Constructing NUMBERS

A layperson's guide to a mathematical landscape

Mike Grannell.

Contents

Preface

What is a number? This question is a rarely asked and seldom addressed outside Mathematics and Philosophy Departments in Universities. I can't speak for Philosophy Departments but, as a mathematician, I know that the issue gets scant attention even within the majority of Mathematics Departments. So what is the layperson, the non-philosopher, non-mathematician, to make of numbers *as numbers*?

The silence on this question is strange, given the prevalence of numbers in modern society. Numbers seem to be all around us from our earliest days. We learn to count with play-bricks and other toys, to add and subtract, to share and to compare. We go to school and learn methods for addition, subtraction, multiplication and division. We learn about negative numbers and fractions, decimals, and using calculators. We probably meet a few strange numbers like $\sqrt{2}$ (the square root of 2), and π (pi). We may learn something about equations and how to solve them. But we almost never get to ask, or get told, what numbers are.

Numbers intrude into our lives in all sorts of ways. We use them to count our money, to weigh and to measure. Our communication systems use them to transmit our messages reliably and securely. Our lives are profoundly influenced by statistics. Can you imagine a society with *no* numbers at all? It would be indescribably different from ours. Except for language, numbers seem to be more fundamental to intellectual activity than any other human trait. Indeed, numbers have proved more enduring than any individual language. But what are numbers? Are they a product of human society, or do they exist independently of us? If we met extra-terrestrial aliens, would they share the same ideas about numbers, even if expressed differently? Who decides what is a number and what isn't?

Of course numbers are abstractions, but so are things like love and beauty, and these get plenty of attention. But we seem to just accept numbers and then spend our time doing more and more advanced computations using them. This book aims to redress the balance for the interested lay reader. It is most definitely *not* a textbook. I will skate over many points. I just want to give you a flavour, perhaps to whet your appetite for more details, but at least to convince you that there is more to numbers than meets the eye at first glance.

The development of numbers is one of humankind's great cultural achievements, something to be proud about even for non-mathematicians, in the same way that we can take delight in Shakespeare's plays, Monet's art, and the Beatles' music, even though we had no hand in making them. We should all have some appreciation of cultural achievements in fields such as literature, art and music, and I want to convince you that the development of numbers also deserves your attention.

The principal aim of this book is to show how all the familiar types of numbers, whole numbers, fractions and decimals, can be developed on sound logical foundations. The book also shows how other types of numbers have been formalized by mathematicians seeking to explore the very concept of *number*. By the end of the book, I hope that you will appreciate why numbers *in themselves* are interesting, quite apart from their utility in measuring and calculating.

This is not a book aimed at mathematicians, nor at historians of mathematics. It is not intended to give you an historical picture about the development of numbers. It does not seek to make you more proficient at performing calculations. It's aimed at people who probably never think about numbers, even though they use them daily. It will try to explain numbers without a lot of complicated mathematics, but it will show how the concept of number can be developed from simple beginnings. I have tried hard to make this book as accessible as possible. However, it is not an "easy read" where you can turn the pages at a rapid pace. Some of the ideas are quite subtle and require some reflection.

I am assuming that you will have seen something about numbers when you were at school or college, maybe even at a university. So you are probably able to use whole numbers and fractions for counting and measuring. As a consequence, and depending on how far you got, you may find the first few chapters cover some familiar topics, but I urge you to linger and look at the view, because our viewpoint is not the usual one. When we get further along I hope you will find this view getting more interesting and not like the one you encountered in your early education.

Acknowledgments
Many people have helped me by reading and commenting on drafts of this book. So I record my thanks to my wife Kathryn, son Christopher, friends and colleagues including Peter Adamson, John Adamson, Gillian Kerr, David Nash, David Parker, Julie Taylor, Patrick Taylor and Peter Taylor. Any deficiencies that remain are entirely my fault.

Chapter 1

Introduction

1.1 Book structure

Our main theme is that of starting small and building bigger. This can be summarized in one word: **construction**. We will show how to construct all the familiar numbers and many other more exotic numbers. This is not how most of these numbers were originally developed. Rather, people adopted a pragmatic approach, creating and using fractions and negative numbers as the needs and their utility became apparent. The desirability of ensuring that these numbers could be manipulated safely, using the familiar and "obvious" properties of the counting numbers, led to the approach that I will describe. Getting ideas onto a rigorous logical basis is part of the job of a mathematician, and it requires both imagination and precision.

This book is divided into two main parts. The first part, comprising Chapters 2 to 6, deals with the numbers you probably met at school such as counting numbers, fractions, negative numbers and decimals. The second part, comprising Chapters 7 to 12, takes you to Complex Numbers and their relatives, and then on up to the realm of the infinite. Section 1.3 below describes the first part of the book in a little more detail, and Section 1.4 does likewise for the second part.

We will not be doing lots of arithmetic or lengthy calculations. However, mathematics does have its own language and I will need to explain relevant bits of this as we go along, starting with some basics in Section 1.2. A Glossary is provided towards the end of the book that collects together definitions and explanations, so that if you need to remind yourself at any point, you can easily locate the item without having to search through the main body of text. An Index provides a further map to the landscape.

1.2 Basic mathematical notation

We will have to use some mathematical notation and terminology. Most of this will be introduced and explained as we go along. If you did mathematics up to the age of 16 at a secondary school, you should have enough to cope even if you are a bit rusty.

We will use single Latin or Greek letters to denote numbers as in "$x = 2$". We will, as normal in mathematics, reserve the Greek letter π for the ratio of the circumference of a circle to its diameter. The other Greek letters used are θ (theta) that is often used to denote an angle, ϵ (epsilon) that is often used for a number that is very small, δ (delta) that is often used in a composite symbol such as δx to denote a change in the value of x, and ω (omega) that we will use to denote the first infinite ordinal as explained in Chapter 10.

As you may already have noticed on the Contents pages, we use bold face symbols such as $\mathbf{N}$ to denote particular collections of numbers, for example $\mathbf{N}$ denotes the collection of Natural Numbers. We will explain each of these, and any other symbols, before we start to use them.

You will not have to do any lengthy calculations or complicated algebra. It will suffice if you can remember enough algebra to multiply $(a+b)\times(c+d)$ and get the answer $a \times c + a \times d + b \times c + b \times d$.

We will use *exponent* notation such as x^2 to mean $x \times x$, and x^3 to mean $x \times x \times x$. Here the 2 and the 3 are called the exponents. We might also refer to x^3 as the third *power* of x. The quantity x^2 is generally called "x squared" and x^3 is generally called "x cubed", but higher powers such as $x^4 = x \times x \times x \times x$ are referred to numerically as in "x to the 4". An object like $2x^2 + 3x - 5$ is called a quadratic expression or polynomial of degree 2 (in the unknown x), and the numbers $2, 3$ and -5 that appear here are called the *coefficients*. An object like $5x^3 + x^2 - 2x + 7$ is called a cubic expression or polynomial of degree 3; here the coefficients are $5, 1, -2$ and 7. Similar objects involving higher powers of x are also called polynomials, and the degree of such a polynomial is the highest exponent of x in the expression.

We will encounter some simple equations like "$3 + x = 5$" (which has solution $x = 2$). The equation $x^2 = 9$ has a solution $x = 3$ because $3 \times 3 = 9$ and, if you remember correctly from school days, it also has another solution $x = -3$ because $(-3) \times (-3) = 9$. If you are happy with these you shouldn't have any trouble with solving any equations in this book.

On the rare occasions when we actually give a complete formal proof of some result, we mark the end of the proof with the symbol $\Box$.

1.3 Plan of Chapters 2 to 6

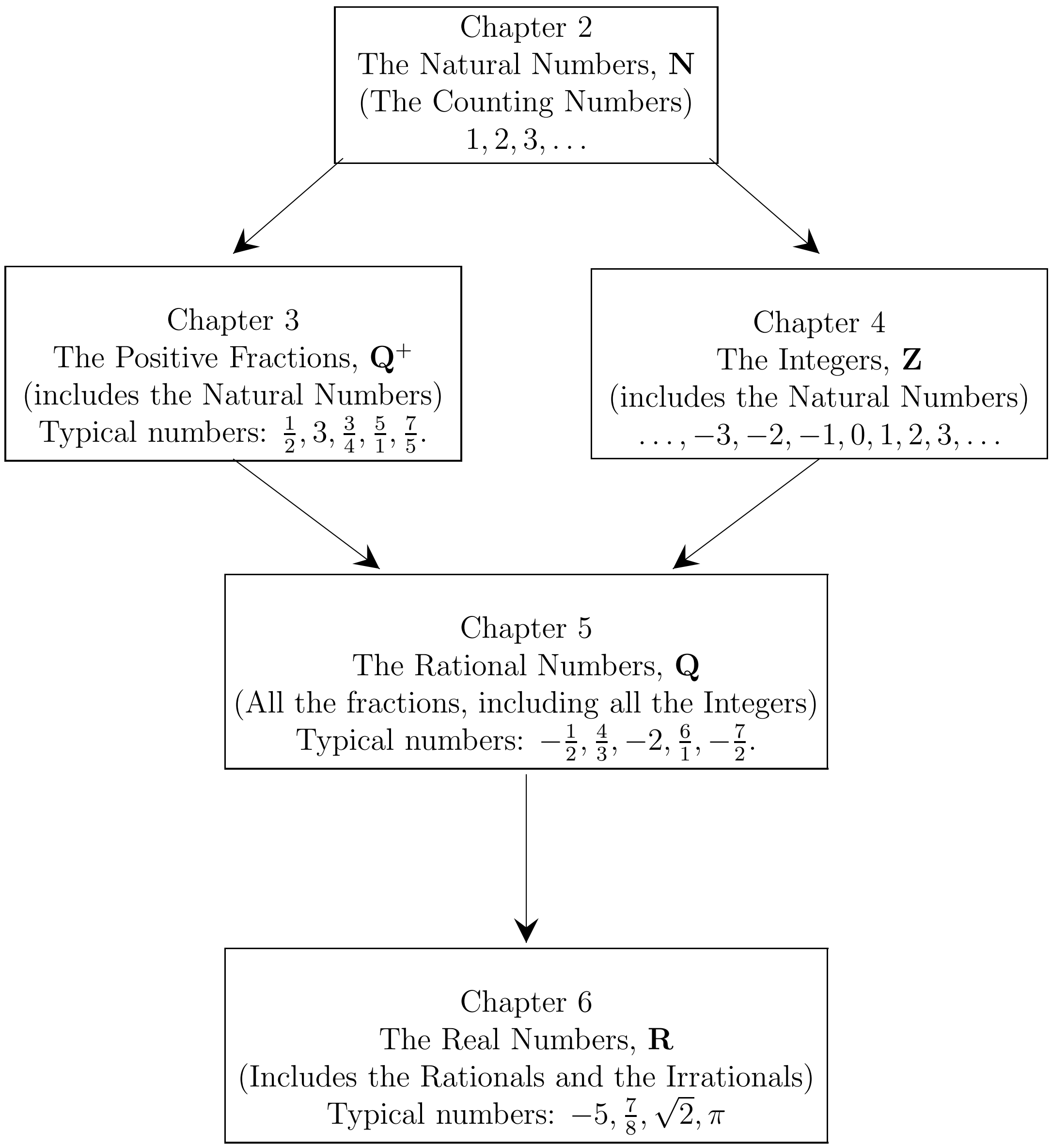
In the first part of the book our starting point will be the counting numbers 1, 2, 3, etc., known in mathematics as the *Natural Numbers*. We will take a first look at these in Chapter 2. Then in Chapter 3 we show how to construct the *Positive Fractions* from pairs of Natural Numbers. As an example, the fraction $\frac{3}{4}$ is clearly constructed in some way from the Natural Numbers 3 and 4. You will have seen this at school but probably not from our viewpoint.

In Chapter 4 we show how to construct negative whole numbers and zero from pairs of Natural Numbers. This construction is unlikely to be familiar to you, so expect a bit of a learning gradient here. But if I write the pair $(1, 3)$ as $1 - 3$ (i.e. as 1 minus 3), you will probably have little difficulty in recognizing it as the negative whole number -2. The whole numbers, positive, negative and zero, are collectively known as the *Integers*. Then in Chapter 5 we put the ideas from the previous two chapters together to construct the *Rational Numbers* that comprise all the fractions, positive and negative, and including all the whole numbers.

Where are decimals in all this? We will start on these when we discuss fractions. Every fraction can be represented as a decimal; for example $\frac{3}{4} = 0.75$. But the big issue, first observed by Greek mathematicians and philosophers around 2,500 years ago, is that some numbers (the square root of 2 is an example) cannot be represented as a fraction. So decimals really need a deeper discussion which we will move onto in Chapter 6. The collection of all possible decimals is known by mathematicians as the set of *Real Numbers*.

The diagram on the next page gives a sort of family tree for the classes of numbers that we will examine in this first part of the book. The terminology will be explained as we move along.

Family Tree for Chapters 2 to 6



1.4 Plan of Chapters 7 to 12

These chapters cover material that you are unlikely to have seen at school. In Chapter 7 we construct the *Complex Numbers* from pairs of Real Numbers by introducing the so-called *imaginary* number i that represents a square root of -1. Actually, and as we will see, it is no more imaginary than, for example, -1. Then a very short Chapter 8 covers a further extension of the Real Numbers to form the *Quaternions.*

The next four chapters all involve the concept of infinity in various ways. In Chapter 9 we explain *Cardinal Numbers.* Here we *prove* that there are different sizes of infinity. For example, there are infinitely many whole numbers and infinitely many decimals, but we will show that the two infinities are of different sizes.

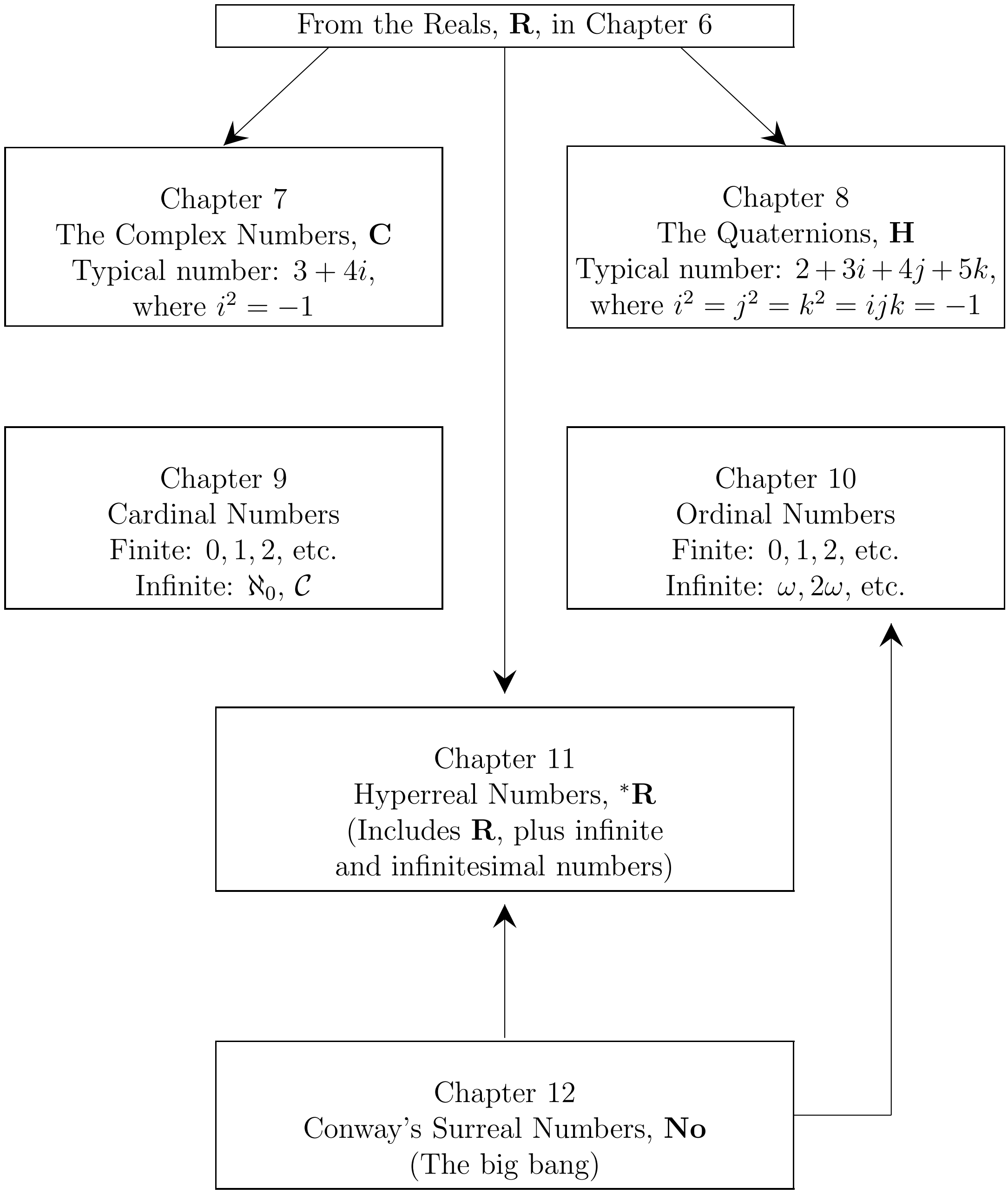
Then in Chapter 10 we define *Ordinal Numbers.* Here we return to basics and try to define the numbers $0, 1, 2, 3$, etc. in terms of simpler objects, providing a partial answer to questions like "'what is 3?". In consequence of this definition we are able to provide a meaning to certain infinite numbers.

Chapter 11 takes us in a different direction involving mathematical logic. Here we encounter an extension of the Real Numbers that includes infinite numbers and infinitesimal numbers that can be manipulated much like the Real Numbers themselves. The driving force behind this concept is to simplify arguments involving numbers that are arbitrarily large or arbitrarily small.

In our final substantive Chapter 12 we describe a construction of great ingenuity that creates from scratch all the Real Numbers as well as versions of the Ordinal Numbers and the infinite and infinitesimal numbers seen in Chapter 11. Not only does it create these numbers but it also provides definitions of order, addition and multiplication of these numbers.

The diagram on the next page gives a sort of family tree for the classes of numbers that we will examine in this second part of the book. Again we will explain the terminology as we move along.

Family Tree for Chapters 7 to 12



1.5 How to read this book

It is said that mathematics is not a spectator sport, and that remark contains a large element of truth. So you are strongly advised to have a pen and paper to hand along with a basic calculator as you read the text. This is not a textbook, so you won't be asked to do a lot of examples, just an occasional one now and again, but you should try to check any calculations for yourself.

You don't need to know any advanced mathematics but some of the arguments are quite subtle. Consequently you may sometimes find yourself stuck on something. There is no shame in this; it happens to us all. A genuine difficulty with presenting mathematics and other technical subjects is that it is impossible to say everything at once. So it becomes necessary to refer to different aspects sequentially. Consequently the explanation of some concept or mathematical device may spread over a few sentences or paragraphs. Getting around an obstacle can sometimes be achieved by reading a bit further ahead in the text to see if that sheds any light. Another useful idea when stuck is to put it down for a day and let your subconscious work on it. It is surprising how often an insuperable difficulty looks much less forbidding the next day. If you have someone you can talk to about it, that is also very beneficial.

Chapters 2 to 8 follow a logical progression, but Chapters 9 and 10 have a large measure of independence from the other chapters. So if you want a rest from the logical progression, try one of those.

There are a few places where a bit more mathematical knowledge is useful. These short sections are clearly marked and you will not lose the thread if you omit them.

I hope that you will enjoy reading this book. It's time to get started and we will begin with the simplest numbers of all.

Chapter 2

The Natural Numbers (N)

2.1 Representation of the Natural Numbers

These are the numbers that you learn as a child. People often call them the Counting Numbers: $1, 2, 3, \ldots$. Here I've sneaked in a bit of mathematical notation: "...", which is meant to indicate these numbers go on indefinitely, a sort of mathematical equivalent of the phrase "and so on". These numbers have been around a long time in human history. Their *names* have varied. The Romans wrote III for our 3, and our numerals are based on Arabic numerals. But we probably all agree that the names aren't important, it's the underlying meaning that is. So whether you write III or 3 or "three" or any translation of these to other languages or cultures, the number remains the same.

It would be hard to have any form of human interaction without at least having knowledge of the first few Natural Numbers ("I'll give you two arrowheads for three coconuts"). So are these numbers a cultural achievement, something created by humans, or are they something external to us? It is certainly true that they seem to be more independent of us than, say, the Mona Lisa painting. There was no *necessity* for the Mona Lisa to be painted[1], but there does seem to be a necessity for the Natural Numbers, at least for the first few of them. It is perfectly possible to imagine a slightly different Mona Lisa, but not a slightly different number 3. Furthermore, many animals seem to have an understanding of the first few Natural Numbers, and it is not hard to imagine why this confers an evolutionary advantage. So I won't claim any great cultural achievement in the *existence* of the Natural Numbers.

However, there is considerable cultural achievement in developing a good

[1]Some philosophers might argue with this. However, the Mona Lisa could be destroyed, but it doesn't seem possible to destroy the number 3.

system for representing these numbers. I am writing this in 2022, which is written in Roman numerals as MMXXII. Now that's not too bad, but 1999 would have been MCMXCIX. Now try multiplying that by MMXXII and you will see why the *place-value system* currently in use is vastly easier for computations. Ancient Romans who regularly had to do calculations would have done them using an abacus rather than pencil and paper, and it seems that these abacuses were arranged rather more like our place-value system with columns for units, tens, hundreds, etc.

What is meant by a "place-value system"? Well, with a number like 1234, we can see it is made up of one thousand plus two hundreds, plus three tens, and plus four units (i.e. four ones). The place each digit occupies determines its value. So the 2 which is in the third place from the right in the number 1234 represents the value 200, i.e. 2 lots of 100. We only need ten symbols $(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)$ to describe all the Natural Numbers because the place each symbol appears in the number tells us whether it counts units, tens, hundreds, thousands, and so on.

Our commonly used place-value system is based on powers of 10. We probably use base-10 because we have 10 fingers (including thumbs). We have units, tens, hundreds, thousands, etc. If we had 12 fingers, we'd probably use 12 symbols and base our place-value system on powers of 12. We write 10^n for 1 followed by n zeros and we call this "10 to the power n" or more simply "10 to the n". It can also be thought of as the product of n 10s. So, for example, 10^3 is 1000, which is $10 \times 10 \times 10$, and which we call a thousand. Similarly $10^6 = 1\,000\,000$, which we call a million, and $10^0 = 1$ (people sometimes get confused by this last one). There's nothing special about 10, so 12^2 means 12×12, and 5^3 means $5 \times 5 \times 5$. More generally m to the n, written as m^n, means n lots of m multiplied together.

The number 3 that appears in the expression 10^3 is called the *exponent*. From this we get the expression *exponential growth*. To illustrate what that means, suppose we have disease with a replication rate of 2 per week, meaning that each infected individual infects two more people in a week before they themselves cease to be infectious. If we start with one person infected, then at the end of week 1 we will have two people infected. Each of these will infect 2 more people, so at the end of week 2 we will have $4 = 2^2$ people infected. At the end of week 3 we will have $8 = 2^3$ people infected, and it will go on growing like this. At the end of week n we will have 2^n people infected, so after 10 weeks that comes to over a million, and after 17 weeks we would have everyone on the planet infected. Even quite low rates of growth can have disastrous consequences if the trend continues for a long time. For example, if the world population were to grow consistently at 1% per annum (in other words a rate of 1.01 per year) then in around 3200 years the mass of human bodies would exceed the mass of the planet Earth! That's exponential growth for you. But exponential growth has its benign side in enabling us to represent large numbers economically with our place-value system based on powers of 10.

Despite our ten-fingered prejudice, other cultures at other times have used different bases for a place-value system instead of our use of 10. Other popular choices have been 60, 20, 16, and 12. Some cultures have even mixed them together. The remnants of some of these systems are still around today - we divide an hour into 60 minutes, and a minute into 60 seconds. Likewise an angle of one degree is also divided into 60 minutes, and a minute into 60 seconds. There are 12 inches in a foot, and there are special words for 20 (a *score*), 12 (a *dozen*) and $12^2 = 12 \times 12 = 144$ (a *gross*).

Digital computers use base-2. This is largely because numbers can then be described with just two symbols (0 and 1) and these can be thought of as electrical switches ("off" and "on") or as logical values ("false" and "true"). The powers of 2 are $2^0 = 1,\ 2^1 = 2,\ 2^2 = 4,\ 2^3 = 8$, etc. As an example, the number represented in base-2 as 1101 may be read from the right as 1 unit, plus 0 twos, plus 1 four, plus 1 eight. In our normal base-10 system this would be $1 + 4 + 8 = 13$.

Whatever place-value base is used, computations can be performed much like you did at school. You may have thought those very tedious, but spare a thought for people trying to do the same calculations with the Roman system. To be fair, a pure base-60 system would require 60 symbols in order to describe each Natural Number, and this would be somewhat unwieldy. Base-2 would not be at all helpful for some purposes - imagine your telephone number expressed in base-2 - it would be about three times as long and consist of a string of 0s and 1s.

How did systems for representing the Natural Numbers develop? We can't be certain, but it is arguably due to the need for representing larger numbers and doing calculations with them. A hunter-gatherer community probably doesn't have any need for numbers above a hundred, and not much need for numbers above ten. Some such communities seem to have managed perfectly well with just one, two and three. But once communities adopted a more settled and agricultural lifestyle, there would be requirements for keeping records of assets such as grain stocks. Construction of large permanent buildings would also require measurements and calculations. Once large scale trading operations became common, accurate numerical records would have become crucial. We should not belittle human beings from 10,000 years ago and imagine ourselves superior to them. They would have superb knowledge of the thousands of stars in the night sky, but they are unlikely to have felt the need to count them. With current mathematical notation, we can easily specify numbers of immense size. For example, 10^{100} (a digit "1" followed by one hundred zeros) exceeds the number of fundamental particles in the Universe multiplied by the number of seconds since the Big-Bang (according to Physicists). Someone from 10,000 years ago might not have conceded that this is a number since their meaning of "number" would probably have been based on enumeration of real tangible objects.

2.2 The philosophical view

Should we adopt the platonic view (named after Plato) that numbers exist outside ourselves, or are they a product of our culture? Philosophers have been arguing this question for centuries. I think it is fair to say that most mathematicians regard numbers as having an external reality, independent of ourselves; they are "out there", waiting to be discovered. A few would disagree, and I have even come across one who maintains that the Natural Numbers are finite, they don't go on indefinitely. I can't believe that many would share this view since we can always play the childhood game of "what's the biggest number that you can think of?". Whatever you opponent says, you can repeat it and then say "plus one". Maybe we shall have to await communication with extra-terrestrial aliens to see if they share our perceptions of numbers. Incidentally, if you are ever abducted by aliens and are worried that no-one will believe you on your return, ask the aliens to write down a proof of (or a counter-example to) the Riemann hypothesis, the most famous and long-standing problem in mathematics. If you come back with a proof, *everyone* will believe you!

The Riemann hypothesis concerns the distribution of prime numbers amongst the Natural Numbers. A Natural Number $n > 1$ is said to be a *prime number* if and only if its only factors are 1 and n. So 5 is a prime number, but $6 = 2 \times 3$ is not. There are many solved and unsolved problems that involve prime numbers and whole books devoted to them. But these are not our focus in this book; here we are considering how to construct numbers in the first place before asking hard questions about them.

I now want to mention an issue that you may have already noticed. In our base-10 place-value system we need ten symbols at our disposal in order to represent any given Natural Number. Nine of these symbols themselves represent Natural Numbers, namely $1, 2, 3, 4, 5, 6, 7, 8$ and 9. But the tenth symbol 0 does not itself represent a Natural Number. The reason is fairly obvious - if we are counting objects it would be perverse to start at zero[2].

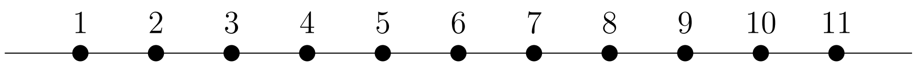
[2]Mathematicians are divided over whether 0 should be regarded as a Natural Number,

In fact the use of zero as a genuine number, on a par with 1, 2 and 3, came surprisingly late although, as a symbol, 0 or something equivalent to it is almost essential to any place-value system. The use of 0 as a number seems to date from about 1500 years ago. So as a number, it is clearly not "natural" and its discovery (invention?) can certainly be regarded as a substantial cultural achievement, even though nowadays it is a commonplace. One might compare it to the invention of the wheel or the discovery of electricity. Now I can almost hear you say that if 0 isn't there in our collection of numbers we should just put it in, and we *will* do that later on and include with it all the negative whole numbers. But before doing that we will deal with (positive) fractions. The reason for proceeding in this order is that you will be familiar with representing Positive Fractions as pairs of Natural Numbers, but you are unlikely to be familiar with representing negative numbers as pairs of Natural Numbers.

2.3 The number line and successors

Although the Natural Numbers are just fine for counting discrete objects, they aren't so useful for measurements of continuously varying quantities such as lengths and weights. For example your height in metres and your weight in kilograms are unlikely to be exact whole numbers. You can see this defect on a ruler that is marked with discrete division points, requiring you to have some way to describe a part of a unit interval. So fractions arose at an early stage, appearing in records as long ago as 4,000 BCE in Sumeria.

As a prelude to discussing fractions and as a help in thinking about the Natural Numbers, we can envisage them marked out along a straight line, as a sort of ruler:



We can see that each Natural Number n has an immediate successor that we can denote as $S(n)$. For example, the successor of 3 is 4, i.e. $S(3) = 4$. And every Natural Number except 1 is the successor of some other Natural Number. We could *define* 2 as the successor of 1: $2 = S(1)$, and likewise define 3 as the successor of 2: $3 = S(2)$, and this could be written as $3 = S(S(1))$, read as "the successor of the successor of 1". We can go on like this defining 4, 5, 6, etc. We might describe the whole collection $2, 3, 4, \ldots$

---

some say yes and some say no, and it makes no difference in the long run. But I am addressing non-mathematicians and I doubt that many of you start counting with zero.

as the repeated successors of 1, reserving the term "the successor" to mean the *immediate* successor, namely the number 2.

All the properties of the Natural Numbers can be deduced using the idea of successors. All that is required to do this is the following five properties, collectively known as the Peano postulates.

1. 1 is a Natural Number.

2. Each Natural Number n has a successor $S(n)$ which is also a Natural Number.

3. 1 is not the successor of any Natural Number.

4. If m and n are Natural Numbers and $S(m) = S(n)$, then $m = n$.

5. If we have a collection C of Natural Numbers that contains the Natural Number 1 and also contains the successor of every Natural Number in the collection, then the collection C contains all the Natural Numbers.

The five properties listed above may look a bit abstract, but they closely follow the earliest known form of recorded counting - the use of tally marks. So in terms of tally marks these properties might be expressed roughly as follows.

1. Here is the first tally mark representing "'one": /

2. If //... / is a tally mark representing a number, then we get the next number (the successor) by adding another tally mark: //... //

3. There is no tally mark for nothing.

4. If two people have the same number of tally marks and they each rub out one tally mark, then they will still have the same number of tally marks.

5. The only numbers we recognize are those given by the tally mark for "'one" and the tally marks for its repeated successors.

The last property deserves some extra explanation. Going back to the original version of the Peano postulates, if property 5 wasn't there we might have additional objects in our collection, such as an additional "starting point", call it $1'$. So the last property ensures that the Natural Numbers comprise just the number 1 along with its repeated successors, $S(1)$, $S(S(1))$, $S(S(S(1)))$, $\ldots$.

It is possible to *define* addition using the notion of successors. In effect, addition is defined by repeatedly adding 1 using the idea of successors. If m is any Natural Number, we can define $m + 1 = S(m)$, $m + 2 = S(m + 1)$, $m + 3 = S(m + 2)$, and generally for each Natural Number n, $m + S(n) = S(m + n)$. For example, once you know the value of $5 + 3$ (which we call 8), then the value of $5 + S(3)$ (that is to say the value of $5 + 4$) must be $S(5 + 3)$, which is the successor of 8, $S(8)$, known more commonly by us as 9.

This probably seems very cumbersome to you, but mathematicians do love defining complicated things (like addition) in terms of simpler things. A consequence of this definition is that $2 + 3 = 3 + 2$. We'll prove that in just a moment. But first I hear you asking why anyone should want to *prove* that $2 + 3 = 3 + 2$ when it is obvious. The first answer is that we have just defined $+$ in terms of successors and so we want to prove that our definition behaves as we would expect. If it didn't, we'd have to go back to the drawing board and find a better definition of $+$. A second answer is that some properties of numbers are far from obvious, so proving simple properties provides good practice for proving more complicated results. So here is our proof that $2 + 3 = 3 + 2$.

Start with the left-hand side and use the definition repeatedly.

$$2 + 3 = 2 + S(2) = S(2 + 2) = S(2 + S(1)) = S(S(2 + 1)) = S(S(S(2))).$$

On the right-hand side we have

$$3 + 2 = 3 + S(1) = S(3 + 1) = S(S(3)) = S(S(S(2))).$$

So both the left-hand side and the right-hand side reduce to the same quantity, and consequently $2 + 3 = 3 + 2$. Phew!

Of course we don't go through this every time we have two numbers to add together and want to be sure that the order of addition doesn't matter. It is possible to prove on the basis of this definition of addition that for any two Natural Numbers m and n, we have $m + n = n + m$. Mathematicians regard such properties with great satisfaction. Of course the Natural Numbers have to behave like this or we wouldn't find them much use for counting things. Imagine the chaos in a shop if, when buying two items, one for \$2 and one for \$3, you were asked for different amounts depending on which one they charged first.

The reason for making a fuss about the order of addition not mattering is that for many things in life, order does matter. Try putting on your shoes and socks in that order. And order often does matter in mathematics. Looking ahead to subtraction, $2 - 3$ is quite different to $3 - 2$.

A further consequence of this definition of addition is that when adding any three Natural Numbers m, n, p, how they are combined in pairs does not matter - in symbols $(m+n)+p = m+(n+p)$. (Here, on the left-hand side we first add m and n, and then add p to the result, and on the right-hand side we first add n and p and then add the result to m.) This rule for removing brackets means that you can just write $m + n + p$ without having to worry about how the numbers are paired.

The idea of successors also allows us to define what we mean by saying that one number is less than (i.e. lies to the left of) another. So we write $m < n$ (and $n > m$) if and only if $n = S(S(\cdots S(m)\cdots))$ for some appropriately long sequence of Ss (at least one S). For example, $1 < 3$ because $3 = S(S(1))$; here $m = 1, n = 3$ and there are two Ss. If you aren't familiar with the symbols $<$ and $>$, just remember that the greater number goes on the big end of the symbol. The symbol $<$ is read as "less than", and the symbol $>$ is read as "greater than". This definition enables us to see that the Natural Numbers come in a definite order: $1 < 2 < 3 < \ldots$.

This is probably a good place to remark on the phrase "if and only if" which we don't tend to use much in everyday English, but which gets a lot of use in mathematics. If A and B represent sentences and we say "A if and only if B", then we are saying that A and B are equivalent sentences - they mean or say the same thing. In mathematics the phrase is often used in a definition, as here where we are defining the precise meaning of "$<$" in terms of successors. It also gets used a lot when two apparently different properties are shown to be equivalent. Looking ahead to Chapter 5, an example of this is the result that a decimal represents a fraction if and only if the decimal recurs or terminates. You do have to be a bit careful to distinguish "if" from "if and only if". So this is true: "$x < 6$ if $x = 2$", but this isn't true: "$x < 6$ if and only if $x = 2$".

Let's get back to definitions. Why are mathematicians so hung up on definitions? Well, it's part of our DNA. We just love precision. If we define something carefully then we can all share a common understanding of it and we can all contribute to exploring its properties. At times this can be an obstacle to non-mathematicians, and a ruthless insistence on total accuracy can lead to simple ideas looking very obscure. I will do my best to avoid the relentless pursuit of precision. This will possibly upset some of my fellow mathematicians and, even so, I may still be seen to be over-fussy by the wider public. Let me try to point to something where even the definition of $=$ is not entirely obvious. Looking ahead for a moment to fractions, are the fractions $\frac{17}{29}$ and $\frac{391}{667}$ equal? They certainly don't *look* equal. How are we to decide? We need a definition of equality for fractions.

At this stage I won't bother any further about a definition of the Natural Numbers; we will just take them as given to us. We will return to how they might be defined in Chapter 10.

Apart from addition, another thing we can do with Natural Numbers is multiplication. This can be regarded as repeated addition. For example, 3×4 means " 3 lots of 4", in symbols $4+4+4=12$. Going back to the definition of addition in terms of successors, it is not hard (although somewhat tedious) to prove that $3\times 4 = 4\times 3$. More generally, $m\times n = n\times m$. And, as with addition, how we bracket terms together does not matter: $(m\times n)\times p = m\times(n\times p)$.

The number 1 has an interesting property when used in multiplication: $1\times m = m$ for any Natural Number m, so the 1 times table is very easy.

Another useful fact connecting $\times$ and $+$ is that $m\times(n+p) = m\times n + m\times p$, often described as multiplying out the brackets. For example, if $m=3, n=4$ and $p=6$ this asserts that $3\times(4+6) = 3\times 4 + 3\times 6$ - something you can easily see is correct since $30 = 12+18$.

2.4 Notation

Rather than keep saying "the Natural Numbers", it's helpful to have a commonly accepted symbol for this collection of interesting objects. The symbol in common use is $\mathbf{N}$, sometimes rendered as $\mathbb{N}$. So we will say that "$\mathbf{N}$ *is the set containing the numbers* $1, 2, 3,$ *etc.*" and write this as $\mathbf{N} = \{1, 2, 3, \ldots\}$. Again I have sneaked in another bit of mathematical notation with the curly brackets $\{,\}$ that mathematicians use to denote a set of objects. One more bit of notation that comes in useful is the symbol $\in$ which is read as "is a member of", or "lies in", or just "in". For example, "$3 \in \mathbf{N}$" is read as "the number 3 is a member of the set of Natural Numbers". And "for every $n \in \mathbf{N} \ldots$ " is read as "for every Natural Number $n \ldots$". We can also use $\notin$ for the opposite purpose. For example, $0 \notin \mathbf{N}$ (the number 0 is not a member of the set of Natural Numbers). This idea of striking through a symbol is often used to negate the original meaning; for example, $n \neq 2$ is read as "n is not equal to 2".

It is probably also worth making a few general remarks about sets. The objects in a set are called its *members* or its *elements*. For the present you can think of a set as any collection of objects. We will see in Chapter 9 that this isn't sufficiently precise, but it will do for us until then. A set may be specified by listing all its members, or by describing all its members, or by giving some precise property that defines all its members. Examples of these three specification methods are

1. $A = \{1, 2, 3\}$,
2. B is the set of all the words in this book,
3. $C = \{n : n \in \mathbf{N} \text{ and } n < 5\}$.

To explain C, it's the set of Natural Numbers that are less than 5, so $C = \{1, 2, 3, 4\}$.

Most of our sets will be sets of numbers of various kinds. A set can only contain one copy of each object - it's either in the set or not in the set, so we consider that something like $\{1, 2, 1\}$ is actually the set $\{1, 2\}$ with two elements. From any set X we can form *subsets* by choosing to omit or to include each element. For example, the sets A and C in the previous paragraph are both subsets of **N**. At one extreme we can omit all the elements of our set X and have a subset which is empty, this is denoted by $\{\}$ or, more usually, by $\emptyset$. At the other extreme we can include all the elements of the original set X, so we regard X itself as a subset of X. Sometimes $\emptyset$ and the set X itself are called *improper* subsets of X and all the other subsets are called *proper* subsets of X.

2.5 Summarizing the properties of N

Now I would like to summarize some of the properties we have discussed. If m, n, p are any members of **N** then

1. $m+n$ and $m \times n$ both lie in **N** (we say that **N** is *closed* under addition and multiplication),
2. $m + n = n + m$ and $m \times n = n \times m$ (called the *commutative* laws),
3. $(m + n) + p = m + (n + p)$ and $(m \times n) \times p = m \times (n \times p)$ (called the *associative* laws),
4. $m \times (n+p) = m \times n + m \times p$ (multiplication is *distributive* over addition),
5. $1 \times m = m$ (we say that the number 1 is a *multiplicative identity*),
6. if $m \neq n$ then either $m < n$ or $n < m$ (we say that **N** is *ordered* by $<$).

Simple equations can sometimes be solved *within* **N** for an unknown quantity x. For example, the equation $3 \times x = 6$ has the solution $x = 2$ (which lies in **N**). However, the system is not without its defects because the apparently similar equation $6 \times x = 3$ has no solution in **N** (i.e. there is no Natural

Number x which, when multiplied by 6, produces 3). The way out of this difficulty is to extend the notion of number to include Positive Fractions, and we will do this in Chapter 3. In a similar way the equation $3 + x = 5$ has the solution $x = 2$ (which lies in $\mathbf{N}$). But the apparently similar equation $5 + x = 3$ has no solution in $\mathbf{N}$ (i.e. there is no Natural Number x which, when added to 5, produces 3). The way out of this second difficulty is to extend the notion of number to include zero and negative numbers, and we will do this in Chapter 4. Historically, fractions were used long before negative numbers and we will follow this precedent. Both Positive Fractions and negative numbers can be constructed from pairs of Natural Numbers, but an additional reason for dealing with fractions first is that you are less likely to have seen negative numbers constructed in this way.

Before moving on to the next two chapters, we remark that it is possible to define limited forms of division ($\div$) and subtraction ($-$) in $\mathbf{N}$. The equation $3 \times x = 6$ has solution $x = 2$ because $6 \div 3 = 2$. In general we can define $m \div n$ for Natural Numbers m and n to be the Natural Number x for which $n \times x = m$ **when such a Natural Number** x **exists**. When there is no Natural Number x for which $n \times x = m$ we leave division of m by n undefined here but, as indicated above, we will deal with this in the next chapter. In a similar way, the equation $3 + x = 5$ has solution $x = 2$ because $5 - 3 = 2$. In general we can define $m - n$ for Natural Numbers m and n to be the Natural Number x for which $n + x = m$ **when such a Natural Number** x **exists**. When there is no Natural Number x for which $n + x = m$ we leave the subtraction of n from m undefined here but, again as indicated above, we will deal with this in Chapter 4.

Chapter 3

The Positive Fractions ($\mathbf{Q}^+$)

3.1 Representation of Positive Fractions

You are probably familiar with some Positive Fractions such as $\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}$, etc. It is clear just by looking at these fractions that each is formed from two Natural Numbers, one on the top of the fraction and one on the bottom. A fraction is occasionally called a *quotient*, so we might call $\frac{3}{4}$ the quotient of 3 and 4. This explains the use of the symbol $\mathbf{Q}^+$ to denote the set of all Positive Fractions, where Q stands for "quotient" and + stands for "positive".

We can list all Positive Fractions in a two-dimensional array as follows

$$\begin{array}{cccccc}
\frac{1}{1}, & \frac{1}{2}, & \frac{1}{3}, & \frac{1}{4}, & \frac{1}{5}, & \dots \\[2ex]
\frac{2}{1}, & \frac{2}{2}, & \frac{2}{3}, & \frac{2}{4}, & \frac{2}{5}, & \dots \\[2ex]
\frac{3}{1}, & \frac{3}{2}, & \frac{3}{3}, & \frac{3}{4}, & \frac{3}{5}, & \dots \\[2ex]
\frac{4}{1}, & \frac{4}{2}, & \frac{4}{3}, & \frac{4}{4}, & \frac{4}{5}, & \dots \\[2ex]
\frac{5}{1}, & \frac{5}{2}, & \frac{5}{3}, & \frac{5}{4}, & \frac{5}{5}, & \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

Every Positive Fraction appears in this array - indeed there is some redundancy because (for example) $\frac{1}{1} = \frac{2}{2} = \frac{3}{3} = \dots$. If you want to locate the Positive Fraction $\frac{m}{n}$ (where m and n are Natural Numbers) in this array, you will find it in the m^{th} row and n^{th} column. For example, $\frac{5}{4}$ is in the 5^{th} row and 4^{th} column. Note that the array contains a copy of every Natural Number because we can write such a number m as $\frac{m}{1}$ (or $\frac{2m}{2}$, or $\frac{3m}{3}$, etc.).

For example, $3 = \frac{3}{1}$. In this sense the Positive Fractions form an extension of $\mathbf{N}$.

Familiarity with Positive Fractions may make it hard to see why anyone should want to explain them in "simpler" terms. And I've put "simpler" in quotes because it refers not to the explanation but to the objects used. Mathematicians do want to do this because we like to start with the simplest possible objects, whose properties we are sure of, and use them to build more complicated objects and derive their properties. In this chapter, the simple objects are the Natural Numbers whose properties we all know and love, and the more complicated objects are the Positive Fractions.

According to Mathematicians, each Positive Fraction is **constructed from** (you might say represented by) an *ordered pair* of Natural Numbers (m, n) that we choose to write as $\frac{m}{n}$. In other words we just replace the comma in (m, n) by the fraction bar - and drop the brackets ( ). The Natural Number m (on the top) is called the *numerator* and the Natural Number n (on the bottom) is called the *denominator*. So $\frac{3}{4}$ has numerator 3 and denominator 4.

I used the term "ordered pair" in the previous paragraph and it is worth remarking on what this means because we will use it again and again in many of the subsequent chapters. All it means is that we have two objects, one of which is regarded as the first, and the other is regarded as the second. In mathematics, such pairs are usually enclosed in round brackets () with a comma between them. If you ever did any 2-dimensional coordinate geometry at school, you will have seen ordered pairs of numbers used to describe the position of a point on a plane. The same thing is done on geographical maps. Of course the order matters: for example, the ordered pairs $(1, 2)$ and $(2, 1)$ are not identical. But in the current chapter and in subsequent chapters, we will give rules, different ones in different chapters as the occasion demands, for regarding non-identical pairs as equivalent, i.e. representing the same number. In this chapter we will, for example, regard the ordered pair $(1, 2)$ as equal to the ordered pair $(2, 4)$, because we are thinking of these as the fractions $\frac{1}{2}$ and $\frac{2}{4}$. The rule in this chapter is that $(m, n) = (p, q)$ (i.e. $\frac{m}{n} = \frac{p}{q}$) if and only if $m \times q = p \times n$. The reason for this is less than obvious so the next couple of pages provide the explanation.

3.2 Equality between fractions

Each fraction has many alternative representations. For example, $\frac{1}{2} = \frac{2}{4}$ and it is easy to appreciate why this is the case. If you cut a pie precisely in half, the size of the half piece is exactly the same as if you cut 2 pies into 4 equal sized pieces. And if you cut a 1 metre ruler in half, each piece is the same size as the pieces you get by cutting a 2 metre ruler into 4 equal lengths. (See Figures 3.1 and 3.2.)

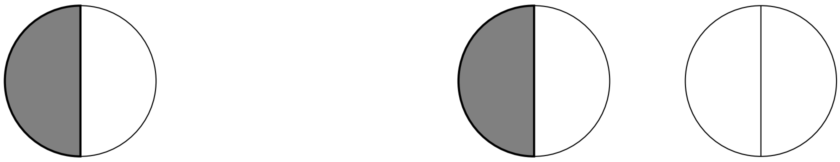


Figure 3.1: Half of one pie is the same size as a quarter of two pies.

Figure 3.2: Half of one metre is the same as a quarter of two metres.

When you were at school you were probably taught that you can cancel common factors that appear in the numerator and the denominator. For example, by cancelling the common factor 3 we get

$$\frac{15}{12} = \frac{5 \times 3}{4 \times 3} = \frac{5}{4}.$$

So $\frac{15}{12}$ and $\frac{5}{4}$ represent the same number. This is all well and good if you can spot common factors. It is not hard to see that 15 and 12 have a common factor 3, but what about 391 and 667? If someone says that $\frac{391}{667} = \frac{17}{29}$, how can you check if they are correct? Well, there are methods to determine the common factor of two numbers, but an even easier way to check is to *cross-multiply*:

$$\frac{391}{667} = \frac{17}{29} \text{ if and only if } 391 \times 29 = 17 \times 667.$$

And yes, in this case both values come to $11,339$, so the two fractions are equal. If you try this out for the case of $\frac{15}{12}$ and $\frac{5}{4}$ you will see that $15 \times 4 = 60$ and $5 \times 12 = 60$, so this proves that $\frac{15}{12}$ and $\frac{5}{4}$ represent the same number, without having to spot the common factor 3 in 15 and 12.

You may remember cross-multiplying without remembering why it works, so here is a little explanation. If you multiply $\frac{1}{2}$ by 2, you get 1. Similarly if you multiply $\frac{1}{n}$ by n you get 1. You can think of this as adding n lots of

$\frac{1}{n}$. Extending this a bit, if we multiply $\frac{m}{n}$ by n, we get m. So faced with the question of whether $\frac{m}{n}$ equals $\frac{p}{q}$, we multiply both by $n \times q$. (You might recognize from your schooldays that $n \times q$ is a common denominator for the fractions $\frac{m}{n}$ and $\frac{p}{q}$; we discuss this further below.) On one hand this gives

$$\frac{m}{n} \times (n \times q) = (\frac{m}{n} \times n) \times q = m \times q.$$

On the other hand it gives

$$\frac{p}{q} \times (n \times q) = (\frac{p}{q} \times q) \times n = p \times n.$$

So the test for equality of $\frac{m}{n}$ and $\frac{p}{q}$ is whether or not $m \times q = p \times n$. But the really revolutionary view of this is to say that the equality $\frac{m}{n} = \frac{p}{q}$ *means* that $m \times q = p \times n$. In other words we can *define* equality between fractions in this way:

$$\text{we say } \frac{m}{n} = \frac{p}{q} \text{ if and only if } m \times q = p \times n.$$

The important thing to note here is that the right-hand side only contains Natural Numbers ($m \times q$ and $p \times n$), so our definition of equality between fractions only involves Natural Numbers. Thus equality of fractions is defined in terms of something we already understand: equality of Natural Numbers.

3.3 Addition and multiplication of fractions

So each Positive Fraction f can be viewed as an *ordered* pair of Natural Numbers (m, n) that we choose to write as $\frac{m}{n}$, and we regard the pair $\frac{m}{n}$ and the pair $\frac{p}{q}$ as defining the same fraction f if and only if $m \times q = p \times n$. This enables us to define addition of two Positive Fractions f and g by the creation of a *common denominator* for f and g. To explain this by means of an example, suppose that we want to add $\frac{5}{7}$ and $\frac{4}{9}$. The two denominators are 7 and 9 so we will use the common denominator $7 \times 9 = 9 \times 7 = 63$. By the definition of equality,

$$\frac{5}{7} = \frac{5 \times 9}{7 \times 9} \text{ because } 5 \times (7 \times 9) = (5 \times 9) \times 7.$$

Similarly

$$\frac{4}{9} = \frac{4 \times 7}{9 \times 7} \text{ because } 4 \times (9 \times 7) = (4 \times 7) \times 9.$$

So we have
$$\frac{5}{7} = \frac{5 \times 9}{7 \times 9} = \frac{45}{63}, \quad \text{and} \quad \frac{4}{9} = \frac{4 \times 7}{9 \times 7} = \frac{28}{63}.$$
Now our two fractions have been expressed with the common denominator 63. The first consists of 45 lots of $\frac{1}{63}$ and the second consists of 28 lots of $\frac{1}{63}$, so altogether we have $45 + 28 = 73$ lots of $\frac{1}{63}$. In other words
$$\frac{5}{7} + \frac{4}{9} = \frac{45}{63} + \frac{28}{63} = \frac{45 + 28}{63} = \frac{73}{63}.$$

There is nothing special about the numbers we chose in this example. We can add $f = \frac{m}{n}$ and $g = \frac{r}{s}$ as follows
$$f + g = \frac{m}{n} + \frac{r}{s} = \frac{m \times s}{n \times s} + \frac{r \times n}{s \times n} = \frac{m \times s + r \times n}{n \times s}.$$
In fact we will take this as a *definition* of addition for Positive Fractions. Strictly speaking we should also check that we get the same answer for $f + g$ irrespective of the representation of f and g, so if we have $f = \frac{m}{n} = \frac{p}{q}$ and $g = \frac{r}{s} = \frac{t}{u}$, we get the same answer whether we use $f = \frac{m}{n}$ or $f = \frac{p}{q}$, and whether we use $g = \frac{r}{s}$ or $g = \frac{t}{u}$. Rest assured this is indeed the case, so this is a *good* definition! As an example, let's try adding $f = \frac{3}{4} = \frac{15}{20}$ and $g = \frac{2}{7} = \frac{4}{14}$. We can do this as $f + g = \frac{3}{4} + \frac{2}{7} = \frac{3 \times 7 + 2 \times 4}{4 \times 7} = \frac{29}{28}$. But equally well we can do it as $f + g = \frac{15}{20} + \frac{4}{14} = \frac{15 \times 14 + 4 \times 20}{20 \times 14} = \frac{290}{280} = \frac{29}{28}$. So, yes, we do indeed get the same answer - bingo!

When you were at school you were probably asked to find the *lowest* common denominator of two fractions. This would have involved finding the highest common factor (HCF) of two Natural Numbers and their lowest common multiple (LCM). I don't intend to repeat the pain here. I just remark that the definition given above for addition always works, but sometimes the resulting fraction is not in its simplest terms, and this will certainly happen if the two denominators have a common factor.

So adding fractions is simple in theory but a bit of a pain in practice if you want to get the answer in its simplest form. We will concentrate on the theory, not the practice! Multiplication of fractions is simpler than addition because we do not need to find a common denominator. For example,
$$\frac{5}{7} \times \frac{4}{9} = \frac{5 \times 4}{7 \times 9} = \frac{20}{63}.$$
Why is this? Let's break the calculation down in stages. The fraction $\frac{4}{9}$ represents 4 lots of $\frac{1}{9}$. If we multiply this by 5, we now have 5 lots of 4 lots

of $\frac{1}{9}$, in other words 20 lots of $\frac{1}{9}$, that is

$$5 \times \frac{4}{9} = \frac{20}{9}.$$

But we really wanted to multiply by $\frac{5}{7}$ rather than by 5, and $\frac{5}{7}$ is a seventh of 5. So we want a seventh of $\frac{20}{9}$. Dividing by 7 gives

$$\frac{20}{7 \times 9} = \frac{20}{63}.$$

Of course this can be generalized to give a *definition* of multiplication for fractions f and g. If $f = \frac{m}{n}$ and $g = \frac{r}{s}$ then we *define*

$$f \times g = \frac{m \times r}{n \times s}.$$

As with addition, if you want to get the answer in its simplest form you may still need to spot common factors in $m \times r$ and $n \times s$. But again this turns out to be a good definition, meaning that it does not depend on how the fractions f and g are represented.

As already mentioned, each Natural Number m can be identified with the Positive Fraction $\frac{m}{1}$ so we may write $m = \frac{m}{1}$ and, in particular, $1 = \frac{1}{1}$.

The properties already mentioned for the Natural Numbers $\mathbf{N}$ are inherited by the Positive Fractions $\mathbf{Q}^+$. We can add and multiply Positive Fractions and the result is always a Positive Fraction. Furthermore, if f, g, h are any Positive Fractions then

1. $f + g = g + f$ and $f \times g = g \times f$ (the commutative laws),
2. $(f + g) + h = f + (g + h)$ and $(f \times g) \times h = f \times (g \times h)$ (the associative laws),
3. $f \times (g + h) = f \times g + f \times h$ (multiplication is distributive over addition),
4. $1 \times f = f$ (the number $1 = \frac{1}{1}$ is a multiplicative identity).

3.4 Multiplicative inverses (reciprocals) and division

But now there is a further interesting property. For each Positive Fraction f there is another Positive Fraction f^{-1}, that is often referred to as the *reciprocal* of f and written as $\frac{1}{f}$, with the property that $f \times f^{-1} = 1$. Mathematicians generally call this the *multiplicative inverse* of f. If $f = \frac{m}{n}$ then

$f^{-1} = \frac{n}{m}$ because $\frac{m}{n} \times \frac{n}{m} = \frac{m \times n}{n \times m} = \frac{1}{1} = 1$. For example, the multiplicative inverse (the reciprocal) of $\frac{5}{7}$ is $\frac{7}{5}$.

By using multiplicative inverses (reciprocals) we can *define* division as follows. If f and g are Positive Fractions, then $f \div g$ is defined as $f \times g^{-1}$, and we can write this in alternative forms as $f \times \frac{1}{g}$ or more briefly as f/g. For example, if we want $6 \div 3$ we take $f = 6 = \frac{6}{1}$ and $g = 3 = \frac{3}{1}$, so that $g^{-1} = \frac{1}{3}$. Then $6 \div 3 = \frac{6}{1} \times \frac{1}{3} = \frac{6 \times 1}{3 \times 1} = \frac{6}{3} = \frac{2}{1} = 2$. I'm not suggesting you should actually do divisions like this in practice, but just illustrating how division is really just multiplication in disguise. In the next chapter we will see that subtraction is really just addition in disguise.

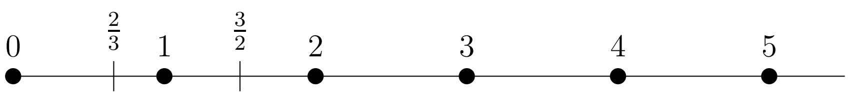
We remarked in Chapter 2 that we could not solve an equation like $6 \times x = 3$ in $\mathbf{N}$. But we can now solve it within the set of Positive Fractions $\mathbf{Q}^+$ and give the answer $x = \frac{1}{2}$. Indeed, the need to solve such equations can be regarded as providing the impetus for defining fractions.

While we are discussing reciprocals, it is appropriate to mention negative powers. For example, 5^{-3} means $(5^{-1})^3$, the cube of $\frac{1}{5}$. In general for any numbers x and n, the number x^{-n} means $(x^{-1})^n$, the n^{th} power of $\frac{1}{x}$. As a consequence,

$$x^{-n} = \left(\frac{1}{x}\right)^n = \frac{1}{x^n} = (x^n)^{-1}.$$

3.5 Order on the number line

We can extend the picture of a number line mentioned in the case of $\mathbf{N}$ to represent Positive Fractions $\mathbf{Q}^+$. It is convenient to start the number line on the left at 0 (even though we haven't included 0 as a Natural Number or as a Positive Fraction). Then as an example, $\frac{3}{2}$ lies halfway between 0 and 3 while $\frac{2}{3}$ lies one third of the way between 0 and 2.



This picture enables us to put Positive Fractions into *order*. We write $a > b$ if a is to the right of b and we write $a < b$ if a is to the left of b. The only other alternative is that $a = b$. From our picture we can see that $\frac{3}{2} > \frac{2}{3}$, and we can also write this as $\frac{2}{3} < \frac{3}{2}$. By cross-multiplying, it can also be seen that for Positive Fractions $\frac{m}{n}$ and $\frac{p}{q}$ we have

$$\frac{m}{n} < \frac{p}{q} \quad \text{if and only if} \quad m \times q < p \times n.$$

3.6 Notation and a warning

On a minor point of notation, fractions are sometimes written and typeset "inline", so that we might write $\frac{3}{2}$ as $3/2$. Also when you were at school you may have been encouraged to avoid so-called "improper" fractions. A *proper* fraction is one where the numerator (on the top) is less than the denominator (on the bottom), while an *improper* fraction is one where the numerator is greater than (or equal to) the denominator. For example, you might have been told to write $\frac{3}{2}$ as $1\frac{1}{2}$ ("one and a half"). But, as we are concerned with theory rather than practice, we will usually be perfectly content with improper fractions.

On a rather more important notational point, mathematicians generally tire quickly of writing excessive symbols. So from here on, we will only use the multiplication sign $\times$ when we want emphasis or to avoid ambiguity. So, if the symbols x and y represent numbers, then xy is taken to be the number $x \times y$. Thus, for example, the distributive property above can be expressed as $f(g+h) = fg+fh$. This saves an awful lot of typing! Of course, if we are actually using numerals like 23 and 71, we can't write 2371 for the product because this would be horribly ambiguous, so we will still write 23×71.

By this point in our discussion, it is becoming increasingly obvious that we need to discuss the role of the number 0 and negative numbers. So our next section covers the Integers, i.e. the positive and negative whole numbers along with 0. But just before moving on to this, note that our Positive Fractions *never* have 0 as a denominator. This is not because mathematicians have an unreasonable hatred of dividing by zero (our hatred of this is perfectly reasonable) and are spitefully stopping people from doing this, but because it just doesn't make sense. You can cut an object into, say, 10 parts each of which is a tenth of the whole, or 3 parts each of which is a third of the whole, or even (stretching English usage a bit) into 1 part comprising the whole of the whole, but you simply can't cut an object into 0 parts. Perhaps this is why 0 was a relative latecomer to our collection of numbers; it just doesn't behave quite like all the others.

3.7 Summarizing the properties of $\mathbf{Q}^+$

We've already noted some of these above, but here is a complete list including the notions of multiplicative inverses and order. If f, g, h are any members of $\mathbf{Q}^+$ then

1. $f+g$ and fg both lie in $\mathbf{Q}^+$ (we say that $\mathbf{Q}^+$ is closed under addition

and multiplication),

2. $f + g = g + f$ and $fg = gf$ (the commutative laws),

3. $(f + g) + h = f + (g + h)$ and $(fg)h = f(gh)$ (the associative laws),

4. $f(g + h) = fg + fh$ (multiplication is distributive over addition),

5. $1f = f$ (the number $1 = \frac{1}{1}$ is a multiplicative identity),

6. for each $f \in \mathbf{Q}^+$, there is a multiplicative inverse (reciprocal) $f^{-1} \in \mathbf{Q}^+$ such that $f(f^{-1}) = 1$,

7. if $f \neq g$ then either $f < g$ or $g < f$ (we say that $\mathbf{Q}^+$s is ordered by $<$).

Having dealt with Positive Fractions in this chapter, we will return to the Natural Numbers in the next chapter and see how they can be used to construct zero and all the negative whole numbers. We will again use ordered pairs of Natural Numbers but define equality between these pairs in a different way to what we did for fractions. So every negative whole number will be defined as a pair of Natural Numbers and many such pairs will yield the same negative whole number. The complete collection of whole numbers, including the positive ones (i.e. the Natural Numbers), the negative ones, and zero, is called the set of Integers and is denoted by $\mathbf{Z}$. Within $\mathbf{Z}$ we will be able to solve equations like $5 + x = 3$ and give the answer $x = -2$ with a clear conscience.

Chapter 4

The Integers (Z)

4.1 Representation of Integers

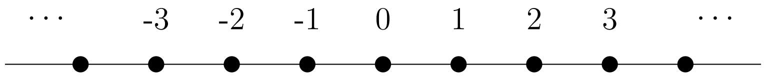
In the previous chapter we showed how the set of Positive Fractions $\mathbf{Q}^+$ could be constructed from the set of Natural Numbers $\mathbf{N}$. We could explain Positive Fractions to extraterrestrial aliens even if they had no prior knowledge of them, provided that they already had an appreciation of the Natural Numbers. We showed that all the properties and operations relating to Positive Fractions, such as equality, addition, multiplication, division and order can be defined using only the Natural Numbers. In the current chapter we will do a similar job for negative whole numbers and zero, by constructing these from the Natural Numbers. We will apply a similar technique, again using pairs of Natural Numbers but in a different way. The set of all whole numbers, positive, negative and zero, is denoted by $\mathbf{Z}$. The letter $\mathbf{Z}$ comes from the German word Zahl for number. We refer to these numbers as the *Integers*.

Unlike fractions, where a number such as $\frac{3}{4}$ clearly "contains" the two Natural Numbers 3 and 4, a negative Integer such as -2 does not obviously "contain" two Natural Numbers. But it can be written as $1 - 3$ (i.e. as 1 *minus* 3) and we can apply this idea to any negative Integer and also to zero. We'll go into the formal details below, but first we'll give an informal description of the Integers and show how to represent them on a number line.

Just like $\mathbf{N}$ we can list the Integers:

$$\mathbf{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

In technological societies nowadays, most people would concede that $0, -1, -2, -3, \ldots$ are indeed numbers. They weren't really used in a consistent fashion until around 500 CE in India. Probably, people feel suspicious about negative numbers because they seem less tangible than the Natural Numbers (you may see two cats, but you'll never see *minus* two cats!). We can picture **Z** using a number line with zero in the middle, the positive Integers to the right and the negative Integers to the left.



This picture immediately suggests an ordering of the Integers. For $a, b \in$ **Z** (i.e. for Integers a and b) we will write $a > b$ if a is to the right of b, and $a < b$ if a is to the left of b. Just beware that this entails things like $-7 < 3$. Even though the quantity 7 is larger than the quantity 3 we take account of the minus sign and reckon -7 as less than 3. If you think of numbers representing a bank balance you won't go wrong. Someone who owes the bank 7 dollars (i.e has -7 dollars in their account) clearly has less money than someone who has 3 dollars in their account. We'll give a formal definition of the ordering after we have shown how to construct the Integers.

We *construct* the set **Z** from the set **N** by identifying an Integer such as -2 with all pairs of Natural Numbers of the form $(a, a+2)$ [**think of this as** a **minus** $(a+2)$]. Because I want you to think of the ordered pair $(a, a+2)$ as a minus $(a+2)$, I will write this pair as $(a \ominus (a+2))$, so the symbol $\ominus$ just replaces the comma. (I won't use the minus sign itself because it carries too much baggage.) For a while I will use red ink to distinguish the new Integers from the old Natural Numbers. If you want a word for the invented symbol $\ominus$ you could call it "oh-minus".

For example, we can represent -2 as $(1 \ominus 3)$ or $(2 \ominus 4)$ or $(3 \ominus 5)$ or even (just to be obscure) as $(165 \ominus 167)$. Similarly 0 is identified with all pairs $(a \ominus a)$, the simplest of which is probably $(1 \ominus 1)$, but we could equally well represent it (for example) as $(5 \ominus 5)$. Rather awkwardly we then have to identify (e.g.) $2 \in \mathbf{Z}$ with all pairs $((a+2) \ominus a)$ (where $a \in \mathbf{N}$). Thus the Integer 2 is represented by a pair of Natural Numbers, and so it is something ever so slightly different from the Natural Number 2.

This way of presenting Integers may well be unfamiliar to you but it is similar to what was done with Positive Fractions where (e.g.) a half can be written as $\frac{1}{2}$ or $\frac{2}{4}$ or $\frac{3}{6}$ or, more generally as $\frac{a}{2a}(a \in \mathbf{N})$. Here too there was an awkwardness about identifying $2 \in \mathbf{N}$ with all "pairs" $\frac{2a}{a}$, where $a \in \mathbf{N}$, and the Positive Fraction $\frac{2}{1}$ was something ever so slightly different from the Natural Number 2.

In general, the pair of Natural Numbers $(m \ominus n)$ represents an Integer. If $m > n$, we write it as $m - n$, if $n > m$ we write it as $-(n - m)$, and if $m = n$ we write it as 0. For example, $(7 \ominus 4) = 3$, $(4 \ominus 7) = -3$ and $(6 \ominus 6) = 0$. Note that the subtraction involved in $(4 \ominus 7)$ is actually the subtraction $7 - 4$ of the Natural Number 4 from the Natural Number 7, and we then put a minus sign in front of the resulting Natural Number 3 to give the Integer -3.

4.2 Equality between Integers

An important question that arose with fractions was how to identify if two apparently different looking fractions are in fact equal. A similar question arises when we represent Integers as pairs of Natural Numbers. Is the Integer $(36 \ominus 921)$ the same as $(245 \ominus 1130)$? Again there is an easy solution once we realize that we are really asking if $36 - 921 = 245 - 1130$, and we can answer this question without using subtraction or minus signs by adding $921 + 1130$ to both sides to give the equivalent question: does $36 + 1130$ equal $245 + 921$? The general question whether $(m \ominus n)$ represents the same Integer as $(p \ominus q)$ really asks if $m - n = p - q$, or equivalently (by adding $n + q$ to both sides) if $m + q = p + n$. Consequently we *define* equality between these ordered pairs of Natural Numbers by saying

$$(m \ominus n) = (p \ominus q) \text{ if and only if } m + q = p + n.$$

The important thing to note here is that the right-hand side only contains Natural Numbers ($m + q$ and $p + n$), so our definition of equality between Integers only involves Natural Numbers. Thus equality of Integers is defined in terms of something we already understand: equality of Natural Numbers. So for our example, we compare $36+1130$ with $245+921$, and since both come to 1166, they are equal and consequently both $(36 \ominus 921)$ and $(245 \ominus 1130)$ represent the same Integer.

Just as a fraction can be written in its simplest form by cancelling common factors between its numerator and denominator, we can come up with a simplest form for the Integer represented by a pair $(m \ominus n)$. How to do this depends on whether or not m is greater than n. If $m > n$, then $(m \ominus n) = ((m - n + 1) \ominus 1)$. On the other hand if $m < n$ or if $m = n$, then $(m \ominus n) = (1 \ominus (n - m + 1))$. So every positive Integer can be represented as a pair $(p \ominus 1)$, where p is a Natural Number. And every negative Integer, along with zero, can be represented as a pair $(1 \ominus q)$, where q is a Natural Number.

Let's see how this works for the Integer represented by $(36 \ominus 921)$. It's easy if you just concentrate on reducing the smaller of 36 and 921 to 1 by subtracting an appropriate amount. Here the appropriate amount is 35 $(= 36 - 1)$, and $921 - 35 = 886$ so we get $(36 \ominus 921) = (1 \ominus 886)$. If you apply the same procedure to $(245 \ominus 1130)$ by subtracting 244 $(= 245 - 1)$, you get the same answer $(1 \ominus 886)$. This represents the (negative) Integer -885.

Here is a second (positive) example. If we have the pair $(27 \ominus 13)$ we reduce the smaller of 27 and 13 to 1 by subtracting 12 $(= 13 - 1)$, giving the equivalent pair $(15 \ominus 1)$, which represents the (positive) Integer 14.

In general if n is a Natural Number, the pair $(1 \ominus (n+1))$ represents the negative Integer $-n$, and the pair $((n+1) \ominus 1)$ represents the positive Integer n. The Integer 0 has the simplest form $(1 \ominus 1)$.

4.3 Addition and multiplication of Integers

Once we've got hold of **Z** we can define the operations $+$ and $\times$. Addition is very easy:

$$(m \ominus n) + (p \ominus q) = ((m+p) \ominus (n+q)).$$

For example, $(-3) + 2 = (1 \ominus 4) + (3 \ominus 1) = (4 \ominus 5) = -1$. You might like to think what happens if you add $(m \ominus n)$ and $(n \ominus m)$.

Multiplication is just a shade more complicated and it helps to remember that we think of $(m \ominus n)$ as $m - n$, so we would expect $(m \ominus n) \times (p \ominus q)$ to come out as a representative of $(m - n) \times (p - q) = mp - mq - np + nq$. So our definition is

$$(m \ominus n) \times (p \ominus q) = ((mp + nq) \ominus (mq + np)).$$

As an example, let's try $(-2) \times 3$. First write -2 as the pair $(1 \ominus 3)$ and 3 as the pair $(4 \ominus 1)$. Then we get

$$(-2) \times 3 = (1 \ominus 3) \times (4 \ominus 1) = ((4+3) \ominus (1+12)) = (7 \ominus 13) = -6.$$

So we get $(-2) \times 3 = -6$. Now you try $(-2) \times (-3)$. You should get 6. Make up a few more examples and give them a try.

As with fractions, we really should check that these definitions of $+$ and $\times$ do not depend of the particular representations of the Integers involved, i.e. if we have $(m \ominus n) = (r \ominus s)$ and $(p \ominus q) = (t \ominus u)$ then we find that $(m \ominus n) + (p \ominus q) = (r \ominus s) + (t \ominus u)$ and $(m \ominus n) \times (p \ominus q) = (r \ominus s) \times (t \ominus u)$. I will check the first of these (the addition) and I invite you to check the second (the multiplication).

Let us give names X and Y to the two Integers involved in the addition, i.e. $X = (m \ominus n) + (p \ominus q)$ and $Y = (r \ominus s) + (t \ominus u)$. According to the definition of addition, $X = ((m+p) \ominus (n+q))$ and $Y = ((r+t) \ominus (s+u))$. So according to the definition of equality of Integers, $X = Y$ if and only if

$$(m+p) + (s+u) = (r+t) + (n+q),$$

i.e (by rearranging slightly) if and only if

$$(m+s)+(p+u)=(r+n)+(t+q).$$

But $(m \ominus n) = (r \ominus s)$ gives $m+s=r+n$, and $(p \ominus q) = (t \ominus u)$ gives $p+u=t+q$. Consequently $(m+s)+(p+u)$ does equal $(r+n)+(t+q)$, and so $X = Y$.

We find the properties mentioned above for $\mathbf{N}$, i.e. the commutative, associative, distributive laws, and the multiplicative identity property of the Integer 1 also hold for $\mathbf{Z}$.

4.4 Zero, additive inverses, subtraction and order

But now there are some interesting additional properties (and now I will drop the red ink).

(a) For each $a \in \mathbf{Z}$, $0+a=a$. We say that the number 0 is an *additive identity*.

(b) For each $a \in \mathbf{Z}$ there is a number $(-a) \in \mathbf{Z}$ such that $a+(-a)=0$. The number $(-a)$ is called the *additive inverse* or negative of a.
If a is represented by the pair of Natural Numbers $(m \ominus n)$, then $(-a)$ is given by $(n \ominus m)$ because $(m \ominus n)+(n \ominus m) = ((m+n)\ominus(n+m)) = 0$.

We can now *define* subtraction by putting

$$a-b=a+(-b) \quad \text{for} \quad a,b \in \mathbf{Z}.$$

As an example, consider the equation $5+x=3$. If we subtract 5 from both sides then on the left-hand side we get $5+x-5=5+(-5)+x=0+x=x$ by properties (a) and (b) above. On the right-hand side we get $3-5=3+(-5)=(4\ominus 1)+(1\ominus 6)=(5\ominus 7)=-2$. So we conclude that $x=-2$. Of course this is terribly pedantic and no-one in their right mind would actually solve the equation $5+x=3$ in this manner. I'm just using it to show that subtraction of Integers as defined above behaves as you would expect.

We have already mentioned informally how order among the Integers can be pictured using a number line. We can define order more precisely by looking at what is meant by saying that $(m \ominus n)$ is greater than zero (positive) or less than zero (negative). As already mentioned, an Integer $(m \ominus n)$ is called positive if the Natural Numbers m and n have $m > n$ and it is called negative if $m < n$. So if x and y are Integers, we will say that x is less than y ($x < y$) if and only if $x - y$ is negative. If x is represented by $(m \ominus n)$ and y is represented by $(p \ominus q)$ then $x - y = x + (-y)$ is represented by $(m \ominus n) + (q \ominus p) = ((m+q) \ominus (p+n))$, and this is negative if and only if $m + q < n + p$. So

$$(m \ominus n) < (p \ominus q) \text{ if and only if } m + q < p + n.$$

The important thing to note here is that the right-hand side only contains Natural Numbers ($m+q$ and $p+n$), so our definition of order between Integers only involves Natural Numbers.

There are three alternatives when comparing the Integers x and y: (i) $x < y$, (ii) $x = y$, (iii) $y < x$. Of course we may write $x < y$ as $y > x$, etc.

4.5 Multiplying negative Integers

You may recall being told at school that “two minuses make a plus”, or words similar. This refers to multiplication (*not* to addition), and is short for saying that if you multiply two negative numbers, you will get a positive number. We can prove that $(-a) \times (-b) = a \times b$ by using the properties listed above as (a) and (b). If you were left wondering at school why on earth two minuses make a plus, this is a proper proof.

Firstly we show that $0 \times x = 0$ for every $x \in \mathbf{Z}$. To see this note $0+1 = 1$ (by property (a) above). Multiplying this by x gives $(0 \times x) + (1 \times x) = 1 \times x$. But $1 \times x = x$ (by the multiplicative identity property), so we get

$$(0 \times x) + x = x.$$

If we now add $(-x)$ to both sides of this equation and use property (b) above we obtain

$$0 \times x = 0,$$

and this is true for any Integer x.

Using this result, for any $a, b \in \mathbf{Z}$ we have

$$\begin{aligned} 0 &= 0 \times (-b) \\ &= [a + (-a)] \times (-b) \quad (\text{using } a + (-a) = 0) \\ &= (a \times (-b)) + ((-a) \times (-b)). \end{aligned}$$

Adding $a \times b$ to both sides gives

$$a \times b = (a \times b) + (a \times (-b)) + ((-a) \times (-b)).$$

But then note that $a \times b + a \times (-b) = a \times [b + (-b)] = a \times 0 = 0$. So we get

$$a \times b = (-a) \times (-b).$$

The proof given above follows from the properties that we have established for **Z**. It is a good proof because it applies whether or not a and b are separately positive, or negative, or even zero.

An alternative proof that the product of two negative Integers is positive can be obtained from the definition of a negative Integer. We can take the two negative Integers to be $(1 \ominus n)$ and $(1 \ominus q)$ where $n > 1$ and $q > 1$, i.e. where $n - 1$ and $q - 1$ are Natural Numbers. By the multiplication rule for Integers

$$(1 \ominus n) \times (1 \ominus q) = ((1 + nq) \ominus (n + q)).$$

To prove that $((1 + nq) \ominus (n + q))$ represents a positive Integer it suffices to show that $(1 + nq) > (n + q)$. But $n - 1$ and $q - 1$ are Natural Numbers, and so therefore is their product, and consequently $(n-1)(q-1) > 0$. This gives $nq - n - q + 1 > 0$. Hence $(1 + nq) > (n + q)$ and consequently $(1 \ominus n) \times (1 \ominus q)$ is positive. Furthermore, the pair $((1 + nq) \ominus (n + q))$ is the same pair that will be obtained from the product $(n \ominus 1) \times (q \ominus 1)$, so we deduce that

$$(1 \ominus n) \times (1 \ominus q) = (n \ominus 1) \times (q \ominus 1).$$

A further comment about "two negatives" follows from the fact that for any $a \in \mathbf{Z}$ we have $a + (-a) = 0$ (property (b) above). Since $(-a)$ is just as good an element of **Z** as a, we also have $(-a) + (-(-a)) = 0$. If we now add a to both sides of this we get $a + (-a) + (-(-a) = a$. Then, using $a + (-a) = 0$ we get $-(-a) = a$ for any $a \in \mathbf{Z}$.

Again there is another way of proving this by going back to the representation of an Integer a in the form $(m \ominus n)$. If $a = (m \ominus n)$ then $-a = (n \ominus m)$, and so $-(-a) = (m \ominus n) = a$.

4.6 Summarizing the properties of Z

Here is a summary of the properties we have discussed. If a, b, c are any members of **Z** then

1. $a + b$ and ab both lie in **Z** (we say that **Z** is closed under addition and multiplication),

2. $a + b = b + a$ and $ab = ba$ (called the commutative laws),

3. $(a+b)+c = a+(b+c)$ and $(ab)c = a(bc)$ (called the associative laws),

4. $a(b+c) = ab+ac$ (multiplication is distributive over addition),

5. $0+a = a$ (0 is an additive identity),

6. $1a = a$ (1 is a multiplicative identity),

7. for each $a \in \mathbf{Z}$ there is an additive inverse $(-a) \in \mathbf{Z}$ such that $a+(-a) = 0$,

8. for each $a \in \mathbf{Z}$ precisely one of the following three alternatives holds
   i) $a > 0$, ii) $a = 0$, iii) $0 > a$,
   and in case i) we say a is positive, in case iii) we say a is negative.

9. If $a, b \in \mathbf{Z}$ and $a > 0, b > 0$ (i.e. a and b are positive), then $a+b > 0$ and $ab > 0$.

We can then define $a > b$ to mean $a - b > 0$, define $b < a$ to mean $a > b$, define $a \geq b$ to mean that either $a > b$ or $a = b$, and define $b \leq a$ to mean $a \geq b$.

The Integers allow more equations to be solved. For example, we can now solve an equation like $5 + x = 3$ that we mentioned in Chapter 2. The solution is simply $x = -2$ (which lies in **Z** but not in **N**). So, to summarize, we have seen how zero and negative whole numbers can be produced from the Natural Numbers **N**. The entire collection of positive whole numbers, negative whole numbers and zero forms the Integers **Z**. In a similar way in the next chapter we describe how the Positive Fractions can be extended by including zero and negative fractions, with the complete collection forming the Rational Numbers **Q**.

Chapter 5

The Rational Numbers (Q)

5.1 Construction of the Rational Numbers

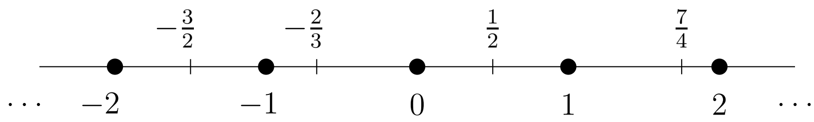
In Chapter 3 we showed how the Positive Fractions could be constructed from pairs of Natural Numbers that we wrote in the form $\frac{m}{n}$ with a rule specifying when two Positive Fractions are equal. In Chapter 4 we showed how the Integers could be constructed from pairs of Natural Numbers that we wrote in the form $(m \ominus n)$ with a (different) rule specifying when two Integers are equal. In the current chapter we explain how to combine the two ideas to produce the set of all fractions, positive, negative, zero, and including all the Integers. The resulting numbers are called the Rational Numbers (rational from "ratio") and the set of these is denoted by **Q**, where the letter Q stands for quotient.

As with **N**, $\mathbf{Q}^+$ and **Z** we can list all the possible fractions as shown in the array below.

$$
\begin{array}{ccccccc}
0, & \frac{1}{1}, & \frac{1}{2}, & \frac{1}{3}, & \frac{1}{4} & \frac{1}{5} & \ldots \\
 & -\frac{1}{1}, & -\frac{1}{2}, & -\frac{1}{3}, & -\frac{1}{4} & -\frac{1}{5} & \ldots \\
 & \frac{2}{1}, & \frac{2}{2}, & \frac{2}{3}, & \frac{2}{4} & \frac{2}{5} & \ldots \\
 & -\frac{2}{1}, & -\frac{2}{2}, & -\frac{2}{3}, & -\frac{2}{4} & -\frac{2}{5} & \ldots \\
 & \frac{3}{1}, & \frac{3}{2}, & \frac{3}{3}, & \frac{3}{4} & \frac{3}{5} & \ldots \\
 & -\frac{3}{1}, & -\frac{3}{2}, & -\frac{3}{3}, & -\frac{3}{4} & -\frac{3}{5} & \ldots \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}
$$

We have written zero as 0, but of course it could be written in the form of a fraction as $\frac{0}{1}$ or indeed as $\frac{0}{n}$ for any Natural Number n. So with that clarification, every possible fraction appears in this listing and, as with the Positive Fractions, there is some redundancy because (for example) $\frac{1}{1} = \frac{2}{2} = \frac{3}{3} = \ldots$. The list also contains a copy of every Integer, for example $-\frac{3}{1}$ is a copy of the Integer -3. So $\mathbf{Q}$ is an extension of $\mathbf{Z}$, and hence also of $\mathbf{N}$.

We can *represent* $\mathbf{Q}$ using the same type of diagram that we used for $\mathbf{Z}$, namely a number line with zero in the middle, the Positive Fractions to the right and the negative ones to the left.



There are at least two ways we can go about *constructing* $\mathbf{Q}$ from previously constructed sets of numbers: we can either start with the set of Integers $\mathbf{Z}$ or start with the set of Positive Fractions $\mathbf{Q}^+$.

Method 1

Starting from the Integers $\mathbf{Z}$, we can define a Rational Number (a fraction) f as an ordered pair of Integers (m, n), where $n \neq 0$, which we choose to write as $\frac{m}{n}$. Here m and n can be negative and m can be zero, so it's more general than what we did when we defined Positive Fractions. As with the Positive Fractions, we define equality by saying that for $m, n, p, q \in \mathbf{Z}$ (i.e. $m, n, p$and q are Integers), with n and q non-zero

$$\frac{m}{n} = \frac{p}{q} \text{ if and only if } mq = pn.$$

And we can then go on and define addition and multiplication of Rational Numbers just like we did for the Positive Fractions.

Method 2

The alternative approach is to start with the set of Positive Fractions $\mathbf{Q}^+$ and define a Rational Number as an ordered pair of these (f, g), which we choose to write as $(f \ominus g)$. Here f and g can be any Positive Fractions, so it's more general than what we did when we defined the Integers. As with the Integers, we define equality by saying that for $f, g, h, i \in \mathbf{Q}^+$

$$(f \ominus g) = (h \ominus i) \text{ if and only if } f + i = h + g.$$

And we can then go on and define addition and multiplication of Rational Numbers just like we did for the Integers.

I am not going to go through all the details of methods 1 and 2 because they repeat almost exactly the reasoning used in Chapters 3 and 4 respectively. Whichever method is used to construct the set of Rational Numbers **Q**, we will find that these numbers have all the familiar properties. Since these properties are so important, we now give a complete list where f, g, h are arbitrary numbers in **Q**.

5.2 Summarizing the properties of Q

1. $f + g$ and fg both lie in **Q** (we say that **Q** is closed under addition and multiplication),

2. $f + g = g + f$ and $fg = gf$ (the commutative laws),

3. $(f + g) + h = f + (g + h)$ and $(fg)h = f(gh)$ (the associative laws),

4. $f(g + h) = fg + fh$ (multiplication is distributive over addition),

5. $0 + f = f$ (0 is an additive identity),

6. $1f = f$ (1 is a multiplicative identity),

7. for each $f \in \mathbf{Q}$ there is an additive inverse $(-f) \in \mathbf{Q}$ such that $f + (-f) = 0$,

8. for each $f \in \mathbf{Q}$, except for $f = 0$, there is a multiplicative inverse (reciprocal) $f^{-1} \in \mathbf{Q}$ such that $f(f^{-1}) = 1$.

9. for each $f \in \mathbf{Q}$ precisely one of the following three alternatives holds
   i) $f > 0$, ii) $f = 0$, iii) $0 > f$,
   and in case i) we say f is positive, in case iii) we say f is negative.

10. If $f, g \in \mathbf{Q}$ and $f > 0, g > 0$ (i.e. f and g are positive), then $f + g > 0$ and $fg > 0$.

We can then define $f > g$ to mean $f - g > 0$, define $g < f$ to mean $f > g$, define $f \geq g$ to mean that either $f > g$ or $f = g$, and define $g \leq f$ to mean $f \geq g$.

5.3 The philosophical view

Let us return briefly to the question of whether numbers exist outside of us or are they actually a product of human culture. I will take counting apples as my subject. If there are 3 apples in my bag, there is absolutely no argument in practice about the three-ness of what is in my bag. I would anticipate that a pony would be able to distinguish a bag with 3 apples from one with only 2, or one containing 4 apples. This sort of argument would seem to apply to any set of real physical objects, although a pony might be less interested in a bag of bricks. There have been experiments showing that many animals and birds can count up to around 7. Primates, dolphins, porpoises, crows and parrots seem to be among the most numerate, but even insects such as bees display some numerical abilities. So there is evidence that positive Integers (Natural Numbers) have an objective existence and that they are not our fabrications.

But the situation is rather different for fractions - even positive ones. Can you cut an apple into *exactly* two equal halves? Even if we agree how to measure equality, perhaps by weight, could you guarantee that your knife would leave the two parts *exactly* equal in weight? A pony could probably observe that there were two part-apples, but be totally unconcerned about the question of whether the weights were exactly the same. So a fraction like $\frac{1}{2}$ seems to be more abstract than a Natural Number like 3. Of course if we had 6 apples we could divide them into two halves, each of three apples, so we might agree that $\frac{1}{2}$ of 6 is fairly concretely 3, but we still have a problem with $\frac{1}{2}$ of 5. So already, fractions seem to be less concrete than the Natural Numbers and more open to the claim that they are our inventions, or at least our abstractions.

Negative numbers are in some ways even worse. One might see something approximating to half an apple, but not to -3 apples. Yes, there are other ways of looking at such numbers, for example as debts, or as measurements leftwards as distinct from rightwards, but these already involve other humanly constructed concepts.

Would the little green man of Science Fiction understand fractions and negative numbers? We may never know. But if he understood the Natural Numbers $\mathbf{N}$, we could show him how to *construct* the Integers $\mathbf{Z}$ and the Positive Fractions $\mathbf{Q}^+$, starting from $\mathbf{N}$. And we could go on to show him how to construct the Rational Numbers $\mathbf{Q}$.

5.4 Decimals

How do Rational Numbers fit in with decimals? To answer this question it suffices to look at positive numbers, since a negative number can be considered as a positive one with a minus sign in front of it. So let's look at an easy example.

Suppose that you pay tax at 20%, then your tax is 0.2 times your taxable income, in other words, $\frac{1}{5}$ times your taxable income. So how do we get from 0.2 to $\frac{1}{5}$ and vice-versa? We have to examine what the figures after a decimal point represent. The first figure to the right of the point counts tenths ($\frac{1}{10}$ s), the next figure to the right (if any) counts hundredths ($\frac{1}{100}$ s), and so on. So in our case 0.2 is just $\frac{2}{10}$, which cancels down to $\frac{1}{5}$.

In general, to unravel the meaning of a decimal, remember that digits to the left of the decimal point (moving leftwards from the point) count units (10^0), followed by tens (10^1), hundreds (10^2), thousands (10^3), and so on. Digits to the right of the decimal point (moving rightwards from the point) count tenths (10^{-1}), hundredths (10^{-2}), thousandths (10^{-3}), and so on. As a consequence, if you multiply a decimal by 10, then the units become tens, the tens become hundreds and so on, and the decimal point moves one place to the right. Similarly, multiplying by 100 moves the point two places to the right. Division by 10, by 100, etc. works in the opposite direction, moving the point an appropriate number of places to the left.

As another example, $32.578 = 32+\frac{5}{10}+\frac{7}{100}+\frac{8}{1000}$, and this could be written as $32\frac{578}{1000}$ or as the improper fraction $\frac{32578}{1000}$. Of course you might want to get this fraction into its simplest form by cancelling common factors, but here we are really only interested in the theory. So it looks like converting decimals to fractions is easy, but there is a snag which we will consider after looking at converting fractions to decimals.

Let's convert $\frac{1}{5}$ to a decimal. The first step is to recognize that the fraction represents a division, in this case of 1 by 5. But, as we learnt at school, 5 into 1 won't go. So don't write 1, instead write 1 followed by a string of zeros - how long a string? - as long as you like, and temporarily forget the decimal point. Now we want to divide $1000\ldots$ by 5, and it is easy; we start with 5 into 10, which goes twice with no remainder, so we get $200\ldots$, and all that remains is to work out where to put the decimal point. That last aspect is easy in this case, so we get $0.200\ldots$ and we can discard all the trailing zeros since they make no contribution to the answer. In general, if we set out the division carefully, we can keep track of where to put the decimal point.

So let's try converting 376/50 to a decimal. We will set out the calculation as a long division. This is to illustrate a point about remainders at each stage of the division. No-one in their right mind should use this method as a practical way of doing this particular calculation because it's very easy: $376/50 = (376 \times 2)/(50 \times 2) = 752/100 = 7.52$.

We start by writing 376 as $376.000\ldots$ and ask how many 50s in 376, to which the answer is 7 with a remainder of $376 - 350 = 26$. We then bring down the digit in the next decimal place, in this case 0, from our numerator $376.\underline{\mathbf{0}}00$ to get 26.0 on the bottom line. So the first stage of our division is as shown in Figure 5.1(a). Note how we keep the decimal point aligned.

In the next stage we think of the 26.0 as 260 and ask how many 50s in 260, to which the answer is 5 with a remainder $260-250 = 10$. We then bring down the digit in the next decimal place, in this case 0, from our numerator $376.0\underline{\mathbf{0}}0$ to get 1.00 on the bottom line. This second stage is shown in Figure 5.1(b).

In the next stage we think of the 1.00 as 100 and ask how many 50s in 100, to which the answer is 2 with a remainder 0. This third stage is shown in Figure 5.1(c). As there is a zero remainder and no more non-zero digits to bring down, we are finished and we can conclude that $376/50 = 7.52$.

$$\begin{array}{rl} & 7. \\ 50 & \overline{)376.000} \\ & \underline{350} \\ & 26.0 \end{array}$$

(a) first stage

$$\begin{array}{rl} & 7.5 \\ 50 & \overline{)376.000} \\ & \underline{350} \\ & 26.0 \\ & \underline{25.0} \\ & 1.00 \end{array}$$

(b) second stage

$$\begin{array}{rl} & 7.52 \\ 50 & \overline{)376.000} \\ & \underline{350} \\ & 26.0 \\ & \underline{25.0} \\ & 1.00 \\ & \underline{1.00} \\ & 0.00 \end{array}$$

(c) third stage

Figure 5.1: Three stages in a division.

This all looks relatively straightforward, even if rather tedious and one might anticipate making errors with the arithmetic or locating the decimal point. It seems tempting to use a calculator and indeed this is generally preferable in practice. But the problem comes when the denominator has a factor other than powers of 2 and 5 (i.e. 2, 4, 8, etc. and 5, 25, 125, etc.). This is because the only prime factors of 10 are 2 and 5. How I wish that we had 12 fingers and counted in powers of 12!

The simplest case where the problem arises is probably converting 1/3 to a decimal. The first three stages are shown in Figure 5.2.

If you follow this through, the problem becomes clear: you never get a remainder 0. In fact the remainder (forgetting the decimal point) is 1 at each

$$\begin{array}{r|l} & 0.3 \\ \hline 3 & 1.0000 \\ & \underline{0.9} \\ & 0.10 \end{array}$$

(a) first stage

$$\begin{array}{r|l} & 0.33 \\ \hline 3 & 1.0000 \\ & \underline{0.9} \\ & 0.10 \\ & \underline{0.09} \\ & 0.010 \end{array}$$

(b) second stage

$$\begin{array}{r|l} & 0.333 \\ \hline 3 & 1.0000 \\ & \underline{0.9} \\ & 0.10 \\ & \underline{0.09} \\ & 0.010 \\ & \underline{0.009} \\ & 0.0010 \end{array}$$

(c) third stage

Figure 5.2: Three stages in converting 1/3 to a decimal.

stage. So we could go on forever: $\frac{1}{3} = 0.33333333333333333333333333333\ldots$.

The cause of the problem is the fact that it is *impossible* to represent $\frac{1}{3}$ as a finite sum of tenths, hundredths, thousandths, etc. We are forced to use an *infinite* sum:

$$\frac{1}{3} = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \ldots.$$

This should be a bit disturbing: a very easy fraction producing a concept like an infinite sum. We'll have more to say about this issue in the next chapter. But before that we just need to look a bit more closely at the decimals that result from fractions (and already there is a bit of a hint in my wording that not all decimals come from fractions).

Get out your calculator and check these:

$\frac{1}{3} = 0.3333\ldots$ (the 3 recurs).
$\frac{2}{3} = 0.666\ldots$ (the 6 recurs).
$\frac{1}{7} = 0.142857142857\ldots$ (the sequence 142857 recurs).
$\frac{2}{7} = 0.285714285714\ldots$ (the sequence 285714 recurs).
$\frac{4}{13} = 0.307692307692\ldots$ (the sequence 307692 recurs).

Perhaps the first issue is how to represent a recurring decimal while specifying exactly the recurring bit. This is done by placing a bar over the recurring section, like so

$$\frac{1}{3} = 0.\overline{3}, \quad \frac{2}{3} = 0.\overline{6}, \quad \frac{1}{7} = 0.\overline{142857}, \quad \frac{2}{7} = 0.\overline{285714}, \quad \frac{4}{13} = 0.\overline{307692}.$$

We can predict how long the recurring section will be by looking at the denominator. To understand this, consider the division of 2 by 7 as shown in detail in Figure 5.3, and look particularly at the remainders at each stage. These are shown separately (without regard to the decimal point).

$$
\begin{array}{r|l}
\multicolumn{1}{r}{} & 0.285714285714\ldots \\ \cline{2-2}
7 & 2.000000000000 \\
\multicolumn{1}{r}{} & \underline{1.4} \\
\multicolumn{1}{r}{} & 0.60 \quad \text{(remainder 6)} \\
\multicolumn{1}{r}{} & \underline{0.56} \\
\multicolumn{1}{r}{} & 0.040 \quad \text{(remainder 4)} \\
\multicolumn{1}{r}{} & \underline{0.035} \\
\multicolumn{1}{r}{} & 0.0050 \quad \text{(remainder 5)} \\
\multicolumn{1}{r}{} & \underline{0.0049} \\
\multicolumn{1}{r}{} & 0.00010 \quad \text{(remainder 1)} \\
\multicolumn{1}{r}{} & \underline{0.00007} \\
\multicolumn{1}{r}{} & 0.000030 \quad \text{(remainder 3)} \\
\multicolumn{1}{r}{} & \underline{0.000028} \\
\multicolumn{1}{r}{} & 0.0000020 \quad \text{(remainder 2)}
\end{array}
$$

Figure 5.3: Converting 2/7 to a decimal.

The last remainder shown is 2, and consequently the next division to be performed is 7 into 20, which is the same division that we started with, so the pattern will now recur. How many possible different remainders are there when dividing by 7? Clearly you can't have a remainder less than 0 or more than 6, so there are at most 7 possible remainders when dividing by 7. When you reach a division that you have encountered previously, the pattern will recur. So, for division by 7, the length of the recurring section is at most 7 (it's actually 6, but let's not worry too much about that).

There's nothing terribly special about $\frac{2}{7}$. In general, the number of possible remainders when dividing by a positive Integer n is at most n. So in a fraction like $\frac{m}{n}$, the length of the recurring section of the corresponding decimal is at most n (it's often less). We can even regard a terminating decimal like $\frac{1}{2} = 0.5$ as a recurring decimal $0.5\overline{0}$. Looked at like this, all Rational Numbers can be viewed as recurring decimals.

That just leaves us to clear up the snag we mentioned concerning conversion of decimals to fractions. One example should suffice, so suppose we are asked to convert $3.56\overline{234}$ to a fraction. It helps if we give this number a name, say x, and write out the repeating section a few times like this: $x = 3.56234234\ldots$. Multiplying a decimal by 10 moves the decimal point one place to the right, similarly multiplying by $100 = 10^2$ moves it two places and multiplying by $1000 = 10^3$ moves it three places. Here the repeating section is of length 3, so we will move the decimal point 3 places by multiplying by 1000. We now have $1000x = 3562.34234234\ldots$ as well as x itself, and we write one above the other, lining up the decimal points:

$$\begin{aligned} 1000x &= 3562.34234234\ldots \\ x &= \quad\;\; 3.56234234\ldots \end{aligned}$$

Then subtract the second line from the first line, observing that the recurring part $234234\ldots$ cancels out, to get

$$999x = 3562.34 - 3.56 = 3558.78.$$

So our decimal $x = 3558.78/999 = 355878/99900$. Yes, it will cancel down (to $19771/5550$) but we are only interested in the theory here so we can write $3.56\overline{234} = 355878/99900$.

The important conclusion to draw from the foregoing discussion is that Rational Numbers correspond to recurring (including terminating) decimals. The disturbing thing is that it is very easy to write down a decimal that neither recurs nor terminates, you just need a pattern. It is very easy to specify such numbers. You can try it for yourself. To get you started, here is an example of a decimal with digits 0 and 1 but with a growing number of 0s between each successive 1:

$$x = 0.101001000100001000001\ldots$$

Is this a number, and if so, what kind? We consider this and some related questions in the next chapter.

Chapter 6

The Real Numbers (R)

6.1 Recurring decimals and infinite series

We saw in the previous chapter that the Rational Numbers correspond to recurring (including terminating) decimals, but that it is easy to specify decimals that do not terminate or recur. The set of all possible decimals forms a set of numbers known as the Real Numbers and denoted by **R**. Unfortunately, although this is true, it leaves a lot to be explained. Do we get a different set of numbers if we use a different number base such as 2 instead of 10 and consider the set of all "binamals"? (The answer is no.) More seriously, a typical Real Number has an infinitely long decimal expansion. How then do we add two such decimals, since addition normally starts on the right hand side with the least significant digits in order to deal with carry digits? Indeed, what is the meaning of an infinitely long decimal expansion? We have seen that it corresponds to a sum of infinitely many terms. How can you add infinitely many terms together? Can we construct the Real Numbers from the Rational Numbers?

We will start with something to which we know the answer by considering the Rational Number $\frac{1}{3} = 0.\overline{3}$. This recurring decimal corresponds to the infinite sum

$$\frac{1}{3} = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \dots.$$

You clearly cannot physically add up an infinite number of terms, so what does this mean? To discuss this we need a bit of terminology.

In everyday English the two words "sequence" and "series" are used pretty well interchangeably. For example, one might say that someone had suffered "a sequence of illnesses" or "a series of illnesses", and these both convey exactly the same meaning. But in mathematics, the two words are used for different purposes. A *sequence* is an ordered list of items, often a list of numbers. The items in the list are called the *terms* of the sequence. A *series* is the result of adding together the terms of a sequence. So here is a sequence

$$\left(\frac{3}{10}, \ \frac{3}{100}, \ \frac{3}{1000}, \ \ldots\right)$$

and here is the resulting series

$$\frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \ldots.$$

It is perhaps unfortunate that the word "series" has been used in this distinct technical sense in mathematics, perhaps "sum" would have been clearer, but we are now stuck with it.

The first term in our sequence is $\frac{3}{10}$, the second is $\frac{3}{10^2}$, the third is $\frac{3}{10^3}$, and so on, with the n^{th} term being $\frac{3}{10^n}$. From the corresponding series we can form the sequence of *partial sums*. The first partial sum is $\frac{3}{10} = 0.3$, the second is $\frac{3}{10} + \frac{3}{10^2} = 0.33$, the third is $\frac{3}{10} + \frac{3}{10^2} + \frac{3}{10^3} = 0.333$, and so on, with the n^{th} partial sum being $0.333\ldots3$ (where there are n 3s). So the sequence of partial sums is

$$0.3, \quad 0.33, \quad 0.333, \quad \ldots.$$

Looking at this sequence of partial sums it is clear that the successive terms get closer and closer to the original number $\frac{1}{3}$. We can make this more precise by observing that however close we would like to get to $\frac{1}{3}$, if we go sufficiently far along the sequence of partial sums, we will get that close and indeed stay that close. For example, if we want to get within $\frac{1}{10}$ of $\frac{1}{3}$ (i.e. between $\frac{1}{3} - \frac{1}{10}$ and $\frac{1}{3} + \frac{1}{10}$) we will be within that envelope from the first partial sum onwards. If we want to get within $\frac{1}{100}$ of $\frac{1}{3}$, we will be within that envelope from the second partial sum onwards. More generally, if we want to get within $\frac{1}{10^n}$ of $\frac{1}{3}$, we will achieve this from the n^{th} partial sum onwards. Figure 6.1 illustrates what is happening with s_n denoting the n^{th} partial sum, so the leftmost dot represents $s_1 = 0.3$, the next dot to the right represents $s_2 = 0.33$, and so on. Of course the picture cannot show tiny quantities, so from $n = 4$ onwards it looks like all the dots (representing s_4, s_5, s_6, etc.) lie on the line representing $\frac{1}{3}$, but of course they are all below it, albeit ever so slightly.

Mathematicians would say that these partial sums *converge* or *tend* to $\frac{1}{3}$. This is expressed in symbols as

$$s_n \to \frac{1}{3} \quad \text{as} \quad n \to \infty.$$

Here again s_n denotes the n^{th} partial sum and the whole sentence is read as "s_n converges to $\frac{1}{3}$ as n tends to infinity". It is important to understand

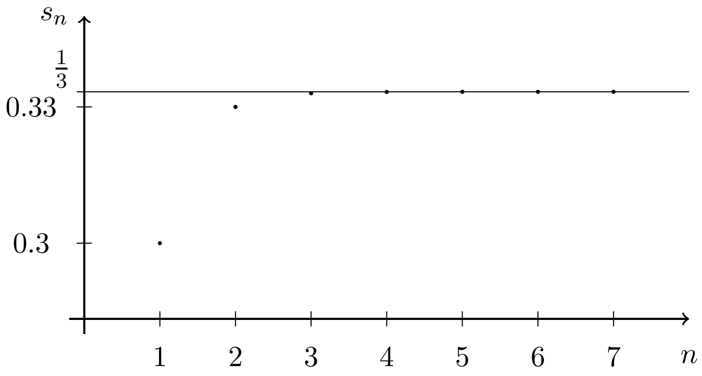


Figure 6.1: Partial sums approaching 1/3.

there is no sense here in which "infinity" (∞) is being treated as a normal number; it might be better if the word had never been used. All we mean here is that n is getting arbitrarily large, and as it does so, s_n is getting arbitrarily close to $\frac{1}{3}$. The number $\frac{1}{3}$ is said to be the *limit* of the sequence of partial sums.

As an aside, I mention that the precise mathematical definition of convergence is omitted here because

1. It is very complicated and took hundreds of years of very careful formulation by very bright mathematicians.

2. Even very bright mathematicians find it hard going at first sight.

3. I think I can get away without it here.

We will more to say about this in Chapter 11 where we will give the precise definition of convergence while discussing how mathematicians have come to grips with the slippery concepts of the infinitely large and the infinitesimally small.

Any recurring decimal can be treated in the same way as $\frac{1}{3} = 0.\overline{3}$, by forming the sequence of partial sums. So when you see a statement like $\frac{2}{7} = 0.\overline{285714}$, the equals sign (=) is subtly different from the same sign in $\frac{4}{2} = 2$. In the latter case there are two numbers, $\frac{4}{2}$ and 2, that we are asserting are equal. But in the former case we are saying something more complicated, namely that if we take the sequence of partial sums s_n from the recurring decimal $0.\overline{285714}$, then this sequence converges to $\frac{2}{7}$. These partial sums are just the decimals formed by truncating the recurring decimal $0.\overline{285714}$, so the sequence in this case is

$$0.2,\ 0.28,\ 0.285,\ 0.2857,\ 0.28571,\ 0.285714,\ 0.2857142,\ \ldots.$$

The equal sign in $\frac{2}{7} = 0.\overline{285714}$ is an assertion that these decimals formed by truncating the recurring decimal get arbitrarily close to $\frac{2}{7}$ as more decimal places are recorded. Another way of putting this is to say that $\frac{2}{7}$ is the limit of the sequence formed by truncating the decimal $0.\overline{285714}$.

Infinite series provide a nice way of resolving certain ancient paradoxes. One of the best known is Zeno's paradox of Achilles and the tortoise. They have a race and since Achilles is faster than the tortoise, he gives the tortoise a head start. When Achilles reaches the starting point of the tortoise, the tortoise has moved on. So Achilles then runs to where the tortoise now is. But once again the tortoise has moved on, and this sequence continues *ad infinitum*. So how does Achilles overtake the tortoise? Let's put some numbers into this to see what happens.

Suppose Achilles runs at 100 metres per minute (well, he is 2500 years old), and the tortoise is unusually agile and runs at 10 metres per minute but has a head start of 100 metres. Start the clock. After 1 minute Achilles has run 100m and the tortoise has run 10m, so the tortoise is 10m ahead of Achilles. After a further $\frac{1}{10}$ minute, the clock reads 1.1 minutes, Achilles has run a total of 110m, and the tortoise has run 11m, so the tortoise is 1m ahead of Achilles. Now move on another $\frac{1}{100}$ minute when the clock reads 1.11 minutes, Achilles has run 111m, and the tortoise has run 11.1m, so the tortoise is 0.1m ahead of Achilles. It is not hard to see that the distance between Achilles and the tortoise is rapidly decreasing and will vanish when the clock reads $1.111\ldots = 1.\overline{1}$ minutes, which represents the fraction $\frac{10}{9}$, that is one-and-a-ninth minutes.

The answer to the paradox is that the *infinite* series $1 + \frac{1}{10} + \frac{1}{100} + \ldots$ converges to a *finite* sum $\frac{10}{9}$.

Now here is an associated issue that has been known to upset people. Consider the recurring decimal $0.\overline{9}$. Let's call it x and apply the procedure that we saw in the previous chapter for converting a recurring decimal to a Rational Number. Since the recurring section is one digit long, we multiply x by 10 and then write down $10x$ with x underneath it, lining up the decimal points.

$$\begin{aligned} 10x &= 9.999999\ldots \\ x &= 0.999999\ldots \end{aligned}$$

Then subtract the second line from the first line, observing that the recurring part .999999... cancels out. We find

$$9x = 9.0$$

and, by dividing both sides by 9, we get $x = 1$. In other words, $0.999999\ldots = 1$. Another way to obtain this is to take the decimal representation of $\frac{1}{3}$, namely $0.\overline{3}$, and multiply it by 3 to get $1 = 0.\overline{9}$. But a lot of people feel that $0.999999\ldots$ is somehow less than 1. Of course all the partial sums are less than 1. The answer to the conundrum is that the "=" sign is really saying that the partial sums converge to the limit 1.

6.2 Irrational numbers

Now what about a non-recurring decimal like $x = 0.101001000100001\ldots$? Its partial sums (formed by truncating the decimal) look like they are converging to *something*, but not to any Rational Number because we know that Rationals correspond to recurring or terminating decimals (see p.46). So there is *nothing* for it to converge to. Just where we expect to find the limit of the sequence of partial sums there is a hole in **Q**! We'll come on to how to fill this hole and all the other similar holes in a moment. We must *invent* or *discover* (depending on your point of view) some new numbers. These new numbers are called *irrational* numbers because they cannot be expressed as ratios of Integers. Because of these holes, the Rational Numbers are said to be *incomplete*. The collection of all Rational and Irrational Numbers forms the set of Real Numbers **R**. It may be worth commenting that the term "irrational number" has unfortunate connotations because of the use of the word "irrational" in everyday English as a near synonym with "crazy" or "mad". There is nothing crazy or mad about these numbers, they are beautiful and well-behaved.

Before looking at filling the holes, we are going to prove that there are easier irrational numbers than our example $x = 0.101001000100001\ldots$. Greek mathematicians around 2500 years ago realized that there was a problem, and they did not use decimals, so how did they come to the realization?

We will try to explain this by looking at equations. For example, the equation $x^2 = 4$ has solutions in **Q** (two in fact, $x = 2$ and $x = -2$). But the similar equation $x^2 = 2$ has no solution in **Q**. Now, you may be quite happy writing $x = \sqrt{2}$ as a solution of this equation (and of course you should say that there are two solutions $x = \sqrt{2}$ and $x = -\sqrt{2}$). I'm not saying this is wrong - I'm simply saying that $\sqrt{2}$ does not lie in **Q**. So I'm saying $\sqrt{2}$ is irrational, i.e. $\sqrt{2} \notin \mathbf{Q}$. In terms of the previous discussion, the decimal expansion of $\sqrt{2}$ does not recur. The irrationality of $\sqrt{2}$ was known to the Ancient Greeks. It is referred to by Aristotle and may have been discovered by Pythagoras or one of his followers around 400 BCE.

Proof of the irrationality of $\sqrt{2}$.

We need a little bit of terminology. An *even* number is an Integer exactly divisible by 2, like $2, 4, 6, 8, 0, -10$, etc. An *odd* number is an Integer not exactly divisible by 2, like $1, 3, 5, 7, -1, -9$, etc. Clearly every Integer is either even or odd, but not both.

Each even number has the form $2x$ for some Integer x. For example, $14 = 2 \times 7$ and $-6 = 2 \times (-3)$.

Each odd number has the form $2y + 1$ for some Integer y. For example, $15 = 2 \times 7 + 1$ and $-11 = 2 \times (-6) + 1$.

If you square an even number $2x$ you get $4x^2$, which is $2 \times (2x^2)$, and so is itself even. If you square an odd number $2y+1$ you get $(2y+1)(2y+1) = 4y^2 + 4y + 1 = 2 \times (2y^2 + 2y) + 1$, and so is itself odd.

Now suppose that $\sqrt{2}$ is a Rational Number. We show that this supposition leads to a contradiction (something ridiculous) and must therefore be false.

Since $\sqrt{2} > 0$, if it is a Rational Number it must be the ratio of two *positive* Integers, i.e. two Natural Numbers, say p and q.

If $\sqrt{2} = \frac{p}{q}$ with $p, q \in \mathbf{N}$, then by cancelling any common factors we can ensure that p, q have no common factors, in particular they are **not both even**.

Squaring $\sqrt{2} = \frac{p}{q}$ gives

$$2 = \frac{p^2}{q^2},$$

so $p^2 = 2q^2$. Hence p^2 is an even Integer. Consequently p cannot be odd and so p **must be even**.

But then we can write p as twice some other Integer, say $p = 2r$ (where r is an Integer). Squaring this gives $p^2 = 4r^2$. However, we already have $p^2 = 2q^2$ and therefore $2q^2 = 4r^2$. This gives $q^2 = 2r^2$. Thus q^2 is even. Consequently q cannot be odd and so q **must be even**.

But now we have deduced that **both** p and q are even, contradicting the fact that they are not both even.

From this we deduce that $\sqrt{2}$ cannot be written as the ratio of any two Natural Numbers p and q, and consequently $\sqrt{2} \notin \mathbf{Q}$, i.e. **the square root of 2 is irrational**. $\square$

You can use the same sort of argument to prove that $\sqrt{3}$ and lots of similar roots are irrational. So it looks like there are lots of holes in **Q** corresponding to Irrational Numbers. Indeed, we will see later on that we can prove (properly) that there are a lot more Irrational Numbers than Rational Numbers (even though there are infinitely many of both). But meanwhile here is an informal demonstration of this.

Think of a sequence of digits chosen at random from the digits 0 to 9. The digits chosen can be used to write down a decimal between 0 and 1. For example you might pick the sequence $3, 7, 0, 1, 2, 9, 5, 1, 7, 2, \ldots$. This will form the decimal $0.3701295172\ldots$. If the digits really are chosen at random and the sequence goes on indefinitely, what's the chance of the decimal recurring? Surely it's *very* unlikely. Well, this isn't a proper mathematical proof (it leaves open what is meant by "random" and how to choose an infinitely long sequence) but it is fairly compelling evidence that *most decimals don't represent Rational Numbers*.

Returning to $\sqrt{2}$, if you evaluate this on a calculator you will get an approximation to maybe 10 decimal places: 1.4142135624. This may have been rounded up or down by the calculator; if it has been rounded up, then it isn't a partial sum of the corresponding series. This example has been rounded up, so the 11^{th} partial sum in this particular instance is 1.4142135623 (it differs in the last digit from the rounded version on the calculator). With that proviso, the decimal expansion given on the calculator[1] provides a sequence of partial sums

$$1,\ 1.4,\ 1.41,\ 1.414,\ 1.4142,\ 1.41421,\ \ldots$$

that converges to $\sqrt{2}$ in the set of Real Numbers **R** (but not in the set of Rational Numbers **Q**).

If we take *any* decimal expansion and write down the sequence of truncated decimals that it produces, then this sequence is composed entirely of Rational Numbers (terminating decimals in fact) and it converges in the set of Real Numbers **R**. But rather more is true. It is possible to prove that each sequence of numbers in **R** (not just in **Q**) whose terms get arbitrarily close together as you move along the sequence must converge to some value $x \in \mathbf{R}$. This property is expressed by saying that the set of Real Numbers is *complete*. In plain English, you can't find a hole in **R**.

Many useful numbers in Mathematics turn out to be irrational. The number π, denoting the ratio of the circumference of a circle to its diameter,

[1]Of course you'd need a calculator with an infinitely long display to get the whole sequence.

is irrational, although it isn't so easy to prove this. Another important irrational number is e, sometimes known as Euler's number, which is defined by the infinite series

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots.$$

Here $1! = 1,\ 2! = 1 \times 2,\ 3! = 1 \times 2 \times 3$, etc. For a Natural Number n, the product $n! = 1 \times 2 \times 3 \times \ldots \times n$ is known as n-factorial . Basically the reason why e is irrational is that the terms in the series have the form $\frac{1}{n!}$ and these get smaller so rapidly as n gets larger that the decimal expansion of e cannot recur. If you form the partial sums of the series defining e you will soon convince yourself that e is approximately 2.72, correct to 2 decimal places. The value of e correct to 5 decimal places is 2.71828.

Before proving that e is irrational, I will just remark that the series given for e may be generalized to define a mathematical function called exp, which is generally known as the exponential function. This is given by the series

$$\exp(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots.$$

This series converges for every Real Number x, and putting $x = 1$ we get $\exp(1) = e$. The truly remarkable property connecting e and exp is that $\exp(x)$ turns out to be e^x, that is to say $\exp(x)$ turns out to be the x^{th} power of e. For example, $\exp(2) = e^2$, and $\exp(\frac{1}{2}) = e^{\frac{1}{2}} = \sqrt{e}$.

Proof of the irrationality of e

We just need one fact before we start and that is the sum of the series $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots$. The partial sums of this series form the sequence $(\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \ldots)$, and it is reasonably obvious that this converges to 1. If you really want to convince yourself of this, take a 1 metre ruler, mark off $\frac{1}{2}$ metre from the left, then a further $\frac{1}{4}$ metre, followed by a further $\frac{1}{8}$ metre , and so on. You will rapidly run out of ruler. So $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$.

We will prove that e is irrational by assuming that it is a Rational Number and showing that this assumption leads to a contradiction, and so must be false. So suppose that $e = \frac{p}{q}$ with $p, q \in \mathbf{N}$. It is fairly obvious that q cannot be 1, 2, 3 or 4 because $2.67 < e < 2.75$. Anyway, we have

$$e = \frac{p}{q} = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{q!} + \frac{1}{(q+1)!} + \frac{1}{(q+2)!} + \frac{1}{(q+3)!} + \dots.$$

If we multiply by $q!$, we obtain

$$\frac{p}{q} \times q! = \Big[q! + \frac{q!}{1!} + \frac{q!}{2!} + \ldots + \frac{q!}{q!}\Big] + \Big[\frac{q!}{(q+1)!} + \frac{q!}{(q+2)!} + \frac{q!}{(q+3)!} + \ldots\Big].$$

The left-hand side $\frac{p}{q} \times q!$ is an Integer because the q in the denominator cancels with the q in the $q!$. The first square bracketed term on the right-hand side is also an Integer because in each of the terms (such as $\frac{q!}{2!}$), the denominator is a factor of the numerator. It therefore follows that the second square bracketed term on the right hand side must also be an Integer. Let us call this remaining term R, i.e.

$$R = \frac{q!}{(q+1)!} + \frac{q!}{(q+2)!} + \frac{q!}{(q+3)!} + \dots.$$

Clearly R is positive so $R > 0$ and since it must be an Integer, we have $R \geq 1$. We will now prove that this isn't the case.

To do this, observe that $\frac{q!}{(q+1)!} = \frac{1}{(q+1)}$ because the only number in the product $(q+1)!$ that is not also in the product $q!$ is the number $(q+1)$. In a similar way we find that $\frac{q!}{(q+2)!} = \frac{1}{(q+1)(q+2)}$, $\frac{q!}{(q+3)!} = \frac{1}{(q+1)(q+2)(q+3)}$, and so on. Hence

$$R = \frac{1}{(q+1)} + \frac{1}{(q+1)(q+2)} + \frac{1}{(q+1)(q+2)(q+3)} + \dots.$$

But $q \in \mathbf{N}$, so q is a positive Integer. Consequently

$$q+1 \geq 2, \quad (q+1)(q+2) > 2^2 = 4, \quad (q+1)(q+2)(q+3) > 2^3 = 8, \quad \text{etc.}$$

Therefore

$$R < \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1.$$

So we now have our contradiction. On the assumption that e is rational we found that $R \geq 1$ and that $R < 1$. Consequently the assumption must be incorrect and so e must be irrational. $\square$

We have already identified **R** as the set of all decimal expansions, but we have also gone to some trouble to see why this "definition" is not very satisfactory as it leads to things like infinite series that are difficult to define and explain. It is possible to give a complete characterization of **R** in terms of its properties. We do this below - you have already met all but one of the properties (called axioms) when we considered the Rational Numbers **Q**. But the collection is so important that we are going to state them again. The additional axiom is, as you might expect, called the completeness axiom.

6.3 Axioms for R

Axioms 1 to 8 are called the field axioms. Any mathematical structure satisfying axioms 1 to 8 is called a *field*, and there are many examples of fields other than **Q** and **R**. Axioms 9 and 10 are called the order axioms, and axiom 11 (the most complicated) is the completeness axiom. All the common results about Real Numbers can be deduced from axioms 1 to 11. In the list of axioms, x, y, z denote arbitrary Real Numbers.

1. $x + y$ and xy both lie in **R** (we say that **R** is closed under addition and multiplication).
2. $x + y = y + x$ and $xy = yx$ (the commutative laws).
3. $(x + y) + z = x + (y + z)$ and $(xy)z = x(yz)$ (the associative laws).
4. $x(y + z) = xy + xz$ (multiplication is distributive over addition).
5. $0 + x = x$ (0 is an additive identity).
6. $1x = x$ (1 is a multiplicative identity).
7. For each $x \in \mathbf{R}$ there is an additive inverse $(-x) \in \mathbf{R}$ such that $x + (-x) = 0$.
8. For each $x \in \mathbf{R}$, except for $x = 0$, there is a multiplicative inverse (reciprocal) $x^{-1} \in \mathbf{R}$ such that $x(x^{-1}) = 1$.
9. For each $x \in \mathbf{R}$ precisely one of the following three alternatives holds
   i) $x > 0$, ii) $x = 0$, iii) $0 > x$,
   and in case i) we say x is positive, in case iii) we say x is negative.
10. If $x, y \in \mathbf{R}$ and $x > 0, y > 0$ (i.e. x and y are positive), then $x + y > 0$ and $xy > 0$. As we did for **Q**, we define $x > y$ to mean $x - y > 0$; $y < x$ to mean $x > y$; $x \geq y$ to mean that either $x > y$ or $x = y$; and $y \leq x$ to mean $x \geq y$.
11. (**Completeness**) If $x_1, x_2, x_3, \ldots$ is an infinite sequence of elements of **R** whose terms get arbitrarily closer together, then the sequence converges to some value $x \in \mathbf{R}$. [The precise meanings of "arbitrarily closer together" and "converges" need careful definition which we omit here.]

Straightaway, here is a major problem. Merely listing a collection of properties does not guarantee the *existence* of a set of numbers satisfying them (after all, they might be contradictory). This will not be a problem here so long as it can be shown that the set of all decimals obeys these axioms, because then the set of all decimals provides a *model* for the axioms.

The good news is that this *can* be proved. Amazingly, it can also be shown that *any set* obeying these axioms can be put into one-to-one correspondence with the set of all decimals. Thus the Real Numbers are unique (apart from their names). There is nothing else quite like them!

Many people will have heard of mathematical or statistical models, but in this book we are using the term "model" with a different meaning. The mathematical models referred to in newspapers, and in most mathematics books and publications, generally give a set of equations by use of which it is possible to make assertions about physical systems. For example we might use equations to predict the behaviour of an aeroplane as it flies at various speeds with various loads in various weather conditions. This will enable designers to make the real aircraft safe. Earlier in this book we described exponential growth of an infectious disease using the equation $N = 2^n$, where n was the number of weeks since the start, and N is the number of people infected during week n. Another example of this sort of model is provided by weather forecasting. Such models are never perfect but, if used carefully, they are extremely helpful to eliminate unsafe conditions and to enable us to prepare for future events. But this is not what we mean by the term "model" here in this chapter and later in Chapters 9 and 11. Here we mean a *logical* model, a precise example that obeys a set of sentences and thereby ensures that the sentences cannot be contradictory. It's the sort of thing you might demand if someone told you they had constructed a time machine to enable us to travel at will through time - a polite response might be "go on, show me one". So for the set of axioms given above, "show me one" could be answered by saying "decimals".

There is a (possibly apocryphal) story about a mathematics PhD candidate who produced a thesis describing a certain algebraic structure and establishing some of its properties. But he neglected to provide a model for his structure and his examiner found a proof that the structure was self-contradictory, so it couldn't exist: end of PhD. Whether or not the story is true, logical models are important.

6.4 An alternative version of the Completeness Axiom

The version of the completeness axiom (axiom 11) given on page 58 looks rather complicated with its mention of "arbitrary closeness" and "convergence". An alternative version of this axiom uses the idea of an upper bound for a set of numbers. A set of numbers, X, is said to be *bounded above* if there is some number m that is greater than or equal to all the numbers in X: in symbols, $x \leq m$ for every $x \in X$. In such a case m is said to be *an upper bound* for X. Note that if m is an upper bound for X, then any larger number M ($M > m$) will also be an upper bound for X.

As an example, consider the set S of all Rational Numbers lying between 0 and 1:

$$S = \{x \in \mathbf{Q} \text{ satisfying } 0 < x < 1\}.$$

We can see that 3 is an upper bound for this set because if $x \in S$ then $x \leq 3$. Of course 2 is also an upper bound for S, as is 1. Clearly no number less than 1 will be an upper bound for the set because if $y < 1$ then $x = (y+1)/2$ lies between y and 1, so x is an element of S that exceeds y, hence y cannot be an upper bound for S. So 1 is the *least* upper bound of S. The least upper bound of a set X is called the *supremum* of X and is denoted by $\sup X$. So in our example $\sup S = 1$. Note that in this example $\sup S \notin S$ because $1 \notin S$. However, if we modify the set S very slightly to form the set T

$$T = \{x \in \mathbf{Q} \text{ satisfying } 0 < x \leq 1\},$$

then $\sup T = 1$ and in this modified example $\sup T \in T$. The moral is that the supremum of a set X which is bounded above may or may not lie in X. If the supremum does lie in the set, then it must be the maximum element in the set. If the set X has a maximum element, that will be the supremum. Of course many sets are not bounded above and such sets do not have a supremum, for example the set $E = \{2, 4, 6, \ldots\}$ of all even Natural Numbers is not bounded above.

To see what this has to do with the completeness axiom, we will look at an example and suppose that we only recognize Rational Numbers as numbers. Consider the set

$$X = \{x \in \mathbf{Q} \text{ satisfying } x^2 < 2\},$$

so that X consists of all the Rational Numbers whose squares are less than 2. Thus (for example) $1.4 \in X$ because $1.4^2 = 1.96 < 2$. However $2 \notin X$ because $2^2 = 4 > 2$. Indeed if $y > 2$ then $y \notin X$ because $y^2 > 4 > 2$, and consequently 2 is an upper bound for X. However there are smaller upper bounds for X, such as 1.5 because if $y > 1.5$ then $y^2 > 2.25 > 2$, so $y \notin X$. Similarly 1.42 is an upper bound for X, as is 1.415, as is 1.4143, etc. What do I mean by "etc"? You'll get the idea if you realize that $\sqrt{2}$ (which lies in **R**, but not in **Q**) is approximately 1.41421356237 (correct to 11 decimal places).

So what is the *least* upper bound (the supremum) for the set X? It's "obviously" $\sqrt{2}$ but only if we recognize $\sqrt{2}$ as a number. So if we only recognize Rational Numbers **Q**, the set X, which is bounded above, has no supremum because the only candidate is $\sqrt{2}$ and $\sqrt{2} \notin \mathbf{Q}$. In a way this is telling us again that **Q** is incomplete - it is full of "holes". If we change our minds and agree to recognize Real Numbers **R** as *the* set of numbers, then life becomes easier. The set X is (still) bounded above, but now it does have a supremum, namely the respectable Real Number $\sqrt{2}$. Speaking more generally, all the "holes" in **Q** are filled in **R** with irrational numbers. This leads to the following alternative formulation of the completeness axiom for the Real Numbers **R**.

Axiom 11'. If X is a set of Real Numbers that is bounded above, then X has a supremum (a least upper bound).

The supremum of a set X that is bounded above may or may not lie in X. If it does lie in X then it is the maximum of all the numbers in X.

It is not entirely obvious that axiom 11 and axiom 11' are indeed equivalent, and we will not *prove* that they are. However the equivalence is plausible because they both resolve the same problem characterized by our example of $\sqrt{2}$. Both axioms ensure that any irrational number (such as $\sqrt{2}$) that is missing from the set of Rational Numbers **Q** must lie in the set of Real Numbers **R**.

Some further insight into the completeness axiom may be gained by considering an infinite sequence (x_n) of Real Numbers that is bounded above. In other words we have Real Numbers $x_1, x_2, x_3, \ldots$ satisfying $x_1 \leq x_2 \leq x_3 \leq \ldots$ and for which there is an upper bound m such that $x_n \leq m$ for $n = 1, 2, 3, \ldots$. It is fairly obvious, and not that difficult to prove, that such a sequence must converge to its supremum $s = \sup(x_n)$. Figure 6.2 illustrates this result with the dots representing the values of x_n, increasing in height as n gets larger, bounded above by m and with supremum s. This convergence property is also equivalent to the completeness axiom.

What we have described for upper bounds can equally well (and equivalently) be done for lower bounds. We just mention the basic definitions. A set of numbers, X, is said to be *bounded below* if there is some number ℓ that is less than or equal to all the numbers in X: in symbols, $\ell \leq x$ for every $x \in X$. In such a case ℓ is said to be *a lower bound* for X. Note that if ℓ is a lower bound for X, then any smaller number L ($L \leq \ell$) will also be a lower bound for X. A consequence of the completeness axiom, and equivalent to it, is the result that any set X of Real Numbers that is bounded below will have a *greatest lower bound.* This is called the *infimum* of X and is denoted by $\inf X$.

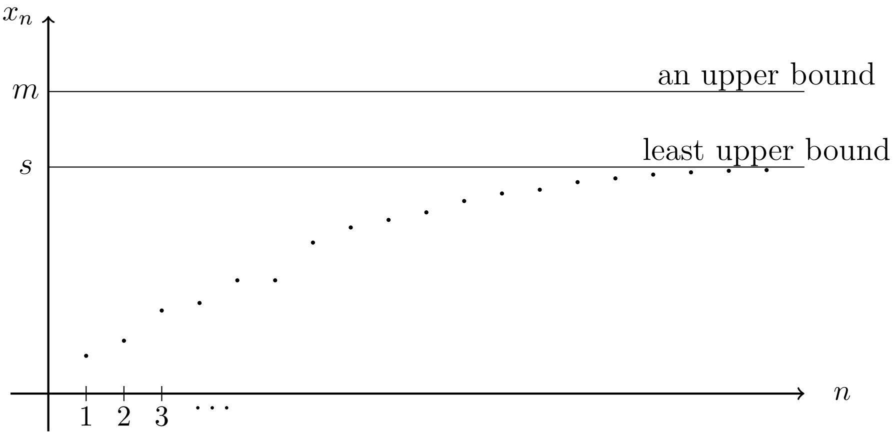


Figure 6.2: $x_n \to s$ as $n \to \infty$.

6.5 Construction of the Real Numbers

Because of the difficulty in working with infinitely long decimals, mathematicians prefer alternative approaches to constructing the Real Numbers **R** from the Rational Numbers **Q**. One approach is by considering sequences in **Q** whose terms get "arbitrarily closer together" (called Cauchy sequences). What is meant by this needs careful specification. Thus $\sqrt{2}$ can be constructed from the sequence

$$1,\ 1.4,\ 1.41,\ 1.414,\ 1.4142,\ 1.41421,\ \ldots .$$

But this is not simple because many other such sequences will also produce $\sqrt{2}$ and we need to check that these sequences are in some way all equivalent to one another.

In earlier chapters we saw how to construct the Positive Fractions $\mathbf{Q}^+$ from the Natural Numbers **N**, how to construct the Integers **Z** from the Natural Numbers **N**, and how to construct the Rational Numbers **Q** from the Integers **Z** or from the Positive Fractions $\mathbf{Q}^+$. In each case the extended system was formed from its precursor as a collection of ordered pairs with a rule for determining when two of the pairs represent the same number (as in $\frac{1}{2} = \frac{2}{4}$). There is such a method for constructing **R**, but rather than pairs of Rational Numbers, it uses pairs of *sets* of Rational Numbers. This method is due to Richard Dedekind c.1858, and the pairs of sets are called *Dedekind cuts*. Using pairs of sets looks like a big complication, but it does have the merit that every Real Number has one and only one representation, unlike in $\mathbf{Q}$ where each number has infinitely many representations such as $\frac{1}{2} = \frac{2}{4} = \frac{3}{6} = \ldots$.

The idea is to represent each $x \in \mathbf{R}$ by the two "sections" of $\mathbf{Q}$ which it produces. For example, $\sqrt{2}$ produces the collection of all Rationals below $\sqrt{2}$ (L, say) and the collection of all Rationals above $\sqrt{2}$ (R, say). In the ordered pair (L, R), L is called the left section and R is called the right section. The two sections L and R determine the position of $\sqrt{2}$ on the number line, so knowing L and R is as good as knowing the value of $\sqrt{2}$. So we *define* $\sqrt{2} = (L, R)$. There is no circularity in this definition because we can test elements of $\mathbf{Q}$ to see if they lie in L or R simply by squaring them. So if $x \in \mathbf{Q}$ is negative (or zero), we place it in L. If $x \in \mathbf{Q}$ is positive and $x^2 < 2$, we also place it in L, but if $x \in \mathbf{Q}$ is positive and $x^2 > 2$, we place it in R.

In general we can envisage the Rational Numbers $\mathbf{Q}$ distributed along the number line. We "cut" the number line by forming two sets L and R *of Rational Numbers* such that

1. every Rational Number lies in either L or R,
2. every Rational Number in L is less than every Rational Number in R,
3. neither L nor R is empty.
4. there is no least element in R

The pair (L, R) may then be used to define a Real Number x corresponding to the point where the cut is made. Figure 6.3 illustrates the line of Rational Numbers partitioned into two sections defining a Real Number x. The left section is denoted by L and shown in green, the right section is denoted by R and shown in red.

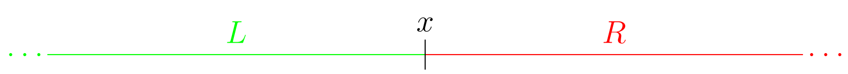


Figure 6.3: Dedekind cut at $x = (L, R)$.

Requirement 4 above deserves some explanation. If we decide to cut the Rationals at $\frac{1}{2}$, then L contains all the Rationals less than $\frac{1}{2}$ and R contains all the Rationals greater than $\frac{1}{2}$, but where to put $\frac{1}{2}$ itself? Requirement 4 tells us not to put it into R because it would then be the least element in R, so we must put it into L. This leaves R containing every Rational greater than $\frac{1}{2}$, and L containing all the remaining Rationals, namely those less than $\frac{1}{2}$, together with $\frac{1}{2}$ itself. So requirement 4 is really just there to tell us which set to use if we decide to cut the Rationals at a Rational Number. If we cut the Rationals at $\sqrt{2}$, there isn't a problem since $\sqrt{2}$ isn't a candidate for either L or R because it isn't a Rational Number ($\sqrt{2} \notin \mathbf{Q}$).

For any Real Number x defined by a cut (L, R), if we are told what R contains, then we can deduce what L contains since L contains all the Rationals that are not in R. So x is actually defined once we know the right section R. (Of course it is also defined once we know L.) Consequently we may identify each Real Number x with just its right section R and omit mention of its left section L. For example, the right section for $\sqrt{2}$ consists of all the positive Rational Numbers whose squares exceed 2. So we get

$$\sqrt{2} = \{\text{Every } f \in \mathbf{Q} \text{ for which } f > 0 \text{ and } f^2 > 2\}.$$

And here I am using red ink to distinguish the new numbers (**R**) from the old numbers (**Q**). However, it is still useful to think of each Real Number as having a left and a right section, and we will return to this in Chapter 12.

When we constructed the Positive Fractions from the Natural Numbers, there was the issue that the Positive Fraction $\frac{2}{1}$ is identified with the Natural Number 2. This didn't cause a problem but, logically speaking, they are slightly different since one lies in $\mathbf{Q}^+$ while the other lies in $\mathbf{N}$. We encountered the same sort of issue when discussing the formation of the set of Integers $\mathbf{Z}$ from the Natural Numbers $\mathbf{N}$. In constructing the Real Numbers we again have this slightly irritating logical difference between $0.5 \in \mathbf{R}$ and $\frac{1}{2} \in \mathbf{Q}$. Although we choose to identify them as the same number, strictly speaking in terms of a right section

$$0.5 = \{\text{Every } f \in \mathbf{Q} \text{ for which } f > \frac{1}{2}\}.$$

If this looks a bit bizarre, remember that the number on the left is regarded as a Real Number, while the f and $\frac{1}{2}$ inside the curly brackets are Rational Numbers. I wrote 0.5 rather than $\frac{1}{2}$ just to emphasize this distinction. But having made the point, I'll now drop using the red ink.

We will briefly examine how Dedekind cuts relate to decimals. First suppose that we have a decimal and we want to determine the cut (L, R). As an example we take our original example of an irrational number $x = 0.1010010001\ldots$. To decide if a given Rational Number f is in L or in R, we just look at the decimal expansion of f and compare it with $x = 0.1010010001\ldots$. If $f < x$ then $f \in L$ and if $f > x$ then $f \in R$. So L and R are easily determined. Thus a decimal determines a cut.

Conversely, suppose that we have the Dedekind cut (L, R) giving a Real Number x, and that we want the decimal expansion of x. We will illustrate the process by considering the case $x = \sqrt{2}$. We have knowledge of what is in L and what is in R, so start by observing that $1.4 \in L$ (because $(1.4)^2 < 2$) and $1.5 \in R$ (because $(1.5)^2 > 2$). Hence $\sqrt{2}$ must be trapped between 1.4 and 1.5, and this tells us that its decimal expansion starts $1.4\ldots$. But we also have $1.41 \in L$ and $1.42 \in R$, so the decimal expansion of $\sqrt{2}$ starts $1.41\ldots$. And we can go on like this, recording the decimal expansion of $\sqrt{2}$ to any desired degree of accuracy.

Of course we have to explain how to add and multiply cuts (i.e. right sections), how to define the order relation ($<$), and then show that axioms 1 to 11 hold good. All this can be done. We will just indicate how to begin. Suppose that R_x and R_y are the two right sections defining Real Numbers x and y respectively.

First we deal with the order relation by defining $x < y$ to mean that R_y is a (proper) subset of R_x. [“proper” means that we exclude the case $R_y = R_x$, which corresponds to $x = y$.] Figure 6.4 illustrates this definition with R_x shown in red and R_y shown in blue. We can then define “positive” and “negative” for Real Numbers.

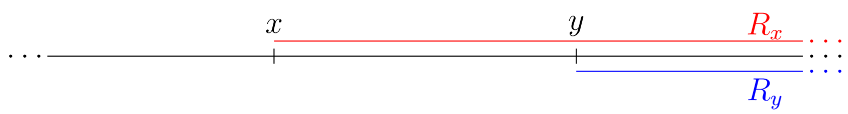


Figure 6.4: $x < y$ means that R_y is a proper subset of R_x.

Next we deal with addition by defining $x + y$ as the (new) right section that contains the sums of all the pairs (s, t) of Rationals for $s \in R_x$ and $t \in R_y$:

$$x + y = \{\text{Every } r = s + t \text{ for which } s \in R_x \text{ and } t \in R_y\}.$$

For example, if the Real Numbers 1 and 2 are to be added then $1 + 2$ has the right section R containing the sum of each Rational Number greater than 1 and each Rational Number greater than 2. So R certainly contains $1.001+2.001 = 3.002$, and it is then easy to see that R contains every Rational Number greater than 3, and no Rational Numbers less than 3. Consequently for Real Numbers, $1 + 2 = 3$; amazing but reassuring!

Multiplication is defined in a similar way although there is a complication caused by the fact that the product of two negative Rationals is positive. Using the same terminology as for addition, we would really like to define

$$x \times y = \{\text{Every } r = s \times t \text{ for which } s \in R_x \text{ and } t \in R_y\}.$$

Although this works fine when x and y are both positive (or zero), it is not good enough when negative numbers are involved. For example, if $x = (-2)$ and $y = (-3)$ then we'd like the right section corresponding to the product to be $R = \{\text{Every } f \in \mathbf{Q} \text{ for which } f > 6\}$. However we have $(-1) \in R_x$ and $(-2) \in R_y$, but $(-1) \times (-2) = 2$, which does not lie in R. So some technical adjustments are needed to deal with this problem. We won't do that here. It is actually easier to stick with the definition above for $x \times y$ when $x \geq 0$ and $y \geq 0$, and then define $(-x) \times y$ and $x \times (-y)$ as $-(x \times y)$, and define $(-x) \times (-y)$ as $x \times y$. But again, I will spare you the details.

It is now relatively easy, but rather tedious, to check axioms 1 to 10 for the Real Numbers **R** as defined by Dedekind cuts. Axiom 11 (completeness) is a bit trickier. For the two reasons, tedium and difficulty, I don't propose to undertake the task here.

6.6 The philosophical view

Let us instead return to the question of what is a number. Are the elements of **R** actually numbers? Pythagoras' followers had their doubts. Indeed we never actually need more than **Q** to do measurements because we can't measure to an infinite number of decimal places. However, it's easy to *construct* $\sqrt{2}$ as the hypotenuse (long side) of a right-angled triangle whose sides adjacent to the right angle are of unit length; see Figure 6.5.

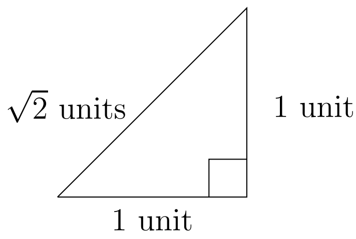


Figure 6.5: Constructing $\sqrt{2}$.

In fact π is also irrational and so the area of circle of radius 1 (which equals π) is therefore in **R** but not in **Q**. Whether we use **R** in preference to just **Q** is really a practical issue. If we want to do precise mathematics then we have to use **R**, but if we are only interested in physical measurements, then **Q** will suffice. But getting an answer $\sqrt{2}$ to a problem is so much more satisfactory than getting 1.4142. The latter might have arisen from an approximation to $\sqrt{2}$ but, on the other hand, it just might be a coincidence.

Since the Real Numbers **R** can be *constructed* (ultimately) from the Natural Numbers **N** that have a good claim to some sort of independent existence outside ourselves, maybe their true status is that they are intermediate between having an independent existence and being totally our inventions. We could certainly explain them to an intelligent extraterrestrial being who understood the Natural Numbers. This sort of fudge will probably not satisfy philosophers, but there is certainly a great cultural achievement in discovering or inventing them. Be proud of what humanity has achieved!

We have now reached the mid-point in our journey across the landscape of numbers. There have been a few steep bits but we are now on a sort of plateau. The number systems presented at school normally stop at the Real Numbers **R**, represented as the set of all decimals. Consequently most people (if they think about the issue at all) consider **R** as *the* set of numbers. However, it is possible to go a lot further and there are several directions to pursue.

Within **R** we can solve equations like $x^2 = 2$ but the similar looking equation $x^2 = -2$ defeats us; the reason being that if x is positive its square is positive and if x is negative its square is again positive - there just aren't *any* Real Numbers with negative squares. To solve this problem we have to extend **R** to a new class of numbers, the *Complex Numbers* **C**. We will look at this in Chapter 7, and a further development along the same lines in Chapter 8.

In Chapter 9 we will start on a different route, looking at how mathematicians handle infinity. This aspect will be examined further in the remaining Chapters 10, 11 and 12.

Chapter 7

The Complex Numbers (C)

7.1 Representation of Complex Numbers

Consider the equation $x^2 = -1$. Solving this in the usual way by taking square roots suggests $x = \pm\sqrt{-1}$. However, $\sqrt{-1}$ cannot be a Real Number because if $x > 0$ then $x^2 > 0$, if $x < 0$ then again $x^2 > 0$, and if $x = 0$ then $x^2 = 0$. So there just aren't *any* **Real** Numbers with negative squares. Whatever $\sqrt{-1}$ may be (if indeed it makes any sense at all) it definitely does not lie in $\mathbf{R}$. We are therefore led to the invention (discovery?) of the so-called imaginary number $i = \sqrt{-1}$. The term "imaginary" is unfortunate because i is no more or less imaginary than, for example, -1; we'll come back to that statement later.

Assuming for the moment that i obeys the usual rules of arithmetic we can evaluate the square root of any negative Real Number without inventing any more numbers. For example,

$$\sqrt{-4} = \sqrt{4} \times \sqrt{-1} = 2i.$$

In fact we can solve any quadratic equation and any quadratic equation now has two solutions (allowing that the two solutions may be equal). Do you remember how to solve quadratic equations? If the equation is $ax^2 + bx + c = 0$, with $a \neq 0$ (so that it really is a quadratic equation), then the two solutions are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

If you've forgotten this, a brief (optional) explanation is given below

****

Solving a quadratic equation

We take the equation as $ax^2 + bx + c = 0$ with $a \neq 0$, and try to find an expression for x in terms of a, b and c. It will help if you first convince yourself that $\left(x + \frac{b}{2a}\right)^2 = x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}$.

We start by doing some algebraic manipulation on $ax^2 + bx + c$.

$$\begin{aligned} ax^2 + bx + c &= a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right) \\ &= a\left(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} - \frac{b^2}{4a^2} + \frac{c}{a}\right) \\ &= a\left(\left(x + \frac{b}{2a}\right)^2 + \left(\frac{c}{a} - \frac{b^2}{4a^2}\right)\right). \end{aligned}$$

So if $ax^2 + bx + c = 0$ and $a \neq 0$, we have the equivalent equation

$$\left(x + \frac{b}{2a}\right)^2 + \left(\frac{c}{a} - \frac{b^2}{4a^2}\right) = 0.$$

This version of the equation can be written as

$$\begin{aligned} \left(x + \frac{b}{2a}\right)^2 &= \frac{b^2}{4a^2} - \frac{c}{a} \\ &= \frac{b^2 - 4ac}{4a^2}. \end{aligned}$$

Then taking square roots gives

$$x + \frac{b}{2a} = \pm\sqrt{\frac{b^2 - 4ac}{4a^2}} = \pm\frac{\sqrt{b^2 - 4ac}}{2a}.$$

So finally by subtracting $\frac{b}{2a}$ from both sides we get to the formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

□

****

Let's try solving $x^2 - 4x + 5 = 0$ using the formula . Here we have $a = 1$, $b = -4$ and $c = 5$, so

$$\begin{aligned} x &= \frac{-(-4) \pm \sqrt{(-4)^2 - 4 \times 5}}{2} \\ &= \frac{-(-4) \pm \sqrt{16 - 20}}{2} \\ &= \frac{4 \pm \sqrt{-4}}{2} \\ &= \frac{4 \pm 2i}{2} = 2 \pm i. \end{aligned}$$

Before we introduced i, we would have become stuck at evaluating $\sqrt{-4}$. In the set of Real Numbers **R** the equation $x^2 - 4x + 5 = 0$ has no solutions. But in the set of Complex Numbers **C** it has two solutions, namely $2+i$ and $2-i$. These solutions correspond to linear factors of the quadratic expression x^2-4x+5. (A linear factor is a factor like $ax+b$ where $a, b \in \mathbf{C}$, e.g. $3x+2i$.)

In general, if we use the quadratic formula to solve the quadratic equation $ax^2 + bx + c = 0$ we will get two solutions for x, say $x = u$ and $x = v$ (which can be equal, but that's OK). Then we find that ax^2+bx+c can be factorized:

$$ax^2 + bx + c = a(x - u)(x - v).$$

So solving $ax^2 + bx + c = 0$ is equivalent to factorizing $ax^2 + bx + c$ into linear factors. If you want to prove this claim, multiply out $a(x - u)(x - v)$ to get $a(x^2 - (u + v)x + uv)$ and then use

$$u = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad v = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

With a bit of patience you'll get to $ax^2 + bx + c$.

From our example above

$$x^2 - 4x + 5 = (x - (2 + i))(x - (2 - i)),$$

i.e. the two linear factors of $x^2 - 4x + 5$ are $x - (2 + i)$ and $x - (2 - i)$. I invite you to verify this by multiplying these two factors. Working in **C**, every quadratic expression can be factorized into two linear factors by using the quadratic formula.

The solution to the equation above produced the numbers $2+i$ and $2-i$. The most general form of a Complex Number is $a + ib$ where a and b are

Real Numbers, so you must expect to see numbers like $3+4i$ and $\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i$. When you have a Complex Number $z = x + iy$ with x and y ordinary Real Numbers, then x is called the *real part* of z and y (NOT iy) is called the *imaginary part* of z. So the real part of $3+4i$ is 3 and the imaginary part is 4, while the real part of $\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i$ is $\frac{1}{\sqrt{2}}$ and the imaginary part is $-\frac{1}{\sqrt{2}}$.

The Real Numbers are embedded in the Complex Numbers by identifying each Real Number x with the Complex Number $x + 0i$, which of course has real part x and imaginary part 0. Similarly, a purely imaginary number such as i can be written as $0 + 1i$ and so has real part 0 and imaginary part 1.

Maybe you already noticed that each Complex Number is really just an ordered pair of Real Numbers. It's just that we chose to write the ordered pair (a, b) as $a + ib$. So the Complex Numbers can be constructed directly from ordered pairs of Real Numbers without ever mentioning i by simply writing $a + ib$ as (a, b). Once again we have to define equality of these ordered pairs and now the definition of equality is easy: $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$. In the other notation, $a + ib = c + id$ if and only if $a = c$ and $b = d$.

Just as we got **Z** from **N**, **Q** from **Z**, and **R** from **Q**, we now get **C** from **R**; all these numbers ultimately rest on **N**. We could share our understanding of these with any extra-terrestrial visitors who were happy with the Natural Numbers.

7.2 Addition and multiplication

It is easy to add Complex Numbers:

$$(a + ib) + (c + id) = (a + c) + i(b + d).$$

For example, $(3 + 4i) + (2 - i) = 5 + 3i$.
In terms of ordered pairs of Real Numbers, the definition of addition is

$$(a, b) + (c, d) = (a + c, b + d).$$

It is also easy to multiply Complex Numbers (remembering that $i^2 = -1$):

$$\begin{aligned}(a + ib) \times (c + id) &= ac + iad + ibc + i^2 bd \\ &= (ac - bd) + i(ad + bc).\end{aligned}$$

For example, $(3 + 4i)(2 - i) = 6 - 3i + 8i - 4i^2 = 6 + 4 + 5i = 10 + 5i$.
In terms of ordered pairs of Real Numbers, the definition of multiplication is

$$(a, b) \times (c, d) = (ac - bd, ad + bc).$$

Using the definition of multiplication in terms of ordered pairs, we get $(0,1)^2 = (0,1) \times (0,1) = (-1,0)$. So denoting $(0,1)$ as i, and $(-1,0)$ as -1, we get $i^2 = -1$.

With $0 = 0 + 0i$ and $1 = 1 + 0i$, it is then easy to check that the set of all Complex Numbers, $\mathbf{C}$, obeys the axioms 1 - 8 specified on page 57 for $\mathbf{R}$ (i.e. if we replace "$\mathbf{R}$" by "$\mathbf{C}$" in statements of the axioms, they are still true). In fact the additive inverse of $a + ib$ is $-(a + ib) = (-a) + i(-b)$.

To find the multiplicative inverse of $a + ib$ look at what happens if you multiply $a + ib$ by $a - ib$ (with $a, b \in \mathbf{R}$).

$$(a + ib)(a - ib) = a^2 - iab + iab - i^2b^2 = a^2 + b^2,$$

and $a^2 + b^2$ is a Real Number. So if $a^2 + b^2 \neq 0$,

$$(a + ib)\left(\frac{a - ib}{a^2 + b^2}\right) = 1.$$

Consequently the multiplicative inverse (the reciprocal) of $a + ib$ is

$$(a + ib)^{-1} = \frac{a - ib}{a^2 + b^2} = \left(\frac{a}{a^2 + b^2}\right) - i\left(\frac{b}{a^2 + b^2}\right)$$

provided $a^2 + b^2 \neq 0$. For Real Numbers a and b, the expression $a^2 + b^2$ can only equal zero if both a and b are zero. So the condition $a^2 + b^2 \neq 0$ is equivalent to $a + ib \neq 0$. Thus every Complex Number apart from 0 has a multiplicative inverse.

Here's an example of finding the multiplicative inverse (i.e. the reciprocal) of $3+4i$. You can use the formula above but it's easier to do as follows. Write this inverse $(3+4i)^{-1}$ as $\frac{1}{3+4i}$. Then multiply top and bottom of the fraction by $3 - 4i$. Note that $(3 + 4i)(3 - 4i) = 3^2 - 12i + 12i - 4^2i^2 = 3^2 + 4^2 = 25$. So we get

$$(3 + 4i)^{-1} = \frac{1}{3 + 4i} = \frac{3 - 4i}{(3 + 4i)(3 - 4i)} = \frac{3 - 4i}{25}.$$

If you really want to, you can write this answer as $\frac{3}{25} + i(-\frac{4}{25})$. If you are not convinced and want to check that the reciprocal of $3 + 4i$ really is $\frac{3-4i}{25}$, just multiply them:

$$(3 + 4i) \times \left(\frac{3 - 4i}{25}\right) = \frac{(3 + 4i)(3 - 4i)}{25} = \frac{25}{25} = 1.$$

Just reflecting on what we did here, we had $3+4i$ in the denominator of $\frac{1}{3+4i}$, so we multiplied by $\frac{3-4i}{3-4i}$, which of course is 1 in disguise. The number $3-4i$ is called the *complex conjugate* of $3+4i$. In general the complex conjugate of $a+ib$ (with a and b Real Numbers) is the number $a-ib$. The trick we have just performed converted the denominator $3+4i$ to the Real Number 3^2+4^2. For Real Numbers a and b you will always have $(a+ib)(a-ib) = a^2+b^2$, so the trick is guaranteed to work.

Considering the remaining axioms that we had for **R** in the previous chapter, we find that the Complex Numbers **C** cannot be ordered in the way described by axioms 9 and 10. This is because i is not positive or negative (or zero). However, it is possible to prove that the Complex Numbers satisfy the completeness axiom (number 11) but again we need to specify what is meant by "arbitrarily closer together" and "converges".

7.3 Argand diagrams

The Real Numbers **R** can be represented on a one-dimensional line. Complex Numbers have two components, a real part and an imaginary part, and they are best represented on a two-dimensional graph (see Figure 7.1). Such a diagram is called an Argand diagram after the person who first came up with the idea of representing Complex Numbers in this way in the early 1800s.

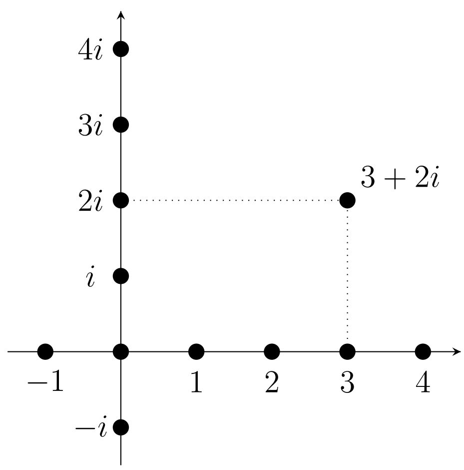


Figure 7.1: An Argand diagram

Looked at like this, it isn't surprising that we can't put these numbers into an obvious ordering. Earlier I remarked that i is really no more imaginary than -1. After all, you can see i on the diagram at one unit up the vertical axis, not a lot different from -1 which lies one unit to the left along the horizontal axis.

It is possible to interpret the operations of addition and multiplication on an Argand diagram. First we look at addition. An example is shown in Figure 7.2 where we are adding the Complex Numbers $2+i$ and $1+3i$ to get $3+4i$.

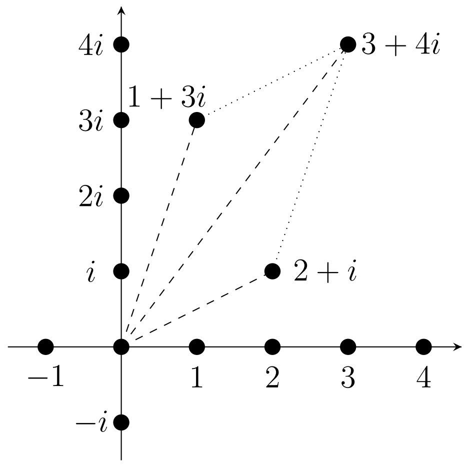


Figure 7.2: $(2+i)+(1+3i)=3+4i$.

In Figure 7.2 you can see how addition of two Complex Numbers can be represented on an Argand diagram by forming a parallelogram or by forming one of the two alternative triangles shown. For those of you familiar with vectors, this is just vector addition in two dimensions.

As you might expect, multiplication is a little more complicated. Figure 7.3 shows the effect of multiplying the Complex Number $z=3+2i$ firstly by i, then by $i^2=-1$, and finally by $i^3=-i$. We have $iz=-2+3i$, $i^2z=-3-2i$ and $i^3z=2-3i$. You can see that the effect of these multiplications is to rotate z anticlockwise by 90^o, by 180^o, and by 270^o respectively about the origin $(0,0)$.

There is nothing special about $z=3+2i$. In general, multiplying any number z by i rotates the point representing z by 90^o about the origin. It is significant that i lies only 1 unit up the vertical axis. If we multiply z by $2i$, not only does it rotate z by 90^o, but it also moves it twice as far away from the origin. More generally, if we multiply z by another Complex Number, say w, there are two effects, one is a rotation of z about the origin, and the other

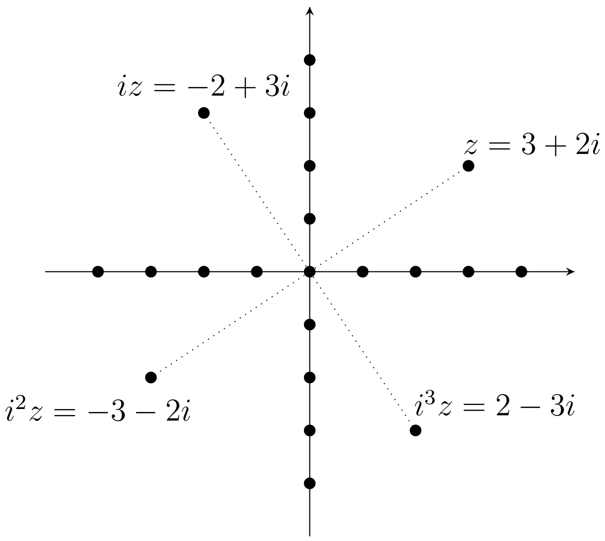


Figure 7.3: $z, iz, -z$ and$-iz$ for $z = 3 + 2i$.

is a magnification (or diminution) of z, and both of these are determined by w. If you are up for a bit of trigonometry I can explain how to determine the angle of rotation and the magnification factor by looking at the *polar form* of a Complex Number w. If you can't face this, never mind, just skip past it.

****

7.4 Polar form (optional)

In Figure 7.4 we have shown a typical point $w = a + ib$. If you look at the heavily drawn triangle you will see that it has horizontal side of length a and vertical side of length b. If we let r denote the length of the third side (the hypotenuse) and let θ denote the angle at the origin, then we have

$$\cos\theta = \frac{a}{r} \quad \text{and} \quad \sin\theta = \frac{b}{r}, \quad \text{also} \quad \tan\theta = \frac{b}{a}.$$

Consequently $a = r\cos\theta$ and $b = r\sin\theta$. So we may write

$$a + ib = r(\cos\theta + i\sin\theta).$$

By Pythagoras' theorem we also have $r^2 = a^2 + b^2$, so $r = \sqrt{a^2 + b^2}$ (of course we take the positive square root). The Real Numbers r and θ form

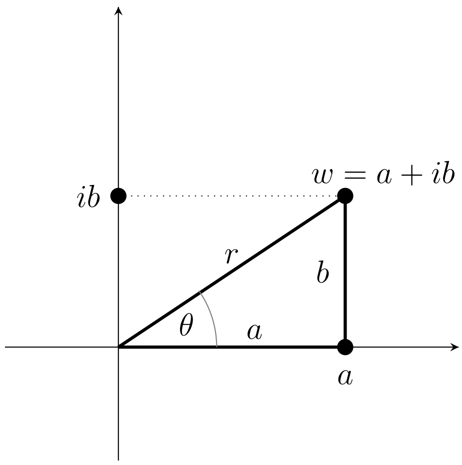


Figure 7.4: Polar form: $a + ib = r(\cos\theta + i\sin\theta)$.

the *polar coordinates* of the Complex Number $w = a + ib$, and we say that $r(\cos\theta + i\sin\theta)$ is the polar form for w.

Let's try an example of converting $w = 1 + i$ to polar form. Here $a = 1$ and $b = 1$, so $r = \sqrt{1^2 + 1^2} = \sqrt{2}$. Also $\tan\theta = \frac{1}{1}$ and so we find that $\theta = 45^o$. Consequently $1 + i = \sqrt{2}(\cos 45^o + i\sin 45^o)$.

Given a Complex Number $w = a+ib$, the quantity $r = \sqrt{a^2 + b^2}$ is called the *modulus* of w; it measures the distance of w from the origin on an Argand diagram and so corresponds to the size of w. It is often written as $|w|$, so in the example above we may write $|1 + i| = \sqrt{2}$. The angle θ is called the *argument* of w; it measures the angle of rotation of w from the positive horizontal axis on an Argand diagram. It is often written as $\arg w$, so in the example above we may write $\arg(1 + i) = 45^o$.

Why have I digressed into polar form? It's because multiplication is very easy if you use polar form. It is not hard to prove that

$$r_1(\cos\theta_1 + i\sin\theta_1) \times r_2(\cos\theta_2 + i\sin\theta_2) = r_1 r_2(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)).$$

In words, we multiply the moduli (r_1 and r_2) and add the arguments (θ_1 and θ_2). The proof just depends on some trigonometry, but this isn't a mathematics textbook, so I'm not going to do it here. But you might look again at Figure 7.3 and note that i has modulus 1 and argument 90^o. So multiplying $3 + 2i$ by i multiplies its modulus by 1 (i.e. it doesn't alter it, and so the new point is the same distance from the origin), and it adds 90^o to its argument (i.e. it rotates it by 90^o about the origin). You can look at multiplication by -1 and by $-i$ in the same way. In general, if you multiply a Complex Number z by a second Complex Number w that has modulus r and argument θ, then the effect on z is to apply a magnification factor r and a rotation θ.

Polar form is very useful. As another example, if you want the square root of i, take the square root of its modulus (1) and halve its argument (90^o). So the square root of i is $(\cos 45^o + i \sin 45^o) = \frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}} = \frac{1+i}{\sqrt{2}}$. You don't believe me? Here's the test:

$$\begin{aligned}\left(\frac{1+i}{\sqrt{2}}\right)\left(\frac{1+i}{\sqrt{2}}\right) &= \frac{(1+i)(1+i)}{2}\\ &= \frac{1+2i+i^2}{2} = \frac{2i}{2} = i.\end{aligned}$$

Honesty compels me to mention a slight snag with this. You would expect to get two square roots, so where has the other one gone? The answer really lies in what we mean by saying that *the* argument of i is 90^o. It would be more accurate to say that *an* argument of i is 90^o. We could write this angle as $(90 + 360)^o = 450^o$. Indeed, we could write it as $(90 + 360n)^o$ for any Integer n. If we choose to use 450 and halve it we get 225. So another square root of i is given by $(\cos 225^o + i \sin 225^o) = -\frac{1}{\sqrt{2}} - i\frac{1}{\sqrt{2}} = -\frac{1+i}{\sqrt{2}}$. You can check that we don't get any different square roots by choosing any of the other possible arguments; for example using $90 + 720$ gives the same answer as using 90. If you pursue this, you will always find three cube roots, four fourth roots, and so on.

A bit of light relief: one of the smartest observations that I ever had from a student came from an Algerian electrical engineering student who pointed out that if you write roots as powers then $z^{\frac{1}{3}}$ denotes a cube root of z, and there are three such cube roots. But if you approximate $\frac{1}{3}$ as $0.3 = \frac{3}{10}$ then you will get 10 possibilities for $z^{0.3}$ (the tenth root of z^3). And the situation gets worse if you refine your approximation to 0.33, when you get 100 possibilities. In fact the better the approximation to $\frac{1}{3}$, the more solutions! You might like to think on this, but then again you might not.

Our final remark on polar form is that it provide a fascinating connection between the cosine, sine and exponential functions. The exponential function exp was described briefly on page 56. It is possible to prove that $\cos\theta + i\sin\theta = \exp(i\theta)$. The proof uses infinite series representations of the three functions and θ must be expressed in radians rather than degrees. If you are prepared to accept this and apply it with $\theta = \pi$ (an angle of 180^o) then $e^{i\pi} = \exp(i\pi) = \cos\pi + i\sin\pi = -1 + 0i$, where e is Euler's number. This gives one of the most beautiful results in the whole of mathematics, known as Euler's identity:

$$e^{i\pi} + 1 = 0.$$

This connects the two numerical constants π and e (both of which are irrational numbers with values approximately 3.142 and 2.718), with the imaginary number $i = \sqrt{-1}$ and the two basic Integers 0 and 1.

At this point we end the digression on polar form and conclude with some general remarks.

****

7.5 Uses of Complex Numbers

One very attractive feature of **C** is that any n^{th} degree polynomial with coefficients in **C** can be factorized into precisely n linear factors in **C** (some of the factors may be equal to one another). This is known as the Fundamental Theorem of Algebra. For example, $(2+i)z^3 + 3z^2 + iz - 5$ is a third degree polynomial (the highest power of the unknown z is z^3) and it will therefore factorize into three linear factors in **C**. The standard of what constitutes a mathematical proof has risen in the past 200 years, but one of the first (more or less) rigorous proofs of the Fundamental Theorem of Algebra was given by Argand in the early 1800s. That is not the same as saying that the factors of such polynomials are easy to determine!

Complex Numbers have turned out to be very useful, particularly in describing complicated physical processes such as electromagnetism, quantum theory and fluid flows. So are these really numbers? The answer depends on your point of view. If you want to solve all quadratic equations (for example) then your answer has to be yes. Mathematicians and Physicists would say yes. Very probably a little green man from another galaxy would have had to use them to get here and, even if he hadn't, we could easily explain them to him just so long as he agreed with us about the Natural Numbers. But they are getting just slightly away from the use of numbers to count and measure. So if all you want to do is count and measure then the answer is probably no. Just one brief warning: engineering textbooks often use j instead of i for $\sqrt{-1}$, the reason being that i is used to denote electrical current.

Having observed that **C** is, essentially, two dimensional, the question arises as to what, if anything, can be done in three or more dimensions. Can we extend **C** still further? The answer is a partial yes. In moving from **R** to **C** we lost some of the nice features of **R** (the ordering). It is possible to extend **C**, but further properties are lost. The simplest extension of **C** is the set of quaternions **H**. The symbol **H** is used to name them in honour of their discoverer, William Hamilton . These are essentially four dimensional numbers. Hamilton spent a lot of time trying to construct a three dimensional extension before realizing that a four dimensional structure was more promising. We cover these very briefly in the next chapter.

Chapter 8

The Quaternions (H)

Quaternions can be constructed from ordered pairs of Complex Numbers or, equivalently, from ordered quadruples of Real Numbers. Taking the latter approach we write the ordered quadruple (a, b, c, d) as $a + bi + cj + dk$ where $a, b, c, d \in \mathbf{R}$. We set up a multiplication table for the "new" numbers i, j, k:

$\times$	i	j	k
i	-1	k	$-j$
j	$-k$	-1	i
k	j	$-i$	-1

The table is read from row to column, so that $i \times j = k$. Using this table we can define the operations of addition and multiplication in a similar way to those in $\mathbf{C}$. We find that axioms 1 - 8 given for the Real Numbers continue to hold *except* that multiplication is no longer commutative - for example $i \times j = -(j \times i)$ from the table. With suitable interpretations of "arbitrarily closer together" and "converges", axiom 11 (completeness) continues to hold.

In case you know something about vectors, Quaternions are closely related to three-dimensional vectors. Each Quaternion $a + bi + cj + dk$ can be split into a *scalar* part (a) and a *vector* part $(bi + cj + dk)$. A pure Quaternion is one with a zero scalar part. If A, B are two pure Quaternions then their Quaternion product $A \times B$ equals $-(A \cdot B) + A \wedge B$, where $A \cdot B$ denotes the scalar product of the vectors A and B and $A \wedge B$ denotes the vector product.

Hamilton discovered what he was looking for on 16th October 1843 in Dublin. He was so excited that he cut into the stonework of one of the Dublin bridges the formula

$$i^2 = j^2 = k^2 = ijk = -1.$$

Hamilton went on to define many of the mathematical operations relating to vectors which are still in use today. In applied mathematics, vectors have largely supplanted the use of Quaternions. But even today Quaternions are used to efficiently describe three-dimensional rotations. In this way they are applied in fields such as computer graphics and robotics.

It is possible to extend $\mathbf{R}$ still further beyond $\mathbf{H}$ to the so-called Cayley Numbers or Octonions which are essentially 8-dimensional. However a further axiom must now be abandoned - the associativity of multiplication is the casualty.

We will not pursue our discussion any further in this direction. It is at least arguable that $\mathbf{H}$ is no longer concerned primarily with counting, measuring and solving equations. We shall therefore return to basics - how to count - before looking at some new types of numbers.

Chapter 9

Cardinal Numbers, finite and infinite

9.1 Problems with infinity

In this chapter we will start to look at how mathematicians handle the infinite. The first thing to emphasize is that "infinity" is not a member of the set of Real Numbers: $\infty \notin \mathbf{R}$. If you try to work with infinity as if it were an ordinary number, you will rapidly come adrift from reality. For example if you think that $\infty + 1 = \infty$, then by subtracting ∞ from both sides you get the nonsense that $1 = 0$. For simple reasons like this, and a whole host of more subtle reasons, mathematicians and philosophers have, for millennia, had a horror of arguments involving the infinite. The real breakthrough came in 1878.

9.2 Finite and infinite cardinals

We will make a start on our journey into the mathematically infinite by looking at *cardinal numbers*. The finite cardinal numbers are the numbers $0, 1, 2, 3, \ldots$. These measure quantity (as in, zero, one, two, three, etc.) rather than order (as in first, second, third, etc.). As an adjective, the word "cardinal" references importance. If we were given a bag of gold coins, the first thing we would be likely to do is to count the contents of the bag and assign a cardinal number. In a sense this is the most important attribute of the contents, given that we already know that each individual item is a gold coin.

Counting objects is done by putting them into one-to-one correspondence with another collection, the reference collection, whose number is known.

The reference collection may be fingers, counters, tally marks, words, or whatever is convenient. In this day and age we tend to use words such as "one", "two", "three", etc. When we have finished the one-to-one matching at, say, "twenty" we claim that there are twenty objects in our collection. In such a case a pedantic mathematician might say that the collection of objects has *cardinality* 20 or that the number 20 is the *cardinal number* of the collection. How many words are there in this paragraph? That is the cardinality of this paragraph.

There is nothing remarkable in such examples. However, in 1878 Georg Cantor realized that the same idea could be extended to *infinite sets*. First we need a word for the number of Natural Numbers, i.e. those in $\mathbf{N}$. "Infinite" is the first word that comes to mind but that isn't sufficiently precise, as we will see. Mathematicians, for less than obvious reasons, use the word "aleph-nought", written as $\aleph_0$.[1] So we say that $\aleph_0$ is the cardinal number of $\mathbf{N}$.

To see how the idea works, let us "count" the positive even numbers $(2, 4, 6, 8, \ldots)$. We will try to put them into one-to-one correspondence with the Natural Numbers $\mathbf{N}$. This is easy. With 2 we count "1", with 4 we count "2", with 6 we count "3", and so on. In general we count the even number $2n$ as number "n". The correspondence can be described by the following table:

Positive even number:	2	4	6	8	10	12	14	...
Natural Number:	1	2	3	4	5	6	7	...

The conclusion is that these two collections have the *same* cardinal number, namely $\aleph_0$. This seems a little disturbing because one tends to feel that there are twice as many positive Integers as there are positive even Integers. Many of Cantor's contemporaries were very disturbed by his work in this area and he faced a lot of opposition from several serious mathematicians. There were even religious objections to his work. That's how revolutionary these ideas were in the late 1800s. Despite the criticism at the time, the theory of cardinal numbers is now well-established. It wouldn't be a gross exaggeration to say that Cantor was the person who had finally conquered the infinite in mathematics.

One can easily show that $\mathbf{Z}$ (the set of Integers) has the same cardinality as $\mathbf{N}$. We count each positive Integer x as number $2x$, each negative Integer $-y$ as number $2y + 1$, and the Integer 0 as number 1:

Integer:	0	1	−1	2	−2	3	−3	...
Natural Number:	1	2	3	4	5	6	7	...

[1] $\aleph$ (aleph) is the first letter of the Hebrew alphabet and the first letter of the Hebrew word for infinity.

It isn't difficult to show that **Q** (the Rationals) also has the same cardinality. Look at the following diagram.

$$
\begin{array}{ccccccccccccccc}
0 & \rightarrow & \frac{1}{1} & \rightarrow & -\frac{1}{1} & & \frac{1}{2} & \rightarrow & -\frac{1}{2} & & \frac{1}{3} & \rightarrow & -\frac{1}{3} & & \cdots \\
 & & & \swarrow & & \nearrow & & \swarrow & & \nearrow & & \swarrow & & \nearrow & \\
 & & \frac{2}{1} & & -\frac{2}{1} & & \frac{2}{2} & & -\frac{2}{2} & & \frac{2}{3} & & -\frac{2}{3} & & \cdots \\
 & & \downarrow & \nearrow & & \swarrow & & \nearrow & & \swarrow & & \nearrow & & \swarrow & \\
 & & \frac{3}{1} & & -\frac{3}{1} & & \frac{3}{2} & & -\frac{3}{2} & & \frac{3}{3} & & -\frac{3}{3} & & \cdots \\
 & & & \swarrow & & \nearrow & & \swarrow & & \nearrow & & \swarrow & & & \\
 & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \cdots
\end{array}
$$

The diagram clearly contains all the elements of **Q** (in fact, apart from 0, they each appear several times). We "count" them by following the path indicated, skipping any previously encountered such as $\frac{2}{2}$ (previously encountered as $\frac{1}{1}$).

Having seen that $\mathbf{N}, \mathbf{Z}$ and **Q** all have the same cardinality $\aleph_0$ one might think that all infinite sets have this cardinality. But this turns out not to be the case. Recall that we showed earlier (very informally on page 55) that the set of Real Numbers **R** contains a *lot* more numbers than **Q**. Most Real Numbers are irrational. We shall now prove that **R** does not have cardinality $\aleph_0$. We do this by assuming that **R** does have cardinality $\aleph_0$ and go on to derive a contradiction.

If **R** does have cardinality $\aleph_0$ then we can arrange the elements of **R** in an ordered list with a first number, a second number, a third, and so on. Let these be given as the decimal expansions shown in the second column of the table below.

Natural Number	Real Number
1	$A.a_1a_2a_3\ldots$
2	$B.b_1b_2b_3\ldots$
3	$C.c_1c_2c_3\ldots$
$\vdots$	$\vdots$

Here $A, B, C, \ldots$ are (signed) Integers and the symbols $a_1, a_2, a_3, \ldots,$ $b_1, b_2, b_3, \ldots,$ $c_1, c_2, c_3, \ldots, \ldots$ are all digits from 0 to 9. To guard against ambiguity in decimal expansions we can exclude decimals with recurring 9's; for example we write "one" as $1.000\ldots$ rather than as the recurring decimal $0.999\ldots$. We are now going to pick a number in **R** which cannot be in the list. Our number can start with "0." and we then choose its decimal digits carefully. Let x_1 be the first decimal digit to be chosen. The basic idea is to pick x_1 so that it doesn't equal a_1, then our number cannot equal the first number in the list. For a technical reason we also choose $x_1 \neq 9$. Similarly, the second digit x_2 is chosen so that it doesn't equal b_2, and again $x_2 \neq 9$. Then our number cannot equal the second number in the list. The third digit x_3 is chosen so that it doesn't equal c_3 and again $x_3 \neq 9$. Proceeding like this we construct

$$0.x_1x_2x_3\ldots$$

which does not appear *anywhere* in the list and doesn't contain recurring 9s (so it isn't a disguised version of a number in the list). But this contradicts the assumption that the list contains all the numbers in **R**. We conclude that the numbers in **R** cannot be put into an ordered list, and so **R** does not have cardinality $\aleph_0$. Thus there are different sizes of infinite sets! Now you know why we didn't use the word "infinite" for the number of (i.e. the cardinality of) the Natural Numbers **N**.

9.3 Power sets

It is possible to prove that, given an infinite set X, the set formed from all subsets of X has a larger cardinality than X. The proof of this is not all that difficult, but to get an idea of why it just might be true, consider what happens if we examine all the possible subsets of a finite set X. As an example, suppose we take a set X with three elements, say $X = \{a, b, c\}$. The subsets of this, starting with the empty set (denoted by $\emptyset$) are

$$\emptyset = \{\}, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\} = X.$$

So you see there are 8 subsets of a 3-element set, including both the empty set $\emptyset$ and the original set X. More generally, if a finite set X has n elements then the different subsets can be formed by specifying whether each element is either in or not in the subset. So there are 2 choices for each element, giving 2^n distinct subsets. At the two extremes, if you specify every element is not in, then you get $\emptyset$, while if you specify that every element is in, you get X itself. In our example, $n = 3$ so there are $2^3 = 8$ subsets. Obviously if n is large, then there will be a very large number of subsets since 2^n is then much larger than n itself. Because n leads to 2^n, the set of all subsets of a set X is sometimes denoted by 2^X and is called the *power set* of X, with alternative notation $\mathcal{P}(X)$. So for a finite set X of cardinality n, the power set $\mathcal{P}(X)$ has cardinality 2^n. This suggests that maybe for an *infinite* set X, the power set $\mathcal{P}(X)$ could have a larger cardinality than X itself. Of course we haven't proved the result for infinite sets, just given an argument to make it look plausible. But we will give a proper proof that applies to all sets, finite or infinite, after discussing Russell's paradox in the next section

Assuming that for an infinite set X its power set $\mathcal{P}(X)$ always has larger cardinality than X itself, we can reapply the result. The power set of an infinite set X is itself a respectable infinite set, and so we can form its power set $\mathcal{P}(\mathcal{P}(X))$. This will have even larger cardinality than that of $\mathcal{P}(X)$. By repeating this argument we can create an infinite sequence of increasingly large infinite cardinal numbers.

The **smallest** infinite cardinal is $\aleph_0$. The set of all subsets of $\mathbf{N}$, namely $\mathcal{P}(\mathbf{N})$, is said to have cardinality $\mathcal{C}$, also written as $2^{\aleph_0}$ and it is possible to prove that this is the cardinality of $\mathbf{R}$. The sets $\mathbf{C}$ and $\mathbf{H}$ also have this same cardinality. The reason that the symbol $\mathcal{C}$ is used is that the set of Real Numbers $\mathbf{R}$ is sometimes called "the continuum", and this denotes the cardinality of $\mathbf{R}$.

We can define arithmetic on infinite cardinals. If X and Y are cardinal numbers then we can define $X+Y$ as the cardinal number of a set formed as the union of two disjoint sets, one of cardinality X and the other of cardinality Y. Similarly we define the product $X \times Y$ as the cardinal number of a set formed as the union of Y disjoint sets, each having cardinality X. But such arithmetic is rather boring because we find

$$X + Y = \max\,(X, Y) = X \times Y.$$

For example $\aleph_0 + \mathcal{C} = \mathcal{C} = \aleph_0 \times \mathcal{C}$. Moreover, there are no additive or multiplicative inverses so the infinite cardinals do not behave in the same way as finite cardinals (i.e. the Natural Numbers).

9.4 The continuum hypothesis

Perhaps the most famous issue concerning the infinite cardinal numbers is the question as to whether or not there is a cardinal number between $\aleph_0$ and $\mathcal{C} = 2^{\aleph_0}$. Cantor himself spent a lot of time trying to find the answer but he didn't succeed. The proposition that there is no such cardinal is called the *continuum hypothesis*. It is now known that the usual rules of set theory are not sufficient to either prove or disprove this proposition. This remarkable fact was proved in two stages by Kurt Gödel (1938) and Paul Cohen (1963).

In 1900, David Hilbert, one of the world's leading mathematicians at the time, proposed a list of problems that he considered to contain the most important questions facing mathematicians at the start of the 20$^{\text{th}}$ Century. The list is certainly famous among mathematicians. The first problem on the list was the continuum hypothesis. To give the vaguest idea of Gödel and Cohen's results I need to say a bit more about what is meant by a logical model of a set of axioms.

As an example, let us take properties 1 to 10 given in Chapter 5 (see p.41) for the Rationals **Q**. We can regard these as a set of axioms for **Q**. We can say that **Q** is a model for these axioms because **Q** obeys them all. But it is also true that **R** obeys the same set of axioms, although it also obeys an additional axiom (completeness) as explained in Chapter 6 (see p.57). And since **R** is not just a copy of **Q** (it contains many more numbers), we see that we have two distinct models for axioms 1 to 10, and that the two models have different properties. In one model (**Q**) the completeness axiom from Chapter 6 does not hold, but in the other model (**R**) the completeness axiom does hold. That's not a surprise because these are two different types of numbers.

The fact that a set of axioms has at least one model ensures that the axioms are consistent and free of contradictions, provided only that the model itself is free of contradictions. So far we have built **Q** and **R** from **N** using a tiny bit of set theory (ordered pairs and sets of numbers), so if **N** is free of contradictions we have every reason to believe that so are **Q** and **R**, and consequently axioms 1 to 10 of Chapter 5 are healthy.

The continuum hypothesis is really a problem about sets of objects. Before trying to resolve it, we need to agree on the "obvious" properties that sets obey, in other words a collection of axioms for set theory. At first sight this looks unnecessary and that we can take a set to be any collection of objects. But in 1903 Bertrand Russell asked if the collection S of sets which don't contain themselves as members is itself a set. If it is a set, and if $S \in S$ then $S \notin S$, while if $S \notin S$ then $S \in S$ (think carefully about this) [2]. Plainly this is crazy, so S cannot be a set. This blew a gaping hole in the previous informal theory and pointed to the need for a carefully formed collection of axioms for sets that everyone (well, almost everyone) was happy with.

There is now a commonly agreed collection of axioms for set theory called the Zermelo-Fraenkel axioms. But there is more than one model for these axioms. Indeed, as Gödel and Cohen showed, there is one model where the

[2] An amusing version of Russell's paradox is given by the story of a small isolated town, perhaps in the Australian Outback, where the barber is known to shave the beards of all the men who don't shave themselves. Which sex is the barber?

continuum hypothesis holds, and another where it does not hold. So there are different types of set theory. Maybe we need another axiom in order to decide which one is "best", but it isn't at all obvious what it should be or even whether "best" makes sense. And the question we are really asking here is whether or not we can have a complete collection of axioms for set theory; a very deep question and one we cannot possibly address in this book.

9.5 Power sets revisited

An idea related to Russell's paradox can be used to provide a proper proof that for any set X, its power set $\mathcal{P}(X)$ has larger cardinality than X itself. The proof is by way of contradiction. We will assume that X and $\mathcal{P}(X)$ have the same cardinality and go on to derive a contradiction, something ridiculous, that shows our assumption must be incorrect.

So suppose that the set X and its power set $\mathcal{P}(X)$ have the same cardinality. This means that there is a one-to-one correspondence between the elements of X and the elements of $\mathcal{P}(X)$. Then to each $x \in X$ there corresponds a unique $A_x \in \mathcal{P}(X)$ (that is a subset A_x of X), and the set formed by all these subsets A_x comprises the whole of $\mathcal{P}(X)$.

Clearly for each $x \in X$, either $x \in A_x$ or $x \notin A_x$. Define the set

$$B = \{x \in X : x \notin A_x\}.$$

Then B is a subset of X and so $B \in \mathcal{P}(X)$. Hence $B = A_z$ for some $z \in X$. Now ask the question whether or not $z \in B$.

- If $z \notin B = A_z$ then, by definition of B, $z \in B$, a contradiction.
- If $z \in B$ then, by definition of B, $z \notin A_z = B$, a contradiction.

So in either case ($z \in B$ or $z \notin B$) we have a contradiction. Consequently our assumption that X and $\mathcal{P}(X)$ have the same cardinality must be incorrect. Therefore these sets have different cardinalities and it only remains to settle which is the larger.

This last point is easy to resolve. Just observe that $\mathcal{P}(X)$ has a subset consisting of all the single-element subsets of X. Clearly this can be put into one-to-one correspondence with X itself by the relationship $x \leftrightarrow \{x\}$. So $\mathcal{P}(X)$ has a subset with the same cardinality as X, consequently it cannot have smaller cardinality than X, and so it must have larger cardinality. $\square$

Chapter 10

Ordinal Numbers, finite and infinite

10.1 Construction of $\mathbf{N}$

We have seen, briefly, how to construct $\mathbf{C}$ and $\mathbf{H}$ from $\mathbf{R}$, $\mathbf{R}$ from $\mathbf{Q}$, $\mathbf{Q}$ from $\mathbf{Z}$, and $\mathbf{Z}$ from $\mathbf{N}$. Thus all these collections of numbers ultimately depend on $\mathbf{N}$. Can we construct $\mathbf{N}$, the set of Natural Numbers, from anything simpler and will this help us to decide what numbers actually are?

Mathematicians can certainly answer the first part of this question, but probably not the second. Rather than answer the question as to the essence of (say) the number 3, we provide an archetypal example of 3-ness and *define* 3 in terms of this example. All the familiar properties of 3 are deducible from this definition. Perhaps the most important property of 3 is that we can add 1 and make 4.

Back in Chapter 2 we spoke of the successor of a Natural Number and used the symbol $S(n)$ to denote the successor of n, which we recognized as $n+1$. We are looking for a definition of each Natural Number n that permits us to form its successor and thereby produce the number $n+1$. We also want our definition to involve only simple mathematical objects. The simplest objects of all are (arguably) sets, and the simplest of all of these is the empty set (denoted by $\emptyset$ or by $\{\}$) which has not got any members at all. We shall use $\emptyset$ as the starting point and, because it has *zero* members, we will find it easier to define $\mathbf{N}_{+0}$ rather than $\mathbf{N}$, where $\mathbf{N}_{+0}$ denotes the set of numbers $\{0, 1, 2, 3, \ldots\}$ that comprises $\mathbf{N}$ together with the number 0.

We start by considering a sequence of sets

$$\emptyset, \quad \{\emptyset\}, \quad \{\emptyset, \{\emptyset\}\}, \quad \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \quad \ldots$$

where **the next member of the sequence is always the set of all the preceding terms in the sequence**. We can then identify

$$\begin{array}{lllll}
0 & = & \emptyset, & & \\
1 & = & \{\emptyset\} & = & \{0\}, \\
2 & = & \{\emptyset, \{\emptyset\}\} & = & \{0, 1\}, \\
3 & = & \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} & = & \{0, 1, 2\}, \\
\vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

Thus, for example, $3 = \{0, 1, 2\}$ is the set of all the preceding terms in the sequence and it contains precisely three elements. So what is 3 according to this definition? It is literally the set containing 0, 1 and 2. And, likewise, 2 is the set containing 0 and 1, 1 is the set containing 0, and 0 is just the empty set $\emptyset$. This just leaves us with the empty set to explain away, but it is certainly simpler than 3 itself. Of course this definition is unlikely to satisfy the philosopher who wants to know the essence of 3. Mathematically it is fairly satisfactory because it bases all numbers on the empty set $\emptyset$ and the construction of larger sets from smaller ones in a particularly simple fashion.

The sequence of numbers $0, 1, 2, 3, \ldots$ as defined above is ordered by set-inclusion, that is to say $m < n$ if and only if m is an element of n (i.e. $m \in n$). The successor of m (namely $m + 1$) is $\{0, 1, 2, \ldots, m\}$, so it is easy to add 1. Thus $3{+}1 = \{0, 1, 2, 3\}$, which we call 4. If you want to add 2 to m, first add 1, then add another 1. So, for example, $3{+}2 = (3{+}1){+}1 = 4{+}1 = \{0, 1, 2, 3, 4\}$, which we call 5. Addition of any two elements of $\mathbf{N}_{+0}$ can be defined in this way by repeated addition of 1.

In mathematical terms it is not hard to show that the set of numbers which results has precisely the properties which we require from $\mathbf{N}$ (or rather $\mathbf{N}_{+0}$). This definition of these numbers gives them an automatic ordering and so when considered in the light of this definition, they are called ordinal numbers - we might call them zero-th, first, second, third, etc., rather than just zero, one, two, three, etc. But we won't actually do that because it gets too cumbersome.

The vertical dots used in the table above do need to be carefully interpreted. I meant to imply that we considered all *finite* sets in the sequence. If we remove this implied restriction we are at once into a much larger collection than $\mathbf{N}_{+0}$.

10.2 Infinite ordinals

Consider the following set.

$$\omega = \{0, 1, 2, 3, \ldots\} \notin \mathbf{N}_{+0}$$

This is an example of an *infinite* ordinal number. (The letter ω is the Greek letter omega.) By contrast, the elements of $\mathbf{N}_{+0}$ are referred to as *finite* ordinal numbers. It is possible to carry out some limited forms of arithmetic with infinite ordinal numbers; for example

$$\begin{array}{rcl} \omega + 1 & = & \{0, 1, 2, \ldots, \omega\}, \\ \omega + 2 & = & \{0, 1, 2, \ldots, \omega, \omega + 1\}, \\ \vdots & \vdots & \vdots \\ \omega + \omega & = & \{0, 1, 2, \ldots, \omega, \omega + 1, \omega + 2, \ldots\}. \end{array}$$

As with finite ordinal numbers, each infinite ordinal number is the set of all its predecessors.

We can define multiplication as repeated addition. But we need to be careful because multiplication involving infinite ordinals turns out to be non-commutative, that is to say, the order matters. Do we refer to $\omega + \omega$ as $2 \times \omega$ or as $\omega \times 2$? The convention is to refer to this as $\omega \times 2$; think of it as "ω repeated twice". So we can write

$$\omega \times 2 = \omega + \omega.$$

But in that case what is $2 \times \omega$? It must be 2 repeated ω times: $2+2+2+\ldots$. Let's evaluate that in easy stages. First $2+2 = 4 = \{0, 1, 2, 3\}$, then $2+2+2 = 4 + 2 = 6 = \{0, 1, 2, 3, 4, 5\}$, and so on. Thus we get

$$2 \times \omega = \{0, 1, 2, \ldots\} = \omega.$$

Thus multiplication is not commutative. For similar reasons the infinite ordinal numbers do not satisfy several other properties of $\mathbf{N}_{+0}$. Particularly critical is the fact that several ordinal numbers (for example ω) have the property of having *no immediate predecessor*. Thus we cannot write (for example) $\omega = \alpha + 1$ for any ordinal number α, and this means we cannot give meaning to $\omega - 1$. In $\mathbf{N}_{+0}$, the only number with this "defect" is 0 (zero). In $\mathbf{N}$ the only number with this "defect" is 1 (one).

Nevertheless, the collection of ordinal numbers does extend arithmetic into the realm of infinite sets. It is possible to define cardinals in terms of ordinals but it is not trivial and the meaning of "+" differs when adding ordinals and when adding cardinals. Different ordinals can have the same cardinal. For example, both ω and $\omega \times 2$ have cardinality $\aleph_0$. The set ω comprises the integers $0, 1, 2, \ldots$, and these can be put into one-to-one correspondence with the Natural Numbers as shown in the following table.

ω :	0	1	2	3	4	5	6	...
Natural Number:	1	2	3	4	5	6	7	...

The set $\omega \times 2 = \omega + \omega$ comprises the integers $0, 1, 2, \ldots$ along with the ordinal numbers $\omega, \omega + 1, \omega + 2, \ldots$, and these can be put into one-to-one correspondence with the Natural Numbers as shown in the next table.

$\omega \times 2 = \omega + \omega$:	0	ω	1	$\omega + 1$	2	$\omega + 2$	3	...
Natural Number:	1	2	3	4	5	6	7	...

We will not pursue this further here but switch attention to a rather more all-embracing extension of the familiar number system **R**.

Chapter 11

Model Theory and the Hyperreal Numbers ($^*\mathbb{R}$)

11.1 Historical background

Much of mathematics was originally developed without regard to what we would now call mathematical rigour. If mathematical rigour had been insisted upon, then much would never have seen the light of day - truly *rigor-mortis* would have set in. One of the main stumbling blocks was in our understanding of infinity. Infinity as a concept is usually conceived as referring to the idea of the indescribably large. But there are also issues regarding what might be called the infinitely small or infinitesimal.

Some of these issues were eventually resolved in a complicated but rigorous fashion by careful definitions such as the meaning of saying that $1 + \frac{1}{n}$ converges to 1 as n tends to infinity; in symbols

$$1 + \frac{1}{n} \to 1 \quad \text{as} \quad n \to \infty.$$

Intuitively one might think of this as saying that when n is infinitely large, then $\frac{1}{n}$ is infinitesimal so $1 + \frac{1}{n}$ is as near as damn it to 1. This is pretty close to saying that $\frac{1}{\infty} = 0$. For centuries mathematicians wrestled with this issue, aware that it was very unsatisfactory to argue in this unrigorous fashion because it sometimes led to erroneous results, but reluctant to give it up because it often worked. It was not until the early 1800s that a satisfactory definition evolved. This definition carefully avoided all mention of infinity, infinitely large numbers and infinitesimals. So what is this mysterious definition?

As applied to the sequence of numbers

$$1+\frac{1}{1},\quad 1+\frac{1}{2},\quad 1+\frac{1}{3},\quad \ldots,\quad 1+\frac{1}{n},\quad \ldots$$

it defines the meaning of

$$\text{“}1+\tfrac{1}{n}\to 1 \quad\text{as}\quad n\to\infty\text{”}$$

as

> "For every $\epsilon > 0$, there exists a positive Integer N such that, for every Integer n greater than N, $1+\frac{1}{n}$ lies within ϵ of 1".

The intent is that ϵ is very small and then the corresponding N is likely to be very large. The Greek letter ϵ is often used in mathematics to denote a positive Real Number when the focus is on small values of that number. By allowing arbitrarily small positive values for ϵ we avoid speaking of infinitesimals, and by allowing arbitrarily large Integers N we avoid speaking of infinite numbers.

In this simple example we can take N to be the first Integer greater than $\frac{1}{\epsilon}$, and here is the proof. If $n > N$ and $N > \frac{1}{\epsilon}$ then $n > \frac{1}{\epsilon}$. Hence $\epsilon > \frac{1}{n}$ and consequently $1 < 1+\frac{1}{n} < 1+\epsilon$, so that $1+\frac{1}{n}$ lies within ϵ of 1.

As you can see, this is a rather complicated definition because it involves three *quantifiers*, namely the words "every", "exists" and a second occurrence of "every", in that order. Working with such a definition, and this is a relatively easy one, can be hard work. In trying to get results, mathematicians still find themselves thinking of "big" and "small" numbers without (possibly) crossing the border into the realms of the infinite.

Most people, including mathematicians, have an urge to argue informally that if n is infinitely large then $\frac{1}{n}$ is infinitesimally small, and so the limiting value of $1+\frac{1}{n}$ (to the nearest Real Number) is 1. But unfortunately **R** does not contain any infinite numbers, nor any infinitesimal numbers. However, this did not stop renowned mathematicians of the calibre of Newton and Leibniz from using such concepts very frequently, although with a certain amount of care.

Newton and Leibniz are credited with the invention of calculus at the end of the 1600s. If you don't know what calculus is, don't worry. Just accept that it was a really fundamental development in mathematics that has had enormous practical applications. Newton and Leibniz had some precursors, several contemporaries and many successors, and almost without exception they all used similar arguments. The justification was that it worked, although people realized that you needed to be careful. If you do know something about calculus, you might like to look at the optional section 11.6 on page 101 before moving on to the next section.

The definition we outlined above and its generalizations to other limiting processes do provide a rigorous basis for the results of Newton, Leibniz, and their successors. But still people hankered after a way of putting infinite numbers and infinitesimals onto a firm footing, primarily because arguments involving such quantities are often more intuitive and look simpler. Over the years there were several attempts to extend **R** to include such quantities, but all proved rather artificial and unsatisfactory until the work of Abraham Robinson around about 1960. Robinson's theory was more general than extending the Real Number system and he called his creation *Non-Standard Analysis*.

11.2 Construction of the Hyperreal Numbers

We have seen in rough outline how one set of numbers may be constructed from another. All the constructions can be made to depend on set theory. Thus we can regard the set of Real Numbers **R** as a *constructed object*, constructed ultimately from $\emptyset$. On the other hand we have listed a *collection of axioms* for **R** and remarked that these axioms allow us to deduce all the known and much loved properties of **R**.

Merely listing desirable axioms does not ensure that they are consistent or that anything satisfies them. The constructed version of **R** ensures that the axioms aren't vacuous (at any rate if we assume that set theory is consistent). As explained in earlier chapters, we say that the constructed object is a *model* for the axioms. In general, if we have any set of sentences and a mathematical object which obeys the sentences (with suitable interpretations of the terms in the sentences) then we say that the object is a model for the sentences.

A rather startling result in mathematical logic asserts that if we have an *infinite* set of sentences (of a suitable form) each *finite* subset of which has a model, then the whole collection has a model. This is known as the compactness theorem. There are some restrictions on the nature of the sentences but they are beyond the scope of this current discussion.

To see how the compactness theorem works, take axioms 1 to 10 from page 57 (these were the first ten axioms for **R**) and, for $n = 1, 2, 3, \ldots$, add the sentences

$$\text{P}(n)\text{: “There is a number } x \text{ satisfying } 0 < x < \tfrac{1}{n}\text{”}$$

Any finite collection of these sentences has a model. Actually, **R** itself provides a model because it certainly obeys axioms 1 to 10, and if N is the largest n in our finite subset of the sentences $\mathrm{P}(n)$, we may take x to be $\frac{1}{N+1}$.

It is much less clear that there is a model which satisfies axioms 1 to 10 and $\mathrm{P}(n)$ for all n, because this implies the existence of a positive "number" less than any positive Real Number (so **R** is *not* a model). However, the compactness theorem asserts that a model does exist - i.e. we can find a mathematical object (we may as well call it a set of numbers) which contains *infinitesimal numbers*. It is possible to specify the construction of the new model directly from **R**, although the level of difficulty in obtaining such a model is considerably greater than any of our previous constructions using things like ordered pairs and Dedekind cuts. If you are anxious to know the name of this construction, it is called an *ultraproduct*; it even sounds hard. The numbers in the enlarged set are called Hyperreal Numbers and the set of these is denoted by $^*\mathbf{R}$.

Informally we can think of zero as being surrounded by infinitesimally small (but non-zero) numbers, both positive and negative, all of which are smaller in size than any positive Real Number. If we add, say, 2 to each of these infinitesimals then we obtain a "cloud" of new numbers infinitesimally close to 2 (and likewise for any other Real Number). The reciprocal of a non-zero infinitesimal is an infinite number which we can think of as lying beyond all the "standard" Real Numbers. The system $^*\mathbf{R}$ can be constructed in such a way that it obeys all the axioms for **R** except the completeness axiom. This latter omission is due to the fact that the completeness axiom cannot be expressed in a sentence of the form required for the compactness theorem. Broadly speaking this is because the completeness axiom relates to *sets* of numbers, whereas all the other axioms deal with properties of individual numbers.

11.3 Properties of the Hyperreal Numbers

To be a little bit more formal, we can summarize the properties of the Hyperreal Numbers $^*\mathbf{R}$ starting with the similarities with the set of Real Numbers **R**.

Firstly $^*\mathbf{R}$ obeys axioms 1 to 10 given for **R** on page 57. Secondly, it contains a copy of **R**, so that all the familiar Real Numbers appear in $^*\mathbf{R}$. The addition (+) and multiplication ($\times$) operations and the order relation ($<$) in $^*\mathbf{R}$ are consistent with those in **R**. So, for example, $1 + 2 = 3$ in **R** and also in $^*\mathbf{R}$, and $-1 < 2$ in **R** and also in $^*\mathbf{R}$.

Every mathematical function defined on $\mathbf{R}$ has a natural extension to ${}^*\mathbf{R}$ that is consistent with its definition on $\mathbf{R}$. For example, $\sin(90^o) = 1$ and $\sqrt{4} = 2$ in both $\mathbf{R}$ and ${}^*\mathbf{R}$. Important subsets of $\mathbf{R}$ have a natural extension in ${}^*\mathbf{R}$. An example of this is the set of Natural Numbers $\mathbf{N}$ which has a natural extension to an enlarged set ${}^*\mathbf{N}$ containing infinite integers.

Of course there are important differences between $\mathbf{R}$ and ${}^*\mathbf{R}$ (the whole purpose of creating the latter). As already remarked, ${}^*\mathbf{R}$ does not obey the completeness axiom (axiom 11 in our collection of axioms for $\mathbf{R}$ given on page 57). However this "disadvantage" is balanced by the existence of numbers in ${}^*\mathbf{R}$ that do not lie in $\mathbf{R}$. In particular, there exist positive numbers ϵ in ${}^*\mathbf{R}$ that satisfy $0 < \epsilon < r$ for every $r \in \mathbf{R}$ with $r > 0$, i.e. these numbers are strictly less than every positive Real Number, but greater than zero. We call such a number a *positive infinitesimal.*

If ϵ is a positive infinitesimal, then its additive inverse $-\epsilon$ satisfies $r < -\epsilon < 0$ for every $r \in \mathbf{R}$ with $r < 0$, and such a number is called a *negative infinitesimal.* The positive and negative infinitesimals, along with zero, are simply called *infinitesimals.*

A non-zero infinitesimal ϵ has a multiplicative inverse (reciprocal) ϵ^{-1} (or $\frac{1}{\epsilon}$) and such an element of ${}^*\mathbf{R}$ is said to be an *infinite* number, positive or negative according to whether ϵ is positive or negative. But you still cannot divide by zero in ${}^*\mathbf{R}$ (!)

We say that two elements of ${}^*\mathbf{R}$, say x and y, are infinitesimally close if and only if $x - y$ is infinitesimal, and we can write this as $x \approx y$. It is easy to see that $x \approx x$, and that $x \approx y$ if and only if $y \approx x$. Furthermore, if $x \approx y$ and $y \approx z$, then $x \approx z$. These properties are expressed by saying that $\approx$ is an *equivalence relation* on ${}^*\mathbf{R}$. If $x \in \mathbf{R}$ (i.e. if x is a Real Number) and $y \approx x$, then we say that x is the *standard part* of y and that y is *finite.*

Because ${}^*\mathbf{R}$ is so much larger than $\mathbf{R}$ we do not have any easy representation of its elements. There is no simple analogue of the decimal expansion of a Real Number. The use of the Hyperreal Numbers is in the simplification they provide for theoretical arguments, not in their use for measuring or even for mathematical precision. The great advantage of ${}^*\mathbf{R}$ is that it enables us to carry out limiting processes in a rigorous but intuitive manner.

11.4 Convergence of sequences

In our example in which we wished to prove that $1 + \frac{1}{n} \to 1$ as $n \to \infty$, we applied the definition mentioned at the start of this chapter. We took an arbitrary positive number ϵ (we are really interested in "small" values of ϵ, but we don't say this because we can't specify how small), and we claimed the existence of a corresponding positive integer N (that depends on ϵ), such that for every integer $n > N$, the number $1 + \frac{1}{n}$ is within ϵ of 1, i.e. $1 + \frac{1}{n}$ lies between $1 - \epsilon$ and $1 + \epsilon$. To prove the result we had to specify N. In this case, we took N to be the first integer greater than $\frac{1}{\epsilon}$. To repeat what we did, if $N > \frac{1}{\epsilon}$ and $n > N$, then $n > \frac{1}{\epsilon}$ and consequently $\epsilon > \frac{1}{n}$. Hence $1 < 1 + \frac{1}{n} < 1 + \epsilon$, so that $1 + \frac{1}{n}$ certainly lies between $1 - \epsilon$ and $1 + \epsilon$.

More generally, if we have an infinite sequence of Real Numbers, $s_1, s_2, s_3, \ldots$, then to prove that the sequence (s_n) converges to some limit $\ell \in \mathbf{R}$, we must show that for every $\epsilon > 0$, there exists a positive Integer N, such that for every Integer $n > N$, s_n lies between $\ell - \epsilon$ and $\ell + \epsilon$. If this is the case we write $s_n \to \ell$ as $n \to \infty$. You can see that this definition of convergence is cumbersome and not exactly "natural".

In the set of Hyperreal Numbers $^*\mathbf{R}$, a more natural definition is possible: the infinite sequence (s_n) converges to the limit $\ell \in \mathbf{R}$ if and only if $s_n - \ell$ is infinitesimal when n is infinite.

11.5 Leibniz' transfer principle

The purpose of constructing the Hyperreal Numbers $^*\mathbf{R}$ is to enable discussion of complicated concepts, such as convergence of sequences, in a more intuitive manner than is possible if we are restricted to the Real Numbers $\mathbf{R}$.

Clearly the whole purpose of creating and using $^*\mathbf{R}$ would be frustrated if the "natural" definition of convergence of a sequence in $^*\mathbf{R}$ was not equivalent to the more time-honoured but complicated definition of convergence in $\mathbf{R}$. What is therefore required is a principle in mathematical logic which says that if we take a sequence of Real Numbers, form its natural extension to a sequence of Hyperreal Numbers, and prove that this converges in $^*\mathbf{R}$, then the original sequence converges in $\mathbf{R}$. And, conversely, convergence of the original sequence in $\mathbf{R}$ ensures convergence of its natural extension in $^*\mathbf{R}$.

Convergence of sequences is only one example of several related and even more complicated concepts such as continuity and differentiability, and of processes such as integration. So the principle required is quite general. It has to transfer results between the contexts of $\mathbf{R}$ and $^*\mathbf{R}$. The precise formulation of this transfer principle was the key step in Robinson's creation of the Hyperreal Numbers

Leibniz clearly had such a principle in his mind 300 years ago, but he was unable to formulate it precisely. To give a precise formulation of the transfer principle is also beyond the scope of our discussion in this book. It requires a lengthy treatment of aspects of mathematical logic involving formal mathematical "languages". However, a precise formulation is possible and it does exactly what we would hope. It enables the transfer of results between formulations in $\mathbf{R}$ and their "natural" extensions to formulations in $^*\mathbf{R}$.

Very briefly, an artificial language $\mathcal{L}$ is constructed, a reduced form of natural languages, that enables restricted discussion of some (but not all) properties of $\mathbf{R}$, but including axioms 1 to 10. Statements in $\mathcal{L}$ may be true or may be false when interpreted as statements about $\mathbf{R}$. If we take such a statement in $\mathcal{L}$ and replace each reference to an object from $\mathbf{R}$ by a reference to its natural extension in $^*\mathbf{R}$, we obtain a statement about $^*\mathbf{R}$. The transfer principle asserts that either both of these statement are true or they are both false. Since $^*\mathbf{R}$ is not the same as $\mathbf{R}$, it is clear that there must be some statements that are true about one of these sets that are not true about the other; such statements do not lie in $\mathcal{L}$ and that is why this restricted language has to be carefully designed and carefully used. Of course if you want to use the Hyperreal Numbers to prove results about the Real Numbers then you do have to come to grips with the details.

Leibniz would be delighted to know that his intuition of a transfer principle has eventually been put onto a sound basis, even though it did take around 300 years.

11.6 Calculus (optional)

This section is addressed to people who already know something about calculus, including differentiation and integration. Differentiation concerns rates of change that are often represented graphically as determining the gradient of a graph at a given point. Suppose that y is a function of some variable quantity x; we may write this as $y = f(x)$. For example, we might have $y = x^2$, i.e. $f(x) = x^2$. In the general case, to estimate the gradient of the graph at an arbitrary point x we examine a small portion of the graph around this value x. This is illustrated in Figure 11.1.

The gradient of the dotted line PQ on the graph is given by $\frac{\delta y}{\delta x}$ where δy is the change in the y-coordinate corresponding to the change δx in the x-coordinate. Note that both δx and δy are composite symbols denoting the changes in x and y; so δx does not mean $\delta \times x$ and δy does not mean $\delta \times y$.

One might reasonably hope that as the size of δx is reduced towards zero, the gradient of the dotted line PQ will approach that of the graph itself at

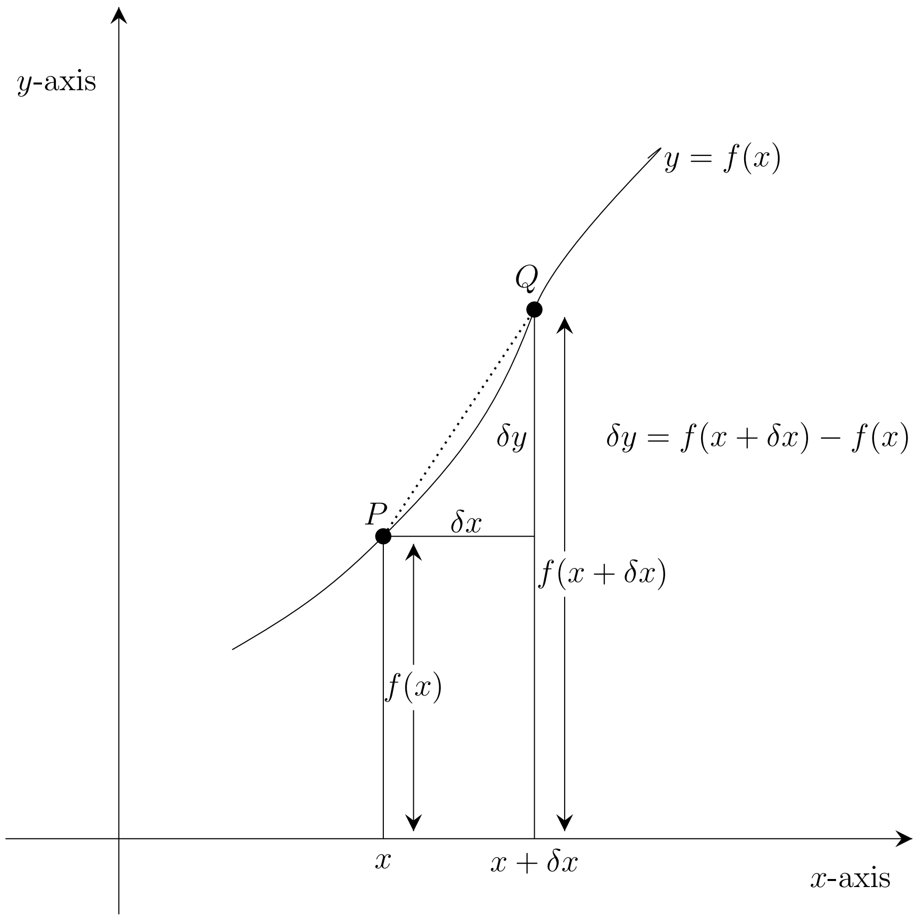


Figure 11.1: Determining the gradient of $y = f(x)$

the point P. So we are led to consider the value of

$$\lim_{\delta x \to 0} \frac{\delta y}{\delta x}, \quad \text{i.e.} \quad \lim_{\delta x \to 0} \frac{f(x + \delta x) - f(x)}{\delta x}.$$

The value of this limit, assuming it exists, is called the derivative of $f(x)$ and may be denoted by $f'(x)$. Figure 11.1 shows $\delta x > 0$, but this is just accidental. The only value of δx that we must not use is zero, for the simple reason that we cannot divide by zero. In Leibniz' notation the limiting value of the quotient $\frac{\delta y}{\delta x}$ is denoted by $\frac{dy}{dx}$. In other words,

$$\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x} = \lim_{\delta x \to 0} \frac{f(x + \delta x) - f(x)}{\delta x} = f'(x).$$

As an example, we will see how this all works for the case when $f(x) = x^2$. So we are considering the graph of $y = x^2$ and trying to get an expression for the gradient at an arbitrary point x. We have $f(x+\delta x) = (x+\delta x)^2 = x^2 + 2x(\delta x) + (\delta x)^2$. So $\delta y = f(x+\delta x) - f(x) = x^2 + 2x(\delta x) + (\delta x)^2 - x^x = 2x(\delta x) + (\delta x)^2$. Hence

$$\frac{\delta y}{\delta x} = \frac{f(x+\delta x) - f(x)}{\delta x} = 2x + \delta x.$$

The limiting value of this as δx tends to zero is clearly $2x$. So we conclude that if $y = x^2$ then $\frac{dy}{dx} = 2x$. In the alternative notation, if $f(x) = x^2$ then $f'(x) = 2x$. The gradient of the graph $y = x^2$ at an arbitrary point x is simply $2x$.

Returning to the general case $y = f(x)$, when you were first introduced to differentiation at school or college it should have been hammered into you (although you may have forgotten) that $\frac{dy}{dx}$ is a single composite symbol, like δy and δx. There are not two separate Real Numbers dy and dx, one of which is divided by the other. So, despite the obvious temptation, dx is *not* the limit of δx as δx tends to zero because, if it were, then it would itself be zero and we could not divide by it. So dx has no separate existence as a Real Number, and neither does dy. If you try to insist that both are zero you get stuck with the undefined (and undefinable) ghastly expression $\frac{0}{0}$.

However, having emphasized that point, it is undeniable that in some ways dy and dx do appear to behave as if they had some separate shadowy non-zero existence. Leibniz clearly saw them as infinitesimal numbers, smaller in size than any non-zero Real Number, but nevertheless non-zero themselves. This is really brought home to your attention when you want to reverse the differentiation process by integration. If we are given a graph whose gradient at an arbitrary point x is $2x$, we can determine the original function up to an arbitrary constant. It is wholly "natural" to start the process by writing (or at least thinking of) $\frac{dy}{dx} = 2x$ in the form $dy = 2x\,dx$, before wrapping integral signs around both sides to get

$$\int dy = \int 2x\,dx.$$

Assuming you know a bit about integration, this gives $y = x^2 + c$, where c is an unknown constant. Although some purists might react with horror at splitting dy from dx in the composite symbol $\frac{dy}{dx}$ (at least, in public), this is what most people, including mathematicians, do in practice.

If we allow ourselves to replace the Real Numbers δx and δy by infinitesimals dx and dy, we can think of the derivative $f'(x)$ as the closest real number to (i.e. the standard part of) the quotient $\frac{dy}{dx}$: genuine division! Leibniz saw it worked even though he couldn't give a precise explanation. The reason we still use his notation today is a testament to its effectiveness.

Here then is how to find the derivative of $f(x) = x^2$ at an arbitrary point $x \in \mathbf{R}$ using infinitesimals. Put $y = x^2$ and let dx be any non-zero infinitesimal. Put $dy = (x + dx)^2 - x^2 = x^2 + 2x(dx) + (dx)^2 - x^2 = 2x(dx) + (dx)^2$. Then

$$\frac{dy}{dx} = 2x + dx,$$

and now this is genuine division of the infinitesimal dy by the infinitesimal dx. The quantity $2x + dx$ has standard part $2x$, i.e. the Real Number $2x$ is the closest Real Number to $2x+dx$ when dx is infinitesimal. So the derivative of $f(x) = x^2$ is $f'(x) = 2x$.

The general process is similar for an arbitrary function f at some point x. Take an arbitrary non-zero infinitesimal dx. Compute $dy = f(x+dx)-f(x)$. If dy is not infinitesimal then f is not differentiable at x. If dy is infinitesimal, compute the quotient $\frac{dy}{dx}$. If the standard part of this quotient (i.e. the closest Real Number to it) exists and does not depend on dx, then f is differentiable at the point x and that standard part is the derivative $f'(x)$.

As a further illustration of the use of infinitesimals in calculus, we consider the chain rule. So suppose that $z = g(y)$ and that $y = f(x)$ are two functions that we determine to be differentiable by computing $\frac{dz}{dy}$ and $\frac{dy}{dx}$ using, in the first case, an arbitrary non-zero infinitesimal dy, and in the second case an arbitrary non-zero infinitesimal dx. We can combine the two functions by substituting $f(x)$ for y in $z = g(y)$. This gives $z = g(f(x)) = h(x)$, say. Thus z is a function of x (i.e. z depends on x) and we can think about determining $\frac{dz}{dx}$, in other words $h'(x)$.

The obvious thing to do is to multiply the quotients $\frac{dz}{dy}$ and $\frac{dy}{dx}$, cancel out the dy terms and get $\frac{dz}{dx}$. This is OK so long as the dy generated by the dx and then fed into $z = g(y)$ isn't zero. But in such a case dz will be zero and so $\frac{dz}{dx}$ is itself zero. The standard part of $\frac{dy}{dx}$ is just $f'(x)$, the standard part of $\frac{dz}{dy}$ is $g'(y)$, so the standard part of $\frac{dz}{dx}$ is $g'(y)f'(x) = g'(f(x))f'(x)$. In other words, $h'(x) = g'(f(x))f'(x)$ or, if you prefer Leibniz' notation, $\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$ where these derivatives are now the familiar ones defined in $\mathbf{R}$, so we are not actually cancelling the dy terms even though we know that can normally be done using infinitesimals in $^*\mathbf{R}$ (with just a slight fudge when the infinitesimal dy generated by dx is zero).

You can develop the whole of calculus using the Hyperreal Numbers, and this approach does lead to some presentational simplicity. Leibniz' notation (dy, dx, etc) is certainly more intuitive than any of the alternatives, as evidenced by the fact that it has persisted for 300 years with only half-hearted attempts to banish it. At least we now know it can be put onto firm foundations.

Chapter 12

Conway's Surreal Numbers (No)

12.1 Overview

Starting with the empty set, $\emptyset$, John Conway's method provides a single construction (a sort of mathematical "Big Bang") that allows us to construct all the Real Numbers $\mathbf{R}$ together with infinite and infinitesimal numbers which continue to obey the usual rules of arithmetic and order (axioms 1 - 10 from page 57).

Conway's construction dates from the 1970s. It has affinities with Dedekind cuts, and with the construction of the Natural Numbers $\mathbf{N}$ and infinite ordinal numbers, both of which we have described in earlier chapters. The numbers produced by Conway's construction were originally simply called Numbers, but subsequent authors have called them *Surreal Numbers*. We will use this latter term to distinguish them from other collections of numbers that we have discussed. The entire collection of Surreal Numbers is denoted by **No**.

Like Dedekind cuts, each Surreal Number is identified with an ordered pair (L, R) of sets of already-constructed Surreal Numbers. The set L is called the left set, and the set R is called the right set. This is like Dedekind cuts, but unlike Dedekind cuts it is not required that all pre-existing Surreal Numbers lie in either L or in R. To get started when we have no pre-existing Surreal Numbers to hand, we must take $L = R = \emptyset$ (the empty set). We'll see how this works in a moment. Roughly speaking (we will be more precise shortly) every element of L must be less than every element of R. This requires a definition of "less than" for **No**. This issue did not arise as a difficulty with Dedekind cuts because the elements of L and R were then

Rational Numbers which came already equipped with "<".

If the restriction that elements of L have to be less than those of R is dropped then many more "objects" are created. Conway called these "Games" and carried out much of his development of Surreal Numbers in terms of Games. Our book is focused on numbers, so we won't cover that aspect here, although it is certainly fascinating.

If we go back for a moment to our discussion of Positive Fractions and Rational Numbers, you might recall that we gave a definition of equality between fractions. We say that the fractions $\frac{m}{n}$ and $\frac{p}{q}$ are equal if and only if the integers mq and pn are equal. So equality between fractions was a defined property, defined in terms of equality between integers. It was necessary because each Rational Number has many different forms, for example $\frac{3}{4} = \frac{6}{8}$. It turns out that each of Conway's Surreal Numbers has many different forms, and so once again we need a definition of equality to determine if two different forms represent the same Surreal Number.

The consequence of the preceding paragraphs is that we require a definition of "less than" ($<$) and a definition of "equals" ($=$) in **No**. Conway's solution to this was to provide a definition of "less than or equals" ($\leq$). Once we have this definition, we can define $=$ by the requirement

$$x = y \text{ if and only if } x \leq y \text{ and } y \leq x.$$

Having then defined $=$, and consequently also $\neq$ ("not equal"), we can define $<$ by the requirement

$$x < y \text{ if and only if } x \leq y \text{ and } x \neq y.$$

Having done all this we can define $x \geq y$ to mean that $y \leq x$, and $x > y$ to mean that $y < x$.
All this hinges on the definition of $x \leq y$.

12.2 Construction of Surreal Numbers (No)

Each Surreal Number x is an ordered pair (L, R), where L and R are themselves sets of Surreal Numbers, and we will write $x = (L, R)$. We denote a typical member of L by x^L and a typical member of R by x^R. We require only that

no x^R is less than or equal to any x^L.

This is the same as saying that every x^L is less than every x^R with the proviso that if either L or R is empty then this condition is considered to be satisfied.

We say that x is less than or equal to y (written as $x \leq y$) if and only if both

(a) **no y^R is less than or equal to x, and**

(b) **y is not less than or equal to any x^L.**

This is the same as saying that x is less than every y^R, and every x^L is less than y, with the proviso that if there is no y^R or no x^L then the corresponding condition is considered satisfied.

These definitions, which cover both "what is a Surreal Number" and what is "$\leq$", are probably as clear as mud at first sight. The intention is that each Surreal Number $x = (L, R)$ is the "simplest" Surreal Number lying between all the x^Ls in L and all the x^Rs in R. It is helpful to think of Surreal Numbers as being formed successively with 0 formed on day 0, with 1 and -1 being formed on day 1, and so on, as we will see.

You need to bear in mind that the Surreal Numbers that are created by Conway's construction, and the order relation $\leq$ cannot be assumed to have any of their normal properties. We will give some of these Surreal Numbers names such as $0, 1, 2, -1, \frac{1}{2}$. This is in anticipation that they will have the usual properties of such numbers. But, for example, we cannot assume that the Surreal Numbers we call 0 and 1 satisfy $0 \leq 1$ unless we prove it by using the definitions of $0, 1$ and $\leq$.

12.3 Construction of $0, 1$ and -1

To form the first Surreal Number we must start with $L = R = \emptyset$ (the empty set) and we call this 0 (zero). Thus $0 = (\emptyset, \emptyset)$. That gets us started with Surreal Numbers and just to be absolutely pedantic we need to get started with $\leq$ by proving that (according to the definition above) $0 \leq 0$. So we observe that

(a) no 0^R is less than or equal to 0 for the simple reason that there is no 0^R because the right set of 0 is empty, and similarly

(b) 0 is not less than or equal to any 0^L because the left set of 0 is also empty.

So yes, indeed, $0 \leq 0$ and consequently $0 = 0$ according to our definition of equality. What a relief! That completes day 0.

Having produced 0 we can consider the possibilities $(\{0\}, \emptyset), (\emptyset, \{0\})$ and $(\{0\}, \{0\})$. The latter object is an example of a Game that is not a Surreal

Number since its only right member (0) is less than or equal to (in fact equal to) its only left member (also 0). However, $(\{0\}, \emptyset)$ and $(\emptyset, \{0\})$ are Surreal Numbers. You can check that both of these satisfy the definition of a Surreal Number. We call them 1 and -1 respectively. These are the Surreal Numbers formed on day 1.

It can be shown that $-1 < 0$ and $0 < 1$, and also that $-1 < 1$. You might think that if $-1 < 0$ and $0 < 1$ then it must follow that $-1 < 1$, however we are working with an entirely new set of "numbers" here and a definition of "$<$" that might not work the way you'd expect. In fact it does work the way you'd expect but that requires a proof. We are not going to prove all three of these inequalities, but here is a proof that $0 < 1$.

To prove $0 < 1$ we must show that $0 \leq 1$, but that it is not the case that $1 \leq 0$, so that $0 \neq 1$, and hence $0 < 1$.

Proving that $0 \leq 1$ means proving that $0 = (\emptyset, \emptyset) \leq (\{0\}, \emptyset) = 1$. So (a) we check there is nothing in the empty set that is less than or equal to 0; clearly this is OK, and (b) we check that 1 is not less than or equal to any member of the empty set; also OK. Hence $0 \leq 1$.

To prove it is not the case that $1 = (\{0\}, \emptyset) \leq 0$, it suffices to observe that condition (b) is violated since there is an element of the left set of 1, namely 0, that is less than or equal to (in fact equal to) 0. It follows that it is not the case that $1 \leq 0$ and consequently $0 < 1$.

If you have followed all that argument, and I could forgive anyone who was confused because it is very pedantic, you will get some inkling why this method of producing numbers is not a good way to introduce numbers to an unsuspecting public. What is needed is a way of automating proofs that ensures the numbers that are produced do in fact obey all the axioms we have come to love. This can be done, but this book is not the place to attempt it. Instead we will henceforth take on trust that the numbers we identify as, for example, 3, $\frac{5}{8}$, $\sqrt{2}$, π, etc. behave as we would expect. So having now got the Surreal Numbers $-1, 0$ and 1, we move on to the construction of further Surreal Numbers.

12.4 Surreal Numbers constructed from $0, 1$ and -1

There are now several possibilities for new Surreal Numbers. For example $(\{0, 1\}, \emptyset)$ qualifies and we'll call this 2. We find that 2 can also be represented as $(\{1\}, \emptyset)$, or as $(\{-1, 1\}, \emptyset)$, or as $(\{-1, 0, 1\}, \emptyset)$. As an example we will check that $(\{0,1\},\emptyset) = (\{1\},\emptyset)$. It helps if we temporarily give these two Surreal Numbers distinct names, say x and y.

First we check that $x = (\{0,1\},\emptyset) \leq (\{1\},\emptyset) = y$. Of the two conditions, (a) is easy because there is no y^R and consequently no y^R less than or equal to x. But (b) is messier as it requires us to prove that $(\{1\},\emptyset)$ is not less than or equal to 0 or to 1, the two elements of the left set of x. However, this is indeed the case because $(\{1\},\emptyset) \leq 0$ would require that 0 is not less than or equal to 1 (but we proved $0 \leq 1$), and likewise $(\{1\},\emptyset) \leq 1$ would require that 1 is not less than or equal to 1 (OK, so we didn't actually prove that $1 \leq 1$, I'll leave that to you). Hence $x \leq y$.

Second we check that $y = (\{1\},\emptyset) \leq (\{0,1\},\emptyset) = x$. Note that the roles of x and y have swapped around. Again condition (a) is easy because there is no x^R and consequently no x^R less than or equal to y. To establish condition (b) we must prove that $(\{0,1\},\emptyset)$ is not less than or equal to 1, the only element of the left set of y. However, $(\{0,1\},\emptyset) \leq 1$ would require that 1 is not less than or equal to either 0 or 1. But you know that $1 \leq 1$ if you did your homework from the previous paragraph, so condition (b) is established. Hence $y \leq x$.

So from the previous two paragraphs we get that $y = x$. In other words the two Surreal Numbers $(\{0,1\},\emptyset)$ and $(\{1\},\emptyset)$ are equal. Both are representations of the same Surreal Number that we called 2.

You might begin to suspect that it is the largest Surreal Number in L and the smallest Surreal Number in R that determines the Number $x = (L,R)$. This is roughly correct. Suppose that $x = (L,R)$ and that we have another Surreal Number $z < x$. Form a new set L' by placing z into L. Then it can be proved that $(L',R) = (L,R) = x$. Similarly, if we have another Surreal Number $y > x$ and we place y into R to form a new set R', then $(L,R') = (L,R) = x$. Although this is phrased as adding extra entries to L and R, it can be used to remove "redundant" entries from L and R.

Other Surreal Numbers that we can get from $-1, 0$ and 1 include $-2 = (\emptyset,\{-1,0\})$, with several other representations. The object $(\{0\},\{1\})$ also qualifies as a Surreal Number and we'll call this $\frac{1}{2}$. This can also be represented as $(\{-1,0\},\{1\})$. Similarly $-\frac{1}{2} = (\{-1\},\{0\}) = (\{-1\},\{0,1\})$. The remaining possibilities for new Surreal Numbers at this stage turn out to be alternative representations for $-1, 0$ and 1. It is possible to prove that

$$-2 < -1 < -\frac{1}{2} < 0 < \frac{1}{2} < 1 < 2.$$

So day 2 saw the construction of the Surreal Numbers $-2, -\frac{1}{2}, \frac{1}{2}, 2$.

12.5 Up to day ω

Moving somewhat faster, on day 3 we construct $3 = (\{0,1,2\},\emptyset) = (\{2\},\emptyset)$, $\frac{3}{2} = (\{1\},\{2\})$, $\frac{3}{4} = (\{\frac{1}{2}\},\{1\})$, $\frac{1}{4} = (\{0\},\ \{\frac{1}{2}\})$, $-\frac{1}{4} = (\{-\frac{1}{2}\},\{0\})$, $-\frac{3}{4} = (\{-1\},\{-\frac{1}{2}\})$, $-\frac{3}{2} = (\{-2\},\{-1\})$ and finally $-3 = (\emptyset,\{-2\}) = (\emptyset,\{0,-1,-2\})$.

On day 4 we construct $4,\ \frac{5}{2},\ \frac{7}{4},\ \frac{5}{4},\ \frac{7}{8},\ \frac{5}{8},\ \frac{3}{8},\ \frac{1}{8}$ along with $-\frac{1}{8},\ -\frac{3}{8},\ -\frac{5}{8},\ -\frac{7}{8},\ -\frac{5}{4},\ -\frac{7}{4},\ -\frac{5}{2},\ -4$. I'll leave you to work out the definitions. Figure 12.1 shows days 0 to 4.

Proceeding in this way on subsequent days we can construct all dyadic rationals (i.e. all fractions of the form $m/2^n,\ m \in \mathbf{Z},\ n \in \mathbf{N}$), which include all the Integers ($\mathbf{Z}$). You might like to try working out on which day a particular dyadic rational will make its first appearance; try the example of 259/16 which can be written as $16\frac{3}{16}$. The integer part of this (i.e. 16,) tells you it can't appear before day 16.

Now suppose that we have reached "day ω". In other words we have completed days $0,1,2,3,\ldots$. So we now have all dyadic rationals at our disposal. Having reached this point we can "fill the gaps" using Dedekind-type cuts of the dyadic rationals to construct all the remaining Real Numbers.

To see how we can do this, think of Real Numbers (as you already know them) expressed not as decimals, but as "binamals", the equivalent of decimals in base-2 notation. Every Real Number has a binamal representation. Dyadic rationals will be those numbers that have a terminating binamal representation. For example, $\frac{3}{8} = \frac{0}{2} + \frac{1}{4} + \frac{1}{8}$ and so is represented as 0.011 in binamal notation. Rational Numbers that are not dyadic rationals will have a recurring (non-terminating) binamal form. For example $\frac{1}{3}$ is represented as $0.010101\ldots = 0.\overline{01}$. Irrational numbers such as $\sqrt{2}$ will have a non-recurring and non-terminating binamal representation. The binamal for $\sqrt{2}$ starts $1.01101010\ldots$, meaning that

$$\sqrt{2} = 1 + \frac{0}{2} + \frac{1}{4} + \frac{1}{8} + \frac{0}{16} + \frac{1}{32} + \frac{0}{64} + \frac{1}{128} + \frac{0}{256} + \ldots.$$

If we truncate this at the n^{th} binamal place, we get a dyadic rational, say x_n less than $\sqrt{2}$, and if we then add $\frac{1}{2^n}$ to x_n we get a dyadic rational greater than $\sqrt{2}$. Thus $\sqrt{2}$ is sandwiched between the dyadic rationals x_n and $x_n + \frac{1}{2^n}$. Since n can be arbitrarily large, and consequently $\frac{1}{2^n}$ an arbitrarily small positive number, we see that $\sqrt{2}$ is the unique Real Number lying between $L = \{$all dyadic rationals less than $\sqrt{2}\}$ and $R = \{$all dyadic rationals greater than $\sqrt{2}\}$, so we define $\sqrt{2}$ in **No** by setting $\sqrt{2} = (L,R)$.

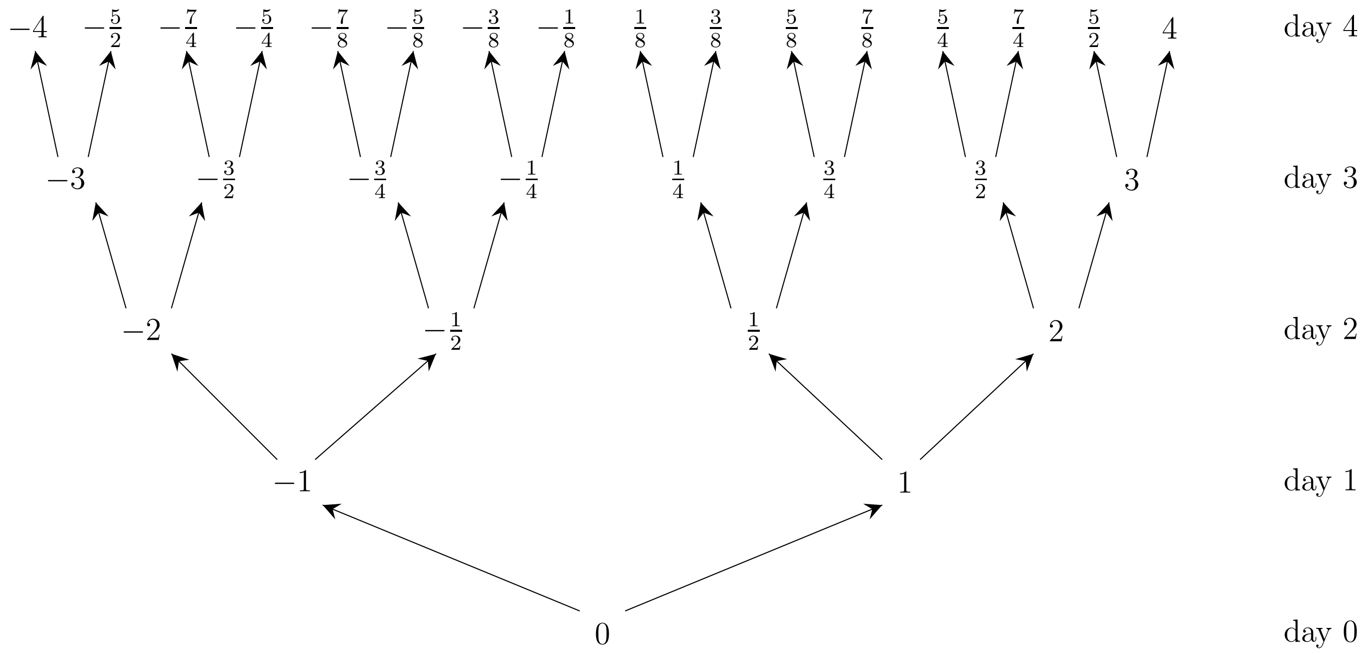


Figure 12.1: The first few Surreal Numbers

We can treat every non-terminating binamal in a similar way. Hence every Real Number not already constructed before day ω (i.e. Real Numbers other than dyadic rationals) can be constructed on day ω by this method. So Conway's construction includes all the Real Numbers.

Our earlier construction of **N** and ordinals are also included in Conway's construction. Back in Chapter 10 we had

$$0 = \emptyset, \quad 1 = \{0\}, \quad 2 = \{0,1\}, \quad 3 = \{0,1,2\}, \ldots, \quad \omega = \{0,1,2,3,\ldots\}.$$

Conway's Surreal Numbers include copies of these:

$$0 = (\emptyset, \emptyset), \quad 1 = (\{0\}, \emptyset), \quad 2 = (\{0,1\}, \emptyset), \quad 3 = (\{0,1,2\}, \emptyset), \quad \ldots$$

and $\omega = (\{0,1,2,3,\ldots\}, \emptyset)$.

However, there is no need to stop there. For example,

$$-\omega = (\emptyset, \{0,-1,-2,-3,\ldots\})$$

is also formed on day ω. Before pursuing this point we need to explain how $+$ and $\times$ are defined.

12.6 Addition and multiplication in No

Suppose that $x = (L, R)$ and $y = (S, D)$ are two Surreal Numbers that we wish to add and to multiply. I've used S and D for the left and right sets of y; S is for sinister and D is for dexter, the Latin words for left and right. We'll deal first with addition.

12.6.1 Addition

If x^L is any element of L and if x^R is any element of R then $x^L < x < x^R$. Similarly if $y^L \in S$ and $y^R \in D$ then $y^L < y < y^R$. Let us assume for a moment that the addition we are trying to define behaves like normal addition. Then adding y to all the terms in the former inequalities and adding x to those in the latter, we see that $x + y$ is greater than both $x + y^L$ and $x^L + y$, and is less than both $x + y^R$ and $x^R + y$. This is all the information that we have about $x + y$. We therefore *define*

$$x + y = (\{x + y^L,\ x^L + y\}, \{x + y^R,\ x^R + y\}).$$

In this definition the left set $\{x + y^L,\ x^L + y\}$ means the set containing every Surreal Number $x + y^L$ for every $y^L \in S$, and every Surreal Number $x^L + y$

for every $x^L \in L$. The right set $\{x + y^R,\ x^R + y\}$ is interpreted in the same way. There is a slight gloss to this definition; when $L = \emptyset$ there is no x^L and so you don't include $x^L + y$, and similarly terms must be omitted whenever R, S or D are empty.

This definition of addition (+) probably looks very puzzling because it appears to be self-referential in that we have $x+y$ on the left defined in terms of things like $x+y^L$ on the right. The point is that $x+y^L$ is in some sense an easier addition than $x + y$ because (hopefully) y^L is an easier number than y. In fact all the definitions, including the definition of a Surreal Number and of $\leq$ have this self-referential appearance. They rely on the fact that on each "day" the constructed Surreal Numbers depend on those constructed on earlier days. So ultimately everything rests on the empty set. Consequently the definitions look self-referential but they are not completely so since we can go back to the start on day 0.

Let's see how this works for some easy additions. First we try $1 + 0$. We have $1 = (\{0\}, \emptyset)$ and $0 = (\emptyset, \emptyset)$. So $1 + 0 = (\{0 + 0\}, \emptyset)$. To resolve this we must decide what $0+0$ is, and that is easy since there are no left and no right members of 0, so $0 + 0 = 0$. Hence $1 + 0 = (\{0\}, \emptyset) = 1$. In a similar way we can prove that $0 + 1 = 1$, so that at least for 0 and 1, addition is commutative.

Now let's get more ambitious and try $1 + 1 = (\{1 + 0, 0 + 1\}, \emptyset)$. Since we just proved that $1 + 0 = 1$ and allowing that $0 + 1 = 1$, this gives $1 + 1 = (\{1\}, \emptyset) = 2$.

You can try a few more if you wish. They are rather tedious but also amusing in a convoluted way. We are not going to prove anything in general here, but this can be done and addition works just like you would expect. In particular, it is commutative ($x + y = y + x$) and associative ($(x + y) + z = x + (y + z)$) with 0 as an additive identity ($0 + x = x$). Moreover addition interacts as you would expect with the order relation, so that if $x > 0$ and $y > 0$ than $x + y > 0$.

Before we leave addition we will mention subtraction. First we define $-x$ for $x = (L, R)$ by setting

$$-x = (-R, -L),$$

where $-R$ denotes the set of all $-x^R$ for $x^R \in R$, and $-L$ denotes the set of all $-y^L$ for $y^L \in L$. Let's check that $-0 = 0$. Since $0 = (\emptyset, \emptyset)$, $-0 = (\emptyset, \emptyset)$. So yes, $-0 = 0$. Slightly more convincingly, we can check that -1 from the definition does correspond with the -1 we had previously. Since $1 = (\{0\}, \emptyset)$, the definition gives $-1 = (\emptyset, \{-0\}) = (\emptyset, \{0\})$, which is the -1 we had previously.

Taking on trust that $-x$ works for every Surreal Number x as you might expect, we define $x - y$ as $x + (-y)$ for any Surreal Numbers x and y. If you have the patience you can prove things like $2 - 2 = 0$. It is possible to use the method of induction to prove that $x - x = 0$ for every Surreal Number x (so $-x$ is the additive inverse of x). If you've never heard of induction, suffice to say that it is a logical principle that enables us to deduce results in an automated step-by-step fashion. It is rather like a ladder from the ground upwards; first you prove that you can get on the bottom step and then you prove that, on the assumption you are on any particular step, you can ascend to the next step. The conclusion is that you can get to any step. All we have done in our examples is to look at the bottom few steps.

12.6.2 Multiplication

Multiplication is a little more tricky than addition. With the same notation as for addition, $x^L < x < x^R$ and $y^L < y < y^R$. In particular $x - x^L > 0$ and $y - y^L > 0$, so multiplying these two we'd expect to find that $(x - x^L)(y - y^L) > 0$ and, if multiplication behaves as we would expect, multiplying out these brackets gives $xy - xy^L - x^Ly + x^Ly^L > 0$. So we get $xy > x^Ly + xy^L - x^Ly^L$. There are three other similar inequalities, so altogether we have

$$\begin{array}{lll}
(x - x^L)(y - y^L) > 0 & \text{i.e.} & xy > x^Ly + xy^L - x^Ly^L \\
(x^R - x)(y^R - y) > 0 & \text{i.e.} & xy > x^Ry + xy^R - x^Ry^R \\
(x - x^L)(y^R - y) > 0 & \text{i.e.} & xy < x^Ly + xy^R - x^Ly^R \\
(x^R - x)(y - y^L) > 0 & \text{i.e.} & xy < x^Ry + xy^L - x^Ry^L
\end{array}$$

These four inequalities give us all the information that we have about xy. We therefore *define*

$$\begin{aligned}
xy = (\{&x^Ly + xy^L - x^Ly^L,\ x^Ry + xy^R - x^Ry^R\}, \\
&\{x^Ly + xy^R - x^Ly^R,\ x^Ry + xy^L - x^Ry^L\}).
\end{aligned}$$

Just like with the definition of addition, $x^Ly + xy^L - x^Ly^L$ stands for every such Surreal Number obtained from $x^L \in L$ and $y^L \in S$, and likewise for the other three entries. And, as with addition, there is again a gloss when one (or more) of the sets L, R, S, D is empty, For example, if R is empty, there is no x^R, so the second and fourth of the four inequalities above do not apply, and the corresponding two terms in the definition of xy must then be omitted.

Let us try a few products. First we try $0 \times y$ for any Surreal Number $y = (S, D)$. With $x = 0 = (\emptyset, \emptyset)$, there is no 0^L and no 0^R, so all four terms drop out of the definition of the product and we get $0 \times y = (\emptyset, \emptyset) = 0$, very satisfying!

Next we try $1 \times y$. With $x = 1 = (\{0\}, \emptyset)$ there is only one x^L, namely 0, and no x^R. The definition gives

$$1 \times y = (\{0 \times y + 1 \times y^L - 0 \times y^L\}, \{0 \times y + 1 \times y^R - 0 \times y^R\}).$$

Bearing in mind we have already shown that multiplying by 0 gives 0, this reduces to

$$1 \times y = (\{1 \times y^L\}, \{1 \times y^R\}).$$

This reduces the calculation of $1 \times y$ to the calculation of $1 \times y^L$ and $1 \times y^R$. At this point I am going to do a little bit of "hand waving". If we assume that every y^L (there may be many or none) was constructed on an earlier day than y itself, and the same for every y^R, then by repeating the process we will eventually end up at day 0 with 1×0 which is of course 0 (we proved that properly). And consequently for all successive numbers (those constructed on later days), we have $1 \times y = y$. The argument is not yet watertight, but it can be made so (by induction). By the same technique we can also prove that $y \times 1 = y$. So 1 is a multiplicative identity.

We'll try one more product: 2×2. We have $2 = (\{1\}, \emptyset)$ so with $x = y = (\{1\}, \emptyset)$ we have $x^L = y^L = 1$ and no x^R or y^R. Thus three of the four terms disappear and we get

$$2 \times 2 = (\{1 \times 2 + 2 \times 1 - 1 \times 1\}, \emptyset).$$

With an extra bit of justification that I am omitting, this reduces to

$$2 \times 2 = (\{2 + 2 - 1\}, \emptyset) = (\{3\}, \emptyset) = 4.$$

Just like addition, it is possible to prove that multiplication is commutative ($xy = yx$) and associative ($(xy)z = x(yz)$). Multiplication is also distributive over addition ($x(y + z) = xy + xz$) and as remarked above, 1 is a multiplicative identity ($1 \times x = x$). Moreover $\times$ interacts as you would expect with the order relation, so that if $x > 0$ and $y > 0$ then $xy > 0$.

Before we finished the discussion of addition we mentioned subtraction and this was fairly simple with $-(L, R)$ defined as $(-R, -L)$. It seems natural therefore to conclude our discussion of multiplication by mentioning division. It would suffice to give a definition of the reciprocal $\frac{1}{x}$ for $x = (L, R)$. Indeed it would suffice if we could do this for $x > 0$. But even this turns out to be very messy. It can be done but it is so complicated I don't think it would leave my intended reader any the wiser, so I am not going to give it. You will just have to accept that it is possible to define the multiplicative inverse x^{-1} (or $\frac{1}{x}$ if you prefer that notation) of the Surreal Number x for every $x \neq 0$, and that $x(x^{-1}) = 1$.

If you are struggling with any of this, take heart. John Conway, a professional and highly experienced mathematician of the first rank remarked that it took a year between his finding the definition of multiplication and finding the definition of a reciprocal, and even multiplication itself took "several weeks' hard thought".

12.7 Beyond day ω

Returning now to the construction of Surreal Numbers, we find

$$(\{0, 1, 2, 3, \ldots, \omega\}, \emptyset) = \omega + 1.$$

More interestingly,

$$(\{0, 1, 2, 3, \ldots\}, \{\omega\}) = v, \quad \text{say,}$$

satisfies $v + 1 = \omega$ or $v = \omega - 1$. Similarly,

$$\omega - 2 = (\{0, 1, 2, 3, \ldots\}, \{\omega, \omega - 1\}).$$

Taking this a stage further, we find

$$(\{0, 1, 2, 3, \ldots\}, \{\omega, \omega - 1, \omega - 2, \ldots\})$$

gives $\frac{\omega}{2}$. It can also be shown that

$$\left(\{0, 1, 2, 3, \ldots\}, \left\{\omega, \frac{\omega}{2}, \frac{\omega}{4}, \frac{\omega}{8}, \ldots\right\}\right)$$

is a square root of ω.

At the other extreme, we find

$$\left(\{0\}, \left\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots\right\}\right) = \frac{1}{\omega}.$$

We can go smaller:

$$\left(\{0\}, \left\{\frac{1}{\omega}\right\}\right) = \frac{1}{2\omega},$$

and so on. For example,

$$\left(\{0\}, \left\{\frac{1}{\omega}, \frac{1}{2\omega}, \frac{1}{4\omega}, \dots\right\}\right) = \frac{1}{\omega^2}.$$

We conclude by emphasizing that **No** obeys axioms 1 - 10 (page 57). However, **No** does not satisfy axiom 11 (the completeness axiom). The $+$ and $\times$ operations on infinite numbers like ω are *not* the same as those mentioned in connection with the arithmetic of the infinite ordinals. For example, the product operation defined on the ordinals in Chapter 10 is not commutative because, as we saw there, $2 \times \omega$ was not the same as $\omega \times 2$. But multiplication in **No** is commutative, so it is different from the multiplication of Chapter 10. It is possible to define certain mathematical functions such as the exponential function exp on **No** and to use certain subcollections as models for the hyperreals $^*\mathbf{R}$.

It is also possible to extend **No** by introducing a number $i = \sqrt{-1}$ in the same way that **R** was itself extended to **C**. Thus we can create Surreal Complex Numbers.

Finally I hope to intrigue you with the observation that **No** cannot be a *set*. We discussed what is a set very briefly in Chapter 2 and again with a deeper look in Chapter 9. It is true that all the other collections to which we have given symbolic names, namely **N**, **Z**, **Q**, **R**, **C**, **H** and $^*\mathbf{R}$ are indeed sets. But if **No** were itself a set, it would be the set of numbers that comprises all Surreal Numbers. In that case $X = (\mathbf{No}, \emptyset)$ would itself be a Surreal Number, greater than any Surreal Number in **No**, a contradiction.

Chapter 13

Concluding Remarks

13.1 A review of previous chapters

In this section we give a brief review of what we have done in the previous chapters where we have seen how to build big from small beginnings. We have seen that collections of "numbers" have grown larger over the centuries. Which objects you choose to regard as numbers is probably a matter of taste. Mathematicians have, for the most part, ceased asking what numbers actually are and now concentrate on defining them in terms of simpler objects and on investigating their properties. I hope that I have convinced you that there's a lot more to this than is immediately apparent.

We started by discussing the Natural Numbers, $\mathbf{N} = \{1, 2, 3, \ldots\}$, in Chapter 2. We began in a very informal way, although in the later Chapter 10 we showed that the Natural Numbers (along with zero), namely $0, 1, 2, \ldots$ could be constructed, with zero represented by the empty set $\emptyset$, and then with each successive number being represented as the set formed from all its predecessors. Thus each Natural Number n has an immediate successor $S(n)$. And in Chapter 2 we saw that the idea of successors could be used to define addition of Natural Numbers, and then also their multiplication as repeated addition.

In Chapter 3 we showed that the Positive Fractions, $\mathbf{Q}^+$, can be constructed from ordered pairs of Natural Numbers in the form $\frac{m}{n}$, where the Natural Number m is called the numerator and the Natural Number n is called the denominator. We gave a definition for equality of fractions that relied only on properties of the Natural Numbers, and we went on to define addition and multiplication of Positive Fractions. We also noted that $\mathbf{Q}^+$ contains a copy of the Natural Numbers $\mathbf{N}$ and that $1 = \frac{1}{1}$ is a multiplicative identity, i.e. $1f = f$ for every $f \in \mathbf{Q}^+$. Moreover every Positive Fraction has a multiplicative inverse (reciprocal) that is also a Positive Fraction; if $f = \frac{m}{n}$ then its multiplicative inverse is $f^{-1} = \frac{n}{m}$ because $ff^{-1} = \frac{m}{n} \times \frac{n}{m} = 1$.

In the following chapter, Chapter 4, we constructed the Integers, **Z**, that include all the negative whole numbers and zero. We again used ordered pairs of Natural Numbers but now in a different manner, essentially representing each Integer as the difference between two Natural Numbers in the form $(m \ominus n)$. We gave a definition for equality of Integers that again only relied on properties of the Natural Numbers, and we went on to define addition and multiplication of Integers. We also noted that **Z** contains a copy of the Natural Numbers **N** and that $0 = (1 \ominus 1)$ is an additive identity, i.e. $0 + x = x$ for every $x \in \mathbf{Z}$. Moreover every Integer has an additive inverse (negative) that is also an Integer; if $x = (m \ominus n)$ then its additive inverse is $-x = (n \ominus m)$ because $x + (-x) = (m \ominus n) + (n \ominus m) = (m + n \ominus m + n) = 0$.

In Chapter 5 we put together the ideas from the two previous chapters to create the set of all fractions, the Rational Numbers, **Q**. We described two possible routes to these; using pairs of Positive Fractions $(f \ominus g)$, or using pairs of Integers $\frac{m}{n}$ (with $n \neq 0$). We listed all the arithmetical properties of **Q** and also described the order relation $<$. The set **Q** contains a copy of the Positive Fractions $\mathbf{Q}^+$ and a copy of the Integers **Z**. We showed how to convert a fraction to a decimal, and that every Rational Number gives rise to a recurring or terminating decimal. Then we showed how to convert a recurring or terminating decimal to a Rational Number. But this gave rise to the important observation that most decimals do not represent Rational Numbers.

We identified the set of all decimals as the set of Real Numbers **R** in Chapter 6. Non-recurring, non-terminating decimals are called irrational because they cannot be expressed as the ratio of two Integers. An important example of an irrational number is the square root of 2. Non-terminating decimals, which include all the irrational numbers and recurring decimals, such as that for the fraction $\frac{1}{3}$, correspond to infinite series. This complicates issues such as how to add and multiply irrational numbers. An alternative presentation of the Real Numbers is given by means of Dedekind cuts. Each such cut partitions the set of Rational Numbers **Q** into an ordered pair of sets (L, R), where every $x \in L$ is less than every $y \in R$, and each partition defines a Real Number. This makes definitions of order, addition and multiplication of Real Numbers easier to handle than the infinite series approach.

Within the set of Real Numbers **R** we can solve equations such as $x^2 = 3$, giving the answer that $x = \sqrt{3}$ or $x = -\sqrt{3}$, and both $\sqrt{3}$ and $-\sqrt{3}$ are Real Numbers. But the equation $x^2 = -3$ cannot be solved in **R** because the square of any Real Number cannot be negative. So in Chapter 7 we discussed the extension of **R** to the set of Complex Numbers **C** which contains a number i satisfying the equation $i^2 = -1$. This extension of The Real Numbers is achieved by the now familiar method of considering ordered pairs of such numbers, but now with a carefully designed definition of multiplication of these pairs. Complex Numbers are essentially 2-dimensional, having the form $x + iy$ where x and y are Real Numbers. Such a number can be represented on an Argand diagram by taking x and y as its coordinates. Sometimes i is called an imaginary number, but this is unfortunate because it is no more imaginary than, for example, -1. The set **C** of Complex Numbers is very satisfactory from a mathematical point of view because every n^{th} order polynomial can be factored into precisely n linear factors in **C**.

There are extensions of the Real Numbers to higher dimensions. In moving from **R** to **C** we lose the order properties that apply to **R**. The set of Quaternions **H**, described briefly in Chapter 5, is a 4-dimensional extension and we lose not only the ordering but also the commutativity of multiplication, so that in general xy is not the same as yx for Quaternions x and y.

After the chapters on what we might call ordinary numbers and their close relatives, we then went really BIG in Chapter 9 by describing Cantor's system of Cardinal Numbers. There are different sizes of infinite sets, each characterized by a Cardinal Number, the smallest of which is $\aleph_0$, the Cardinal Number of the set of Natural Numbers. We showed that the cardinal Number $\mathcal{C}$ of the set of Real Numbers is larger than $\aleph_0$. In fact there is a whole hierarchy of infinite Cardinal Numbers of increasing sizes. We then discussed the continuum hypothesis (that there is no Cardinal Number between $\aleph_0$ and $\mathcal{C}$) which has now been shown to be undecidable on the basis of the commonly accepted axioms for set theory.

Chapter 10 introduced Ordinal Numbers. The definition of the finite Ordinal Numbers took us back to the idea of successors that was discussed in Chapter 2 in the context of the Natural Numbers. But the idea behind Ordinal Numbers extends to infinite sets, enabling us to define the first infinite ordinal $\omega = \{0, 1, 2, \ldots\}$ and to give meaning to numbers like $\omega + 1$, and hence to some limited forms of arithmetic.

Chapter 11 saw something of a return to the familiar Real Numbers but augmented with many extra numbers including infinitesimal and infinite numbers. This extension from **R** to the hyperreal Numbers ***R** can be achieved using results from mathematical logic. The central purpose of constructing such an extension is to simplify arguments involving convergence that involve arbitrarily small Real Numbers, and arbitrarily large Real Numbers, replacing them (respectively) with infinitesimal and infinite Hyperreal Numbers.

Finally in Chapter 12 we described Conway's creation of Surreal Numbers **No** that include not only all the Real Numbers but also infinitesimal and infinite numbers with a fully functioning system of order and arithmetic. The construction is startlingly simple, again using ordered pairs of sets, similar to Dedekind cuts. However, proving that the numbers created by this construction can be identified with numbers we recognise and that they have their familiar properties turns out to be rather more complicated. If you got as far as this chapter, you would probably have found it the toughest of them all. (But well worth the challenge!)

In several chapters we have, somewhat light-heartedly, argued that if we ever came into contact with extra-terrestrial visitors who understood the Natural Numbers and some basic ideas about sets, then we could explain to them all the numbers that appear in this book. It would be fascinating to see if they shared our perceptions about numbers. But maybe we are being over-confident. They might be so advanced that they would regard our numbers as little more advanced than we regard tally marks made on a wooden stick or on a cave wall. There could be a lot more to numbers than meets our eyes at the present stage of our development.

13.2 Notes and further reading

This book collects together material from diverse sources. Many difficulties have been glossed-over in my account, which I acknowledge is very incomplete. My aim was just to give you the flavour. If you want to read more about an individual aspect, my first recommendation is that you search for that topic on the internet where Wikipedia is a good first place to look. It is true that some items on the internet are inaccurate or downright false, but in mathematics that seems to be largely limited to people who claim to have proved impossible things or have a short proof of a very hard result. These can usually be detected with ease because they are accompanied by claims that the author has been ignored by the mathematical community who have failed to recognise their genius. However, to get you started on further reading, a few suggestions are given below.

Some of these books do not focus on the construction of numbers, but on their properties and applications. The material on Natural Numbers, Rational Numbers and Real Numbers is covered to varying extents in many

University-level textbooks on Introductory Algebra or Introductory Real Analysis, but it is often dismissed in a few lines and these sets of numbers are then taken as the starting point.

Perhaps the closest to my book in its outlook is the one by Higgins [9] in the Oxford *Very Short Introduction* Series. This gives a good informal treatment outlining various properties of numbers, especially prime numbers and questions about them, including their use in cryptography. It describes certain counting problems, specific types of irrational numbers, continued fractions, and the representation of Complex Numbers by matrices. Its focus is rather more on these aspects than on the construction of numbers *per se.*

However, Körner [13] gives an extensive treatment of Natural Numbers, Rational Numbers, Real Numbers and Complex Numbers with many historical asides. This is done in textbook style with Definitions, Lemmas, Theorems and Exercises. So if you want to read more on these topics, this is as good a place to start as any other. Körner gives the construction of the Real Numbers from Cauchy sequences of Rational Numbers rather than the Dedekind cut approach that we have used. His book also includes material on quaternions as well as on prime numbers, modular arithmetic and finite fields.

Another textbook that covers most of my material in greater detail, is the one by Ebbinghaus et al [7], but this is a further step up in the hierarchy of difficulty. This book gives both the Dedekind cut construction of the Real Numbers as well as the Cauchy sequence construction, and a further construction using nested intervals. Stillwell [17] also gives a detailed account of many aspects but his treatment would be found very hard going by non-mathematicians. Dedekind's original account is given in his book [5].

Perhaps because the Complex Numbers are often not covered at high school, many books on Complex Analysis give a complete treatment of the construction of the Complex Numbers $\mathbf{C}$ from the Real Numbers $\mathbf{R}$. One such is the book by Brown and Churchill [1], but there are many others. These books cover the construction but their real focus is on calculus and applications.

As regards Cardinal and Ordinal Numbers, the interested reader really needs to learn something about set theory. Two reasonably accessible books are those by Devlin [6] and by Machover [14], but there are many more of a similar nature. The book by Cantor [2] is a translation of his original work. For historical reasons I include the book by Gödel [8] and the papers by Cohen [3] in the references because they contain the proof that the continuum hypothesis is independent of the usual axioms of set theory. But it is very unlikely that you would find these readable.

The first book by Keisler [10] develops elementary calculus using the

Hyperreal Numbers $^*\mathbf{R}$ presented axiomatically. The second book by Keisler [11] is the “Instructors Manual” for the first and gives more details of how the Hyperreal Numbers are constructed. Similar material is given in the book by Stroyan and Luxemburg [18]. Robinson’s original development of non-standard analysis is described in his book [16], another tough read.

The book by Conway [4] covers his construction of Surreal Numbers and Games in considerable detail and is rather more accessible. The term “Surreal Numbers” was introduced by Knuth in his book of the same title [12]; unusually for a book with real mathematical content, this is written as a novel, and is an easy read.

You might be tempted to think that a book with a title such as “Number Theory” would be likely to contain the material presented in this book. However, that would be unusual. Books on Number Theory are plentiful, but they usually focus on properties of numbers, and especially prime numbers. We did mention the Riemann hypothesis in passing, which is almost certainly the most famous unsolved problem in mathematics. It was first described by Bernhard Riemann in 1859 and it concerns the distribution of prime numbers amongst the Natural Numbers. The readable book by Mazur and Stein [15] will give you some insight into this topic.

References

[1] Brown, J. W. and Churchill, R. V. **Complex Variables and Applications (9th ed.)**, McGraw-Hill Education, 2013, ISBN:9780073383170.

[2] Cantor, G. **Contributions to the Founding of the Theory of Transfinite Numbers**, Reprint of the 1915 translation, Dover Publications, 1955, ISBN:9780486600451.

[3] Cohen, P. J. *The Independence of the Continuum Hypothesis*, Proc. Nat. Acad. Sci. (USA), 50, 1963, p1143–1148 and 51, 1964, p105–110.

[4] Conway, J. H. **On Numbers and Games**, A. K. Peters/CRC Press, 2000, ISBN:9781568811277.

[5] Dedekind, R. **Essays on the Theory of Numbers**, Reprint of the 1901 translation, Dover Publications, 1963, ISBN:9780486210100.

[6] Devlin, K. **The Joy of Sets -Fundamentals of Contemporary Set Theory (2nd ed.)**, Springer, 1993, ISBN:9780387940946.

[7] Ebbinghaus, H.-D., Hermes, H., Hirzebruch, F., Koecher, M., Mainzer, K., Neukirch, J., Prestel, A. and Remmert, R. **Numbers**, *Graduate Texts in Mathematics*, Springer, 1991, ISBN:9781461210054.

[8] Gödel, K. **The Consistency of the Continuum Hypothesis**, Princeton University Press, 1940, ISBN:9780691079271.

[9] Higgins, P. M. **Numbers: A Very Short Introduction**, Oxford University Press, 2011, ISBN:9780199584055.

[10] Keisler, H. J. **Elementary Calculus: An Infinitesimal Approach (2nd ed.)**, Dover Publications, 2011, ISBN:9780486484525. Also available in pdf format for free download from the author at `https://people.math.wisc.edu/ keisler/calc.html`

[11] Keisler, H. J. (1976), **Foundations of Infinitesimal Calculus**, Prindle Weber & Schmidt, 1976, ISBN:9780871502155. Also available in pdf format for free download from the author at `https://people.math.wisc.edu/ keisler/foundations.html`

[12] Knuth, D. E. **Surreal Numbers**, Addison-Wesley, 1974, ISBN:9780201038125.

[13] Körner, T. W. **Where do Numbers come from?**, Cambridge University Press, 2020, ISBN:9781108738385.

[14] Machover, M. **Set theory, logic and their limitations**, Cambridge University Press, 1996, ISBN:9780521479981.

[15] Mazur, B. and Stein, W. **Prime Numbers and the Riemann Hypothesis**, Cambridge University Press, 2016, ISBN:9781107101920.

[16] Robinson, A. **Non-Standard Analysis**, Princeton University Press (reprint) 1996, ISBN:9780691044903.

[17] Stillwell, J. **Real Numbers**, Springer, 2013, ISBN:9783319015767.

[18] Stroyan, K. D., and Luxemburg, W. A. J. **Introduction to the Theory of Infinitesimals**, Academic Press, 1976, ISBN:9780126741506.

Glossary

- **Argand Diagram** A 2-dimensional representation of Complex Numbers. The number $x + iy$ is represented as the point with Cartesian coordinates (x, y).

- **Argument** (of a Complex Number). See Polar Form.

- **Associative** We say that addition $+$ is associative for a particular collection of numbers if $(x + y) + z = x + (y + z)$ for every triple of numbers x, y and z. Likewise we say multiplication is associative if $(x \times y) \times z = x \times (y \times z)$ for every triple of numbers x, y and z.

- **Axioms** A collection of basic properties that we accept as the foundation for obtaining further results. For example, on page 57 we give eleven axioms for the Real Numbers, from which all their properties may be deduced.

- **Base** As in base-10, base-2; see Number Base.

- **Cardinal Number** A number used to denote the number of elements in a set. The finite Cardinal Numbers are just the Natural Numbers $1, 2, 3, \ldots$, along with 0. The entire set of Natural Numbers, and any set that can be put into one-to-one correspondence with it, is said to have Cardinal Number $\aleph_0$ (aleph-nought). This is the first (lowest) infinite Cardinal Number. The set of all subsets of the Natural Numbers, and the set of Real Numbers each have the larger Cardinal Number $\mathcal{C}$, also denoted as $2^{\aleph_0}$. See also Continuum Hypothesis and Ordinal Number.

- **Commutative** We say that addition $+$ is commutative for a particular collection of numbers if $x + y = y + x$ for every pair of numbers x and y. Likewise we say multiplication is commutative if $x \times y = y \times x$ for every pair of numbers x and y.

- **Complex Numbers** These are numbers formed from a pair of Real Numbers by adjoining a new "number" i that satisfies $i^2 = -1$. A typical Complex Number has the form $x + iy$ where x and y are Real Numbers. The set of all Complex Numbers is denoted by **C** (C for Complex). A Complex Number of the form $x + i0$ may be identified with the Real Number x, so the Complex Numbers include a copy of every Real Number.

- **Conjugate** The complex conjugate of the Complex Number $x + iy$ (where x and y are Real Numbers) is the Complex Number $x - iy$.

- **Continuum Hypothesis** The hypothesis that the Cardinal Number $\mathcal{C}$ (also known as $2^{\aleph_0}$) is the next largest Cardinal Number to $\aleph_0$. In other words, it is the second infinite Cardinal Number. It is now known that this hypothesis can neither be proved nor disproved on the basis of the usual axioms of set theory. See also Cardinal Number.

- **Counting Numbers** See Natural Numbers.

- **Decimal** A Real Number expressed in the base-10 place-value system. Rational Numbers correspond to decimals that recur or terminate. For example, $\frac{1}{3} = 0.3333\cdots = 0.\bar{3}$, where the bar over the 3 indicates that this part of the decimal recurs indefinitely. A terminating decimal such as 0.25 (representing the Rational Number $\frac{1}{4}$) can be regarded as a recurring decimal $0.25\bar{0}$. Most decimals do not recur or terminate and so most decimals represent irrational Real Numbers. See also Irrational Number and Real Numbers.

- **Distributive** In any particular collection of numbers, we say that multiplication ($\times$) is distributive over addition ($+$) if $x \times (y + z) = x \times y + x \times z$ for every triple of numbers x, y and z.

- **Empty Set** The set without any members. It is denoted by $\{\}$ or (more usually) by $\emptyset$.

- **Exponential notation** For a natural number n, x^n denotes the product of n copies of x. For a negative integer n, x^n is taken to be $\frac{1}{x^{-n}}$, and x^0 is defined as 1 for all $x \neq 0$. (0^0 is undefined). The definition is extended to fractional powers. For example if $x > 0$, $x^{\frac{1}{2}}$ is taken to be the (positive) square root of x.

- **Factorial** For a Natural Number n, the number $n!$ (n-factorial) is defined to be the product $1 \times 2 \times 3 \times \ldots \times n$. For example $3! = 1 \times 2 \times 3 = 6$. By convention $0!$ is taken to be 1 (there are good reasons for this).

- **Hyperreal Numbers** These form an extension of the Real Numbers that includes infinite and infinitesimal numbers. The set of Hyperreal Numbers is denoted by $^*\mathbf{R}$. They facilitate more intuitive arguments about convergence than is possible with the Real Numbers alone. The Hyperreal Numbers may be constructed from the Real Numbers using Model Theory and results from Mathematical Logic. See also Leibniz' Transfer Principle, Model, and Standard Part.

- **Identity element** We say that 0 is an additive identity in a particular collection of numbers if $0 + x = x$ for every number x. Likewise 1 is called a multiplicative identity if $1 \times x = x$ for every number x.

- **Integers** These are the whole numbers, positive, negative and zero: $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$. The set of all Integers is denoted by $\mathbf{Z}$ (Z for Zahl).

- **Inverse element** In any particular collection of numbers, we say that $-x$ is the additive inverse of x if $x + (-x) = 0$. Similarly we say that x^{-1} is the multiplicative inverse (or reciprocal) of x ($x \neq 0$) if $x \times (x^{-1}) = 1$.

- **Irrational Number** A Real Number that is not a Rational Number is called an irrational number. See also Decimal, Rational Number, Real Number.

- **Leibniz' Transfer Principle** A logical rule for transferring results about Hyperreal Numbers to corresponding results about Real Numbers, and vice-versa.

- **Model** A structure that obeys a collection of axioms when a suitable interpretation is given to the linguistic terms that appear in the axioms. This ensures that the axioms are consistent (i.e. they are not contradictory) provided that the model itself is sound. A collection of axioms may have more than one model. See also Axioms.

- **Modulus** (of a Complex Number) See Polar Form.

- **Natural Numbers** These are the positive whole numbers: $1, 2, 3, \ldots$. They are sometimes known as the Counting Numbers. The set of all Natural Numbers is denoted by $\mathbf{N}$ (N for Natural).

- **Number Base** The decimal system uses base-10, meaning that the digits in a number count powers of 10. The first digit to the left of the decimal point counts units (10^0), the next to the left counts tens (10^1), the next to the left counts hundreds (10^2), and so on. On the other side of the decimal point we have tenths (10^{-1}), then hundredths (10^{-2}), and so on. The binary system (base-2) is similar but based on powers of 2. See also Place-value system.

- **Ordered pair** A list of two elements with a first member and a second member. For example $(1, 2)$ is an ordered pair with first member 1 and second member 2. Such a pair is generally regarded as distinct from the pair $(2, 1)$. It is also allowed that the two members may be the same as in $(2, 2)$. One use of ordered pairs is representing points on a plane with Cartesian coordinates. Ordered pairs are used in several different ways to build various collections of numbers from "smaller" collections of numbers.

- **Ordinal Number** A number used to denote the position of an element in a list, as in first, second, third, etc. The finite Ordinal Numbers are just the Natural Numbers (although it is convenient to start with 0 rather than 1). So the finite Ordinal Numbers are essentially the same as the finite Cardinal Numbers. But there are infinite Ordinal Numbers, the first of which is usually denoted by ω (omega). The infinite Ordinal Numbers are not the same as the infinite Cardinal Numbers. See also Cardinal Number.

- **Place-value system** Numbers are represented using a small set of symbols in a string where the position of each symbol in the string determines its value. An example is our base-10 system of decimal representation which uses ten symbols $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$. So 1234.56 represents one thousand plus two hundreds plus three tens plus four units plus five tenths plus six hundredths. See also Number Base.

- **Polar Form** The Polar form of the Complex Number $z = x + iy$ is $z = r(\cos\theta + i\sin\theta)$ where $r = \sqrt{x^2 + y^2}$ is the modulus of z and θ is the argument of z given by $\cos\theta = \frac{x}{r}, \sin\theta = \frac{y}{r}$ and $\tan\theta = \frac{y}{x}$.

- **Positive Fractions** These are the numbers that can be represented as $\frac{m}{n}$ where m and n are Natural Numbers. So typical Positive Fractions are $\frac{1}{2}$, $\frac{3}{4}$, $\frac{15}{7}$, $\frac{6}{3}$, etc. The set of all Positive Fractions is denoted by $\mathbf{Q}^+$ (Q for Quotient and + for positive). Since every Natural Number m can be expressed as $\frac{m}{1}$, the Positive Fractions include a copy of every Natural Number.

- **Quaternions** These are numbers formed from a quadruple of Real Numbers by adjoining three new "numbers" i, j, k where $i^2 = j^2 = k^2 = ijk = -1$. A typical Quaternion has the form $a + ib + jc + kd$ where a, b, c and d are Real Numbers.

- **Rational Numbers** These are the numbers that can be represented as $\frac{m}{n}$ where m and n are Integers and n is not zero. They are sometimes known as fractions. Typical Rational Numbers are $-\frac{7}{8}$, $\frac{2}{3}$,$-\frac{4}{2}$,$\frac{0}{5}$, etc. The set of all Rational Numbers is denoted by **Q** (Q for Quotient). Since every Integer m can be expressed as $\frac{m}{1}$, the Rational Numbers include a copy of every Integer.

- **Real Numbers** These are the numbers that can be expressed as decimals. The set of all Real Numbers is denoted by **R** (R for Real). Since every Rational Number can be expressed as a recurring or terminating decimal, the Real Numbers include a copy of every Rational Number. But (in a precise sense) most decimals do not recur or terminate so there are many more Real Numbers than Rational Numbers, and these additional numbers are called irrational numbers. The Real Numbers $\sqrt{2}$, $\sqrt{3}$ and π are examples of irrational numbers.

- **Sequence** An ordered list of objects. For example $(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots)$ is an infinite sequence. Sequences are usually enclosed in round brackets. An infinite sequence of numbers may or may not converge to a limiting value. This particular sequence does converge to the limit 0. Beware that the terms "sequence" and "series" mean different things in math-speak. See also Series.

- **Series** The result of adding together the numbers in a sequence of numbers. For example $1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots$ is an infinite series. Such a series may or may not converge to a limiting value. This particular series does converge to the limit 2. To determine if such a series converges, we examine the sequence formed from its partial sums. In this case the partial sums form the sequence $(1, 1 + \frac{1}{2}, 1 + \frac{1}{2} + \frac{1}{4}, 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots)$ that reduces to the sequence $(1, \frac{3}{2}, \frac{7}{4}, \frac{15}{8}, \ldots)$. Beware that the terms "sequence" and "series" mean different things in math-speak. See also Sequence.

- **Set** The naive definition of a set is as any collection of objects. The objects may be specified by listing, by description, or by a rule that they satisfy. Curly brackets, also known as braces, are used to indicate a set; for example $\{2, 4, 6\}$ denotes the set whose members are the numbers 2, 4 and 6. The symbol $\in$ is used to indicate membership, so $4 \in \{2, 4, 6\}$ is read as "the number 4 is a member of the set $\{2, 4, 6\}$". Members of a set are often called its elements. The naive definition proves unsatisfactory when it involves self-reference as in Russell's paradox. Consequently more precise rules (axioms) are needed to specify what exactly is a set. An example of such a collection of rules is the Zermelo-Fraenkel axiom system.

- **Standard Part** (of a finite Hyperreal Number) The nearest Real Number to a finite Hyperreal Number x is called the standard part of x.

- **Surreal Numbers** These are numbers created using John Conway's construction, as ordered pairs of sets of numbers. The collection of all Surreal Numbers is denoted by **No**. The construction is inductive, as are the definitions of addition and multiplication. Starting with the empty set, the construction produces a collection of numbers that includes all Real Numbers, along with infinite and infinitesimal numbers.

Index