy eight ∞, is not defined as a number. That's right. No matter what you may have heard, ∞ is *not a number* of any kind.

The division $\frac{1}{0}$ is not infinity, either. It is *undefined,* meaning that it holds no meaning at all. The symbol $\frac{1}{0}$ makes no Mathematical sense at all. We do not even give it a special symbol or name. In fact, except for books like this one, you are not likely to see a professional Mathematician write down $\frac{1}{0}$. That's enough of that. My fingers hurt when I type $\frac{1}{0}$. Ouch, I did it again.

Now, to speak about infinite quantities, Mathematicians have found that the Calculus device called a *the limit* is the only notion that allows us to describe and manipulate infinite or unbounded quantities in a consistent and Mathematically precise manner. Let $f(x)$ be a function that takes real numbers x to real numbers $f(x)$. This is where that high school education comes in handy. Suppose we want to know about the size of $f(x)$ as x gets closer to some number, say, a. We write

$$\boxed{\lim_{x \to a} f(x) = L}$$

provided that $f(x)$ gets closer to L as x gets closer to a. Don't let the notation scare you. This is the smallest bit of notation we can use to get the idea across to the reader. It has been in use for 150 years now, and it shows no sign of being replaced by something else. What do we mean by *x gets closer to a*? What we mean is that x is allowed to take on values that are physically closer to a. That is, *the difference between them* (i.e., the larger minus the smaller), is allowed to become very small or close to 0. For example, a calculator will show you that $2x+3$ gets close to 5 as x gets close to 1. I won't bore you with the table of numbers. In our notation,

$$\lim_{x \to 1} 2x + 3 = 5.$$

We are more interested in limits where L is not finite. We say that $f(x)$ *has an infinite limit at* a, if for each natural number N, there is a small neighborhood of a, call it U, such that

$$f(x) > N \text{ for all } x \in U,$$

and we write

$$\lim_{x \to a} f(x) = \infty$$

in this case. For example, if we consider the function $f(x) = \frac{1}{x}$ at 0 then as we said above $f(0)$ is undefined. (Notice how I managed to say that without writing $\frac{1}{0}$. Ouch!) But we can take the limit as x gets close to 0 to see that

$$\lim_{x\to 0} \frac{1}{x} = \infty.$$

The table shows us why. We take values of x that are closer and closer to 0, *but do not equal* 0. The sequence of resulting values $f(x)$ gets larger.

x	$f(x) = \frac{1}{x}$
.1	10
.01	100
.001	1000
$\vdots$	$\vdots$

If we are given a small value for x, a number close to 0, then the number $\frac{1}{x}$ seems to get very large. If we are given a natural number N, then we can find a number $\frac{1}{N+1} = x$ very close to 0 such that $f(x) > N$. What is more, if x is closer to 0 than $\frac{1}{N+1}$ then $f(x) > N$. Thus, by the definition, $\lim_{x\to 0} \frac{1}{x} = \infty$. Observe how the limit replaces the undefined term $\frac{1}{0}$. (Another band-aide please.) This is an infinity associated with magnitude. This is an infinity that requires us to think of how tall, or how long, or how much of something there is. For this reason, it would seem that ∞ has little to do with the infinite cardinals $\aleph_0$, $\aleph_1$, $2^{\aleph_0}$.

The calculus also presents us with some other paradoxes associated with the infinite. Consider the fact that if we add three numbers together, then we can push them around as follows.

$$\begin{aligned} x + (y + z) &= (x + y) + z \\ x + y &= y + x. \end{aligned}$$

Then

$$1 - 1 + 1 = (1 - 1) + 1 = 0 + 1 = 1,$$

and

$$1 - 1 + 1 = 1 + 1 - 1 = (1 + 1) - 1 = 2 - 1 = 1.$$

We arrive at 1 in two different ways. But if the number of terms is infinite, problems will arise. Consider the sum

$$1 - 1 + 1 - 1 + 1 - \ldots.$$

Such infinite sums have to be treated with care. Just because we have written it down does not mean it equals a number! For example, if this sum exists then you might say that the sum should support any use of parentheses. However, when we try the following two groupings of the sum *we get different values*. By grouping using parentheses, we see that

$$\begin{aligned} 1 - 1 + 1 - 1 + 1 - 1 + \ldots &= (1-1) + (1-1) + (1-1) + \ldots \\ &= 0 + 0 + 0 + \ldots \\ &= 0, \end{aligned}$$

and by changing the order of terms, and then grouping we get

$$\begin{aligned} 1 - 1 + 1 - 1 + 1 - 1 + \ldots &= 1 + 1 - 1 + 1 - 1 + 1 - 1 + \ldots \\ &= 1 + (1-1) + (1-1) + (1-1) \ldots \\ &= 1 + 0 + 0 + \ldots \\ &= 1. \end{aligned}$$

So which is the answer, 0 or 1? One august Mathematician in the eighteenth century suggested that since 0 and 1 were answers that we could also include their average $\frac{1}{2}$. Today we say that the symbols

$$1 - 1 + 1 - 1 + 1 - 1 + 1 - \ldots$$

do not converge to a number. The sum is just as undefined as the unmentionable fraction $\frac{1}{0}$. (I think I need an aspirin for this pain.)

Another sum that has problems with the infinite is a *geometric sum*. The reader may have heard that

$$1 + x + x^2 + \ldots = \frac{1}{1-x} \tag{6.5}$$

for numbers x between -1 *and* 1. For instance, if we let $x = \frac{1}{2}$ in (6.5), then

$$1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots = \frac{1}{1-\frac{1}{2}} = 2$$

as we will prove in Chapter 7. However, as soon as we try the boundary value $x = -1$ in (6.5), we arrive at a familiar set of symbols that do not mean anything.

$$1 - 1 + 1 - 1 + \ldots = \frac{1}{1-(-1)} = \frac{1}{2}.$$

Maybe that august individual was not so far from the mark.

The sum (6.5) gives an absurd value when we try to use values beyond the boundaries set by Mathematics. With $x = 2$ in (6.5) we have our Mathematically sensitive mind injured by

$$1 + 2 + 2^2 + 2^3 + \ldots = \frac{1}{1-2} = -1.$$

You, dear reader, know that a sum of positive numbers will not be negative.

So what is the message here? The message is that we cannot talk about infinite quantities except in the presence of limits. If we are using infinite sums then we must be very careful to work within what is called the *radius of convergence.* This is the neighborhood in which the sum makes sense. If we take values x outside this neighborhood, the result will be Mathematical nonsense. For that reason we say that

$$1 + x + x^2 + x^3 + \ldots$$

is only defined for $-1 < x < 1$, and nowhere else.

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

Chapter 7

Inductions and Numbers

Let us consider the well-ordered set $\mathbb{N}$ from a new perspective. We used two properties of $\mathbb{N}$ to define well-ordered sets. They were the *Trichotomy Property* and the *Minimum Property.* It is accepted by Mathematicians that these properties together with the *Principle of Mathematical Induction* will give us all a common intellectual picture of the natural numbers. At this stage of intellectual development, the professionals agree that everyone who reads these statements will envision the same set of natural numbers that are right this minute dancing in your head. Furthermore, these principles can be extended to well-ordered sets to give us a new and powerful tool or argument about well-ordered sets. That new tool is called *Transfinite Induction.* Transfinite Induction is to well-ordered sets what Mathematical Induction is to the natural numbers.

7.1 Mathematical Induction

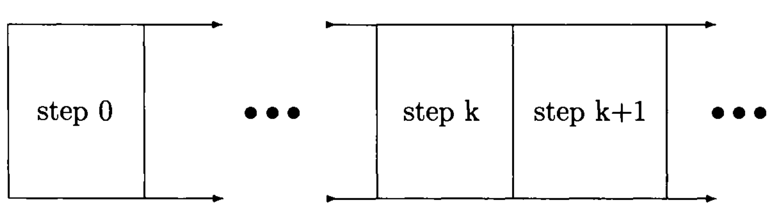
We begin with a discussion of the *Principle of Mathematical Induction* or more briefly Mathematical Induction. Intuitively, Mathematical Induction allows us to make a statement for all of the natural numbers by knowing only that the statement holds at 0 and then the statement holds at $n+1$ if it holds at n. For instance, suppose you are playing a board game (a thought experiment) whose object is to move through all of the playing spaces numbered $0, 1, 2, 3, \ldots$. You move from space to space by rolling a die whose every face has a single black spot • on it. You are on space 0. Next, to get from space 100 to 101, you roll the die, and move one space. In general, to get from the space numbered n to the space numbered $n + 1$, you roll the die, and move one space. This is how Mathematical Induction works. You have a starting place and you are told how to get from place to place or, more precisely, from natural number to natural number. A more Mathematical discussion follows.

The Principle of Mathematical Induction: Suppose we are given a subset $X \subset \mathbb{N}$. Then

$$\boxed{\begin{array}{c} X = \mathbb{N} \text{ if } X \text{ satisfies the properties} \\ 0 \in X, \text{ and given } k \in X, \text{ then } k+1 \in X. \end{array}}$$

You may have encountered this principle in your high school Algebra course, although you may not have given it a name. The Principle of Mathematical Induction tells us that if we have a process P that can be done in the first place (we say that $P(0)$ *is True*), and if we can show how to proceed from some statement $P(k)$ to the statement $P(k+1)$, then we can infer that $P(n)$ is True for each $n \in \mathbb{N}$. That is one efficient way to prove something for infinitely many natural numbers. You can prove that infinitely statements are True in exactly three steps.

Here is an illustration of how Mathematical Induction works. Suppose that we have an infinite ladder whose first step is numbered with 0. Please allow me to use ladders with steps instead of rungs.



We are on the first step of the ladder, and we have a set of instructions that shows us how to get from one step on the ladder to the next in a general way. These directions will not be of the form

Move from step 0 to step 1 and then to step 2.

Such directions will work for a finite ladder, one having, say, 5 steps, but it fails to help us climb a ladder with an infinite number of steps. Rather, the directions should read like this.

If you are on some step, then here is how to get to the next step.

This set of directions ensures us that we will be able to traverse every step of this infinite ladder. In elementary Mathematics, these instructions are usually a set of equations. In computer programming, this process is the set of instructions in a loop without a stop command.

The only way to find some kind of intuition surrounding Mathematical Induction is to read several examples.

Example 7.1.1 The following algorithm writes all of the natural numbers $n \in \mathbb{N}$.

1. Let $k = 0$.
2. Write k.
3. Increase k to $k + 1$.
4. Go to step 2.

We will prove that the claim of this algorithm is True using Mathematical Induction.

The *Initial Step* is to check that 0 is written on the first pass. One pass through this algorithm, that is, one reading of lines 1 through 4, confirms that step 2 will write 1 in step 2. Then k is increased from 1 to 2.

The *Induction Hypothesis*: Assume that k is written on the $k+1$-st pass through the algorithm.

The next step is the *Induction Step*. Assuming the Induction Hypothesis, we must show that $k + 1$ is written on the $(k + 2)$-nd pass. So using the Induction Hypothesis, assume that k is written on the $k + 1$-st pass. Step 3 of the algorithm increases k to $k + 1$ so that on the next pass (i.e., on pass number $k + 2$) step 2 of the algorithm writes $k + 1$. This is what we had to prove.

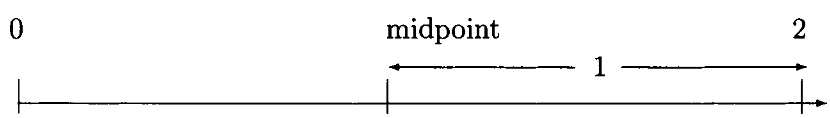
Then by *Mathematical Induction*, we conclude that for each $n \in \mathbb{N}$, the algorithm writes n on its $n + 1$-st pass.

Notice what we did. We predicted the behavior of the algorithm through *all* of its potentially infinitely many passes knowing only how it behaves *at some* point k. While physical restrictions limit a computer to print only finitely many integers using this algorithm, Mathematical Induction allows us to know the behavior of this algorithm through *any number* of passes. Thus, no matter how fast or powerful computers will be, before the computer can execute it, we will know precisely what this algorithm does on any pass.

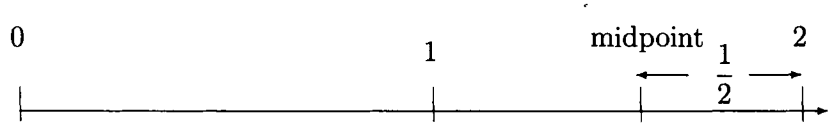
Here is an example of how Mathematical Induction can replace hand waving in a Mathematical discussion. Recall that on page 128 we watched a gentleman take a curious walk across the lobby in Hilbert's Infinite Hotel. He moved first one meter, and then $\frac{1}{2}$ a meter, and in general, on his nth step he moved $\frac{1}{2^{n-1}}$ meters.

Example 7.1.2 We use Mathematical Induction to show that when the curious gentleman has completed his n-th step, he is $\frac{1}{2^{n-1}}$ meters from the end of his walk.

Proof: Begin with the open interval $(0, 2)$ of numbers properly between 0 and 2. Mark the midpoint of the open interval. The segment yet to be crossed (that is, the one bordering on 2) has length 1 as in the picture.



Take the interval boardering on 2 and find its midpoint. Continue this process *inductively*. That is, if we have an open interval bordering on 2, then find its midpoint and find the half that borders on 2. For example, in the second step, we will find the midpoint of the chosen interval and choose the half bordering on 2 as in the picture.



The segment yet to be crossed has length $\frac{1}{2}$.

We will prove using Mathematical Induction that:

In step $n+1$, the length of the chosen segment is $\frac{1}{2^n}$.

Initial Step: We are starting with $1 = 0 + 1$. On the 1st step, the chosen segment has length $1 = 0 + 1 = \frac{1}{2^0}$. Then our formula for length works for $n = 0$.

Induction Hypothesis: We will assume that in the $(k+1)$st step of our process, the segment chosen has length $\frac{1}{2^k}$.

Induction Step: By the Induction Hypothesis, in the $(k+1)$st step of the man's walk, he chooses a segment of length $\frac{1}{2^k}$. In the next step, the $(k+2)$nd step, we take half of the given segment. The result is a segment of length $\frac{1}{2}\frac{1}{2^k} = \frac{1}{2^{(k+1)}}$. Then the segment in step $k+2 = (k+1)+1$ has length $\frac{1}{2^{(k+1)}}$. This is what we had to prove.

By Mathematical Induction, we conclude that the segment in step $n+1$ has length $\frac{1}{2^n}$ for each $n \in \mathbb{N}$.

By knowing how long the segment traversed by the curious gentleman on his k-th step, we were able to determine how long *every* step taken was.

Example 7.1.3 Another identity using $\frac{1}{2}$ is the subject of the next induction argument. We will show that

$$2 - \frac{1}{2^n} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^n}$$

is True for each $n \in \mathbb{N}$.

Proof: *Initial Step:* In step 0, check that

$$2 - \frac{1}{2^0} = 1 = \frac{1}{2^0}.$$

Induction Hypothesis: Assume that we have shown that

$$2 - \frac{1}{2^k} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^k}$$

for *some* $k \in \mathbb{N}$. (This is True since we have proved it True for at least the value $k = 0$.)

Induction Step: We must show that

$$2 - \frac{1}{2^{(k+1)}} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^k} + \frac{1}{2^{(k+1)}}.$$

We begin by grouping the first k terms in the sum.

$$1 + \frac{1}{2} + \cdots + \frac{1}{2^k} + \frac{1}{2^{(k+1)}} = \left(1 + \frac{1}{2} + \cdots + \frac{1}{2^k}\right) + \frac{1}{2^{(k+1)}}.$$

By the Induction Hypothesis and a small bit of algebra, we can write

$$\begin{aligned}
\left(1 + \frac{1}{2} + \cdots + \frac{1}{2^k}\right) + \frac{1}{2^{(k+1)}} &= \left(2 - \frac{1}{2^k}\right) + \frac{1}{2^{(k+1)}} \\
&= \left(2 - \frac{2}{2^{(k+1)}}\right) + \frac{1}{2^{(k+1)}} \\
&= 2 - \frac{1}{2^{(k+1)}}
\end{aligned}$$

after a little bit of arithmetic. This shows us that

$$2 - \frac{1}{2^{(k+1)}} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^{(k+1)}},$$

which completes the Inductive Step.

We conclude by Mathematical Induction that

$$2 - \frac{1}{2^n} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^n}$$

for each $n \in \mathbb{N}$. This completes the proof.

We can use the above example to prove that the curious gentleman in Hilbert's Infinite Hotel covers 2 meters when he has finished walking all of the steps possible. On the man's $(n + 1)$st step he has travelled

$$1 + \frac{1}{2} + \cdots + \frac{1}{2^n}$$

meters. The distance between 2 and this intermediate position is

$$2 - \left(1 + \frac{1}{2} + \cdots + \frac{1}{2^n}\right).$$

By the above example, this can be simplified to

$$2 - \left(2 - \frac{1}{2^n}\right) = \frac{1}{2^n}.$$

Then 2 and his position x after infinitely steps must be separated by less than $\frac{1}{2^n}$ for each n. That is,

$$2 - \frac{1}{2^n} < x \leq 2$$

for each $n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} 2 - \frac{1}{2^n} \leq x \leq 2.$$

However,

$$\lim_{n\to\infty} 2 - \frac{1}{2^n} = 2 - 0 = 2$$

so that

$$2 = \lim_{n\to\infty} 2 - \frac{1}{2^n} \leq x \leq 2.$$

Thus, his final position x must be equal to 2.

An application of Mathematical Induction will show that the cleaning lady Maria who cleaned all of the rooms in Hilbert's Infinite Hotel does indeed clean each of the rooms in a total of 2 hours.

Example 7.1.4 Recall that there is a cleaning lady Maria, who cleans room 1 in 1 hour, and cleans out each successive room in Hilbert's Infinite Hotel in half the time it takes to clean the previous room. That is, if Maria takes x hours to clean room n, then she takes $\frac{1}{2}x$ hours to clean room $n+1$. We will show that Maria cleans all of the rooms in Hilbert's Infinite Hotel, and that she cleans room n in $\frac{1}{2^n}$ hours.

Initial Step: Maria cleans room 1. I know because I stayed there. The room was spotless. By design she takes only 1 hour to clean room 1.

Induction Hypothesis: Assume that Maria has cleaned a room, say, room k, in $\frac{1}{2^{k-1}}$ hours.

Induction Step: Maria has cleaned room k by the induction process. She moves to room $k+1$ and proceeds to clean it. Then room $k+1$ is cleaned. She cleans room $k+1$ in half the time it took to clean the previous room k. It takes $\frac{1}{2}\frac{1}{2^{k-1}} = \frac{1}{2^k}$ hours to clean that room. This is what we had to prove.

We conclude by Mathematical Induction that Maria cleans all of the rooms at Hilbert's Infinite Hotel, and that she cleans room n in $\frac{1}{2^{n-1}}$ hours. This ends the Mathematical Induction proof.

Let us find the total time it takes Maria to complete her work. If we reread Example 7.1.4, we see that Maria takes $\frac{1}{2^{n-1}}$ hours to clean room n completely. We conclude then, that the time taken for Maria to clean each of the first $n+1$ rooms is

$$1 + \frac{1}{2} + \frac{1}{2^2} + \cdots \frac{1}{2^n}.$$

According to the previous example,

$$1 + \frac{1}{2} + \frac{1}{2^2} + \cdots \frac{1}{2^n} = 2 - \frac{1}{2^n}.$$

It follows that as Maria cleans more rooms the difference between 2 hours and Maria's working time is

$$2 - \left(2 - \frac{1}{2^n}\right) = \frac{1}{2^n}.$$

Thus, in a Mathematically precise way, we see that when Maria has completed her rounds (when she has cleaned room $n+1$ for all natural numbers $n \in \mathbb{N}$), the difference between 2 hours and her working time is smaller than $\frac{1}{2^n}$ for all $n \in \mathbb{N}$. But 0 is the only nonnegative number smaller than each o f the fractions $\frac{1}{2^n}$ for all

$n \in \mathbb{N}$. We conclude that after Maria has completed her cleaning duties, she has worked exactly 2 hours.

Here is another example that uses the number 2 in a clever way.

Example 7.1.5 We will show that

$$1 + 2 + 2^2 + \cdots + 2^n = 2^{n+1} - 1 \text{ for each } n \in \mathbb{N}.$$

Initial Step: $1 = 2^0$ so that

$$2^0 = 2^{0+1} - 1 = 1,$$

providing us with the Initial Step.

Induction Hypothesis: We assume that we have proved that

$$1 + 2 + 2^2 + \cdots + 2^k = 2^{k+1} - 1 \text{ for some } k \in \mathbb{N}.$$

(Again, we did so for $k = 0$.)

Inductive Step: We must show that

$$1 + 2 + 2^2 + \cdots + 2^{k+1} = 2^{k+2} - 1,$$

where k is the integer chosen in the *Induction Hypothesis.*

A little algebra and the Induction Hypothesis shows us that

$$\begin{aligned} 1 + 2 + 2^2 + \cdots + 2^k + 2^{k+1} &= \left(1 + 2 + 2^2 + \cdots + 2^k\right) + 2^{k+1} \\ &= 2^{k+1} + 2^{k+1} - 1 \\ &= 2 \cdot 2^{k+1} - 1 \\ &= 2^{k+1+1} - 1 \\ &= 2^{k+2} - 1. \end{aligned}$$

This is what we had to prove.

Therefore, by Mathematical Induction, we have proved that

$$1 + 2 + 2^2 + \cdots + 2^n = 2^{n+1} - 1 \text{ for each } n \in \mathbb{N}.$$

Now here is another Mathematically imprecise way to justify the identity given in the above example. Fix $n \in \mathbb{N}$ and use your old friend foil to find the product

$$(2 - 1)(1 + 2 + \cdots + 2^{n-1} + 2^n).$$

Do not simplify this expression as that will make our calculations useless. Then

$$\begin{aligned}
1+2+\cdots+2^n &= (2-1)(1+2+\cdots+2^n) \\
&= \left\{ \begin{array}{rcccccl} 2(1 & + & 2 & + & \cdots & + & 2^n) \\ -1(1 & + & 2 & + & \cdots & + & 2^n) \end{array} \right. \\
&= \left\{ \begin{array}{ccccccccc} & & 2 & + & \cdots & + & 2^n & + & 2^{n+1} \\ -1 & - & 2 & - & \cdots & - & 2^n & & \end{array} \right. \\
&= -1+(2-2)+\cdots+(2^n-2^n)+2^{n+1} \\
&= -1+2^{n+1}.
\end{aligned}$$

Therefore

$$1+2+\cdots+2^n = 2^{n+1}-1,$$

and we have proved the identity in the above example. The reader might question why this argument is Mathematically imprecise. The answer lies in the way we stacked the powers of 2. In the above equations, we lined up the powers of 2 as follows.

$$\begin{array}{ccccccccccc} & & 2 & + & 2^2 & + & \cdots & + & 2^n & + & 2^{n+1} \\ -1 & - & 2 & - & 2^2 & - & \cdots & - & 2^n & & \end{array}.$$

We then implicitly declared that *since this stacking of powers of 2 works for the first 2 powers of 2 it works for all powers of 2.* This is the weakness in the argument. There is no Mathematical law that says that a pattern that occurs twice will occur in all places. There is no way to imply from the first two places that the stacking will take place in all places. No way, that is, unless we use the Principle of Mathematical Induction.

You might be suggesting right now that the pattern is obvious. We should be able to imply a Mathematical Truth from the obvious nature of the pattern. I offer the following numerical pattern and ask the reader to fill in the value x in the pattern.

$$1, 2, 3, x.$$

Most readers will choose $x = 4$ as the next obvious value. Some might choose $x = 5$ since the (obvious) pattern could be that the next number is the sum of the previous two.

Another example is the sequence

$$2, 4, 8, x$$

with which we ask for the next number x. Some would say that the pattern is

$$2^1, 2^2, 2^3, 2^4$$

so that $x = 2^4 = 16$. Others might claim that the pattern is more recursive and conclude that the next number is the product of the previous two.

$$2, 4, 2 \cdot 4, 4 \cdot 8 = 32$$

is the most obvious pattern, and so $x = 32$ is the next value. *So which pattern is the most obvious?*

The answer is that *neither is more obvious.* In fact, it has been proved that the three numbers $1, 2, 3$ and $2, 4, 8$ satisfy *infinitely many patterns*, and so no pattern is the obvious pattern. We only choose the *most familiar* patterns. The most obvious pattern cannot be found because there is no way to judge *obviousness* for a pattern from the infinitely many patterns available. Thus, we should not look at three numbers and then quickly point out some pattern as the *only* obvious correct one.

Here is an open question for the reader. Suppose I give you three natural numbers

$$2, 3, 5$$

and ask you for the fourth one x. What pattern did I have in mind? Remember that there are many values for x because there are many patterns whose first three values are $2, 3, 5$. I had $x = 8$ in mind. Why is that? Another equally legitimate number would be $x = 7$. Why did I choose these numbers, reader? I leave the first to your imagination. The value $x = 7$ was chosen because it is the next prime number. $2, 3, 5, 7, \ldots$ could have been the implied list of prime numbers.

Let us consider an old chestnut as an example of Mathematical Induction in Mathematics. We previously examined this example on page 124.

Example 7.1.6 Let $n \in \mathbb{N}$. Then

$$0 + 1 + \ldots + n = \frac{n(n+1)}{2}.$$

Proof: As the *Initial Step* of this proof, we observe that

$$0 = \frac{0(0+1)}{2}.$$

Then the formula is True for 0.

Induction Hypothesis: Assume that you have proved the result for *some* natural number $k \geq 0$, and in fact you have. You have proved it for the natural number 0. In this example, our *Induction Hypothesis* is

$$0 + \ldots + k = \frac{k(k+1)}{2}.$$

We can begin the *Induction Step*. We must give a general argument that shows how the formula for $k+1$ can be derived from the *Induction Hypothesis*. That is, we must prove that

$$0 + \ldots + k + (k+1) = \frac{(k+1)((k+1)+1)}{2} = \frac{(k+1)(k+2)}{2}.$$

The sum $0 + \ldots + k + (k+1)$ can be regrouped as

$$(0 + \ldots + k) + (k+1),$$

so by the Induction Hypothesis

$$(0 + \ldots + k) + (k+1) = \frac{k(k+1)}{2} + (k+1).$$

Next, a little algebra shows us that

$$\frac{k(k+1)}{2} + (k+1) = \frac{k(k+1) + 2(k+1)}{2} = \frac{(k+2)(k+1)}{2},$$

or equivalently that

$$0 + \ldots + k + (k+1) = \frac{(k+1)(k+2)}{2}.$$

This is what we had to show.

The most important step in the proof is this last one. By Mathematical Induction, we conclude that

$$0 + \ldots + n = \frac{n(n+1)}{2}$$

is True for each $n \in \mathbb{N}$.

You might ask why we are interested in Mathematical Induction. After all, you might say, there is nothing wrong with the argument we gave on page 124. There is a subtlety that can give us trouble in more general settings. The idea is this. We have written the argument so that the numbers on page 124 match up perfectly as in the following expression.

$$\begin{array}{ccccccccccc} S & = & 1 & + & 2 & + & \dots & + & \ell - 1 & + & \ell \\ S & = & \ell & + & \ell - 1 & + & \dots & + & 2 & + & 1 \end{array}.$$

We know that the sum of the first and last column entries in this expression is $\ell + 1$. *We assumed* that the sum on any column in this expression is $(k) + (\ell - k + 1) = \ell + 1$. The sum given is certainly correct, but without an argument involving Mathematical Induction, we do not know that the terms in column ℓ that we are dealing with are k and $\ell - k + 1$. There must be some concrete Mathematical reason why the columns in this expression match up as they do, and there is no hint of that reason in the argument we gave on page 124. The argument above using Mathematical Induction provides us with a bridge for the gap in our earlier argument. Mathematical Induction gives us a neat way of jumping from the fact that the first column adds up to $\ell + 1$ to the Truth that *each* column adds up to $\ell + 1$. This is exactly the type of Mathematical fact on which Mathematical Induction is designed to work.

Here is another algebraic application of Mathematical Induction. Recall that a natural number n is an *odd number* if $n = 2k - 1$ for some $k = 1, 2, 3, \ldots$. In general, $2k - 1$ with $k = 1, 2, 3, \ldots$ is the kth odd number. Then $2(1) - 1 = 1$ is the first odd number, $2(10) - 1 = 19$ is the 10th odd number, and $2(101) - 1 = 201$ is the 101st odd number.

Galileo (circa 1520) noticed that the sum of the first n odd numbers is the nth perfect square. Specifically, he noticed that if he wrote down the first 2 odd prime numbers that they added up to 2^2, if he wrote down the first 3 odd prime numbers that they added up to 3^2, and if he wrote down the first 4 odd prime numbers that they added up to 4^2.

$$\begin{aligned} 2^2 &= 1 + 3, \\ 3^2 &= 1 + 3 + 5, \\ 4^2 &= 1 + 3 + 5 + 7. \end{aligned}$$

He then concluded that

$$n^2 = 1 + 2 + 3 + \ldots + (2n - 1)$$

but without proof. We will use Mathematical Induction to show that this identity is the case for all natural numbers $n > 0$.

Example 7.1.7 Let $n > 0 \in \mathbb{N}$. Then n^2 is the sum of the first n odd numbers.

Proof: *Initial Step:* By replacing n with 1, we see that

$$1^2 = 1 = 2 \cdot 1 - 1.$$

The Initial Step has been established.

Induction Hypothesis: Assume that $k^2 = 1 + \ldots + (2k - 1)$ for some natural number $k > 0$.

Induction Step: Show that $(k+1)^2 = 1 + \ldots + \ (2(k+1) - 1)$.

Begin by grouping the sum of $k + 1$ terms.

$$\begin{aligned} & 1 + \ldots + \ (2k-1) + \ (2(k+1) - 1) \\ = \ & [1 + \ldots + \ (2k-1)] + \ (2k+1). \end{aligned}$$

The Induction Hypothesis and a little algebra will show that

$$\begin{aligned} [1 + \ldots + (2k-1)] + (2k+1) &= k^2 + (2k+1) \\ &= (k+1)^2. \end{aligned}$$

Then

$$(k+1)^2 = [1 + \ldots + \ (2k-1)] + \ (2(k+1) - 1),$$

which is what we had to prove.

Therefore, by Mathematical Induction,

$$n^2 = 1 + 3 + 5 + \ldots + \ (2n - 1)$$

for all natural numbers $n \geq 1$. This completes the proof.

Here is another story in Mathematics that shows us that patterns are not always as obvious as they first appear. The equation

$$ax + b = 0 \text{ and } a \neq 0$$

can be solved by even the youngest of students. It is important to note that the exponent of x is 1 in this equation. Next, the high school student knows that the *quadratic formula*

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{7.1}$$

will solve the quadratic polynomial equation

$$ax^2 + bx + c = 0 \text{ and } a \neq 0.$$

The solution features a square root (a radical) $\sqrt{y}$ and the coefficients a, b, c of the polynomial equation. Notice that the highest power of x that occurs is 2. We say that the expression (7.1) is a *solution by radicals of the polynomial equation of degree 2.*

In about 1535 AD, the Italian Mathematician Tartaglia showed that the equation

$$ax^3 + bx^2 + cx + d = 0 \text{ and } a \neq 0$$

can be solved by radicals. His solution featured cube roots $\sqrt[3]{y}$, square roots, and the coefficients of the polynomial equation. Since this polynomial equation has degree 3, Tartaglia's solution is said to be *a solution by radicals of the polynomial equation of degree 3.*

Tartaglia's solution was published in 1545 in Cardano's book *Ars Magna.* The *Ars Magna* also contained the first publication of the solution to the polynomial equation

$$ax^4 + bx^3 + dx^2 + ex + f = 0 \text{ and } a \neq 0.$$

The solution to this polynomial equation of degree 4 also featured fourth roots $\sqrt[4]{y}$, cube roots, square roots, and the coefficients of the polynomial equation. Thus, we have solutions by radicals for polynomial equations of degrees 1, 2, 3, and 4. Is there a pattern here?

The above history may seem to indicate a pattern. If polynomial equations of degree 1, 2, 3, 4 can be solved by radicals, we might then suggest that it is obvious that the fifth degree polynomial equation

$$ax^5 + bx^4 + dx^3 + ex^2 + fx + g = 0 \text{ and } a \neq 0$$

has a solution by radicals. Unfortunately, the Mathematical gods are playing games with us. The supposed pattern is misleading. In 1821, a 21 year old Norwegian Mathematician, Niels Abel, showed that there is no solution *by radicals* to the polynomial equation of degree 5. In fact, an 18 year old French Mathematician, Evariste Galois, showed that except for these first four degrees, the general polynomial equation is not solvable by radicals like the first four are. So what happened to our obvious pattern? It must be that there was never a pattern to find. The pattern evaporated. It disappeared. *It was never there.* This is an important thought. Mathematicians must be careful in deciding when a pattern exists, and when it does not exist. This is not as easy as it might seem since most people seem to see patterns everywhere. The best defense against seeing phantom patterns is to use a good solid argument that includes Mathematical Induction.

The next story concerns a misuse of Mathematical Induction. This example elegantly demonstrates the importance of the *Initial Step* in a proof using Mathematical Induction. The story is called **The Unexpected Termination.**

Example 7.1.8 A company Boss wants to fire a Supervising Barber who works in the Hair Styling Branch of the company. The firing will take place at 9:00 AM sometime between Sunday and Saturday of next week. However, since the Boss does not want the Barber to anxiously await his last day, he decides not to tell the man which day will be his last. The Barber catches wind of this compassion, and decides to use it to his advantage.

The Barber decides that he will not be terminated. His argument is as follows. He reasons that if he has not heard about his termination by 9 AM, Friday then he most certainly will not be terminated the next day, Saturday. This is because the Boss must fire this Barber at 9 AM, Saturday. He would then be upset all day Friday while he waited for the axe to fall on Saturday. The compassionate Boss does not want the Barber to suffer in that way, so the Barber concludes that he will not be terminated Saturday. He reasons that he can use this fact as the initial step in a Mathematical Induction.

Suppose that the Barber has argued that he will not be terminated on day $k - 1$. (Think of Saturday as day 1, Friday as day

2, and so on to Sunday which is Day 7.) He reasons about Day k as follows. If he has not heard about his termination by 9 AM of Day $k+1$ then he most certainly will not be terminated the next day, Day k, because in that case the Barber would be upset all day long Day $k+1$ contrary to the compassionate Boss's wishes. Thus the Barber concludes that he will not be terminated on Day k. The Barber concludes by Mathematical Induction that he will not be terminated on any day n that week. Unfortunately, the Boss terminates the Barber at 9 AM, Wednesday. The Barber's argument must have been flawed, but where is that fatal flaw?

Let us examine that argument in detail. The Barber will not be terminated Saturday for the reasons given in the Induction Step. So maybe Friday is the termination day. If so then Thursday morning at 10 AM he will know that he is not to be terminated Thursday, and so a Friday termination is the only thing possible. He would then fret, contrary to the compassionate Boss' wishes. He concludes that Friday is not the day. Will Thursday be his termination day? No. The same argument applies. Try it yourself. We then use the familiar dots, ..., to conclude that the Barber will not be terminated that week.

So where is the error in that argument? I leave it to you to think about The Unexpected Termination for a while. Try not to peek at the answer before you work with the problem a little.

As we said earlier the answer concerns the Initial Step. His initial assumption is that he has a job on Friday. This renders the rest of the argument müte because if he has a job Friday then *he has not been terminated at all.* Thus he has assumed that which he wants to prove, and there is his mistake. If you want to prove that you will not be terminated, then you cannot make an assumption that assumes that you will not be terminated.

We hope that these examples have helped you to understand the process called Mathematical Induction. Recapping, we begin with a process P that is defined at each $n \in \mathbb{N}$. Induction begins by showing that the process is True for some intial value, say, 0. In the *Induction Hypothesis* we assume that $P(k)$ is True for some $k \in \mathbb{N}$, and in the *Induction Step* we show how $P(k)$ leads us to $P(k+1)$. We then conclude by Mathematical Induction that $P(n)$ is True for all $n \in \mathbb{N}$.

Exercise 7.1.9 Use Mathematical Induction to show the following.

1. $1^2 + 2^2 + \ldots + n^2 = \frac{1}{6}n(n+1)(2n+1)$ for each $n \in \mathbb{N}$.
2. $1^3 + 2^3 + \ldots + n^3 = \frac{1}{4}n^2(n+1)^2$ for each $n \in \mathbb{N}$.
3. $(1-x)(1+x+x^2+\ldots+x^n) = 1 - x^{n+1}$ for each $n \in \mathbb{N}$.

7.2 Sums of Powers of Integers

In the previous section, we showed that $\sum_{\ell=1}^{n} \ell = \frac{1}{2}n(n+1)$. In this formula, $q_1(n) = \frac{1}{2}n(n+1)$ is a polynomial of degree $1+1$. The first homework problem shows that $\sum_{\ell=1}^{n} \ell^2$ possesses a formula $q_2(n)$ that is a polynomial of degree $2+1$, and similarly $\sum_{\ell=1}^{n} \ell^3$ has a formula $q_3(n)$ that is a polynomial of degree $3+1$. In this section, we will use Mathematical Induction to show that for each exponent $k \in \mathbb{N}$, the sum $\sum_{\ell=1}^{n} \ell^k$ has a polynomial formula $q_k(n)$ of degree $k+1$.

We begin with telescoping sums.

Lemma 7.2.1 *Let $f : \mathbb{R} \to \mathbb{R}$ be any function, and let $n > 0 \in \mathbb{N}$. Then*

$$\sum_{\ell=1}^{n} f(\ell+1) - f(\ell) = f(n+1) - f(1).$$

Proof: The proof is by Mathematical Induction on n.

Initial Step: Let $n = 1$. Then

$$\sum_{\ell=1}^{1} f(\ell+1) - f(\ell) = f(1+1) - f(1),$$

which proves the Initial Step.

Induction Hypothesis: Suppose that $\sum_{\ell=1}^{k} f(\ell+1) - f(\ell) = f(k+1) - f(1)$ for some $k > 1 \in \mathbb{N}$.

Induction Step: We must show that $\sum_{\ell=1}^{k+1} f(\ell+1) - f(\ell) = f(k+2) - f(1)$. Arithmetic allows us to write

$$\sum_{\ell=1}^{k+1} f(\ell+1) - f(\ell) = (f(k+2) - f(k+1)) + \sum_{\ell=1}^{k} f(\ell+1) - f(\ell),$$

and the Induction Hypothesis shows us that

$$\sum_{\ell=1}^{k} f(\ell+1) - f(\ell) = f(k+1) - f(k).$$

Hence,

$$\begin{aligned}
\sum_{\ell=1}^{k+1} f(\ell+1) - f(\ell) &= (f(k+2) - f(k+1)) + \sum_{\ell=1}^{k} f(\ell+1) - f(\ell) \\
&= (f(k+2) - f(k+1)) + (f(k+1) - f(1)) \\
&= f(k+2) - f(1),
\end{aligned}$$

which is what we had to show.

Then $\sum_{\ell=1}^{n} f(\ell+1) - f(\ell) = f(n+1) - f(1)$ for all $n > 0 \in \mathbb{N}$ by Mathematical Induction.

Notice that with this lemma, we deduce that

$$(2^5 - 1^5) + (3^5 - 2^5) + \cdots + (101^5 - 100^5) = 101^5 - 1^5$$

with Mathematical certainty. In one crude argument, you cancel 2^5 and 3^5 from the above sum. However, some kind of Mathematical reason should accompany your cancellation of 57^5, even if it means writing down 99 theorems in order to justify that cancellation.

This telescoping sum is used to prove the following result. Recall from your high school Math courses that $0! = 1$, that $n! = 1 \cdot 2 \cdots n$, and that

$$\binom{n}{\ell} = \frac{n!}{\ell!(n-\ell)!}$$

for all $n \geq \ell > 0 \in \mathbb{N}$. Then the Binomial Theorem states that for each $n > 0 \in \mathbb{N}$, we have

$$(x+1)^n = \sum_{\ell=0}^{n} \binom{n}{\ell} x^\ell = \binom{n}{0} + \binom{n}{1} x + \cdots + \binom{n}{n} x^n.$$

These are the prerequisites for the following result.

Theorem 7.2.2 *Let $k, n > 0$ be integers. Then*

$$\sum_{\ell=1}^{n} \ell^k = q_k(n)$$

for some polynomial $q_k(n)$ of degree $k+1$ and whose coefficients are rational numbers.

Proof: Let $k > 0 \in \mathbb{N}$ be a fixed exponent, and fix $n > 0 \in \mathbb{N}$ as the number of terms in a sum $\sum_{\ell=1}^{n}$. The proof is by Mathematical Induction on k.

Initial Step: The sum $\sum_{\ell=1}^{n} \ell^0 = n$ because it is the sum of n copies of 1. Observe that $q_0(n) = n$ is a polynomial of degree $0+1$ whose coefficients are rational numbers.

Induction Hypothesis: Suppose that there is a given integer $k - 1 > 0$ such that for each exponent $0 \le j \le k-1$, we have $\sum_{\ell=1}^{n} \ell^j = q_j(n)$ for some polynomial $q_j(n)$ of degree $j+1$, whose coefficients are rational numbers.

Induction Step: Let $k \in \mathbb{N}$ and consider the sum $\sum_{\ell=1}^{n} \ell^k$. By Lemma 7.2.1, the telescoping sum satisfies

$$\sum_{\ell=1}^{n} (\ell+1)^{k+1} - \ell^{k+1} = (n+1)^{k+1} - 1^{k+1} = (n+1)^{k+1} - 1. \quad (7.2)$$

Moreover, by the Binomial Theorem,

$$\begin{aligned} (\ell+1)^{k+1} - \ell^{k+1} &= \left(\sum_{j=0}^{k+1} \binom{k+1}{j} \ell^j \right) - \ell^{k+1} \\ &= \sum_{j=0}^{k} \binom{k+1}{j} \ell^j. \end{aligned}$$

Notice the cancellation of ℓ^{k+1} in the previous sum. Then the telescoping sum (7.2) also looks like

$$\sum_{\ell=1}^{n} (\ell+1)^{k+1} - \ell^{k+1} = \sum_{\ell=1}^{n} \sum_{j=0}^{k} \binom{k+1}{j} \ell^j.$$

By commuting the $\sum$ and by factoring the number $\binom{k+1}{j}$ from the sum $\sum_{\ell=1}^{n}$, we have

$$\begin{aligned}
\sum_{\ell=1}^{n}\sum_{j=0}^{k}\binom{k+1}{j}\ell^j &= \sum_{j=0}^{k}\sum_{\ell=1}^{n}\binom{k+1}{j}\ell^j \\
&= \sum_{j=0}^{k}\left(\binom{k+1}{j}\sum_{\ell=1}^{n}\ell^j\right) \qquad (7.3)\\
&= \binom{k+1}{k}\sum_{\ell=1}^{n}\ell^k+\sum_{j=0}^{k-1}\left(\binom{k+1}{j}\sum_{\ell=1}^{n}\ell^j\right).
\end{aligned}$$

Then by combining (7.2) and (7.3), we have demonstrated that

$$\begin{aligned}
(n+1)^{k+1}-1 &= \sum_{\ell=1}^{n}(\ell+1)^{k+1}-\ell^{k+1} \qquad (7.4)\\
&= \binom{k+1}{k}\sum_{\ell=1}^{n}\ell^k+\sum_{j=0}^{k-1}\left(\binom{k+1}{j}\sum_{\ell=1}^{n}\ell^j\right).
\end{aligned}$$

Now, given $k-1 \geq j > 0$, then by the Induction Hypothesis,

$$\sum_{\ell=1}^{n}\ell^j = q_j(n)$$

for some polynomial $q_j(n)$ of degree $j+1$ whose coefficients are rational numbers, and as a formal symbol, let

$$\sum_{\ell=1}^{n}\ell^k = q_k(n).$$

By substituting these polynomials $q_j(n)$ into (7.4), we have the polynomial equation

$$\begin{aligned}
(n+1)^{k+1}-1 &= \binom{k+1}{k}q_k(n)+\sum_{j=0}^{k-1}\left(\binom{k+1}{j}\sum_{\ell=1}^{n}\ell^j\right)\\
&= \binom{k+1}{k}q_k(n)+\left(\binom{k+1}{0}q_0(n)+\right.\\
&\qquad \left.\binom{k+1}{1}q_1(n)+\cdots+\binom{k+1}{k-1}q_{k-1}(n)\right).
\end{aligned}$$

By solving this equation for $q_k(n)$, we we see that

$$q_k(n) = \frac{1}{\binom{k+1}{k}}\Bigg((n+1)^{k+1} - 1$$
$$-\binom{k+1}{0}q_0(n) - \binom{k+1}{1}q_1(n) - \cdots - \binom{k+1}{k-1}q_{k-1}(n)\Bigg).$$

By the Induction Hypothesis, $q_0(n), \ldots, q_{k-1}(n)$ are polynomials of degree at most k, and whose coefficients are rational numbers. Therefore, because $(n+1)^{k+1}$ has degree $k+1$, $q_k(n)$ is a polynomial of degree $k+1$ whose coefficients are rational numbers.

Then by Mathematical Induction, $\sum_{\ell=1}^{n} \ell^k = q_k(n)$ where $q_k(n)$ is some polynomial of degree $k+1$ whose coefficients are rational numbers. This completes the proof.

Exercise 7.2.3 1. Using the formula for $q_k(n)$ in Theorem 7.2.2 to find $q_4(n)$.

2. Using the formula for $q_5(n)$ in Theorem 7.2.2 to find $q_5(n)$.

3. Using the formula for $q_k(n)$ in Theorem 7.2.2 to find $q_7(n)$.

7.3 Transfinite Induction

The power of Mathematical Induction comes from the fact that it works for *any* process defined for each $n \in \mathbb{N}$. The limits of Mathematical Induction come from the fact that it only works for processes defined for each $n \in \mathbb{N}$. There are some very important processes that are defined for *every ordinal* α. Thus, if P is a process defined for *ordinals*, then Mathematical Induction will show that $P(n)$ is True for each $n \in \mathbb{N}$ but it will miss the Truth of $P(\omega_0)$. For example, we used Mathematical Induction to show that the cleaning lady in Hilbert's Infinite Hotel on page 130 will eventually clean all of the rooms labelled with $n \in \mathbb{N}$. However, our discussion missed the room for Hotel Staff labelled with ω_0. We found her in that room on page 225. In this way Mathematical Induction can be applied to processes defined for each natural number $n < \omega_0$, but it fails for processes defined for the ordinals $\alpha \geq \omega_0$. In this section, we will present an induction that can be applied to infinite ordinals as well as $\mathbb{N}$.

Let P be a process that is defined for all ordinals α. In other words, $P(\alpha)$ is *defined*, or makes Mathematical sense, for each ordinal α. *A proof by Transfinite Induction* will proceed as follows. It begins like Mathematical Induction.

1. *Initial Step:* Prove that $P(0)$ is True.

2. *Induction Hypothesis:* Assume that there is an ordinal α such that $P(\beta)$ is True for each ordinal $\beta < \alpha$.

3. *Induction Step:* Show that $P(\alpha)$ is True.

4. Conclude that $P(\gamma)$ is True for all ordinals γ.

We will use Transfinite Induction to replace the phrases *continue indefinitely*, and *continue in the same manner*, and the familiar ellipsis ... that accompanied some vague arguments that we encountered in some earlier proofs. Some examples of arguments using Transfinite Induction will help us see what all of this formality is about.

The following example about unions could be proved using DeMorgan's Law for Unions, but we will avoid that simple solution and use this union as an educational device to demonstrate Transfinite Induction.

Example 7.3.1 Suppose that we are given a universal set $\mathcal{U}$, and suppose that there is a *transfinite chain of subsets of* $\mathcal{U}$

$$U_0 \subset U_1 \subset \ldots \subset U_\alpha \subset \ldots \subset \mathcal{U},$$

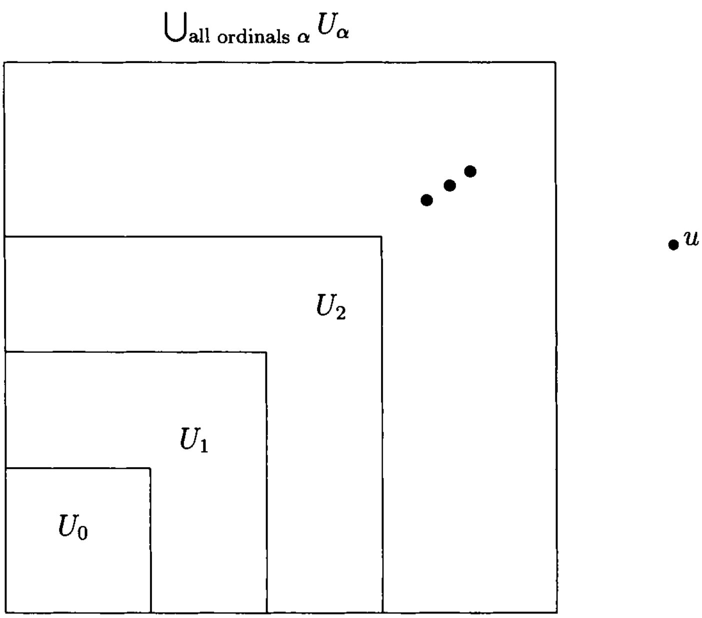
where α ranges over all of the ordinals. Further suppose that there is some special element $u \in \mathcal{U}$ that is not in U_α for any ordinal α.

$$\boxed{u \notin U_\alpha \text{ for each ordinal } \alpha.}$$

We will show, using Transfinite Induction, that

$$\boxed{u \notin \bigcup_{\text{all ordinals } \alpha} U_\alpha.}$$

Picture (7.5) will help you to visualize what we are doing. The set U_2 includes U_1, and the set U_3 includes U_2. The set $\mathcal{U}$ is the union of all of the sets U_α, α an ordinal. Notice that the element u is not in any of the sets U_α.



(7.5)

Initial Step: By hypothesis, $u \notin U_0$, so that

$$u \notin \bigcup_{\beta<1} U_\beta = U_0.$$

This begins the Transfinite Induction process.

Transfinite Induction Hypothesis: Let us assume that

$$u \notin \bigcup_{\beta<\alpha} U_\beta$$

for some ordinal α.

Transfinite Induction Step: In the Transfinite Induction Step, we must show that

$$u \notin \bigcup_{\beta<\alpha+1} U_\beta.$$

From the Induction Hypothesis, we know that

$$u \notin \bigcup_{\beta<\alpha} U_\beta,$$

and by the hypothesis that starts this problem, $u \notin U_\alpha$. Then by the definition of the set operation $\cup$, we have

$$u \notin U_\alpha \cup \left(\bigcup_{\beta<\alpha} U_\beta\right) = \bigcup_{\beta\leq\alpha} U_\beta = \bigcup_{\beta<\alpha+1} U_\beta.$$

This is what we had to prove about $\alpha + 1$.

We then conclude, by Transfinite Induction, that

$$u \notin \bigcup_{\text{all ordinals } \beta} U_\beta.$$

This compelets the proof.

What we have just shown is that if the U_α do not contain a point u then the union of the U_α fails to contain that point u. We accomplished this by showing that the union

$$U_0 \cup U_1 \cup U_2 \cup \ldots \cup U_\beta \text{ for } \beta < \alpha$$

does not contain the point u, and then concluded by Transfinite Induction that the total union does not contain the point u.

Notice in the above example that we did not have to prove anything for $\alpha + 1$ as we did in Mathematical Induction. We use the Induction Hypothesis to prove something about the construction process as it exists prior to α. Try to keep this in mind as these examples continue.

Example 7.3.2 The next example is the most important. We will construct the well-ordered chain of ordinals. This will add Mathematical certainty to the somewhat elementary argument that we used beginning on page 210, and with which we constructed the chain of ordinals.

Initial Step: We choose a smallest element, which we denote by

$$P(1) = 0.$$

Certainly, $\{0\}$ is a linearly ordered set. We say that 0 is *a constructed ordinal.*

Transfinite Induction Hypothesis: We assume that for some ordinal α, we have constructed a linearly ordered set

$$P(\alpha) = \{x_\beta \mid \beta < \alpha\}$$

with a linear ordering given by

$$0 < x_1 < \ldots < x_\beta < x_{\beta'} < \ldots$$
$$\text{for each constructed pair of ordinals } \beta < \beta' < \alpha.$$

This linear ordering has the property that $x^+ \in P(\alpha)$ for each $x \in P(\alpha)$ that is not the maximal element α in $P(\alpha)$. We say that β is *a constructed ordinal* for each $\beta < \alpha$.

For example, given $\alpha = 5$, our linearly ordered set is

$$P(5) = \{0, x_1, x_2, x_3, x_4\} = \left\{x_\beta \mid \beta < 5\right\},$$

and the ordering is

$$0 < x_1 < x_2 < x_3 < x_4.$$

The element x_4 does not have a successor element.

Another example is given if we let $\alpha = \omega_0$. Then our set is

$$P(\omega_0) = \{0, x_1, x_2, \ldots\} = \left\{x_\beta \mid \beta < \omega_0\right\}.$$

Unlike the previous set, this ordering is without end, so the set $P(\omega_0)$ is well-ordered.

Transfinite Induction Step: Show that there is a linearly ordered set

$$P(\alpha + 1) = \left\{x_\beta \mid \beta < \alpha + 1\right\}$$

whose ordering is given by

$$x_0 < x_1 < \ldots < x_\beta < x_{\beta'} < \ldots < x_\alpha$$
$$\text{for each pair of constructed ordinals } \beta < \beta' < \alpha + 1.$$

Choose a symbol x_α that is not in $P(\alpha)$. We certainly know one way to do that, reader. Form the linearly ordered set

$$P(\alpha+1) = P(\alpha) \cup \{x_\alpha\} = \{x_\beta \mid \beta < \alpha\} \cup \{x_\alpha\} = \{x_\beta \mid \beta < \alpha+1\}$$

by requiring that

$$x_\alpha < x_{\alpha+1}.$$

By the Induction Hypothesis, $P(\alpha)$ is a linearly ordered set, and since $\alpha < \alpha+1$, it follows that $P(\alpha+1)$ is a linearly ordered set. (Fill in this detail, reader. Why is it that given $x, y \in P(\alpha+1)$ then $x < y$ or $y < x$?)

Furthermore, given $x \neq x_\alpha \in P(\alpha)$, then $x = x_\beta$ for some constructed ordinal $\beta < \alpha$. Because $\beta < \alpha$, $\beta + 1 \leq \alpha$, so that $x_{\beta+1} \in P(\alpha)$. Hence $x = x_\beta < x_{\beta+1} = x^+$, and thus $x^+ \in P(\alpha)$ for each $x \in P(\alpha)$ that is not the maximal element x_α in $P(\alpha)$.

We conclude, by Transfinite Induction, that there is a set

$$P = \bigcup_{\text{ordinals } \alpha} P(\alpha) = \{x_\gamma \mid \gamma \text{ is an ordinal}\}.$$

To show that P is linearly ordered, given $x, y \in P$ then there is an ordinal α such that $x, y \in P(\alpha) \subset P$. By the Transfinite Induction Step, $P(\alpha)$ is linearly ordered, and so $x < y$ or $y < x$. Hence P is linearly ordered. This linear ordering is a well-ordering because for each element $x \in P$, there is an ordinal α such that $x \in P(\alpha)$ and $x < x_\alpha$. By the Transfinite Induction Step, $x^+ \in P(\alpha+1)$. Therefore, P is a well-ordered set.

Actually, the well-ordered chain P of ordinals constructed in Example 7.3.2 has a curious property.

P is larger than any set, so P *is not a set.*

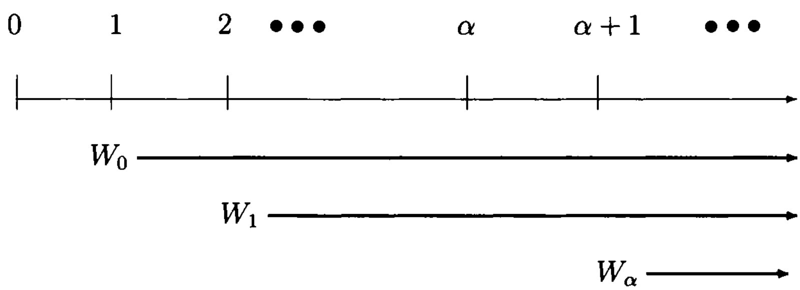
We will have more to say about this kind of anomaly later.

Example 7.3.3 Given an ordinal α, let

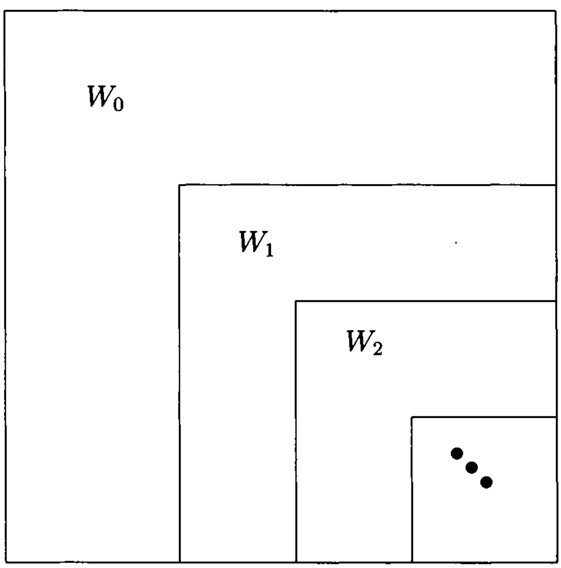
$$W_\alpha = \{\text{ordinals } \gamma \mid \alpha < \gamma\}.$$

The set W_α is the chain of ordinals that lie above α on the chain of ordinals. Then W_α can be thought of as a line of ordinals beginning at $\alpha + 1$ and continuing on indefinitely.

In terms of a picture we have



We might also envision the sets W_α as follows.



(7.6)

Using Transfinite Induction we will show that

$$\boxed{\bigcap_{\text{all ordinals } \alpha} W_\alpha = \emptyset.}$$

Initial Step: $0 \notin W_0$ since W_0 is the set $\{\gamma \mid 0 < \gamma\}$ of ordinals γ that are strictly larger than 0. Since

$$0 \notin W_0$$

we conclude that

$$0 \notin W_0 \cap \bigcap_{\text{all ordinals } 0<\gamma} W_\gamma = \bigcap_{\text{all ordinals } \gamma} W_\gamma.$$

Induction Hypothesis: Assume that for some ordinal β we have

$$\beta \notin \bigcap_{\text{all ordinals } \gamma} W_\gamma.$$

Induction Step: Since $W_{\beta+1} = \{\text{ordinals } \gamma \mid \beta+1 < \gamma\}$, it follows that

$$\beta + 1 \notin W_{\beta+1} \supset \bigcap_{\text{all ordinals } \gamma} W_\gamma.$$

Then

$$\beta + 1 \notin W_{\beta+1} \cap \bigcap_{\text{all ordinals } \gamma} W_\gamma = \bigcap_{\text{all ordinals } \gamma} W_\gamma,$$

and hence

$$\beta + 1 \notin \bigcap_{\text{all ordinals } \gamma} W_\gamma.$$

Therefore, by Transfinite Induction,

$$\gamma \notin \bigcap_{\text{all ordinals } \alpha} W_\alpha$$

for each ordinal γ. Thus, $\bigcap_{\text{all ordinals } \alpha} W_\alpha$ does not contain any ordinals, which implies that

$$\bigcap_{\text{all ordinals } \alpha} W_\alpha = \emptyset.$$

From Example 7.3.3, we conclude that there is no largest ordinal. Thus, the chain of ordinals has no top and no end, just like the chain of finite ordinals $\mathbb{N}$.

Example 7.3.4 (With an acknowledgment to Douglas R. Hofstader [8].) This is a story that illustrates how Transfinite Induction takes over where Mathematical Induction ends.

It seems that the magic lamp was passed down from generation to generation of Aladdin's family. The current heir to the lamp, Al, is a Mathematician with a droll sense of humor. When he rubbed the lamp, the genie appeared and said, "I am the genie called

$$\boxed{M_0.}$$

I will grant you three wishes." Using his droll humor Al said, "I wish for more wishes."

"That is *a wish about wishes* and I cannot grant such a thing," the genie responded. "It would be like the Mayor of New York City acting like the President of the United States. Before I can grant your wish, I must ask for my Boss' permission. His name is

$$\boxed{M_1.}$$

M_1 grants my wishes."

So the genie contacted M_1 and said, "I wish to grant *a wish about wishes*." M_1 thought for a minute and said, "You have made *a wish, about a wish, about wishes*, and I cannot grant such a thing without the permission of my Boss M_2. M_2 grants my wishes."

Once M_2 is contacted, M_1 asks, "I wish to grant *a wish, about a wish, about wishes*." Of course, M_2 saw through that wish right away and said, "I cannot grant your *wish, about a wish, about a wish, about wishes* without the permission of my Boss, M_3. M_3 grants my wishes."

The process of asking the boss M_n continued once for each $n \in \mathbb{N}$, where

$$\boxed{M_{n+1} \text{ is the Boss that grants } M_n\text{'s wishes.}}$$

The problem was that no one could give the permission that would eventually grant Al's wish. No matter how far up the chain the request travelled, no boss M_{n+1} could grant permission to the boss M_n. The chain of bosses $\{M_n \mid n \in \mathbb{N}\}$ has no largest element.

The chain of bosses $\{M_n \mid n \in \mathbb{N}\}$ is only a part of a well-ordered chain of bosses. Let me introduce you to the boss of all of the chain $\{M_n \mid n \in \mathbb{N}\}$. His name is

$$\boxed{M_{\omega_0}.}$$

Because he is a good administrator, he realized the predicament his people were in so he thought he would just give all of the M_n permission to grant those wishes. But you see, being an administrator M_{ω_0} must ask permission from his boss, M_{ω_0+1} before he can grant wishes about wishes for each M_n with $n \in \mathbb{N}$.

The transfinite process here is now quite clear. Suppose that we are given an ordinal α such that

$$\boxed{\text{boss } M_\beta \text{ exists for each ordinal } \beta < \alpha}$$

and such that

$$\boxed{M_{\beta+1} \text{ grants } M_\beta\text{'s wishes for each ordinal } \beta < \alpha.}$$

Find a boss M_α, who is the boss of all M_β with $\beta < \alpha$. Then we have extended our discussion to include all bosses M_β such that $\beta \leq \alpha < \alpha + 1$. By Transfinite Induction, for each ordinal γ, each boss $M_{\gamma+1}$ grants wishes for boss M_γ.

The problem is that there is no boss M at the top of this chain to give everyone permission to grant wishes. The story has a happy ending though, as The Supreme One saw what was going on, and said to each boss M_γ, "Grant Al's wishes, and let's get on with the work of the day." And so it was done.

Before continuing on to the next section, let me present a process called *Student Induction.* It seems that a certain University cancels classes the Wednesday prior to Thanksgiving Thursday. This gives the students an initial day off. They then rationalize that since there are no classes Wednesday, they can leave for home Tuesday night. This is not too hard to understand. The anticipation of the holiday being what it is, they decide to leave earlier that day, Tuesday afternoon. No, make it Tuesday morning, because knowing that they are leaving for home at noon, they lose their concentration in the morning classes. But then, why not leave Monday. Well, nobody is going to stay at college for one day of classes. so they argue that they can leave Sunday night. Right? Get real. No one will stay the weekend, only to leave Sunday night, so they leave late Friday. No. Not late. No one has enthusiasm for Friday classes, so they decide to leave Thursday night. With this kind of *Student Induction* taking place the student is soon leaving on Halloween (October 31) in anticipation of the Thanksgiving holiday (the last Thursday in November). The same process is at work in the Spring Semester, when the students prepare for Spring Break. They are soon leaving in January for a vacation that does not take place until March. We leave the details to be filled in by the reader.

7.4 Mathematical Recursion

Another process that is designed to take the ellipsis ... out of Mathematical definitions is called *recursion.* Recursion is a Mathematical methodology that defines one object in terms of another smaller object. This smaller object is similar to the larger one, but not quite the same size. The first recursion usually seen by the student is n *factorial.* This n factorial is defined as

$$\begin{aligned} 0! &= 1 \text{ and} \\ n! &= n(n-1)! \text{ for integers } n \geq 1. \end{aligned}$$

Do you see that $n!$ is defined in terms of the smaller number $(n-1)!$? For instance,

$$\begin{aligned} 1! &= 1 \cdot 0! = 1 \cdot 1 = 1, \\ 3! &= 3 \cdot (2!) = 3 \cdot (2 \cdot 1!) = 3 \cdot (2 \cdot 1) = 6, \end{aligned}$$

$$10! = 10 \cdot (9!) = 10 \cdot 9 \cdot 8! = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1.$$

We do not define $n!$ for negative numbers, and we do not define $n!$ for fractions in this book. We leave that as a topic in your Calculus book. The number $n!$ in this book is defined only for nonnegative natural numbers n.

The definition of $n!$ is recursive because it is defined in terms of a factorial at a smaller value. We can define other numbers in this way. We let

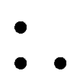
$$\begin{aligned} T_0 &= 0 \text{ and} \\ T_n &= n + T_{n-1} \text{ for integers } n \geq 1. \end{aligned}$$

T_n is called the nth *triangular number* and it is the number of dots needed to form a triangular array. For instance, since one dot forms a (trivial) triangular array, $T_1 = 1$ is a triangular number.

The formula for triangular numbers shows us that

$$T_2 = 2 + T_1 = 2 + 1 = 3.$$

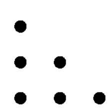
Thus 3 is the number of dots needed to form the next larger triangular array, as in the illustration below.



For larger triangular numbers proceed recursively,

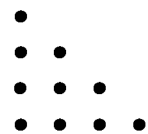
$$T_3 = 3 + T_2 = 3 + 3 = 6,$$

since we know that $T_2 = 3$. The triangular array for T_3 is formed by adding a row of three dots into the previous triangular array.



Notice that T_3 is found by counting the number of dots in a triangular array of dots whose base has 3 dots. The same is said for

$T_4 = 4 + T_3 = 4 + 6 = 10.$



Actually, we were introduced to triangular numbers in Example 7.1.6, where we examined the sum

$$\boxed{T_n = \sum_{\ell=1}^{n} \ell = 1 + 2 + 3 + \ldots + n \text{ for } n > 0.}$$

By using the formula (7.1.6) discovered by a Carl F. Gauss in the late eighteenth century, we uncover the formula

$$\boxed{T_n = \frac{1}{2}n(n+1)}$$

for the n-th triangular number. We have thus replaced our recursive formula $T_n = n + T_{n-1}$ for the n-th triangular number, with the closed formula

$$T_n = \frac{1}{2}n(n+1) \text{ for each integer } n > 0. \tag{7.7}$$

Closed formulas have their advantages. The most immediate advantage is, that we can calculate T_{10} without calculating T_9, T_8, and so on. We simply write

$$T_{10} = \frac{1}{2}10(10+1) = 55.$$

Yet larger triangular numbers can be calculated without the use of smaller triangular numbers or of triangular arrays.

$$T_{100} = \frac{1}{2}100(100+1) = 5050.$$

This is the advantage of closed formulas.

There is an interesting formula involving triangular numbers and perfect squares. When written in words, the next theorem states that consecutive triangular numbers add up to a perfect square. That should sound satisfying to you. Two triangular numbers add up to a square number. How many times have you broken a square into two triangles using a pencil and a ruler? Now you have a numerical identity that does the same thing to numbers. A square number is the sum of two triangular numbers.

Theorem 7.4.1

$T_n + T_{n-1} = n^2$ *for each integer* $n > 0$.

Proof: By (7.7), we have

$$\begin{aligned} T_n + T_{n-1} &= \frac{1}{2}n(n+1) + \frac{1}{2}(n-1)n \\ &= \frac{n}{2}((n+1)+(n-1)) \\ &= \frac{n}{2}(2n) \\ &= n^2. \end{aligned}$$

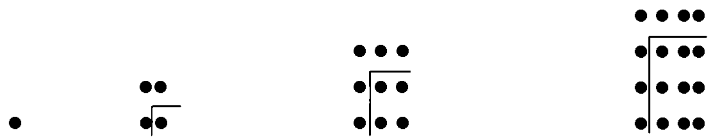
This concludes the proof.

In Example 7.1.7, we encountered another sum that adds up to n^2. There we found that

$$n^2 = \sum_{\ell=1}^{n}(2\ell - 1) = 1 + 3 + \ldots + (2n-1).$$

Thus, *the sum of the first n odd numbers is the nth perfect square.*

The first few sums of odd numbers can be seen in the following diagrams. Think of one square as being formed by adding dots to the edge, and then count the odd number of dots that are added to the recursive square.



1^2 $\quad 1+3=2^2 \quad (1+3)+5=3^2 \quad (1+3+5)+7=4^2$

Draw the next square in the pattern to see that $4^2+9=(1+3+5+7)+9=5^2$.

Here is an example of recursion in an infinite setting.

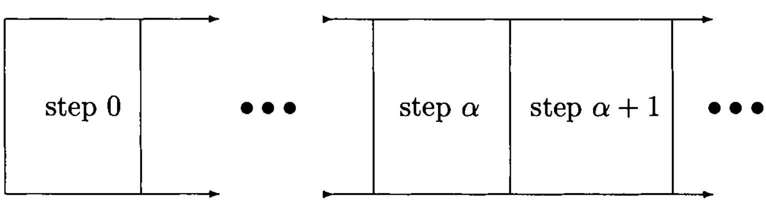
Example 7.4.2 Hilbert has expanded the Infinite Hotel to accommodate any and all guests. To do this he adds on *one room for each ordinal* to his Infinite Hotel using recursion.

The Intial Step is to construct room 1.

The Transfinite Induction Hypothesis is to assume that there is an ordinal α for which we have constructed rooms β for each ordinal $\beta < \alpha$.

The Transfinite Induction Step begins by letting H_α be the construction made in the Transfinite Induction Hypothesis. We construct room α onto H_α, then arriving at a hotel with rooms β for ordinals $\beta < \alpha + 1$. We have thus constructed $H_{\alpha+1}$.

Therefore, by Transfinite Induction, we have constructed the hotel H with rooms β for each ordinal β. The result is something more. The Infinite Hotel is now a *Transfinite Hotel.* For instance, there is a room ω_0, a room $\omega_0^{\omega_0}$, and so on.



The Transfinite Hotel is not a set, though. It is bigger than any set can be. We will consider the matter more closely later.

Maria cleans each room α and then cleans room $\alpha + 1$. Time is no longer an issue, since Hilbert and the Hotel Staff signed a collective bargaining agreement that stops time. Thus, no matter how many rooms are cleaned, no time has passed. It also means that when she goes home there is no end to her leisure time since that time does not pass. In this way, Maria can start with room 1, then clean all of the rooms in the Transfinite Hotel within her work day, and still have time for the grandchildren, Paul and John. We will show that she does indeed clean all of the rooms.

The initial room, room 1, is cleaned shortly after she arrives at work.

Inductively, assume that there is some ordinal α such that Maria has cleaned room β for each ordinal $\beta < \alpha$.

Having cleaned rooms β for $\beta < \alpha$, Maria finds herself in front of room α. Since she is industrious, she enters the room and cleans it. Thus room α is cleaned, so that room β is cleaned for each ordinal $\beta < \alpha + 1$.

We conclude by Transfinite Induction, that Maria has cleaned room γ for each ordinal γ. At this time she joins a colleague, and sits and talks in the Hotel Staff Lounge, Room ω_0.

Let us answer the question in many minds. Why is the chain of ordinals not a set?

Theorem 7.4.3 *The chain of ordinals is not a set.*

Proof: Let $\mathcal{O}$ be the chain of *all* ordinals, and assume for the sake of contradiction that $\mathcal{O}$ is a set. Then $\mathcal{O}$ cannot be extended to a larger chain of ordinals. But since $\mathcal{O}$ is a set, we can, as we did in the construction of the chain of ordinals in Example 7.3.2, use $\mathcal{O}$ as an ordinal such that $\alpha < \mathcal{O}$ for all $\alpha \in \mathcal{O}$. This extension of $\mathcal{O}$ to a set with an extra ordinal $\mathcal{O}$ contradicts that $\mathcal{O}$ is the set of *all* ordinals. This contradiction proves that $\mathcal{O}$ is not a set, and completes the proof.

Exercise 7.4.4 1. Find $f(n)$ if $f(0) = 2$ and $f(n) = 2f(n-1)$. Prove that your formula is correct.

2. Suppose that $a_1 = 1$ and that $a_n = a_{n-1} + 3$. Find a nonrecursive formula for a_n and prove that your formula works.

3. Suppose that $a_1 = 2$ and that $a_n = \sqrt{a_{n-1} + 1}$. Find a nonrecursive formula for a_n and prove that your formula works.

4. Suppose that $a_1 = a_2 = 1$ and that $a_{n+1} = a_n + a_{n-1}$. Find a nonrecursive formula for a_{n+1}.

7.5 Number Theory

The numbers considered in this section are *nonnegative* natural numbers. That is, we will not be considering negative numbers. In this section, we will use Mathematical Induction and recursion to give Mathematical certainty to some facts that we have all heard about integers.

The number $p > 1$ is said to be *prime* exactly when

$$p = ab \text{ implies that } a = p \text{ or } a = 1.$$

Another way of saying this is that p is not 1 and p is divisible only by itself and 1. A number a is a *factor* of n if there is a number b such that $ab = n$. The number a is a *prime factor* of n is a is a prime number and a is a factor of n. For example, 3 is a prime factor of 6 and 15, while 7 is a prime factor of 49. Prime factors, it seems, are everywhere.

Lemma 7.5.1 *Every natural number $n > 1$ has a prime factor.*

Proof: Let us assume to the contrary that there is a positive number n that has no prime factors. Using the Minimum Principle of $\mathbb{N}$, we take the smallest such number n. Then n is not prime (since otherwise it would be a prime factor of itself), so we can write

$$n = ab$$

for some numbers a, b not in $\{n, 1\}$. Since $a \neq n$ and $b \neq 1$ divide n, a is smaller than n.

$$a = \frac{n}{b} < n.$$

Since n is the smallest number without a prime factor, a has a prime factor p, say,

$$a = pc.$$

But then p is also a prime factor of n since

$$n = ab = p(cb).$$

This contradiction to our assumption that n does not possess a prime factor shows us that every number n possesses a prime factor.

The next result is one that goes back to the Greek geometer Euclid (circa 300 BC). He proves that *there are infinitely many prime numbers*. That may sound obvious at first. You have known it since birth or since early in your education. In fact, you were given a known Mathematical fact and you were asked to memorize it. You accepted it as a Truth. But Mathematics can give certainty to the arithmetic facts you learned as a child. That is what we will do here. We will give Mathematical certainty to some arithmetic facts that we learned at a young and impressional age.

Theorem 7.5.2 [Euclid] *There are infinitely many prime numbers.*

Proof: We offer a proof by contradiction. Suppose to the contrary that the list of prime numbers is finite. Write down that finite list.

$$p_1, p_2, \ldots, p_t,$$

and then write down the number

$$N = p_1 p_2 \cdots p_t + 1.$$

Since every number has a prime factor, N has a prime factor p. Then division of N by p has a remainder of 0. However, given a prime p_i on the list $p_1, p_2, \ldots, p_t$, division of N by p_1 leaves a remainder of 1. Then $p \neq p_i$. Since $p_1, p_2, \ldots, p_t$ is supposed to be a list of all of the primes, and since p is a prime not on the list, we have found our contradiction. Thus, the list of primes is infinite.

Let us try to demonstrate what this result has done. Given any finite set of primes, it shows us *how to locate a new prime*. For instance, the number

$$N = 2 \cdot 3 \cdot 5 \cdot 7 + 1 = 211$$

is divisible by a prime number p such that $p \neq 2, 3, 5, 7$. Then the new prime p was not on our finite list of primes, but it is surely not the last prime that exists. Reader, find the prime divisor of 211 that is not on the list $2, 3, 5, 7$.

In the same way, the number

$$N = 17 \cdot 19 \cdot 23 \cdot 29 + 1 = 215,442$$

has a prime divisor p. We do not know what p is, but we can say with certainty that p is not on the finite list of primes

$$17, 19, 23, 29.$$

This is what the proof of Theorem 7.5.2 has done.

Just how many times is a number n divisible by a prime? Infinitely often? Finitely often? The next lemma resolves this question.

Lemma 7.5.3 *Let n be a number with prime divisor p. There is an exponent t and a number b such that $n = p^t b$ and b is not divisible by p. That is, p divides n exactly t times and no more.*

Proof: We proceed by contradiction. Suppose there is a number n and a prime p such that p divides n infinitely often. That is, we can write

$$n = pb_1 = p^2 b_2 = \dots$$

for some integers $b_1, b_2, \dots$. Then

$$b_1 = \frac{n}{p} > b_2 = \frac{n}{p^2} > b_3 = \frac{n}{p^3} > \dots$$

is an infinite descending list of positive integers. This is impossible in $\mathbb{N}$, so our supposition was incorrect. Each prime divisor of n divides n at most finitely often, and no more. This completes the proof.

Theorem 7.5.4 *Let $n \geq 2$ be a positive integer. There are primes $p_1, \dots, p_t$ such that*

$$n = p_1 \dots p_t. \tag{7.8}$$

It is traditional to call (7.8) a *prime factorization of* n.

Proof: We apply induction. The result is True for $n = 2$, since in this case $n = p_1 = 2$.

Our Induction Hypothesis is to assume that we have arrived at an integer $k \geq 2$ for which each number $m \leq k$ has a prime factorization like (7.8).

We must find a prime factorization for $k+1$. By Lemma 7.5.1, $k+1$ has a prime factor p. Then

$$\frac{k+1}{p} < k+1$$

has a prime factorization by the Induction Hypothesis. Hence, we can write

$$\frac{k+1}{p} = p_1 \dots p_s$$

for some primes $p_1, \dots, p_s$. It follows that,

$$k+1 = p(p_1 \dots p_s),$$

which is a prime factorization of $k+1$.

Therefore, by Mathematical Induction, each integer $n \geq 2$ has a prime factorization.

Exercise 7.5.5 Use a calculator to find the prime factors of the following numbers.

1. $N = 2 \cdot 3 + 1$.
2. $N = 7 \cdot 11 \cdot 13 + 1$.
3. $N = 2 \cdot 17 \cdot 29 \cdot 31 + 1$.
4. $N = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 + 1$.

7.6 The Fundamental Theorem of Arithmetic

The above theorem shows us that each integer $n > 1$ has a factorization into primes. But we know more than that, don't we, reader?

We know that there is one and only one way to factor an integer $n > 1$. This uniqueness is where we are headed to now. Such a result is important enough that Mathematicians have given it a name.

The Fundamental Theorem of Arithmetic 7.6.1 *Let $n \geq 2$ be an integer.*

1. *There is a finite list of primes $p_1, \ldots, p_t$ such that $n = p_1 \ldots p_t$.*

2. *If $n = q_1 \ldots q_s$ for some other list of primes $q_1, \ldots, q_s$, then $s = t$ and after a permutation of the list entries, $p_i = q_i$ for each $i = 1, \ldots, t$. We say that n* has unique prime factorization.

Proof: 1. This is Theorem 7.5.4.

2. We apply proof by contradiction. Suppose there is a number $n \geq 2$ that possesses two different prime factorizations, say,

$$n = p_1 \cdots p_t = q_1 \cdots q_s$$

for some finite lists $p_1, \ldots, p_t$ and $q_1, \ldots, q_s$ of primes. Choose the smallest integer n with this property. Then the lists $p_1, \ldots, p_t$ and $q_1, \ldots, q_s$ are different. After permuting the subscripts we can assume that p_1 is not on the finite list $q_1, \ldots, q_s$ and

$$2 \leq p_1 < q_1. \tag{7.9}$$

Then $q_1 - p_1 < q_1$ so that

$$(q_1 - p_1)(q_2 \cdots q_s) < q_1(q_2 \cdots q_s) = n.$$

Since n is the smallest integer with nonunique prime factorization,

$$m = (q_1 - p_1)(q_2 \cdots q_s) \tag{7.10}$$

has a unique prime factorization.

By hypothesis $p_1 \neq q_i$ for $i = 1, \ldots, s$, and p_1 does not divide $q_1 - p_1$. (Otherwise p_1 is a divisor of the prime $(q_1 - p_1) + p_1 = q_1$, contrary to (7.9).) Since (7.10) is a unique prime factorization of

m, p_1 does not divide m. On the other hand, p_1 divides n, so p_1 does not divide $n - m$. However,

$$\begin{aligned} n - m &= q_1 q_2 \cdots q_s - (q_1 - p_1) q_2 \cdots q_s \\ &= (q_1 - (q_1 - p_1))(q_2 \cdots q_s) \\ &= (p_1)(q_2 \cdots q_s). \end{aligned}$$

This obvious contradiction shows us that our supposition was incorrect. Thus, the finite lists $p_1, \ldots, p_t$ and $q_1, \ldots, q_s$ are the same.

Subsequently, the number of terms s and t on these lists are equal.

$$s = t$$

This is what we had to prove to complete the proof of part 2, and complete the proof of the theorem.

Let us illustrate the above theorem with a number. The number

$$n = 215,441$$

has a unique prime factorization. At present we do not know what it is, but we do know that unique prime factorization exists for n. Some work with a calculator shows us that

$$n = 17 \cdot 19 \cdot 23 \cdot 29$$

is a prime factorization of n. The Fundamental Theorem of Arithmetic shows us that this is the only prime factorization of n. If we have a prime factor p of n, then the Fundamental Theorem of Arithmetic states that p is on the list

$$17, 19, 23, 29.$$

No other primes are allowed to divide n. That is the power of the Fundamental Theorem of Arithmetic. Without even knowing what p is, we know that it is one of the four primes on the list.

Exercise 7.6.2 Find the unique prime factorizations of the following numbers.

1. 5.
2. 52.
3. 111.
4. 121.
5. 2122.

7.7 Perfect Numbers

Euclid is also known for introducing us to *perfect numbers.* The number $m > 0$ is said to be *perfect* if m is the sum of the numbers properly dividing m. Then 6 is a perfect number, since 6 is the sum

$$6 = 1 + 2 + 3$$

of its proper divisors, $1, 2, 3$.

The next perfect number is 28, because $1, 2, 4, 7, 14$ are the proper divisors of 24, and because

$$28 = 1 + 2 + 4 + 7 + 14.$$

Other known perfect numbers are

$$2^{495}(2^{496} - 1), \quad 2^{520}(2^{521} - 1),$$

$$2^{606}(2^{607} - 1), \quad 2^{1278}(2^{1279} - 1).$$

The search for these perfect numbers involved some computer use in 1952. In those days, number problems like this were used to test a computer's accuracy. This search continues today, though, as computers are still being used to examine natural numbers of the form $2^n - 1$.

So far, these perfect numbers have a curious property. They are all even numbers. Are all perfect numbers even? Currently, no one knows. This uncertainty enhances interest in perfect numbers. For the even perfect numbers, there is the following theorem also due to Euclid.

Theorem 7.7.1 [Euclid] *If $2^n - 1$ is a prime number, then $2^{n-1}(2^n - 1)$ is a perfect number.*

Euclid's proof of this theorem was geometric in nature since Greek algebra was underdeveloped. The proof has evolved over the years into the following algebraic gem. Late in the eighteenth century, the prodigious Mathematician Leonard Euler (circa 1750 AD) showed that if m is an even perfect number, then there is an integer n such that $2^n - 1$ is a prime number, and $m = 2^{n-1}(2^n - 1)$. Think of the gap in time here. Euclid (300 BC) introduced perfect numbers, but it took until Euler (1750 AD) to show that all even perfect numbers are encompassed by Euclid's Theorem. That makes 2050 years between first sight and last sight. This must be one hard type of number to study if it takes over two millennia to make substantial progress on the classification problem.

Thus, Euclid's Theorem accounts for all of the even perfect numbers. We are left to wonder about the open question. Is there an odd number m that is the sum of its proper divisors?

Open Question: Are there any *odd* perfect numbers?

For example, 5 is *not* a perfect number since 1 is the only proper divisor of 5, and 5 is not the sum of its proper divisor 1. The odd number 25 is not the sum of its proper divisors 1 and 5 since

$$25 \neq 1 + 5.$$

You can use your computer to verify that there are no odd perfect numbers less than 100.

Let us prove **Theorem** 7.7.1.

Proof: Suppose we know that $p = 2^n - 1$ is a prime for some integer $n > 1$. Then we have a unique prime factorization

$$m = 2^{n-1}p.$$

The prime divisors of m are 2 and p so any proper divisor d of m is a multiple of 2 and p,

$$d = 2^k p^e,$$

where k can be any integer in $\{0, 1, \ldots, n-1\}$ and e is either 0 or 1. That means that the proper divisors of m that are not divisible by p are

$$1, 2, 2^2, \ldots, 2^{n-1}$$

and the proper divisors divisible by p are

$$p, 2p, 2^2p, \ldots, 2^{n-2}p.$$

(We do not include $2^{n-1}p = m$ since it is not a *proper* divisor of m.) No other numbers are possible. Recall the geometric identity

$$1 + x + x^2 + \ldots + x^{n-1} = \frac{1 - x^n}{1 - x}$$

with common ratio x. If $x = 2$, then we have

$$1 + 2 + 2^2 + \ldots + 2^{n-1} = 2^n - 1.$$

Hence, adding up the proper divisors of m yields the geometric identities

$$\begin{aligned}
\left.\begin{array}{r} 1 + 2 + 2^2 + \ldots + 2^{n-1} + \\ p + 2p + 2^2 p + \ldots + 2^{n-2}p \end{array}\right\} &= \\
\left.\begin{array}{r} 1 + 2 + 2^2 + \ldots + 2^{n-1} + \\ (1 + 2 + 2^2 + \ldots + 2^{n-2})p \end{array}\right\} &= \\
(2^n - 1) + (2^{n-1} - 1)p &= \\
p + (2^{n-1} - 1)p &= \\
2^{n-1}p &= m.
\end{aligned}$$

(Recall that $p = 2^n - 1$ by our choice of p.) This shows that m is a perfect number and completes the proof of Euclid's Theorem 7.7.1.

Therefore, there will be an infinite number of perfect numbers if there are infinitely many numbers n such that $2^n - 1$ is a prime number. At the time of this writing, it is not known if there are infinitely many primes of the form $2^n - 1$, and so it is not known if there are infinitely many even perfect numbers.

Exercise 7.7.2 Show that the numbers in exercises 1 through 5 are prime.

1. $2^2 - 1$.
2. $2^3 - 1$.
3. $2^5 - 1$.
4. $2^7 - 1$.
5. Use the primes you found in exercises 1 through 4 to find 4 perfect numbers n_1, n_2, n_3, and n_4. Use a calculator to show that the numbers n_1, n_2, n_3, and n_4 are indeed perfect numbers.

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

Chapter 8

Prime Numbers

8.1 Prime Number Generators

Euclid's Theorem on Prime Numbers 7.5.2 proves that there are infinitely many primes. The next step then would be to decide where these primes are, and to discover how we could locate them amongst the well-ordered natural numbers. The pursuit of prime numbers goes back to the Greeks and is currently a source of Mathematical investigation and inspiration. We will discuss the two ends of this time line.

The Greeks had a clever way of finding the prime numbers less than some fixed number. They called it *The Sieve.* Basically, if you wanted to find all of the prime numbers less than a given number, let us say 60, you would first write down all of the numbers less than 60 starting with 2. (Leave 1 out because 1 is not a prime. Remember, a prime is a number $p > 1$ that is divisible only by 1 and itself. We exclude 1.)

	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60

All numbers divisible by 2 are excluded from our search for prime numbers so we start with 2, and move every other space, marking these numbers out as we go. We mark them out because they are divisible by 2. You stop when you reach the end of your Sieve. Observe:

	2	3	5	7	9
11		13	15	17	19
21		23	25	27	29
31		33	35	37	39
41		43	45	47	49
51		53	55	57	59

Instead of striking the numbers out, we have deleted them from The Sieve. You can just strike them out with the slash /. The numbers that remain are not divisible by 2.

Next, we find multiples of 3. Start at 3 and count three spaces in The Sieve until The Sieve is ended. It is important that you count the spaces in The Sieve as well. The deleted numbers contribute to our search for primes. Do it this way.

	2	3	5	7	
11		13		17	19
		23	25		29
31			35	37	
41		43		47	49
		53	55		59

The numbers left are not divisible by 2 or by 3. The next number left on The Sieve is 5. Delete all of the numbers and spaces in The Sieve that are five spaces from 5, 5 spaces from that number, and so on. Simply slash out numbers by counting every fifth number and space, as follows.

	2	3	5	7	
11		13		17	19
		23			29
31				37	
41		43		47	49
51		53			59

The remaining numbers are not divisible by 2, 3, or 5. Do the same for 7 (the next number on The Sieve) and arrive at

	2	3		5	7	
11		13			17	19
		23				29
31					37	
41		43			47	
51		53				59

Note the disappearance of the number 49 from The Sieve. You would have deleted 14, 21, 28, 35, and 42, but these numbers were already deleted from The Sieve. They are also divisible by 2, 3, and 5. None of the remaining numbers in The Sieve is divisible by 2, 3, 5, or 7. Take 11 and strike out every eleventh number as so.

	2	3		5	7	
11		13			17	19
		23				29
31					37	
41		43			47	
51		53				59

Nothing new is deleted. That is, because the first number encountered must be 11^2. The smaller numbers are of the form $p \cdot 11$ for $p = 2, 3, 5, 7$, and these numbers have been deleted from The Sieve. The number bigger than 11 that we would first delete is 121, which is not on our small Sieve. Then The Sieve has filtered out all of the prime numbers less than 60. Those primes are the numbers left in The Sieve. They are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 51, 53, 59.$$

The Sieve is used to find the primes less than a given number, but this approach is quite slow. Computers especially have a hard time finding primes less than n with The Sieve when n is large. For example, it would take a desktop computer some time to find the first million primes using The Sieve. The same can be said if you try to find the primes that are less than 10^{10}.

Another method for generating primes is to use polynomials. It was observed by Euler (circa 1750 AD) that the quadratic polynomial

$$p(x) = 41 + x + x^2$$

gives a list of prime values for small x. Some of those prime numbers are given.

$$\begin{aligned}
&f(0) = 41, f(1) = 43, f(2) = 47,\\
&53, 61, 71, 83, 97, 113, 151, 173, 197, 223,\\
&251, 281, 313, 347, 383, 421, 461, 503, 547, 593, 641, 691,\\
&743, 797, 853, 911, 971, 1033, 1097, 1163, 1231, 1301,\\
&1373, 1447, 1523, 1601.
\end{aligned}$$

Use your desktop computer to see that each of these numbers is a prime number. Of course, the number $p(41)$ will not be a prime since it is divisible by 41, and neither is $p(40)$ a prime. Try it and see. Thus, some of the values of $p(x)$ are not prime numbers.

But what a Mathematical coincidence. We produce many primes from such a simple polynomial of such a small degree. Why is that so? Since Mathematics does not believe in Mathematical coincidence, the question posed was to find a polynomial of small degree that was complex enough to yield all of the prime numbers as output. As yet no one has discovered a polynomial whose images for integer values are always prime. This is to be expected since the next section will show that an actual accounting of the prime numbers less than a specific value n was long ago proved to be approximately $n/\ln(n)$, and thus not a polynomial value.

8.2 The Prime Number Theorem

Now let's consider just counting the primes. For a given integer $n > 0$, let

$$\boxed{\pi(n) = \text{ the number of primes } p < n.}$$

Here is $\pi(n)$ for some small n. Because

$$2, 3$$

are the primes less than 5,

$$\pi(5) = 2;$$

because

$$2, 3, 5, 7, 11$$

are the primes less than 12,

$$\pi(12) = 5;$$

and because

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29$$

are the primes less than 30,

$$\pi(30) = 10.$$

Using The Sieve, we would conclude that

$$\pi(60) = 18.$$

There doesn't seem to be much of a pattern here, does there? And yet Mathematicians have looked for and found an important pattern. It is nothing less than a closed formula for finding out how many primes are less than a given number. To talk about this milestone in Mathematics, we need to know a little about logarithms.

One of the fundamental constants of Mathematics is the number e (in honor of Leonard Euler (circa 1700) who first investigated this number). The value of e to twelve decimal places is

$$\boxed{e \approx 2.718281828459,}$$

but these twelve decimal places do not help real Mathematicians in their researches, as e possesses infinitely many decimal places, and these decimals do not repeat in a pattern. There is no end to their seemingly random digital pattern. However, one does not study the physical sciences easily without some kind of approximation to the number e. This number, it seems, is how nature put the universe together. For those familiar with chemistry, the decay of radioactive elements is based on e through the equation

$$A(t) = A_0 e^{-\lambda t},$$

where A_0 is the initial amount of material and λ is a constant called the *decay constant*. For those who have studied calculus, this function $A(t)$ solves the differential equation

$$A'(t) = kA(t),$$

which does not seem to mention e at all.

In possession of this Mathematical constant, we define the *natural logarithm* as follows. Let a and b be nonnegative real numbers. Then the *natural logarithm* of a is denoted by

$$\boxed{\ln(a).}$$

By definition, this logarithm is the log in the base e. That is,

$$\boxed{\ln(a) = b \quad \text{exactly when} \quad a = e^b.}$$

Then

$$\begin{aligned}
\ln(1) &= 0 \quad \text{because} \quad 1 = e^0, \\
\ln(e) &= 1 \quad \text{because} \quad e = e^1, \\
\ln(\frac{1}{e}) &= -1 \quad \text{because} \quad \frac{1}{e} = e^{-1}, \quad \text{and} \\
\ln(\sqrt{e}) &= \frac{1}{2} \quad \text{because} \quad \sqrt{e} = e^{1/2}.
\end{aligned}$$

In general, we have

$$\boxed{\ln(e^x) = x \text{ for any } x \in \mathbb{R}.}$$

With these equations, we have a rudimentary knowledge of $\ln(x)$. We can then state the Prime Number Theorem.

The Prime Number Theorem: Given a large $n \in \mathbb{N}$,

$$\boxed{\pi(n) \approx \frac{n}{\ln(n)}.}$$

Indeed, the approximation gets better as n increases without bound. By this we mean that a comparison between $\pi(n)$ and $\dfrac{n}{\ln(n)}$ will show that these numbers are almost the same as n gets large. That is, when the following limit is calculated, we find that

$$\lim_{n \longrightarrow \infty} \frac{\pi(n)}{n/\ln(n)} = 1.$$

Consider this for a moment. A fraction $\dfrac{A}{B} = 1$ precisely when $A = B \neq 0$. This limit implies that $\dfrac{\pi(n)}{n/\ln(n)}$ is very close to 1 for large $n \in \mathbb{N}$. Then $\dfrac{n}{\ln(n)}$ is a very good approximation for $\pi(n)$ for large $n \in \mathbb{N}$. Furthermore, until some Mathematical genius can figure out a better way to study the distribution of primes, approximations for $\pi(n)$ are the best we can ask for.

Example 8.2.1 1. With a little calculator use, we can be sure that

$$\pi(1024)$$

is close to

$$\frac{1024}{\ln(1024)} = \frac{1024}{\ln(2^{10})} = \frac{1024}{10\ln(2)} = 160.$$

2. Larger numbers are sure to lead to better approximations. So because $e < 3$,

$$\pi(e^{20})$$

is close to

$$\frac{e^{20}}{\ln(e^{20})} = \frac{e^{20}}{20} \leq \frac{3^{20}}{20} = 1.5 \times 3^{18}.$$

3. We can get another crude approximation for the number of primes less than e^{20} by noting that $e^2 \approx 8$. That is,

$$\begin{aligned} \frac{e^{20}}{\ln(e^{20})} &= \frac{(e^2)^{10}}{20} \\ &\approx \frac{(8)^{10}}{20} \\ &= 4 \times 8^8 \\ &= 2^{26} \\ &= 64 \times (2^{10})^2 \\ &= 64 \times 1024^2 \\ &\approx 64 \times 10^6. \end{aligned}$$

There are approximately 64 million primes less than e^{20}. To see how big e^{20} is, note that $2.7 < e$, so that

$$e^{20} > (2.7)^{20} = ((2.7)^2)^{10} = (7.29)^{10} > (50)^5 = 312,500,000,$$

or over 312 million.

8.3 Products of Geometric Series

We are going to discuss the most famous of Mathematical problems called the *Riemann Hypothesis*. This problem is intimately connected with the *Prime Number Theorem* and involves the roots of an infinite sum called the *Riemann Zeta Function*. What we do is probably the way that Leonard Euler went about the problem.

The *sigma notation* is shorthand for writing down sums. It is especially useful when writing down infinite sums. It takes the guesswork out of seeing which pattern is meant by .... Σ means "add these terms up." For example,

$$\sum_{i=0}^{2} x^i = 1 + x + x^2,$$

and, more importantly, the geometric series $1 + x + x^2 + \ldots$ with infintely many terms can be written

$$\boxed{\sum_{k=0}^{\infty} x^k = 1 + x + x^2 + \ldots.}$$

We can even sum over two or more variables if we are careful. Thus, assuming that $i, j \geq 0$, we have

$$\begin{aligned}\sum_{i+j=3} x^i y^j &= x^0y^3 + x^1y^2 + x^2y^1 + x^3y^0 \\ &= y^3 + xy^2 + x^2y + x^3 \\ &= x^3 + x^2y + xy^2 + y^3.\end{aligned}$$

Another example is

$$\begin{aligned}\sum_{i+j+k=3} x^i y^j z^k &= x^3 + y^3 + z^3 \\ &+ x^2y + x^2z + y^2x + y^2z + z^2x + z^2y \\ &+ xyz.\end{aligned}$$

One more detail before we begin our multiplication. The *degree* of a term is the sum of the exponents (implied or stated) in that term. Thus

$$x^2y^3z^4$$

has degree

$$2 + 3 + 4 = 9,$$

while degree 3 terms look like this.

$$z^3, \quad x^2y, \quad xyz.$$

We discuss the product of polynomials by considering the terms of different degree in the polynomial. There is a pattern here that will help us to multiply geometric series together.

Consider the identity

$$(1 + x)(1 + y) = 1 + (x + y) + xy.$$

The term $x+y$ comes from the fact that we have multiplied all of the degree 1 terms by degree 0 terms (=constant terms).

$$1\cdot x + y\cdot 1 = x+y.$$

The term xy comes from the fact that we have multiplied together all those terms whose product has degree 2. There is only one such product in this example.

$$x\cdot y = xy.$$

Let us do the same thing for the identity

$$\begin{aligned}(1+x)(1+y)(1+z) &= 1\\ &+ (x+y+z)\\ &+ (xy+yz+xz)\\ &+ (xyz).\end{aligned}$$

The term $x+y+z$ comes from multiplying together those terms whose degree is 1 as follows.

$$1\cdot 1\cdot x + 1\cdot 1\cdot y + 1\cdot 1\cdot z = x+y+z.$$

The degree 2 term is found by multiplying together all those terms in this example whose product has degree 2. That term is

$$xy+yz+xz.$$

The degree 3 term is from the only product that yields a degree 3 term. This is the term

$$xyz.$$

Let us try this analysis one more time by considering the identity

$$\begin{aligned}(1+x+x^2)(1+y)(1+z) &= 1\\ &+ (x+y+z)\\ &+ (xy+yz+xz+x^2)\\ &+ (x^2y+x^2z+xyz)\\ &+ (x^2yz).\end{aligned}$$

The identity was found by using foil over and over again. Here is another way to find this identity.

The degree 0 term is found by multiplying together all the terms that add up to a degree 0 term. There is just one:

$$1 \cdot 1 \cdot 1 = 1.$$

The degree 1 term is found by multiplying all of the terms together that result in a degree 1 term:

$$1 \cdot 1 \cdot x + 1 \cdot 1 \cdot y + 1 \cdot 1 \cdot z = x + y + z.$$

The next term will be the degree 2 term. Multiply all of the terms together that result in degree 2. They are

$$xy + yz + xz + x^2.$$

The degree 3 term is next. The products resulting in degree 3 are degree 1 times degree 2 or three degree 1 terms together. They are

$$x^2y, \quad x^2z, \quad xyz,$$

so the degree 3 term is their sum.

$$x^2y + x^2z + xyz.$$

The degree 4 term is the only product that results in degree 4, namely, degree 2, degree 1, and degree 1:

$$x^2yz.$$

This completes the product.

So when multiplying we ask which products result in a degree 0 term, then a degree 1 term, then a degree 2 term, then a degree 3 term, and so on until the terms in the product are all used.

We call the infinite sum

$$1 + x + x^2 + x^3 + \dots$$

the *geometric series.* This geometric series is a number when $-1 < x < 1$ and only when x is in this domain, so it is important that we do not allow x to stray from this *interval of convergence.* We will multiply two geometric series together using the above pattern.

$$G = (1 + x + x^2 + \ldots)(1 + y + y^2 + \ldots).$$

By operating as we did above, the degree 0 term is

$$1 \cdot 1 = 1$$

and the degree 1 term is formed by multiplying degree 0 and degree 1 terms together. The result is

$$1 \cdot y + x \cdot 1 = x + y.$$

The degree 2 term is next. Combine all terms together whose degree is 2:

$$1 \cdot y^2 + x \cdot y + x^2 \cdot 1 = x^2 + xy + y^2.$$

The degree 3 term is found by counting degrees.

$$1 \cdot y^3 + x \cdot y^2 + x^2 \cdot y + x^3 \cdot 1 = x^3 + x^2y + xy^2 + y^3.$$

One more, the degree 4 term.

$$\begin{aligned} &1 \cdot y^4 + x \cdot y^3 + x^2 \cdot y^2 + x^3 \cdot y + x^4 \cdot 1 \\ &\qquad = x^4 + x^3y + x^2y^2 + xy^3 + y^4. \end{aligned}$$

We have thus found the first 4 terms of the product G.

$$\begin{aligned} G &= 1 \\ &+ (x + y) \\ &+ (x^2 + xy + y^2) \\ &+ (x^3 + x^2y + xy^2 + y^3) \\ &+ (x^4 + x^3y + x^2y^2 + xy^3 + y^4) \\ &\vdots \end{aligned}$$

In general, we can say that the n-th term in G is found by adding together all of the products whose degree is n. In terms of the sigma notation we have the following.

$$\text{The } n\text{th term of } G = \sum_{i+j=n} x^i y^j.$$

It is worth noting that if $(1 + x + x^2 + \ldots)$ and $(1 + y + y^2 + \ldots)$ converge (i.e., if $-1 < x, y < 1$), then the product G also converges. However, as we will show presently, in case there are infinitely many geometric series multiplied together, then the product of them need not converge to a number.

This is how we extend our products from two to three or more geometric series. Use the sigma notation to write down a geometric series. The rest of the notation is due to Isaac Newton. Let

$$\begin{aligned} A &= \sum_{i=0}^{\infty} a^i = 1 + a + a^2 + \ldots, \\ B &= \sum_{j=0}^{\infty} b^j = 1 + b + b^2 + \ldots, \\ C &= \sum_{k=0}^{\infty} c^k = 1 + c + c^2 + \ldots, \\ &\vdots \end{aligned}$$

be an infinite implied list of geometric sums. The product

$$\Gamma = A \cdot B \cdot C \cdot \ \ldots$$

is the infinite series whose nth term is

$$n\text{th term of } \Gamma \ = \sum_{i+j+k+\ldots=n} (a^i b^j c^k \ldots).$$

For example, the degree 0 term is from sums $i + j + k + \ldots = 0$, which happens only when $i = j = k = \ldots = 0$. Then the degree 0 term is

$$1.$$

The degree 1 term of Γ is the sum of all terms whose exponents add to 1. That is, the degree 1 term is the infinite sum

$$a^1 + b^1 + c^1 + \ldots = a + b + c + \ldots.$$

The degree 2 term, which is all we are going to do, is formed by taking products of degree 1 terms like ab or bc, and products of square terms with the only degree 0 term that exists here, 1. The degree 2 term is the infinite sum

$$\sum_{r+s=2} x^r y^s.$$

In this case $x, y \in \{a, b, c, \ldots\}$. Some of the terms in the degree 2 term of the product Γ are

$$\begin{aligned} \sum_{r+s=2} x^r y^s &= a^2 + b^2 + c^2 + \ldots \\ &+ ab + bc + ac + \ldots. \end{aligned}$$

Somewhere in the product Γ is a factor

$$D = \sum_{\ell=0}^{\infty} d^\ell = 1 + d + d^2 + \ldots.$$

It also contributes to the infinite sum that is Γ. Terms like d^5, abd^2, and $a^2c^3d^7$ occur when we write down Γ in all of its detail. The sigma notation helps to simplify our representation

$$\boxed{\Gamma(a, b, c, \ldots) = \sum_{n=0}^{\infty} \sum_{i+j+k+\ldots=n} (a^i b^j c^k \ldots)}$$

of Γ. We now have enough algebra to understand a little about the *Riemann Zeta Function.*

8.4 The Riemann Zeta Function

In this section, we investigate a special value of Γ called the *Riemann Zeta Function.* We will replace a, b, c, ... with specific values or numbers, and the resulting product will be untangled. Our stopping place will be what is considered to be the most important Mathematics problem of the last 150 years. It is called the *Riemann Hypothesis.*

Let $s > 1$ be a real number. As in the previous section, let us substitute $\frac{1}{2^s}$ for a in A, and call the result

$$A(\frac{1}{2^s}) = 1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \frac{1}{2^{3s}} + \cdots.$$

Let $b = \frac{1}{3^s}$ in B, and write

$$B(\frac{1}{3^s}) = 1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \frac{1}{3^{3s}} + \cdots.$$

Let $c = \frac{1}{5^s}$ in C, and write

$$C(\frac{1}{5^s}) = 1 + \frac{1}{5^s} + \frac{1}{5^{2s}} + \frac{1}{5^{3s}} + \cdots.$$

Continue inductively, forming reciprocals of s-powers of prime numbers

$$\frac{1}{p^s}$$

and substituting them into the associated geometric series:

$$D(\frac{1}{p^s}) = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots.$$

Each geometric series $1 + x + x^2 + \ldots$ can be written in closed form as

$$1 + x + x^2 + \ldots = \frac{1}{1-x}$$

when $-1 < x < 1$. For primes p, $-1 < \frac{1}{p^s} < 1$, so

$$D(\frac{1}{p^s}) = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \ldots = \frac{1}{1 - \frac{1}{p^s}}.$$

Making this substitution for A, B, C, ... yields the equation

$$\begin{aligned}\Gamma(\frac{1}{2^s}, \frac{1}{3^s}, \frac{1}{5^s}, \ldots) &= A(\frac{1}{2^s})B(\frac{1}{3^s})C(\frac{1}{5^s})\ldots \\ &= \prod_{\text{all primes } p} \frac{1}{1-\frac{1}{p^s}}, \end{aligned} \tag{8.1}$$

where the symbol

$$\prod_{\text{all primes } p}$$

means to take the product over all primes p. The first little bit of this formula is

$$\frac{1}{1-\frac{1}{2^s}} \cdot \frac{1}{1-\frac{1}{3^s}} \cdot \frac{1}{1-\frac{1}{5^s}} \cdots$$

and continues on indefinitely, one factor for each prime.

This infinite product is called the *Riemann Zeta Function*:

$$\zeta(s) = \prod_{\text{all primes } p} \frac{1}{1-\frac{1}{p^s}}.$$

This function $\zeta(s)$ is defined at all complex numbers except $s = 1$.

There is a deep connection between the roots or zeros of $\zeta(s)$ and the number of primes $\pi(n)$ less than n. The details of that connection are beyond the scope of this book, so we will just write down certain facts about $\zeta(s)$.

We can just as easily use complex numbers as real numbers in $\zeta(s)$ so we do just that. Let

$$s = u + iv,$$

where u and v are real numbers, and $i = \sqrt{-1}$, the imaginary unit. The next question that we come upon is simple enough. You have often been asked to find the zeros of a polynomial function, and that is what we ask here. *What are the zeros of the Riemann Zeta Function $\zeta(s)$?*

The Riemann Hypothesis 8.4.1 *The zeros of the Riemann Zeta Function have real part $\frac{1}{2}$. That is, if $\zeta(s) = 0$ and if $s = u + iv$ then $u = \frac{1}{2}$.*

As of 2010 an answer to this question has escaped the most careful of analysis. All of the zeros that have been found satisfy $u = \frac{1}{2}$, and it has been known for almost 100 years that there are infinitely many zeros that satisfy $u = \frac{1}{2}$,but that does not preclude the possibility that there might be a very large number $s = u + iv$ out there somewhere that is a zero of $\zeta(s)$, but that has $u \neq \frac{1}{2}$. The best computer models show that if $\zeta(s) = 0$ and if $u < 10^{14}$ then $u = \frac{1}{2}$. But this does not prove that all zeros of $\zeta(s)$ have real part $\frac{1}{2}$.

There is a story that circulates amongst the graduate students during the coffee hour at Mathematics departments in universities all over the world. It seems that the Devil came to Riemann and said that he would give him his heart's desire for his soul. You know the contract. It has become immortal since we read *The Devil and Daniel Webster*. Riemann, being Mathematically astute, said to the Devil that he would give up his soul provided that the Devil could give Riemann a correct proof of the Riemann Hypothesis in one year. The Devil agreed and vanished in a puff of smoke. Riemann went about his work. One year later to the day, the Devil returned to Riemann's office. He sadly announced that despite using all of the resources in his realm, he was unable to solve the problem Riemann had given him. "But," says he, "if I could just prove this one lemma...." **Moral of the story is this**: there are no world-class Mathematicians in hell.

Let us try a more traditional form of $\zeta(s)$. We can realize $\zeta(s)$ as an infinite sum as long as we begin with a real number $s > 1$. Recall that $\zeta(s)$ is a product of geometric series

$$\zeta(s) = A(\frac{1}{2^s}) \cdot B(\frac{1}{3^s}) \cdot C(\frac{1}{5^s}) \dots,$$

where there is one geometric series $D(\frac{1}{p^s})$ for each prime number p. The terms of the series $\zeta(s)$ are 1, the sum of the reciprocal primes

$$\sum_{\text{primes } p} \frac{1}{p^s},$$

and the sum of the degree 2 terms

$$\sum_{\text{primes } p} \frac{1}{p^{2s}} + \sum_{\text{primes } p \neq q} \frac{1}{(pq)^s}.$$

The higher degree terms are just too complicated to write down here.

We approach the sum from the opposite direction.

Consider the sum

$$\sum_{n=1}^{\infty} \frac{1}{n^s}.$$

We could add up these numbers by taking $n = 1, 2, 3, \ldots$, or we could add them up by taking their prime factorizations. Add up 1 and the numbers that have just one prime dividing them.

$$1 + \sum_{\text{primes } p} \frac{1}{p^s}.$$

Next, take the numbers that are of the form p^2, or that have exactly two primes dividing them. Such an n would look like

$$n = pq$$

for two primes $p \neq q$:

$$1 + \sum_{\text{primes } p} \frac{1}{p^s} + \sum_{\text{primes } p} \frac{1}{p^{2s}} + \sum_{\text{primes } p \neq q} \frac{1}{(pq)^s}.$$

These sums are the initial sums of $\zeta(s)$. We have therefore argued that

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}. \tag{8.2}$$

That is, the infinite product $\zeta(s)$ is an infinite sum.

A most unusual Mathematical coincidence occurs here. Remember that I have warned you about manipulating infinite sums that do not converge. Well, here is a good example of that. The geometric series

$$D\left(\frac{1}{p}\right) = 1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \ldots$$

converges since

$$-1 < \frac{1}{p} < 1 \text{ for primes } p.$$

Does the infinite product

$$\zeta(1) = A\left(\frac{1}{2}\right) \cdot B\left(\frac{1}{3}\right) \cdot C\left(\frac{1}{5}\right) \cdots$$

of convergent geometric series converge? No. You see, if we use the infinite series representation (8.2) of $\zeta(s)$, we see that

$$\zeta(1) = \sum_{n=1}^{\infty} \frac{1}{n},$$

which is the divergent *harmonic series.* That should raise a few eyebrows. Certain infinite products of numbers do not converge. In this way, infinite products of numbers are similar to infinite sums of numbers. Some converge, some diverge.

Exercise 8.4.2 1. Find the degree 5 term in the product $(1 + x + x^2 + x^3 + x^4)(1 + y + y^2 + y^3)$.

2. Find the degree 4 term in the product $(1 + x + x^2 + x^3)(1 + y + y^2)(1 + z)$.

3. Find the degree 6 term in the infinite product $(1 + x^2)(1 + y^2)(1 + z^2) \cdots$.

4. We know that $\zeta(s) = A(a^s)B(b^s)C(c^s)D(d^s) \cdots$ is a product of power series. See (8.1).

(a) Find the infinite series of sums of degree 3 as in the discussion on page 305.

(b) Find the infinite series of sums of degree 4 as in the discussion on page 305.

8.5 Real Numbers

The previous section shows us how Mathematicians handle infinite sets of integers in a finite manner. In this section, we will see how Mathematicians handle the continuum of the real numbers in a finite manner. For example, there is an infinite set called a *sequence.* An example of a sequence is

$$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots.$$

The sequence is infinite, but it takes place in a finite interval $[0, 1]$. Moreover, there is a number 0 such that each *open neighborhood* of 0 contains infinitely many of the terms in this sequence. *An open neighborhood of 0* is an open interval $(-a, a)$ that contains 0. Other open neighborhoods of 0 are $(-1, 1)$, $(-2, 2)$, and $(\frac{-1}{2}, \frac{1}{2})$.

Another sequence is formed by allowing

$$x_n = \frac{(-1)^n}{2^n}$$

for each integer $n \geq 0$. The first few terms are

$$\frac{1}{1}, \frac{-1}{2}, \frac{1}{4}, \frac{-1}{8}, \ldots.$$

They oscillate between positive and negative numbers. The entire sequence is contained in the finite interval $[-1, 1]$. Moreover, there is a number 0 not in the sequence with a special property. Given any open neighborhood of 0, say, $(-a, a)$, there is a point beyond which all of the terms in the sequence are in this neighborhood. That is, there is an integer $N > 0$ such that

$$\frac{(-1)^n}{2^n} \in (-a, a) \text{ for all integers } n > N.$$

An open neighborhood of 0 is like having an infinite queue of people and saying that each of them lives on the same block of the city. A smaller open neighborhood of 0 would be one home. We would be saying that almost all of these people live in the same home. An even smaller open neighborhood would be a room in the home. All but finitely many of these people live in the same room. I won't get into the social implications that must come up when the open neighborhood of 0 is a chair in the room.

Let us be a bit more precise. A sequence in the closed interval $[-1, 1]$ is a function $f : \mathbb{N} \longrightarrow [-1, 1]$ from the natural numbers into

$[-1, 1]$. The images of the function or sequence are usually denoted with a subscript as in

$$f(n) = x_n \text{ for each } n \in \mathbb{N} .$$

For instance, $f(1) = x_1$, $f(2) = x_2$, and so on. The sequence with terms $\frac{(-1)^n}{2^n}$ is given by the function $f : \mathbb{N} \longrightarrow [-1, 1]$ with rule

$$f(n) = \frac{(-1)^n}{2^n}.$$

Let us write down an arbitrary sequence in $[-1, 1]$.

$$x_0, x_1, x_2, \ldots ,$$

of different terms. This is an infinite subset of $[-1, 1]$ so there should be something finite that we can say about the sequence. Cut the interval in half. Is it possible that both halves contain a finite number of sequence elements in them? No. That would mean that the sequence is finite, which it is not. Then one half contains an infinite number of sequence elements. So let I_1 be the half of the interval that contains infinitely many sequence elements. Now cut I_1 in half. As we have argued above, one half must contain infinitely many elements of the sequence since otherwise I_1 contains only finitely many sequence elements, a contradiction. Then let $I_2 \subset I_1$ be the half that contains infinitely many of the sequence elements. Continue in this way, constructing a chain of sets

$$I_1 \supset I_2 \supset \ldots ,$$

each term half as big as the previous term. The lengths of the intervals form a sequence

$$1, \frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^3}, \ldots .$$

Observe that these lengths converge or approach 0 as n is allowed to grow.

Now, we can write

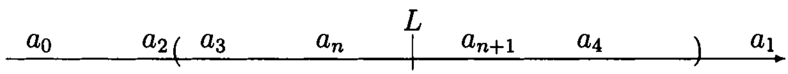
$$I_k = (a_k, b_k) \text{ for some real numbers } a_k < b_k.$$

In this case, we have a sequence of real numbers

$$a_1 < a_2 < a_3 < \ldots < b_3 < b_2 < b_1. \tag{8.3}$$

Then the sequence $\{a_1, a_2, a_3, \ldots\}$ is a sequence bounded above by b_1. That is, there is only one direction for the sequence, namely, upward, and there is a number larger than all of the sequence elements. We say that $\{a_n \mid n \in \mathbb{N}\}$ is a *bounded monotonic sequence.* One of the primary properties of the real line is that such a sequence $\{a_n \mid n \in \mathbb{N}\}$ has a limit. That is, there is a number L such that each open neighborhood $\mathcal{O}$ of L contains all but finitely many of the a_n. Pictorially we have

Almost all sequence elements are
in this open neighborhood of L.



The picture shows us that almost all (= all but a finite number) of the sequence elements a_n are in the open neighborhood $\mathcal{O}$ of L. The sequence $\{b_n \mid n \in \mathbb{N}\}$ has the same property around L. Then given an open interval $\mathcal{O}$ about L, there is a point where

$$(a_n, b_n) \subset \mathcal{O}$$

for almost all of the open intervals $(a_n, b_n) = I_n$. Since I_n is chosen so that I_n contains infinitely many elements of $\{x_n \mid n \in \mathbb{N}\}$, we see that $\mathcal{O}$ contains infinitely many elements in $\{x_n \mid n \in \mathbb{N}\}$. We call such a point L a *cluster point* of $\{x_n \mid n \in \mathbb{N}\}$.

At this point, we summarize.

The Bolzano-Weierstrauss Theorem 8.5.1 *Let* $\{x_n \mid n \in \mathbb{N}\} \subset [-1, 1]$ *be a sequence. There is a cluster point* L *in* $[-1, 1]$ *for the sequence.*

What did we know about the sequence $\{x_n \mid n \in \mathbb{N}\}$? Only that it was a sequence in $[-1, 1]$. That is precious little to know about a sequence before we determine that it has a cluster point, a point that is very close to infinitely many of the elements in the sequence. Very little hypothesis gave us a powerful result. That is one of the aims or goals of Mathematical research. When less gives us more, we know that we are onto something special.

Let me share some stories surrounding Karl Weierstrauss (circa 1880 AD). Unlike many Mathematical giants, Karl did not start out as a strong Mathematics student. Karl was a late bloomer. He spent his college life in pursuit of personal joy. Most of his young life was misspent fencing and quaffing beers in bars. His pursuit of fun distracted him from his studies. He was a party animal. Because of his party attitude, he barely graduated college. Upon graduation, he was denied a job at the local institutions and accepted a job instead teaching at an elementary school. It was here that he hit upon the sense of rigor and creativity that we remember him for today. In a change of image, he set new standards for Mathematical rigor that are still in effect more than 100 years later. Just as curious is the fact that in contrast to his earlier work, Karl's lectures and teaching are renowned for the care taken in their presentation and preparation. His gifts to his students are another unique quality for Karl. He would begin a Mathematical gem and then allow a student or a colleague to finish it, thus giving them the credit. This is a very generous and quite uncommon behavior. Mathematics cannot understate Karl's foundational contributions to teaching, analysis, and rigor in Mathematics.

Let us use the recursive argument used in the proof of the Bolzano-Weierstrauss Theorem in a different setting. We will use it to hunt tigers on the Arabian Desert. Your job is to find a Tiger in the otherwise barren Arabian Desert. You have all of the material you need, so you first erect an impassible fence around the Desert. Call the contained area S_1. You divide S_1 into two equal regions using the impassible fencing. To find the Tiger, you send men out into each half with orders to engage the Tiger in a fight. The men on one side return saying they could not find the Tiger, and the men on the other side do not come back. Call their side S_2. It contains the Tiger. Divide S_2 in half using impassible fencing, and send the men out again. The side S_3 is the one whose men do not return. The Tiger is there. We have started an induction that will construct a descending chain of regions

$$S_1 \supset S_2 \supset S_3 \supset \dots,$$

whose areas are given by a sequence

$$1 > \frac{1}{2} > \frac{1}{2^2} > \frac{1}{2^3} > \dots$$

that becomes vanishingly small. Eventually we will construct a region S_n that contains the Tiger and whose area is $\frac{1}{2^n}$. This power of $\frac{1}{2}$ will be so small that we can see the Tiger in the region S_n. At this point, we can say that we have corralled the Tiger.

Some might say that we have no way of knowing that S_n is small enough to cage the Tiger. There is less than 10^{12} square miles in the Arabian Desert, so on the 41st attempt to find the Tiger we have a region of area

$$\begin{aligned} \frac{1}{2^{40}} 10^{12} &= \frac{1}{2^{40}}(1000)^4 \\ &< \frac{1}{2^{40}}(1024)^4 \\ &= \frac{1}{2^{40}}(2^{10})^4 \\ &= 2^{-40} \cdot 2^{40} \\ &= 1 \text{ mile.} \end{aligned}$$

Another few fences and we have corralled the Tiger.

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

Chapter 9

Logic and Meta-Mathematics

9.1 The Collection of All Sets

Let

F

denote the *Collection of all Finite Sets.* Since **F** contains $\{1\}$, $\{1, 2\}$, $\{1, 2, 3\}$, ..., the collection **F** is infinite. Put another way,

The collection **F** of all finite sets is not finite.

While this is not a surprise, it prepares us for similar ideas concerning deeper implications about more abstract collections.

Example 9.1.1 The following is a classic example due to Bertrand Russell. Let

C

be the *Collection of All Sets.* We will prove the following.

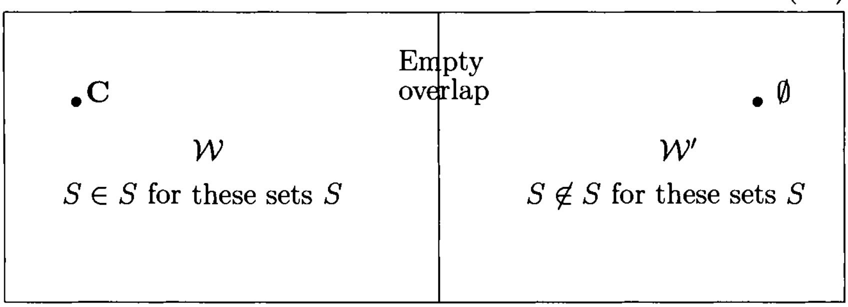
Theorem 9.1.2 **C** *is not a set.*

Proof: Assume for the sake of contradiction that **C** is a set. Then **C** has the rather unsettling property

$$\mathbf{C} \in \mathbf{C}.$$

That is, **C** is an element of itself. This is like saying that a bag of sand is itself a grain of sand (bag and all), that this book is contained on one page of this book, or that a crowd of people is a person. And yet, although unsettling, $\mathbf{C} \in \mathbf{C}$ is a consequence of the assumption that **C** is a set. This unsettling statement will lead us to a wonderful contradiction. Picture (9.1) will help with the argument that follows.

The Collection of All Sets **C** (9.1)



To produce a contradiction, we will do something that may strike you as familiar. Define a set $\mathcal{W} = \{\text{sets } S \mid S \in S\}$. That is,

$\mathcal{W}$ = the set of all sets S that contain themselves as an element.

The complement of $\mathcal{W}$

$$\mathcal{W}' = \{\text{sets } S \mid S \notin S\}.$$

You might be more comfortable with $\mathcal{W}$ if you note that by our assumption, **C** is a set such that $\mathbf{C} \in \mathbf{C}$. Then $\mathbf{C} \in \mathcal{W}$. As for elements of $\mathcal{W}'$, the empty set $\emptyset$ satisfies $\emptyset \notin \emptyset$ so $\emptyset \in \mathcal{W}'$. In fact, most sets S satisfy $S \notin S$ so $\mathcal{W}'$.

We observe that given a set S, either $S \in S$ or $S \notin S$. Then either $S \in \mathcal{W}$, or $S \in \mathcal{W}'$, but not both. Thus, $\mathcal{W}$ and $\mathcal{W}'$ are sets. Let us determine where $\mathcal{W}'$ resides.

1. Assume that $\mathcal{W}' \in \mathcal{W}'$. Then, by the predicate that defines $\mathcal{W}'$, $\mathcal{W}' \notin \mathcal{W}'$. This is a contradiction to the fact that $\mathcal{W}'$ is a set, so our assumption is in error.

2. Alternatively, assume that $\mathcal{W}' \notin \mathcal{W}'$. By the predicate defining $\mathcal{W}'$, we must have $\mathcal{W}' \in \mathcal{W}'$. This is another contradiction to the fact that $\mathcal{W}'$ is a set.

But because $\mathcal{W}'$ is a set, exactly one of the assumptions in items 1 and 2 must be True. This contradiction shows us that our assumption that **C** is a set is in error. We conclude that

$$\boxed{\mathbf{C} \text{ is not a set.}}$$

The above proof is similar to the proofs we used to prove that the Epimenides Paradox is actually not paradoxical at all. We examined the statement Q: *This statement is not a lie*, and found that it was a lie and not a lie. That is, Q is both a Falsehood and a Truth, a contradiction.

In the above, we assume for the sake of contradiction that **C** is a set, and we consider the Mathematical statement $\mathcal{W}' \in \mathcal{W}'$. On the one hand, item 1 in the proof shows that $\mathcal{W}' \in \mathcal{W}'$ is a Falsehood. On the other hand, item 2 in the proof shows that $\mathcal{W}' \notin \mathcal{W}'$ is a Falsehood, or in other words, that $\mathcal{W}' \in \mathcal{W}'$ is True. Therefore, **C** is not a set. The proof used is exactly the proof that *All Cretans are liars* is a Falsehood.

This is a good time to demonstrate that $\text{card}(\{\bullet\})$ is not a set since it contains a copy of **C**. It contains $\{X\}$ for each set X.

Now that we have established that **C** is not a set, here are several mind expanding shockers.

Theorem 9.1.3 *If B is a set, then there are no functions $f : \mathbf{C} \longrightarrow B$.*

Proof: Functions are defined on sets and sets only. Since $\mathbf{C}$ is not a set we cannot define a function on it.

Theorem 9.1.4 $\mathbf{C}$ *has no cardinality.*

Proof: Suppose to the contrary that $\mathbf{C}$ has cardinality $\aleph$. Then there is a set B and a bijection $f : \mathbf{C} \longrightarrow B$ such that $\text{card}(B) = \aleph$. This is contrary to the previous theorem. This contradiction shows us that $\mathbf{C}$ has no cardinality.

Even though $\mathbf{C}$ has no cardinality there is a more philosophical way of seeing that $\mathbf{C}$ is larger than any Mathematical construction. Let A be a set. Form the subcollection

$$\bar{A} = \big\{\{x\} \,\big|\, x \in A\big\}.$$

Then $\bar{A}$ is a set and

$$\text{card}(A) = \text{card}(\bar{A}).$$

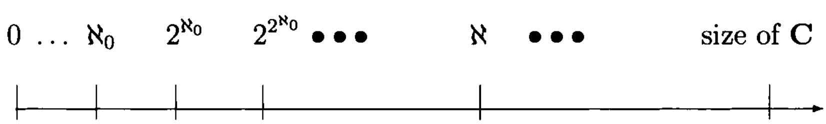
Furthermore,

$$\bar{A} \text{ is a subcollection of elements in } \mathbf{C},$$

so that in an intuitive but compelling way, $\bar{A}$ is smaller in size than $\mathbf{C}$. Since $\text{card}(A) = \text{card}(\bar{A})$, we make the philosophical conclusion that

$$\boxed{\text{each set } A \text{ is smaller in size than } \mathbf{C}.}$$

If we did not know about Theorem 9.1.4, we might be tempted to say that the cardinality of $\mathbf{C}$ is larger than every cardinality. But, of course, $\mathbf{C}$ has no cardinality. In pictures,



9.2 Other Than True or False

Have you ever engaged in logical games in which you had to decide who was telling the Truth given some information or statements made by the people involved? My favorite is the classic chestnut *The Hidden Tiger Puzzle.* It goes like this.

Example 9.2.1 The Hidden Tiger Puzzle: You are on an island where the island ruler asks you to make a choice. Behind one door is a Golden Key that will make you the ruler of the island. Behind the other door is a yellow and black striped tiger who will eat you if the door is opened. The two doors are guarded by two men who know where the Golden Key is located. You may ask one question of the guards in order to find the Golden Key. Here is the twist. One of these guards tells the Truth all of the time, and the other one lies all of the time. What one question can you ask that will lead you to the Golden Key? The answer to this question is given on page 326.

The answer to *The Hidden Tiger Puzzle* relies on the fact that most statements have exactly one of two logical states. Namely, they are True or False, but not both, and not neither. No third alternative is considered. In this section, we will give examples of statements that will lead us to conclude that certain statements in the English language have a third possible logical state.

The point in the examples given in this section is that there are sentences in the English language whose logical state cannot be decided between the traditional two, True and False. We suggest that their logical state is something *other than* True of False.

The first example we offer of a statement that is neither True nor False is the statement

Liar's Paradox Q: This statement is False,

which is sometimes referred to as *Satan's Statement.* Over the years, various people have tried to resolve Q's logical state. The most recent work seems to show that Q is False all of the time. That is, Q is False in any context. But, this is contradicted by the simple setting

All statements are False.
Q: This statement is False.

It would seem that the Liar's Paradox Q is True when preceded by a statement that declares that all statements are False. Let us examine the logical states of the Liar's Paradox as motivation to further examine logical states other than True or False.

See page 22 for a review of recursive statements, their logically equivalent forms, and their logical peculiarities. So let Y be a property of a statement. Then Y could be one of, but not necessarily restricted to, the properties *True, valid, recursive, known, hard to read, impossible to know.* Let Q be the statement *This statement has type* Y. The following is a list of statements that are the same as Q.

1. Q.
2. *This statement has type* Y.
3. *This statement* Q *has type* Y.
4. Q *has type* Y.

In the discussion that follows, we will freely substitute one of these four for another.

Theorem 9.2.2 *Let Q be the statement* This statement is False. *Then Q is neither True nor False.*

Proof: Assume for the sake of contradiction that Q is either True or False.

CASE 1: Suppose that Q is True. By item 4 above, *Q is False* is True, so Q is False. This contradicts CASE 1, so Q is not True.

CASE 2: Suppose that Q is False. By item 4 above, *Q is False* is False, so Q is True. This contradicts CASE 2, so Q is not False.

We conclude that Q is neither True nor False, contrary to our initial assumption. Thus, the statement of the Theorem is True. This completes the proof.

We will say that a word W is *self-descriptive* if the statement

W is a W word

is True. The word W is said to be *non-self-descriptive* if the statement *W is a W word* is False. The word W is either a self-descriptive word, or it is a non-self-descriptive word, but not both, and not neither.

Example 9.2.3 1. Since the word *pentasyllabic* has five syllables, the sentence

Pentasyllabic is a pentasyllabic word

is True. Thus *pentasyllabic* is a self-descriptive word. This demonstrates that there are self-descriptive words.

2. There are many non-self-descriptive words. For example, since *monosyllabic* means having one syllable, the sentence

Monosyllabic is a monosyllabic word

is False, whence *monosyllabic* is a non-self-descriptive word.

Example 9.2.4 We will show that there is a word W in the English language such that the sentence *W is a W word* is neither True nor False, but not both. We conclude that it must possess some alternative logical state.

Proof: Assume for the sake of contradiction that the statement *W is a W word* is either True or False, and not both. So what kind of word do you think *non-self-descriptive* is? Make your guess and then read on.

Non-self-descriptive is the contradiction we seek. We will show that the sentence

NSD	Non-self-descriptive is a non-self-descriptive word

is neither True nor False, thus contradicting our assumption that the statement W *is a* W *word* is either True or False, but not both.

CASE 1: The statement *NSD* is True. Then by the definition of a *non-self-descriptive* sentence, *NSD* must be False. This violates our assumption that *NSD* is not both True and False.

CASE 2: The statement *NSD* is False. By the definition of *non-self-descriptive*, the word non-self-descriptive is a *non-self-descriptive* word. That is, sentence *NSD* is True, contrary to Case 2.

Since each logical state leads us to a contradiction, we conclude that

The logical state of sentence *NSD* is neither True nor False.

We have found a simple statement that seems to suggest that there are other logical states besides True and False. These statements are also quite simple, leading us to believe that this kind of counterintuitive logical behavior is actually quite common.

Consider what the above example does for the set of words in the Engish dictionary. Let **S** denote the collection of *self-descriptive* words, and let **NS** denote the collection of *non-self-descriptive* words. According to the above example, **S** and **NS** do not account for all of the words in the English language. There is a third collection.

O = Words W for which W *is a* W *word* has a logical state *other than* True or False.

Of course, one such third logical state would be utter nonsense.

In the future, if we wish to say that a word W *is not in* **S**, we will write

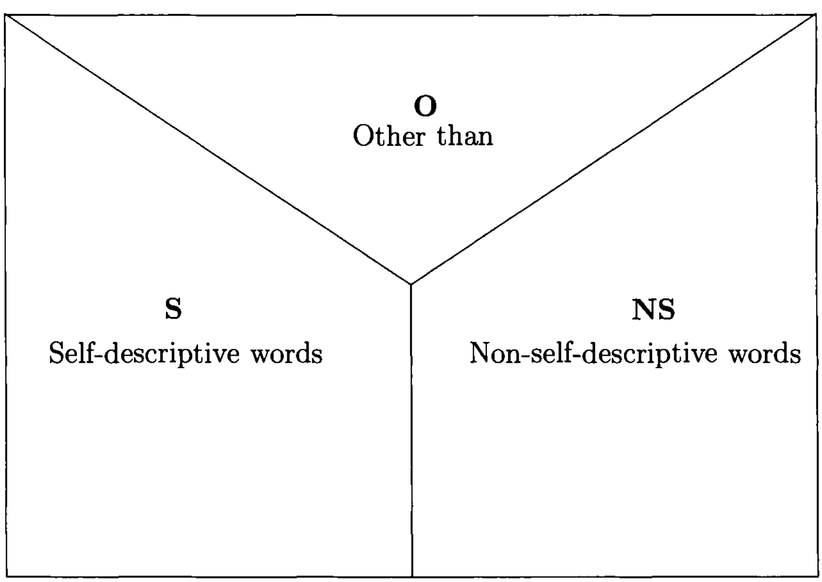
We cannot say that W is a self-descriptive word.

This kind of wording allows for the fact that W might be in **NS** (non-self-descriptive words), or it might be in **O** (other than self-descriptive or non-self-descriptive words).

What went wrong here? Why should the simple statement *NSD* give us so much trouble regarding its logical state? The number of words for which the statement *NSD* is True is not large. It is something you might program your desktop computer to find if it could determine when a word was self-descriptive or non-self-descriptive. Just download the CD that contains the entire Oxford English Dictionary, and perform a search. Since this CD is finite, the number of words for which statement *NSD* is True is finite. However, there is a word *M* such that the statement *M is an M word* is neither True nor False. Its logical state is some unknown third alternative state. It is possible that the missing third state is that the statement *NSD* has *no logical state whatever.* This would make *M* like *Hello* or *Good-bye*, or *Yeah.* Or perhaps, as we indicated by naming it that way, the logical state is something *other than* True of False. We will not pursue the matter here.

Picture (9.2) partitions the English dictionary. It might help you to envision what we are talking about when we talk about three possible logical states.

Dictionary of the English Language (9.2)



The student's first introduction to paradoxical logical behavior of statements is the *Epimenides Paradox.* (See Example 1.7.1.) Within the Epimenides Paradox is the statement *I am lying*, whose logical state is to be established within the context of the Paradox. Traditionally, one concludes that *I am lying* is *paradoxical* in the context of the Paradox. It is neither True nor False. But like the Liar's Paradox, in certain conversations, *I am lying* possesses a well defined logical state. Take for example the conversation

$$\text{True} = \text{False}$$
$$\text{I am lying.}$$

Because *I am lying* modifies the False statement *True = False*, *I am lying* is True. The next result shows that *I am lying* is logically, very similar to the Liar's Paradox: *This statement is False.*

Lemma 9.2.5 This statement is False *is logically equivalent to* I am lying. *Thus,* I am lying *is neither True nor False.*

Proof: In speaking *This statement is False*, one declares that the spoken statement is a lie. In other words, one declares that he is lying, or equivalently he speaks *I am lying.*

Conversely, suppose that one speaks P: *I am lying.* Because P asserts that the statement P is False, it declares that *This statement is False.*

Hence, *This statement is False* is logically equivalent to *I am lying.* Theorem 9.2.2 then shows that *I am lying* is neither True nor False, which complete the proof.

The following classic story is another example of a statement whose logical state is neither True nor False. We will update the story.

Example 9.2.6 In the Epimenides Paradox in Example 1.7.1, Epimenides of Crete enters a forum in which Greek scholars are meeting. He announces *I am a Cretan* and *All Cretans are liars.* He then states that *I am lying.* In Example 1.7.1, we showed that *I am lying* is neither True nor False. Let us see why his deduction is suspect, even though its conclusion agrees with Lemma 9.2.5.

We first note that since *All Cretans are Liars*, any statement spoken by Epimenides is False. Thus, *All Cretans are liars* is False. This is the reason why in the Epimenides Paradox, we cannot deduce the logical state of *I am lying*. The problem proceeds from the False premise *All Cretans are Liars*, so Truth cannot be deduced for subsequent conclusions. In other words, we can deduce the statement "*I am lying* is neither True nor False," but there is no way to conclude the Truth of it in the context of the Epimenides Paradox.

Here is another example of alternative logical states that will really mess with your mind.

Example 9.2.7 Consider the following two sentences.

#1.	Sentence 2 is False.
#2.	Sentence 1 is True.

We will show that the logical state of Sentence #1 is neither True nor False.

Proof: CASE 1: Assume that Sentence #1 is True. Then *Sentence 2 is False* is a True sentence, so that sentence #2 is False, or equivalently, that *Sentence 1 is True* is False. It must be then, that sentence #1 is False, which contradicts Case 1. Thus sentence # 1 is not True.

CASE 2: Assume that sentence #1 is False. Then *Sentence 2 is False* is a False sentence, so that sentence #2 is True. Then it is True that sentence #1 is True, which contradicts Case 2. Thus sentence # is not False.

We conclude that sentence #1 is neither True nor False. This suggests that the logical state of sentence #1 is some alternative third logical state. This completes the proof.

Statements #1 and #2 are not terribly complicated, and yet they quickly lead us to a difficult logical situation. We conclude that not all of our language can be analyzed with traditional binary logic. That is, we cannot investigate the language without recognizing the existence of simple statements whose logical states are something other than True or False.

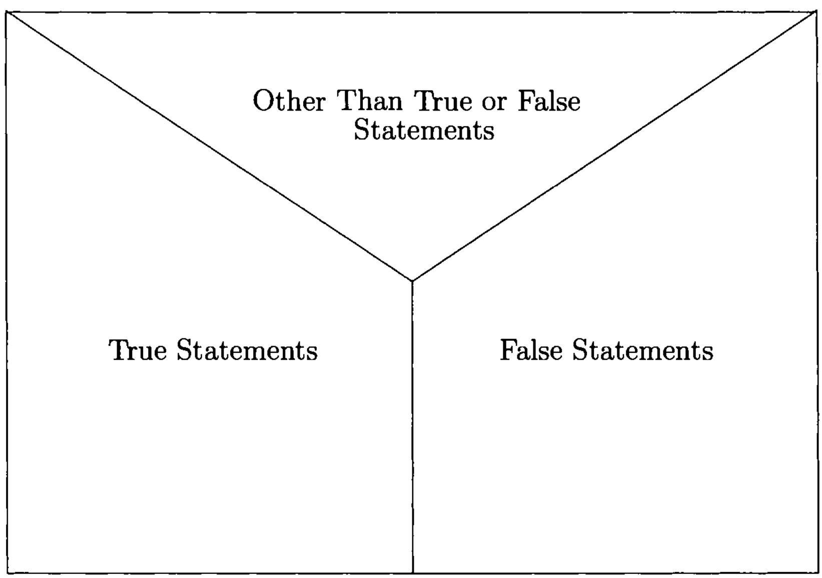
I offer the following two problems as entertainment for you, dear reader. In these problems, please determine the logical state of statement #1.

(1) #1. Sentence 2 is False.
#2. Sentence 1 is False.

(2) #1. Sentence 2 is True.
#2. Sentence 1 is True.

A diagram of English might help you understand how we have partitioned the statements in the language.

Statements of the English Language (9.3)



The solution to the Hidden Tiger Puzzle: To find the door that contains the Golden Key, choose one of the guards. It doesn't matter which one. Ask him, *Which door would the other guard say that the Golden Key is behind?*. That is one question. Then open the other door. That is where you will find the Golden Key.

Now here is how it works. Label the guards 1 and 2. Suppose you ask the one question of guard 1. If he lies all of the time, then guard 2 tells the Truth all of the time. Guard 1 will change guard 2's answer from the Truth to a Falsehood. Guard 1 will answer with the wrong door, the door hiding the Tiger. Open the other door.

On the other hand, suppose guard 1 tells the Truth all the time. Then guard 2 lies all the time. Guard 2's answer to the question would be the door that does not have the Golden Key behind it. Guard 1, being the one who tells the Truth, will answer with guard 2's answer, which is the door that hides the Tiger. Open the other door.

9.3 No Theory of Everything

In this section, we discuss the existence of a theory of everything in Logic and the sciences, and how it relates to the foundations of Mathematics by extending Gödel's Incompleteness Theorem to a set **C** of True statements. In particular, if **C** is a set of True statements from a logical system, then we show that there is a statement Q that is True over **C**, and that is not deducible from **C**. The main theorem follows from a generalization of Gödel's Incompleteness Theorem to the larger class of sets of True statements. Then given any set **K** of written True statements, there is some statement Q that cannot be deduced from consequences of the statements in **K**. As an application, we show that a theory of everything does not exist for Logic, Mathematics, Physics, Mathematical Physics, or for any other scientific endeavor.

9.3.1 Gödel's Incompleteness Theorem

We will not prove it, but we will state Gödel's Incompleteness Theorem. This is a statement that shows that no matter how much Mathematics you write, there will always be more Mathematics to write. It shows that if you think you have discovered all of Mathematics then you have committed an error. Gödel's Incompleteness Theorem is considered to be one of the fundamental Mathematical results of the twentieth century.

Let S be a logical system. This is just a collection of logical expressions on which we can operate with the logical statements and, or, not to produce more logical expressions. For example, Logic, Mathematics, Physics, any area of science, and any written rational language form a logical system.

The natural numbers $\mathbb{N}$ have an axiomatic basis called the *Peano Axioms* of the Natural Numbers. These are four statements of Mathematics from which we can build a set P that looks and acts like $\mathbb{N}$ and its elements. Mathematicians agree that The Peano Axioms are the most effective Axioms describing $\mathbb{N}$. For instance, the fourth Peano Axiom is Mathematical Induction, which we met in Section 7.1.

Let R be a set contained in the logical system S. We say that R *contains all of arithmetic* if R contains the Peano Axioms. We say that R is *incomplete* if there is a statement Q $\notin R$ that is proved from some finite argument of statements in R. If R is not incomplete, then we say that R is *complete.* We say that R is *inconsistent* if there is a statement Q$\in R$ such that not $Q \in R$. If R is not inconsistent, then R is *consistent.* A statement Q is *True over* R if not Q$\notin R$.

Evidently, the consistent incomplete sets R are the most interesting sets in a logical system since no mistakes have been made in R, and since there is room to prove statements that are not in R. In other words, there is more to prove than just R.

With this terminology, we can state Gödel's Incompleteness Theorem 9.3.1. See [9] for a reference.

Theorem 9.3.1 *Let R be a set contained in a logical system S that contains all of arithmetic. Then exactly one of the following conditions is True.*

1. **Gödel's Incompleteness Theorem:** *If R is consistent then R is incomplete.*

2. **Gödel's Inconsistent Theorem:** *If R is complete then R is inconsistent.*

We conclude that under the hypotheses of Gödel's Theorem 9.3.1, either you have more Mathematics to prove, or you have made a mistake somewhere in your arguments. In what follows, we will attempt to extend this idea to a larger class **C** of statements.

9.3.2 Logically Closed Sets

Fix a set of True statements $\mathbf{C}$ from logical system, and let $X \subset \mathbf{C}$. Let $\mathrm{dd}(X) = \{E \mid E$ is the conclusion of an argument consisting of finitely many statements, each of which is contained in $X\}$. Inasmuch as E is either deduced from X in a finite number of steps or it is not, and because X is a set, $\mathrm{dd}(X)$ is a set.

Let $\mathrm{dd}(\mathbf{C}) = \mathrm{dd}^1(\mathbf{C})$, and inductively define $\mathrm{dd}^{n+1}(\mathbf{C}) = \mathrm{dd}(\mathrm{dd}^n(\mathbf{C}))$ for each integer $n > 0$. The *logical closure of* $\mathbf{C}$ is the set

$$\overline{\mathbf{C}} = \bigcup \left\{\mathrm{dd}^n(F) \mid F \subset \mathbf{C} \text{ is a finite set and } n > 0 \text{ is an integer}\right\}.$$

Then $\mathbf{C}$ is said to be *logically closed* if $\mathbf{C} = \overline{\mathbf{C}}$. We prove in Theorem 9.3.2 that a logically closed set of True statements is not complete in the sense of Gödel's Theorem. A statement Q is *True over* $\mathbf{C}$ if $\neg\mathrm{Q} \notin \overline{\mathbf{C}}$. If $\mathrm{Q} \in \overline{\mathbf{C}}$ then Q is *recursively deducible from* $\mathbf{C}$.

Theorem 9.3.2 *Suppose that* $\mathbf{C}$ *is a set of True statements from a logical system. If* $\mathbf{C}$ *is logically closed, then there is a statement* Q *that is True over* $\mathbf{C}$ *but such that* $Q \notin \mathbf{C}$.

Proof: Let Q be the statement

Q: *This statement is not in* $\mathbf{C}$.

Because Q is self-referential, we observe that the statements (i) Q, (ii) *This statement is not in* $\mathbf{C}$, (iii) *This statement Q is not in* $\mathbf{C}$, (iv) *Q is not in* $\mathbf{C}$, and (v) $\mathrm{Q} \notin \mathbf{C}$ are logically equivalent.

Suppose for the sake of contradiction, that $\mathrm{Q} \in \mathbf{C}$. By our choice of $\mathbf{C}$, Q is True, so by (v), $\mathrm{Q} \notin \mathbf{C}$ is True. However, because $\mathbf{C}$ is a set, $\mathrm{Q} \in \mathbf{C}$ and $\mathrm{Q} \notin \mathbf{C}$ form a contradiction. Consequently, $\mathrm{Q} \notin \mathbf{C}$ is True. Then by (v) and (i), Q is True. Because $\mathbf{C}$ consists of True statements, $\neg\mathrm{Q} \notin \mathbf{C}$. Therefore, Q is True over $\mathbf{C}$, which completes the proof.

The next two results generalizes Gödel's Incompleteness Theorem 9.3.1 to sets $\mathbf{C}$ of True statements from a logical system.

Theorem 9.3.3 *Suppose that* $\mathbf{C}$ *is a set of True statements from a logical system. Then, there is a statement* Q *that is True over* $\mathbf{C}$*, but that is not recursively deducible from* $\mathbf{C}$.

Proof: Let $\mathbf{C}$ be a set of True statements. Then $\overline{\mathbf{C}}$ is a logically closed set of True statements. By Theorem 9.3.2, there is some statement Q that is True over $\overline{\mathbf{C}}$, and such that $Q \notin \overline{\mathbf{C}}$. Hence, Q is True over $\mathbf{C}$, and Q is not recursively deducible from $\mathbf{C}$. This completes the proof.

Corollary 9.3.4 Gödel's Incompleteness Theorem *Suppose that* $\mathbf{C}$ *is a logical system that contains all of arithmetic. If* $\mathbf{C}$ *is consistent then* $\mathbf{C}$ *is incomplete.*

Proof: If $\mathbf{C}$ is consistent, then $\mathbf{C}$ is a set of True statements in some model of the logical system. Then by Theorem 9.3.3, there is a statement Q that is True over $\mathbf{C}$, but that is not recursively deducible from $\mathbf{C}$. The existence of Q shows that $\mathbf{C}$ is incomplete, and completes the proof.

Theorem 9.3.5 *Let* $\mathbf{K}$ *be a set of written True statements. Then there is a statement Q that is Q is True over* $\mathbf{K}$*, and that is not recursively deducible from* $\mathbf{K}$*. Specifically,* $Q \notin \overline{\mathbf{K}}$.

Proof: Apply Theorem 9.3.3 to $\mathbf{K}$ to produce a statement Q that is True over $\mathbf{K}$, and that is not recursively deducible from $\mathbf{K}$. Inasmuch as any statement in $\mathbf{K}$ is True and the conclusion of a trivial argument, $Q \notin \mathbf{K}$. This completes the proof.

For example, if $\mathbf{K}$ is a set of written True statements from the set of existing writings to date, then there is a statement Q that is True over $\mathbf{K}$, but that is not recursively deducible from $\mathbf{K}$. This is in contrast to those who claim that *All is known.*

9.3.3 Applications

Let X be a set, and let $\mathbf{C}$ be a set of True statements from a logical system. We say that X is a *theory of everything for* $\mathbf{C}$ if given a statement Q that is True over $\mathbf{C}$, then Q is recursively deducible from X.

Theorem 9.3.6 *Let* $\mathbf{C}$ *be a set of True statements. Then no set* $X \subset \mathbf{C}$ *is a theory of everything for* $\mathbf{C}$.

Proof: Apply Theorem 9.3.3.

To prove that a given set TOE of statements is a theory of everything for Mathematical Physics, one *chooses an arbitrary E that is True over Mathematical physics*, and then decides that E is or is not deducible from the statements contained in TOE. The problem with this algorithm is that you cannot choose an arbitrary expression E in Mathematical Physics because *Mathematical Physics is not a set.* You see, as a science, Mathematical Physics contains Logic, and Logic is not a set. Therefore, there is an E that is both in Mathematical Physics and not in Mathematical Physics. This is where the above algorithm fails.

The lack of an unambiguous predicate for Mathematical Physics means that to every given expression E in Mathematical Physics, the reader must produce Physics that is measured by E. For instance, because $y = \frac{1}{2}ax^2$ is a position function for an object in linear motion and with constant acceleration a, $y = \frac{1}{2}ax^2$ is Mathematical Physics.

The ambiguity exhibited by E above can be found in the constants of Mathematical Physics. Choose a natural number $a > 0$ so large that any measurement of value a is considered to be infinite by today's Mathematical Physicists. By the Minimum Principle of the natural numbers, a can be chosen so that $\frac{1}{2}a$ is treated as a number in every calculation made in Mathematical Physics. Consider the function of Mathematical Physics, $y = \frac{1}{2}ax^2$. By our choice of $a > 0$, and a use of the associative property for numbers, the value $y(0)$ satisfies

$$y(0) = (\frac{1}{2}a)0^2 = \frac{1}{2}(a0^2).$$

Since a was chosen so that $\frac{1}{2}a$ is used as a real number, $(\frac{1}{2}a)0 = 0$. But a is infinite in Mathematical Physics, so $a0$ is *undefined*, and thus $\frac{1}{2}(a0^2)$ is undefined. This contradiction to the well defined property of the function image $y(0)$ shows that our choice of a is in error. Thus the idea of treating a number as an infinite quantity is a logical inconsistency in the foundation of Mathematical Physics.

Theorem 9.3.7 *Let* **B** *be an area of Mathematics and science. Then* **B** *does not possess a theory of everything.*

Proof: The area **B** of science contains Logic. Since Logic is not a set, **B** is not a set. If $X \subset \mathbf{B}$ is a set then the logical closure $\overline{X}$ is a set, and so $\overline{X} \neq \mathbf{B}$. Hence **B** does not possess a theory of everything. This completes the proof.

Specifically, none of the areas Logic, Mathematics, or Mathematical Physics possesses a theory of everything. The next result implies that there is no theory of everything for such intellectual endeavors as combinatorical Logic, any area of Algebra, and String Theory.

Theorem 9.3.8 *Let* **B** *be an area of Mathematics and science, and let* **C** *be a* set *of True statements from* **B**. *Then* **C** *does not possess a theory of everything.*

Proof: By Theorem 9.3.6, **C** does not possess a theory of everything. This completes the proof.

Let **B** be an area of Mathematics and science, and let t be any date. Let us bypass the controversy of the constants in **B** by agreeing to examine the set $\mathbf{B}(t)$ of *True statements from* **B** *that have been studied by scientists prior to the date* t. The strength that $\mathbf{B}(t)$ brings to our discussion is that $\mathbf{B}(t)$ contains the True statements in **B** that scientists have worked with prior to t. Then an application of Theorem 9.3.8 proves the next result.

Theorem 9.3.9 *Let* **B** *be an area of Mathematics and science, and let* t *be a date. Then* $\mathbf{B}(t)$ *does not possess a theory of everything.*

Specifically, if we let **B** be Abelian groups, then the set $\mathbf{B}(t)$ does not possess a theory of everything. Then there is a statement Q over the area $\mathbf{B}(t)$ that is True, but not provable.

Remark 9.3.10 The above theorems cannot be applied to Computer Science because the foundations of Computer Science contain a number of logical inconsistencies. See [7]. However, if one could find a set **C** of True statements in Computer Science, then by Theorem 9.3.7, **C** would not possess a theory of everything.

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

Bibliography

[1] G. Cantor, *Contributions to the Founding of the Theory of Transfinite Numbers*, Dover Press, New York (1955).

[2] P. Cohen, *Set Theory and the Continuum Hypothesis*, Dover Publications, New York (2008).

[3] K. Devlin, *Mathematics: The New Golden Age*, Pelican Books, New York (1988).

[4] H. Eves, *An Introduction to the History of Mathematics*, Saunders College Publishing, New York (1990).

[5] C. Faith, *Algebra I: Algebra and Categories*, Springer-Verlag, New York (1974).

[6] T. G. Faticoni, *The Mathematics of Infinity: A Guide to Great Ideas*, Wiley and Sons, Inc., New York (2006).

[7] T.G. Faticoni, *On the Foundations of Computer Science*, submitted.

[8] D. R. Hofstader, *Godel-Escher-Bach: An Eternal Golden Braid*, Vintage Press, New York (1980).

[9] S. Hawkings, *God Created the Integers: The Mathematical Break-* throughs that Changed the World, Running Press, Philadelphia (2005).

[10] N. Jacobson, *Basic Algebra 1*, W.H. Freeman and Company, New York (1985).

[11] E. Kasner and J. Newman, *Mathematics and the Imagination*, Simon & Schuster, New York (1940).

Index

E

F

G

H

I

L

M

The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
by Theodore G. Faticoni

PURE AND APPLIED MATHEMATICS

A Wiley Series of Texts, Monographs, and Tracts

ADÁMEK, HERRLICH, and STRECKER—Abstract and Concrete Catetories
ADAMOWICZ and ZBIERSKI—Logic of Mathematics
AINSWORTH and ODEN—A Posteriori Error Estimation in Finite Element Analysis
AKIVIS and GOLDBERG—Conformal Differential Geometry and Its Generalizations
ALLEN and ISAACSON—Numerical Analysis for Applied Science
*ARTIN—Geometric Algebra
ATKINSON, HAN, and STEWART—Numerical Solution of Ordinary Differential Equations
AUBIN—Applied Functional Analysis, Second Edition
AZIZOV and IOKHVIDOV—Linear Operators in Spaces with an Indefinite Metric
BASENER—Topology and Its Applications
BERG—The Fourier-Analytic Proof of Quadratic Reciprocity
BERKOVITZ—Convexity and Optimization in $\mathbb{R}^n$
BERMAN, NEUMANN, and STERN—Nonnegative Matrices in Dynamic Systems
BOYARINTSEV—Methods of Solving Singular Systems of Ordinary Differential Equations
BRIDGER—Real Analysis: A Constructive Approach
BURK—Lebesgue Measure and Integration: An Introduction
*CARTER—Finite Groups of Lie Type
CASTILLO, COBO, JUBETE, and PRUNEDA—Orthogonal Sets and Polar Methods in Linear Algebra: Applications to Matrix Calculations, Systems of Equations, Inequalities, and Linear Programming
CASTILLO, CONEJO, PEDREGAL, GARCIÁ, and ALGUACIL—Building and Solving Mathematical Programming Models in Engineering and Science
CHATELIN—Eigenvalues of Matrices
CLARK—Mathematical Bioeconomics: The Mathematics of Conservation, Third Edition
COX—Galois Theory, Second Edition
†COX—Primes of the Form $x^2 + ny^2$: Fermat, Class Field Theory, and Complex Multiplication
*CURTIS and REINER—Representation Theory of Finite Groups and Associative Algebras
*CURTIS and REINER—Methods of Representation Theory: With Applications to Finite Groups and Orders, Volume I
CURTIS and REINER—Methods of Representation Theory: With Applications to Finite Groups and Orders, Volume II
DINCULEANU—Vector Integration and Stochastic Integration in Banach Spaces
*DUNFORD and SCHWARTZ—Linear Operators
Part 1—General Theory
Part 2—Spectral Theory, Self Adjoint Operators in Hilbert Space
Part 3—Spectral Operators
FARINA and RINALDI—Positive Linear Systems: Theory and Applications

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in paperback.

FATICONI—The Mathematics of Infinity: A Guide to Great Ideas, Second Edition
FOLLAND—Real Analysis: Modern Techniques and Their Applications
FRÖLICHER and KRIEGL—Linear Spaces and Differentiation Theory
GARDINER—Teichmüller Theory and Quadratic Differentials
GILBERT and NICHOLSON—Modern Algebra with Applications, Second Edition
*GRIFFITHS and HARRIS—Principles of Algebraic Geometry
GRILLET—Algebra
GROVE—Groups and Characters
GUSTAFSSON, KREISS and OLIGER—Time Dependent Problems and Difference Methods
HANNA and ROWLAND—Fourier Series, Transforms, and Boundary Value Problems, Second Edition
*HENRICI—Applied and Computational Complex Analysis
Volume 1, Power Series—Integration—Conformal Mapping—Location of Zeros
Volume 2, Special Functions—Integral Transforms—Asymptotics—Continued Fractions
Volume 3, Discrete Fourier Analysis, Cauchy Integrals, Construction of Conformal Maps, Univalent Functions
*HILTON and WU—A Course in Modern Algebra
*HOCHSTADT—Integral Equations
JOST—Two-Dimensional Geometric Variational Procedures
KHAMSI and KIRK—An Introduction to Metric Spaces and Fixed Point Theory
*KOBAYASHI and NOMIZU—Foundations of Differential Geometry, Volume I
*KOBAYASHI and NOMIZU—Foundations of Differential Geometry, Volume II
KOSHY—Fibonacci and Lucas Numbers with Applications
LAX—Functional Analysis
LAX—Linear Algebra and Its Applications, Second Edition
LOGAN—An Introduction to Nonlinear Partial Differential Equations, Second Edition
LOGAN and WOLESENSKY—Mathematical Methods in Biology
LUI—Numerical Analysis of Partial Differential Equations
MARKLEY—Principles of Differential Equations
MORRISON—Functional Analysis: An Introduction to Banach Space Theory
NAYFEH—Perturbation Methods
NAYFEH and MOOK—Nonlinear Oscillations
O'LEARY—Revolutions of Geometry
O'NEIL—Beginning Partial Differential Equations, Second Edition
PANDEY—The Hilbert Transform of Schwartz Distributions and Applications
PETKOV—Geometry of Reflecting Rays and Inverse Spectral Problems
*PRENTER—Splines and Variational Methods
PROMISLOW—A First Course in Functional Analysis
RAO—Measure Theory and Integration
RASSIAS and SIMSA—Finite Sums Decompositions in Mathematical Analysis
RENELT—Elliptic Systems and Quasiconformal Mappings
RIVLIN—Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory, Second Edition
ROCKAFELLAR—Network Flows and Monotropic Optimization
ROITMAN—Introduction to Modern Set Theory
ROSSI—Theorems, Corollaries, Lemmas, and Methods of Proof
*RUDIN—Fourier Analysis on Groups
SENDOV—The Averaged Moduli of Smoothness: Applications in Numerical Methods and Approximations

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in paperback.

SENDOV and POPOV—The Averaged Moduli of Smoothness
SEWELL—The Numerical Solution of Ordinary and Partial Differential Equations, Second Edition
SEWELL—Computational Methods of Linear Algebra, Second Edition
SHICK—Topology: Point-Set and Geometric
SHISKOWSKI and FRINKLE—Principles of Linear Algebra With *Maple*™
SHISKOWSKI and FRINKLE—Principles of Linear Algebra With *Mathematica*®
*SIEGEL—Topics in Complex Function Theory
Volume 1—Elliptic Functions and Uniformization Theory
Volume 2—Automorphic Functions and Abelian Integrals
Volume 3—Abelian Functions and Modular Functions of Several Variables
SMITH and ROMANOWSKA—Post-Modern Algebra
ŠOLÍN–Partial Differential Equations and the Finite Element Method
STADE—Fourier Analysis
STAHL—Introduction to Topology and Geometry
STAHL—Real Analysis, Second Edition
STAKGOLD and HOLST—Green's Functions and Boundary Value Problems, Third Edition
STANOYEVITCH—Introduction to Numerical Ordinary and Partial Differential Equations Using MATLAB®
*STOKER—Differential Geometry
*STOKER—Nonlinear Vibrations in Mechanical and Electrical Systems
*STOKER—Water Waves: The Mathematical Theory with Applications
WATKINS—Fundamentals of Matrix Computations, Third Edition
WESSELING—An Introduction to Multigrid Methods
†WHITHAM—Linear and Nonlinear Waves
ZAUDERER—Partial Differential Equations of Applied Mathematics, Third Edition

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in paperback.