LOGIC and MATHEMATICS

A Review for Non-Mathematicians.

Alun Wyn-jones

Revised March 15, 2014

Logic and Mathematics.

SETS and LOGIC.

Chapter 1.
Introduction and Motivation.

1.1 Aims of the Book.

This text is intended as a review for people who wish to learn enough mathematics to understand some of the more commonly quoted parts of mathematics. It is an extremely condensed introduction to some key mathematical theories, mostly algebraic, that permeate the whole of modern mathematics. The intended readership is adults with an interest in philosophy, and adults who have used mathematics in their studies who desire some understanding of foundations.

I do not recommend this text for younger readers, and I do not recommend that you use this text to learn the subjects of this text. For that I recommend the books by Birkoff and MacLane [BM]. My notation is conventional, but unfortunately conventions differ. Therefore, even if you are already familiar with a subject treated here, you might glance at the section to familiarize yourself with the notation. For subsequent reference, there is a glossary of symbols in an appendix.

I assume that you have already met mathematical sets. I shall review simple set theory, really quickly. A set consists of members or elements. An explicit set is written with its elements enclosed in curly braces, for example: { apple, orange, banana } which is a set of three elements, which is to say, it is a set having three members. Apple and orange are in the set, pear is not. The set { Alice, John, Nell, John } contains three names, not four—repetitions are ignored. See Chapter 2 for a more thorough description of set theory.

1.2 Group Theory.

As a motivation I shall start with an account of one of the most popular subjects in mathematics, Group Theory and its applications. Groups arose originally in the study of solutions to polynomial equations. The famous Évariste Galois[1]derived the basic principles of Group Theory (which he named) to prove that, in general, the roots of polynomials could not be computed using only the extraction of roots and the usual operations of arithmetic (adding, subtracting, multiplying, dividing). This had been previously proved by a Norwegian mathematician, Niels Henrik Abel, but his proof was so specific to the task that there was little prospect of generalizing the result, whereas Galois' approach was the progenitor of vast areas of modern mathematics and theoretical physics.

The advent of Group Theory was overdetermined. The theory also has roots in the Theory of Invariants, Special Functions, and Geometry. Indeed the reason it became a study in its own right was that many results had been proved several times independently by researchers in different fields, published in different journals, until mathematicians woke up to the reality that they were using the same underlying techniques, and deriving the same results, but in different guises. Hence, Group Theory was born as an act of synthesis which placed these common ideas into a single corpus. Axioms were laid down for what constituted a group. There are variations, all equivalent, those which follow are adapted from the text by Rotman [Rot].

1.3 Axioms for a Group.

A set, G, is a group if it is endowed with a binary operator "$\circ$", called the group product, satisfying the following:

(i) **Closure:** If a and b are in the set G, then so is $a \circ b$.

(ii) **Associativity:** If a, b, and c are in the set G, then $a \circ (b \circ c) = (a \circ b) \circ c$.

(iii) **Identity:** There is an element, call it e, in G which satisfies $e \circ x = x$ for every x in G.

(iv) **Inverse:** Given any x in G, there exists x' also in G such that $x' \circ x = e$, where e is an element satisfying axiom *(iii)*.

The above axioms are spare indeed. So I shall expand upon each of them and try to convey what is behind them.

[1] Galois led a tumultuous life. He died as a result of a duel at age 20. See [WG].

Axiom *(i)* There is not much to add to Axiom *(i)* except to note that as a general principle in mathematics, anything not explicitly prohibited is allowed. So in this axiom, there is nothing said about a and b being different members of G, therefore $a = b$ is allowed. So, in particular, this axiom says that if a is any member of G then $a \circ a$ is in G.

Axiom *(ii)* This axiom is rather difficult to explain; it seems somehow pedantic to differentiate $a \circ (b \circ c)$ from $(a \circ b) \circ c$. So I shall provide an example where these are not the same. Let G be the integers (the whole numbers, positive and negative, including zero, which is usually denoted by the special symbol $\mathbb{Z}$), and take the product to be subtraction. So "$\circ$" is "$-$". Take a, b, c to be almost any three numbers, $7, 3, 2$, say.

$$\begin{aligned} 7 \circ (3 \circ 2) &= 7 - (3 - 2) = 6, \quad \text{whereas} \\ (7 \circ 3) \circ 2 &= (7 - 3) - 2 = 2 \end{aligned}$$

The two associations do not give the same answer. So here is an example of a product on a set which satisfies Axiom *(i)* but not Axiom *(ii)*, and therefore is not a group. However, please note that same set, the integers, $\mathbb{Z}$, with addition as the product is a group. So the choice of product is just as important as the choice of the set.

Axiom *(iii)* The element e mentioned in this axiom is a special member of G; it is called a **left identity** element ("left" because it appears on the left of the product in the axiom.) It turns out that it is unique; that is to say, there can be none other in G satisfying these axioms. Later, we shall actually prove this! But wait, there is more! Not only is it unique it also serves as a right identity: $x \circ e = x$ for every x in G! So once we have proved all this we shall drop "left" or "right" and just call e the identity of the group.

Axiom *(iv)* Having defined a left identity element, a natural question is whether the product of two members of the group can produce this identity. This axiom says that not only is this possible, but that given any element there is a second element whose product with the first is the left identity. The second element is called a **left inverse** of the first. It is called a left inverse because to produce the identity it appears on the left of the product with the first element. We shall later show that the second element also serves as a right inverse of the first, so we can call it just an **inverse**. What is more the inverse is unique: there none other whose product with the first gives the identity.

What strikes most people when they first encounter Group Theory is the simplicity and paucity of axioms. Yet the Group Theory so defined is an extensive field, rich in insights, full of beauty, and has powerful application in physics and chemistry.

1.4 Deductions from the Axioms.

Let us now get down to the business of proving some of the things claimed above. We start with a simple result whose significance will not be obvious until we use it to prove our main result. Since this first statement is only required to prove the second, we shall call it a "lemma." Both this lemma and the theorem which follows it are taken from Rotman[Rot].

1.4.1 Lemma

Let G be a group (that is, G satisfies Axioms *(i)*-*(iv)* above). If an element x of G satisfies the equation $x \circ x = x$, then $x = e$, a left identity of G.

Proof. We have only four axioms to work with, so we need to figure out how we can use them to prove the lemma. The trick is find a product of three elements that we can simplify in two different ways using the associativity axiom.

To start with, we assume what we are given, namely, that we have an element x of the group G satisfying $x \circ x = x$.

We first use Axiom *(iv)* which tells us that x has a left inverse, let it be y; so $y \circ x = e$. We consider the product $y \circ (x \circ x)$. Associating it one way we get

$$y \circ (x \circ x) = y \circ x = e$$

Associating it the other way, we get

$$(y \circ x) \circ x \;=\; e \circ x \;=\; x$$

In the last equation we used Axiom *(iii)*. We now use Axion *(ii)* (associativity) to deduce that the two associations are equal. Hence, $x = e$. □ (← This is the modern form of "QED.")

Now we are ready for our theorem. This will actually prove everything I claimed above.

1.4.2 **Theorem** Let G be a group and let e be a left identity of G. Then, e is also a right identity, that is, $g \circ e = g$ for every member g of G. Also, if h is a left inverse of g, then it is also a right inverse; that is, $h \circ g = g \circ h = e$.

Furthermore, e is unique: none other in G satisifies the axioms. Lastly, h is a unique inverse of g: g has no other inverse.

Proof. Let g be any member of G as in the theorem statement. Let h be a left inverse for g (Axiom *(iv)*). So $h \circ g = e$.

Consider $(g \circ h) \circ (g \circ h)$. We shall use the associativity axiom to simplify this product.

$$(g \circ h) \circ (g \circ h) \;=\; g \circ ((h \circ g) \circ h) \;=\; g \circ (e \circ h) \;=\; g \circ h \tag{1}$$

We now use the lemma. The lemma stated that if $x \circ x = x$, then $x = e$. So if we set $x = g \circ h$ in Equation (1) we deduce that $g \circ h = e$. This proves that h is a right inverse as well as a left inverse of g. (See a discussion of this point after the end of the proof.)

Let us now show that e is a right identity. Since g is an arbitrary member of G, we need only show that $g \circ e = g$. We can assume what we have already proved, namely, that h is both a left and right inverse of g. Now $e = h \circ g$, therefore $g \circ e = g \circ (h \circ g) = (g \circ h) \circ g = e \circ g = g$ proving e is also a right identity.

Lastly, we need to show uniqueness. The lemma directly gives us the uniqueness of the identity as follows: let f be another identity, then $f \circ f = f$ by Axiom *(iii)*. But the lemma then says that $f = e$. So e is unique.

We now need to show that g has only one inverse. Suppose g has another inverse i, say, $i \circ g = e$.

$$i \;=\; i \circ e \;=\; i \circ (g \circ h) \;=\; (i \circ g) \circ h \;=\; e \circ h = h \qquad \square$$

Just a reminder: the square indicates the end of the proof.

Some might have felt a little vertigo at the point where Lemma 1.4.1 was invoked after Equation (1). The lemma states that $x \circ x = x$ implies that $x = e$. You might not have expected you could apply this fact when x is a product like $g \circ h$. But, what is not prohibited is allowed, and Lemma 1.4.1 placed no restriction on x, only that it belonged to a group.

1.4.3 **Comments** When we regard the group product as analagous to multiplication, we usually abbreviate repeated products such as $x \circ x \circ x \circ \cdots \circ x$ (x occurring n times) to x^n. The associativity rule (Rule *(ii)*) assures us that x^n is unambiguous. But if associativity does not hold, then powers like x^3 are not well-defined: for example, if we again take product to be subtraction, then x^3 can mean $(x - x) - x$ which equals $-x$, or it could mean $x - (x - x)$ which equals x.

So we have ascribed a well-defined meaning to x^n in groups when n is a positive whole number. Can we extend the definition to negative n? Yes, we can. When the group product is regarded as akin to multiplication it is customary to write the inverse of an element as the element to the power of -1. Thus, the inverse of x would be denoted by x^{-1}. By thus identifying the inverse of x with x^{-1} we naturally have that x^{-n} means $x^{-1} \circ x^{-1} \circ \cdots \circ x^{-1}$ with x^{-1} occurring n times. It is easy to prove that the usual law of exponents works: $x^m \circ x^n = x^{m+n}$, $x^m \circ x^{-n} = x^{-n} \circ x^m = x^{m-n}$, and of course, $x^0 = e$, the identity element.

I have used the circle, "$\circ$" for the group product. This symbol is meant to denote an arbitrary, general operation between two objects. You have all met "variables" in school algebra like x, y, etc. which stood for unknown, numerical quantities. The circle symbol "$\circ$" similarly stands for an unknown operation on two objects. The actual operation will depend on the group. In practice, group theorists only occasionally use this notation, preferring to use the notation most commonly used for the particular group under study. When there is none commonly in use, they generally use juxtaposition for general groups and the plus sign for abelian groups (abelian groups will be defined below).

1.5 Examples of Groups.

As you can imagine from the historical introduction above, there are many examples of groups. Here are a few.

1.5.1 Example: The Integers

Let $G = \mathbb{Z}$ be the integers. The integers consist of zero, the positive and negative whole numbers: $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$. We let $\circ$ be the addition of two integers. That is, we identify "$\circ$" with "+". Then, the unique identity element is 0. The inverse of an integer x is its negation.

$1 \circ 3 = 4 = 3 \circ 1$ because $1 + 3 = 4 = 3 + 1$;

$7 \circ 0$ which means $7 + 0 = 7 = 0 + 7$. That is, $7 \circ 0 = 7 = 0 \circ 7$; So, 0 is the identity of this group;

$5 \circ {}^{-}5 = 0$ the inverse of 5 is ${}^{-}5$ since the two when added give 0 which is the identity.

There are an infinity of integers, so this is an example of an **infinite** group.

In this group, $x \circ y = y \circ x$ is always true (since $x + y = y + x$ for any two integers x and y). The property $x \circ y = y \circ x$ is called **commutivity** of x and y. If we take commutivity of every pair as an additional axiom to those for a group we get what is called an **abelian** group. (Abel, after whom this type of group is named, has been dead long enough that his name is no longer capitalized.)

The integers is the archetypical abelian group, and as such the group operation in general abelian groups is usually written "+" instead of "$\circ$," and the group inverse is usually written as "$-$" (minus).

1.5.2 Example: Rotations of a Triangle in Space.

Let G be rotations of a rigid, equilateral, triangular hoop in three dimensions which return the hoop to its original position. See Diagram 1.5.2 below for a pictorial representation of this group.

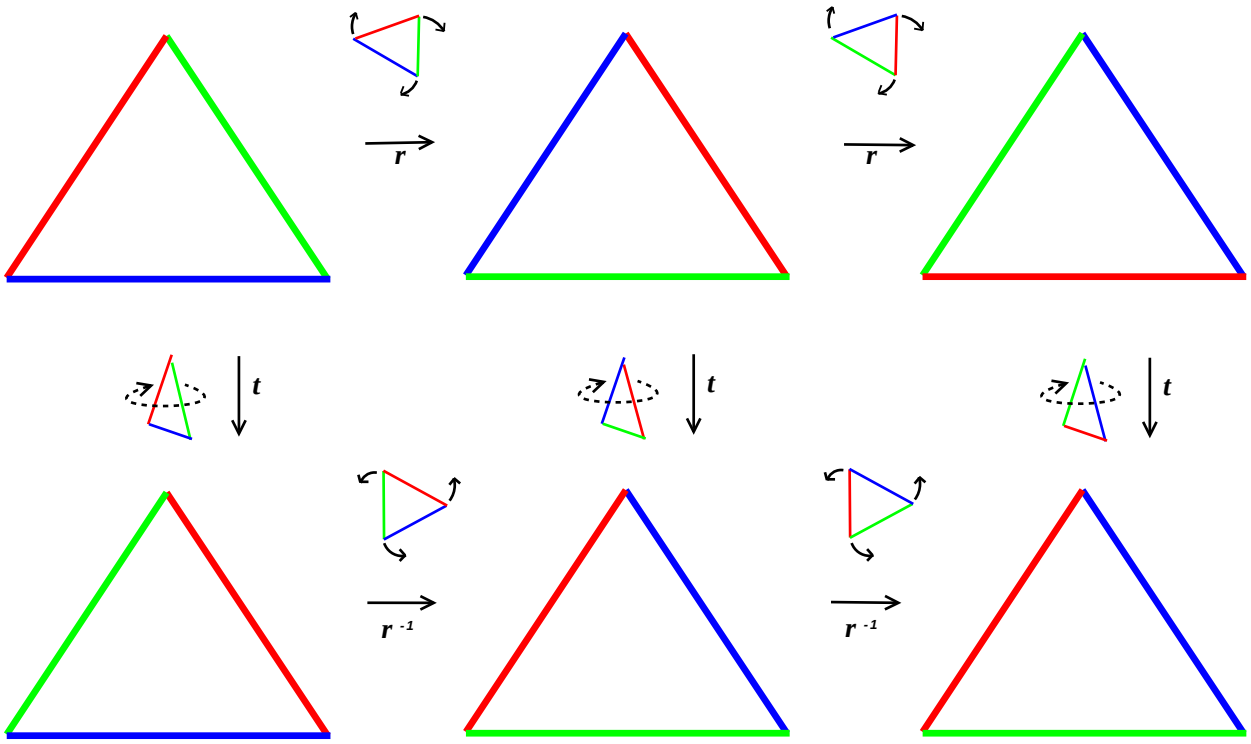


Diagram 1.5.2. Rotations of a Triangle

The group consists of six possible operations which return the triangle into its original position:

(i) Do nothing, rotate the triangle by 360° in its own plane, or by any multiple of 360°. This is the identity of the group.

(ii) Rotate the triangle by 120° clockwise in its own plane, which I denoted by the letter r in the diagram.

(iii) Rotate the triangle by 240° clockwise, or what amounts to the same thing, 120° counter-clockwise, which is the same as executing r twice. I denote this operation by r^{-1}.

(iv) Twisting the triange around a vertical axis by 180°. This twist is denoted by t in the diagram.

(v) Rotating by 120° followed by a twist; and

(vi) Rotating by 240° followed by a twist.

The product in this group thought of as the phrase "followed by" as in:

$r \circ t$ is a rotation by 120° followed by a twist.

Likewise r^2 means a rotation by 120° followed by another, and so on.

Since there are only six objects in this group, this is an example of a **finite** group. Just by looking at the diagram, we can see relationships between rotations in this group. For example, $r \circ t = t \circ r^{-1}$. Note that $r \circ t \neq t \circ r$. This group is **non-abelian**.

1.5.3 **Example: A Permutation Group.** In some ways (when generalized) this is an archetypical example of a finite group. Take five cards out of a deck, and set the remainder of the pack aside. This group consists of all possible shuffles on the five cards you selected. Let us suppose you picked out the following hand.

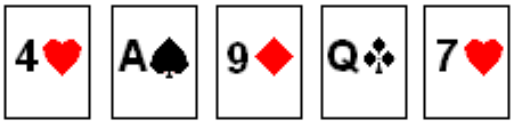


Diagram 1.5.3a. Five Cards.

After the first shuffle, let us suppose that this becomes:

Diagram 1.5.3b. Five Cards Shuffled.

Consider all possible shuffles of this mini-pack of five cards: Is it a group?

Clearly, a shuffle followed by a second is another shuffle. So the closure axiom is satisfied.

How can we verify the associativity in the case of shuffles? The associativity axiom has the reputation of being the most schizophrenic axiom in mathematics–it is either obvious and requires only a momentary reflection to verify, or it is extremely difficult and tedious to verify. Fortunately, in all cases where the group consists of actions which can be imagined as operating temporally on something or somethings, the associativity axiom is trivial because the order of execution in time solely determines the outcome, not whether we do the last two actions together or the first two together. So in the case of shuffles, associativity is satisfied.

The identity is the shuffle which returns the cards to their starting position before the shuffle began, in other words, the shuffle that effectively does nothing. It is obvious (at least with only five cards) that every shuffle can be undone by another shuffle. So inverses exist, and we have a group.

1.5.4 **Abstract Group** This example of shuffling cards, in common with the previous two, is an example of a concrete realization of a group–a group which is manifest as operations on objects (in this case, playing cards) which are not part of Group Theory itself. In the previous example the objects were triangles. Both playing cards and triangles are external to group theory.

In contrast to a concrete group, an **abstract** group refers only to the group-theoretic aspects of its relationships. The abstract group consists of the essence of the group–that which is purely group-theoretic. If we use the letters $a, b, c, \ldots$ to denote members of an abstract group, then the meaning of the letters is unspecified, being merely labels for members of a group. Thus, in an abstract group, $a \circ b$ denotes a group product between two members of the group, and nothing more.

To see the significance of the abstract group, consider another group, the group of permutations on the sequence of symbols 1, 2, 3, 4, 5. Here is one possible permutation.

$$\overbrace{1\ \ 2\ \ 3\ \ 4\ \ 5}$$
$$\downarrow$$
$$\underbrace{4\ \ 2\ \ 1\ \ 5\ \ 3}$$

Diagram 1.5.4. A Permutation of the Digits 1 2 3 4 5.

The top line shows the starting arrangement, the second line the arrangement after the permutation. Compare this rearrangement with the shuffle on the five cards in Diagram 1.5.3a and Diagram 1.5.3b: the first card went to the third position, the second card stood still, the third went to the last position, the fourth went to the first position, and last card went to the fourth position. It is the same permutation as in Diagram 1.5.4 the only difference is one involves playing cards, the other numbers. I hope this convinces you that the possible permutations on five digits forms another group which exactly emulates the possible shuffles on five cards.

Abstractly, the permutations on 1, 2, 3, 4, 5 is the same group as the shuffle group on five cards. We can identify the digit 1 with the card 4♡, 2 with A♠, 3 with 9♢, 4 with Q♣, and 5 with 7♡. This provides

a correspondence between two different concrete groups, but the abstract group, whether exemplified by shuffles of five playing cards, or by permutations of the five digits 1,2,3,4,5, is the same. One can draw a correspondence between every shuffle and every permutation. Such an equivalence of two groups is called an **isomorphism**, and two groups which are realizations of the same abstract group are called **isomorphic**.

It may seem that the abstract group in this example is rather trivial. But it is not, and in fact, it contains 120 distinct permutations or shuffles. What is more, Évariste Galois, after analyzing this group, concluded that the general quintic could not be solved in radicals. That is he proved that the equation

$$ax^5 + bx^4 + cx^3 + dx^2 + ex + f \;=\; 0$$

need not have a solution that could be expressed in terms of the coefficients a, b, c, d, e, f using square roots, cube roots, 4^{th} roots, 5^{th} roots, or indeed n^{th} roots, together with the usual operations of arithmetic: addition, subtraction, multiplication, and division. (On the other hand, the roots of polynomials of degree less than 5 can be so represented.)

1.5.5 **Example: The Affine Group.** Suppose space is infinite in all directions. Consider a rigid body, a space vessel, say, in this space, located far away from any other object. In this example the group is all possible rigid relocations of this space vessel–rigid here means that the space vessel is not to be bent, distorted, or broken in any way. It is an infinite, non-abelian group. The group consists of 3-dimensional rotations, 3-dimensional movements from one point to another, and combinations of these.

In Diagram 1.5.5, there are three movements shown, a, b, and c. The third movement, c, is the same as applying a followed by b. That is, $c = a \circ b$. The circular arrows are meant to show that b and c re-orient the spaceship as well as shifting it through space. Incidentally, the movements of the spaceship are specified as instructions to a pilot sitting in the vessel, and are interpreted from the pilot's point of view: turn left 72°, up 43°, move forward 1,500 miles, etc. There is no reference to any external landmarks or pointers.

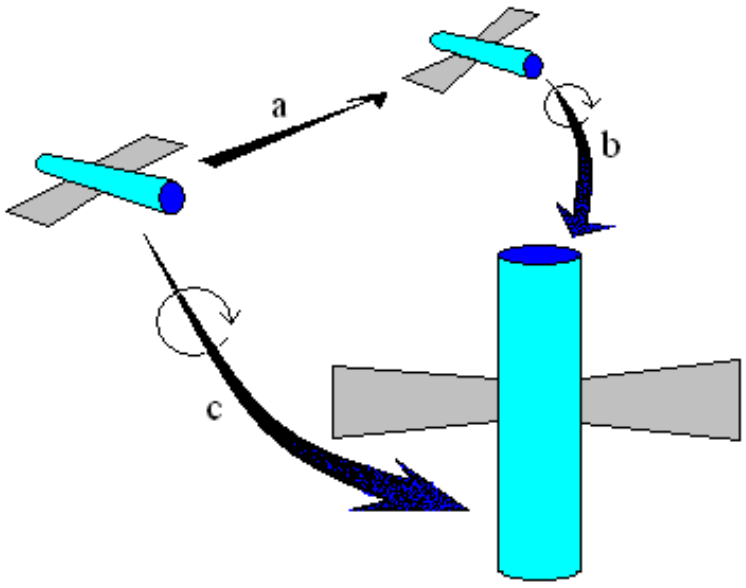


Diagram 1.5.5. Movements of a Spaceship.

1.5.6 **Invariance.** The laws of physics (in both Newtonian and relativistic theories) are invariant under the affine group. That is, the laws stated in one position and orientation are equally valid in any other position and orientation. There is a famous theorem due to Emma Noether which states that if a law of physics is invariant with respect to a group operation, then there exists a corresponding conservation law.

In this case, the conservation laws implied by the symmetries of the affine group are conservation of linear and angular momenta.

1.6 Rings.

Before we leave Group Theory and undertake a more formal introduction to mathematics, I shall take the opportunity to introduce the most common extension of groups.

Think of the integers. Not only are they a group (with + as the group product), they can also be multiplied. Could the integers be a group with multiplication as the product? Let us write multiplication as the dot. Firstly, we see that multiplication is associative: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. So that's good. There is an identity, namely the number 1: $1 \cdot x = x$. So that's good too. But, there are almost no integers with multiplicative inverses which are integers. For example, the multiplicative inverse of 2 is ½ which is not an integer. So, the integers with multiplication do not form a group. Just the same, the multiplication of integers is almost a group; it fails only the inverse axiom.

This situation where a set is a commutative group under one operation and which is close to being a group under a different product occurs frequently in mathematics. Such sets are called **rings**.

1.6.1 Axioms for a Ring.

Let R be a set endowed with two products "+" and "$\circ$". R is said to be a **ring** if it satisfies axioms *(i)* through *(vi)* below.

(i)	Addition:	R is an abelian group under the + operator with identity 0. (It satisfies all the group axioms of §1.3 and also the commutivity axiom: $a + b = b + a$.)
(ii)	Closure: of $\circ$	If a, b are in the set R, then so is $a \circ b$.
(iii)	Associativity of $\circ$:	If a, b, and c are in the set R, then $a \circ (b \circ c) = (a \circ b) \circ c$.
(iv)	Nilpotency of zero:	For every a in R, $0 \circ a = a \circ 0 = 0$.
(v)	Left distributivity:	If a, b, and c are in the set R, then $a \circ (b + c) \ = \ a \circ b \ + \ a \circ c$
(vi)	Right distributivity:	If a, b, and c are in the set R, then $(b + c) \circ a \ = \ b \circ a \ + \ c \circ a$
		If in addition we also have
(vii)	Commutivity of $\circ$:	For any a, b in the set R, $a \circ b = b \circ a$

then the ring is said to be a **commutative ring** (rarely, abelian ring) If the ring does not satisfy Axion *(vii)*. then it is called a **non-commutative ring**.

A property that is easily deduced is $(-x) \circ y = x \circ (-y) = -x \circ y$.

If additionally there is a multiplicative identity, that is an $e \in R$ satisfying $e \circ a = a \circ e = a$ for every a in R, then R is said to be a **ring with identity**. The vast majority of rings studied by mathematicians do have such an identity, and it is usually written as "1" not "e".

1.6.2 Examples

(i) The archetypical example of a commutative ring is the integers with addition for "+", and multiplication for "$\circ$".

(ii) The set of all fractions is a commutative ring with the usual addition and multiplication.

(iii) For any positive integer n, let $\mathbb{Z}_n$ denote the congruence class modulo n. This is the set of integer remainders after dividing by n. This is another commutative ring. The set of remainders is usually (but not always) taken to be $\{0, 1, 2, \ldots, n-1\}$. This is an example of a finite ring.

We denote the remainder of x after division by n with the notation $x \bmod n$. That $\mathbb{Z}_n$ satisfies the axioms of a commutative ring follows from the following two simple facts.

$$(x + y) \bmod n \ = \ (x \bmod n \ + \ y \bmod n) \bmod n$$

$$\text{and,} \quad (xy) \bmod n \;=\; (\,(x \bmod n)(y \bmod n)\,) \bmod n$$

In other words, whether we take remainders before or after performing one of the ring operations we get the same result.

If we wish to indicate that two integers, x, y, have the same remainder after division by n, we write $x \equiv y \pmod{n}$.

I shall demonstrate with n equal 10,

$$14(128+45) \;=\; 14 \times 173 \;=\; 2422 \;\equiv\; 2 \pmod{10}$$

If we take remainders first, we get

$$14(128+45) \;\equiv\; 4(8+5) \;=\; 52 \;\equiv\; 2 \pmod{10}$$

Let us return to Example *(i)*, the integers. They satisfy the cancellation law:

$$\text{if } ax = bx \text{ and } x \text{ is non-zero, then } a = b,$$

Many rings do not satisfy this law. In fact, the Example *(iii)*, the remainders modulo 10, demonstrates this possibility: $1 \times 5 \equiv 3 \times 5 \pmod{10}$, but it is false that $1 \equiv 3 \pmod{10}$– the 5 cannot be cancelled. If a commutative ring does obey the cancellation rule, then it is called an **integral domain**.

Division of two integers is an integer only when the denominator divides the numerator evenly, without remainder. However, in Example *(ii)*, a fraction can always be divided into another provided it is non-zero. In other words, the non-zero fractions are a multiplicative group. A commutative ring with this property is called a **field**. (Beware, that the word "field" is used with completely different meaning in physics.) Returning once again to the story of Galois. It was his study of the properties of number fields formed by adding the roots of polynomials to the fractions that led him to his great discovery of the connection between groups and solvability of polynomials using radicals.

Chapter 2.
Sets.

We now adhere to a more formal course. The Group Theory of the last section ran ahead of fundamental concepts which we must now broach.

2.1 Set Operations.

I assume that you have met set theory before. I shall review only naïve set theory; this is the theory that most people meet in introductory courses in logic or mathematics.

Sets consists of "members", also sometimes called "elements." Many people who have come across sets in popular expositions have the impression that sets may have any kind of elements. For example, a set might consist of the chair over there, the thoughts of James Joyce, the number 12, and all electrons in the universe. However, such extravagance in the formation of sets can lead to contradictions. Therefore, I shall be more parsimonious, requiring that set elements be specifiable by "well-defined steps" in terms of "well-understood" items. A well-defined step will be a process that all present should understand, and well-understood items shall not be exotic or strange to anyone present [2]. We shall allow at least the natural numbers $(0, 1, 2, 3, 4, \ldots)$, symbols written in the Roman and Greek alphabets, and groupings of these such as labels.

Lastly, we take for granted the existence of the **empty set**, the set which contains nothing, and we denote such a set by the special symbol $\emptyset$. The empty set has a unique property: it is a subset of every set.

Even with this parsimony, we can still construct all the usual structures of mathematics, including transfinite arithmetic, provided we are a little liberal in what we allow as well-defined steps.

The two defining characteristics of a set are

(i) the set is entirely specified by its constituents, its members, and

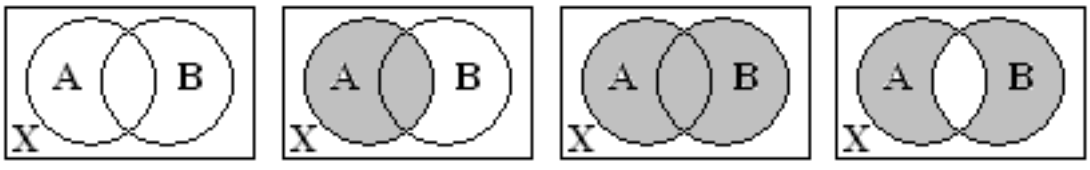
(ii) either the set has a member or it does not–sets are divorced from any concept of multiplicity.

Sets are written with the elements enclosed in curly braces. Thus, $\{0, 9, ½, 0.5893\}$, $\{1, x, -94, 4t + 3\}$, {apple, orange, apple, lemon, pear}. The last set would appear to contain two apples, but as the second defining characteristic says, only existence matters, not multiplicity. Hence, {apple, orange, apple, lemon, pear} is the same set as {apple, lemon, orange, pear}.

When we join two sets together, an operation called "set union", we get a set consisting of every element from both sets. Whether an element occurs in both of the original sets, or just in one and not the other is immaterial, the result is the same, it occurs in the union.

The parsimony principle requires that we construct sets only using well-defined steps. There are many such steps. Set union that we just discussed is one; another is "set intersection", which forms the set consisting of those elements which occur in both original sets; a third example, is "set difference" which consists of all members of the first set which are not members of the second. A set complement refers to all elements not in the set. Set complementation presumes we are given a universe of elements enclosing all possible elements under consideration otherwise it is undefined.

The following diagrams show possible operations on two sets, A and B, within a universe called X. The dark regions denote elements in the result, white regions denote elements not in the result. These type of idealized pictures of sets are called Venn diagrams.



$\emptyset$ A $A \cup B$ $A \star B$

[2] Indeed many mathematicians allow only the empty set as a starting point!

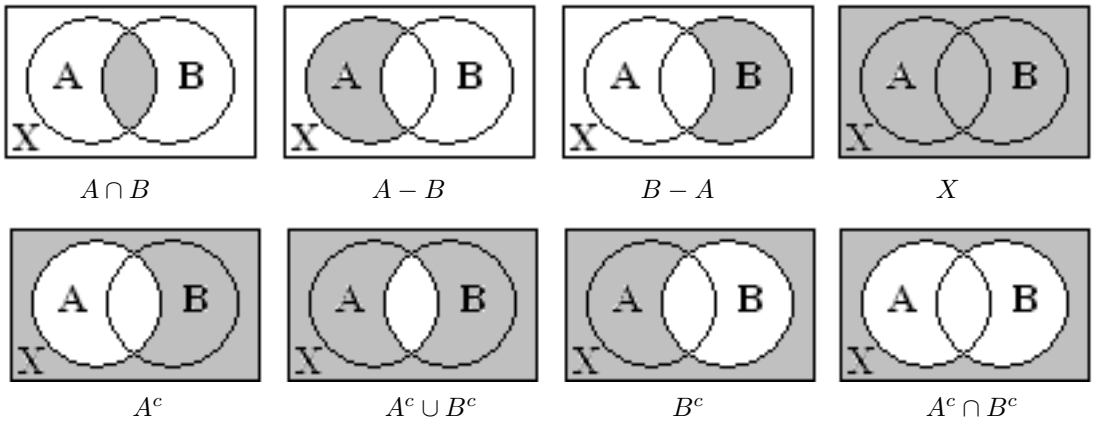


Diagram 2.1. Venn Diagrams for Two Sets.

The operations shown are $A \cup B$ (union), $A \star B$ (symmetric difference, equal to $(A - B) \cup (B - A)$), $A \cap B$ (intersection), $A - B$ (set difference), A^c (set complement, equal to $X - A$).

The table below lists set relationships.

Expression	English Meaning
$a \in A$	a is a member, element, or constituent of the set A.
$A \subset B$	A is a subset of B, possibly equal to B; that is, each element of A is also an element of B.
$A \subsetneq B$	A is a subset of B but is not equal to B; A is said to be a **proper** subset of B.
$A \not\subset B$	A is not a subset of B; that is, at least one element of A is not in B.
$A = \emptyset$	A is empty, has no members
$A = B$	A and B are the same sets. A has the same members as B.

The reader may be curious as to why I used capital letters for sets, and a lowercase letter for a member of a set (as in "$a \in A$" above). This is a useful convention in the set theory typically used in the sciences, where sets of sets are rare, and where there is therefore a clear distinction between sets and their members. But the reader should be warned that sets of sets are perfectly good sets, and we shall be considering such things. So, I make no promise to adhere to this convention in this exposition.

2.2 **The Principle of Substitution.** There is a subtlety in set equality, $A = B$. The statement that A and B have the same members is in fact the definition of set equality, and equality of the two sets means not just that they have elements in common, but also that wherever one set can be used so can the other. So, for instance, suppose we are given two sets which are differently defined, again let us call them A and B. Upon investigation, we find that a sentence $S(A)$ involving the set A is true. Subsequently, we prove that $A \subset B$, so now we know that every element of A is also an element of B. We then find out that $B \subset A$, so that every element of B is also a member of A. We have shown that A and B have the same elements. So, of course, we conclude that $A = B$. But what is really powerful about equality is that we can now also conclude that $S(B)$ is true, no matter what the sentence S might be, so long as $S(A)$ is true.

This principle of substitution of equal objects is the essence of mathematics, and is a reason why it is separate from pure logic. Many non-mathematicians are surprised by this, some because they think mathematics is a developent of logic, and others, perhaps more sophisticated, because they believe that mathematics is distinct from logic by virtue of its incorporation of arithmetic. But I believe, that mathematics parts ways with logic primarily because of the principle of substitution:

If $A = B$, and $S(A)$ holds for a statement S, then so does $S(B)$.

The substitution principle cannot be stated in full in a formal, symbolic system because the only restriction on S is that we can comprehend it[3].

[3] Actually, we do need to exclude statements which refer to the name of the object being substituted.

There have been attempts to base mathematics on logic. Probably the most famous is "Principia Mathematica" by Bertrand Russell and Alfred North Whitehead [RW]. This work attempted to demonstrate that mathematics was just a development of logic, requiring no new axioms, merely additional definitions. However, this work failed for two reasons. One reason is that the authors were forced to propose axioms for logic which were quite unintuitive. I remember that in reading Principia, I found that one such axiom not only lacked an immediacy but was also difficult to grasp. An axiom of logic must surely be self-evident, otherwise it must be at best an axiom for a formal system based on logic.

However, the death knell of all attempts to incorporate mathematics into logic, or axiomatize it based on logic, was Kurt Gödel's proof of the incompleteness of any system incorporating predicate calculus, which certainly included arithmetic. Incompleteness means that there are true statements in such a formal system which cannot be proved within the system. Since arithmetic incorporates such a formal system, the result must apply to it, and so there are true statements in arithmetic which cannot proved within arithmetic. Such statements are called **undecidable** in the system.

You may wonder whether the existence of undecidable systems is a deficiency in the axioms. Perhaps we can add such an undecidable statement as an axiom of arithmetic. After all, there are precedents for this; for instance, an axiom called The Axiom of Choice is now fairly routinely incorporated into mathematics. If such statements are added to arithmetic as axioms, then they do indeed become theorems in the augmented arithmetic. However, Gödel's theorem applies also to the new system, and again there are unprovable but true statements. We can add these also, obtaining a second generation arithmetic so to speak. Proceeding thus we will never stop. The conclusion is inescapable: if arithmetic is to be complete and founded on logic, then it must have an infinity of axioms. All attempts to make arithmetic logical and definitive will come up short, infinitely short.

2.3 **Exercises** Before inflicting exercises on you, I would like to impart a word to the wise. I once talked to a teacher of mathematics, and asked if there was an age by which certain mathematical concepts must be taught or else they would never be grasped. He replied "No", there was no age limit to the understanding of concepts in mathematics, but that there did seem to be an age limit to acquiring proficiency in doing mathematics. Since this text is meant for educated adults, I am therefore conscious that the exercises might be somewhat daunting to readers. The purpose served by exercises in this text is to ensure that you have understood the concepts. If you do understand the concept, and you find the exercise impossible, skip it.

Which of these sets are the same? That is, how many distinct sets are there in *(i) -(vi)*?[4]

(i) $\{10, 4, 1, 7\}$
(ii) $\{1, 4, 4, 7, 10\}$
(iii) $\{1, 10\} \cup \{4, 7\}$
(iv) $\{1, 4, 7, 10\} - \{5\}$
(v) $\{1, 4, 4, 7, 10\} - \{4\}$.
(vi) $\{1, 4, 4, 7, 10\} - \{4, 5\}$.

2.4 **Creating New Types of Sets** The next table lists set operations which create objects of a different type than the original sets.

Expression	English Meaning
$\{A\}$	the set consisting a single element, namely. the set A
$\|A\|$	the number of elements in A
$A \times B$	the Cartesian product of the sets A and B (see below, §2.5)

Beware!

(i) $\{A\}$ is not the same as A. In particular, and very importantly, $\{\emptyset\}$ is not empty!

(ii) $x \in A$ is not the same as $x \subset A$.
For example, if $A = \{1, 2, 3\}$, then $1 \in A$ is correct, $1 \subset A$ is wrong, but $\{1\} \subset A$ is correct.

[4] *(i)=(ii)=(iii)=(iv)≠(v)=(vi)* (the last two are missing the number 4).

(iii) Nonetheless, $x \in A$ and $x \subset A$ is possible. Here is an example,

$$A = \big\{\{1,2\},1,2\big\}, \quad \text{and} \quad x = \{1,2\}$$

2.5 **Cartesian Product** The Cartesian set product of sets A and B can be thought of as a set consisting of pairs of elements taken from A and B. Consider such a pair, (a,b), say, where $a \in A$, and $b \in B$. The parentheses around "a,b" is to remind us that the two elements are appearing together as a pair, and the comma is to remind us that the two elements are not being combined in any way except by being paired with a being first, and b being second – order matters, (a,b) is not the same as (b,a). The Cartesian product of A and B is written $A \times B$ and is the set of all pairs (a,b) with $a \in A$, and $b \in B$.

The term "Cartesian" honors René Descartes who had the idea of analyzing plane geometry by constructing two perpendicular lines in the plane which we call axes. He observed that each point in the plane is uniquely specified by its distances from the two axes. Call the two axes the x-axis and y-axis, and let P denote a general point in the plane. We write its distance from the x-axis as y, and its distance from the y-axis as x (see Diagram 2.5). Then, the pair (x,y) uniquely identifies the point P.

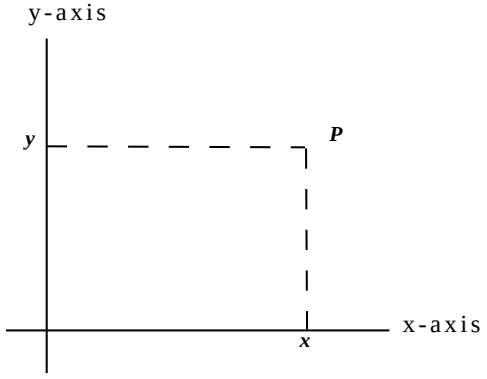


Diagram 2.5. Cartesian Plane.

Descartes showed that the plane can be regarded as a Cartesian product of two perpendicular lines. Thus was born the first ever translation of geometry into algebra.

2.6 **Some Examples**

2.6.1 **Example** Let $A = \{1,2,5\}$, and $B = \{11,32\}$. Then,

$$A \times B \;=\; \{\ (1,11),\ (1,32),\ (2,11),\ (2,32),\ (5,11),\ (5,32)\ \}$$

which can be arranged more prettily as

$$A \times B \;=\; \left\{ \begin{matrix} (1,11), & (2,11), & (5,11), \\ (1,32), & (2,32), & (5,32) \end{matrix} \right\}$$

This should make plain that $|A \times B| = |A| \times |B|$–the number of elements in $A \times B$ equals the number of elements in A times the number of elements in B.

This is an example of a cartesian product of finite sets.

2.6.2 **Example** In this example we solve a simple problem in Euclidean geometry but using Descarte's representation of the Euclidean plane.

Let us find an expression for the area of $\triangle ABC$ in Diagram 2.6.2.

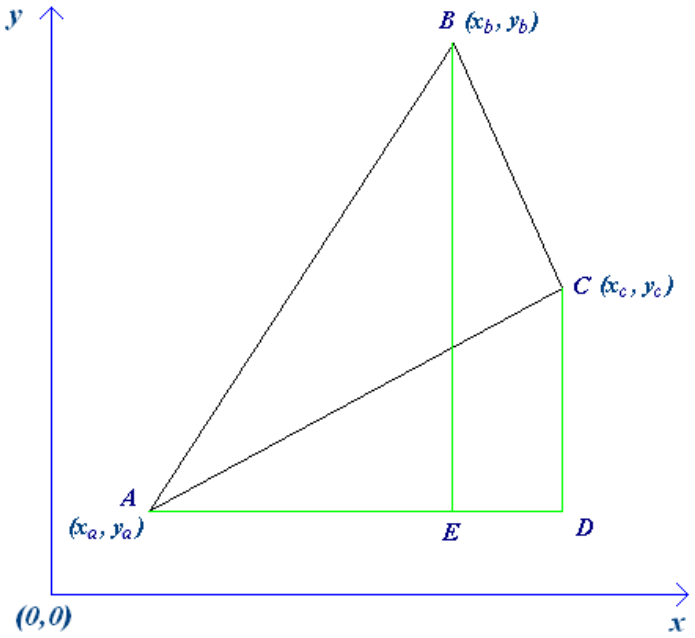


Diagram 2.6.2. Area of a General Triangle in the Plane.

The vertices of the triangle are at the points A, B, C whose coordinates are (in the same order) (x_a, y_a), (x_b, y_b), (x_c, y_c). We shall calculate the area of $\triangle ABC$ as follows. Construct lines BE and CD parallel to the y-axis where AED is a straight line parallel to the x-axis. Then, the area of $\triangle ABC$ is given by

$$\triangle ABC \;=\; \triangle ABE \;+\; \square\, BCDE \;-\; \triangle ACD$$

The area of $\triangle ABE$ and $\triangle ACD$ are easy to calculate since they both have right angles. The area of the quadrilateral $BCDE$ is also easy to calculate because it had two sides parallel.

The area of $\triangle ABE = \frac{1}{2} BE \cdot AE$. Now, BE is vertical so its length is $y_b - y_a$, the difference in the y-coordinates of the points B and E. Similarly, AE is horizontal, its length is the difference in x coordinates of the points A and E which is $x_b - x_a$. So, $\triangle ABE = \frac{1}{2}(x_b - x_a)(y_b - y_a)$.

The area of $\triangle ACD = \frac{1}{2} AD \cdot CD$. Following the same reasoning we used before, we get $\triangle ACD = \frac{1}{2}(x_c - x_a)(y_c - y_a)$.

Finally, the area of $\square\, BCDE = \frac{1}{2}(BE + CD)DE$ (average of the lengths of the parallel sides times the perpendicular distance between them). BE and CD are vertical, and DE is horizontal, so we can calculate their lengths as differences of coordinates. Specifically, $BE = x_b - x_a$, $CD = x_c - x_a$, and $DE = y_c - y_b$ giving $\square\, BCDE = \frac{1}{2}(x_c + x_b - 2x_a)(y_c - y_b)$.

Putting all this together we get

$$\begin{aligned} \triangle ACD \;&=\; \tfrac{1}{2}\,\left((x_b - x_a)(y_b - y_a) \;+\; (x_c + x_b - 2x_a)(y_c - y_b) \;-\; (x_c - x_a)(y_c - y_a)\right) \\ &=\; \tfrac{1}{2}\,\left(x_a y_b - x_b y_a \;+\; x_b y_c - x_c y_b \;+\; x_c y_a - x_a y_c\right) \end{aligned} \tag{2}$$

Look how pretty is Formula (2). Take the first monomial $x_a y_b$, switch the subscripts $a \leftrightarrow b$ and reverse the sign, and you get the second term. Now take both these terms and cycle the subscripts $a \to b$, $b \to c$, and $c \to a$, and you get next two terms. Apply this cycle again, you get the last two terms. If you were to cycle once more, you would arrive back at the first pair of terms.

Does this sound familiar?

Let us review again these two transformations. In Formula (2) when we swap $a \leftrightarrow b$ thoughout we get back Formula (2) but negated. (Check it out.) The same thing happens if we switch $b \leftrightarrow c$ or $a \leftrightarrow c$. So, it

follows that if we switch $a \leftrightarrow b$ and then switch $a \leftrightarrow c$ then logically we should get back the same expression negated twice, which means we get the same expression exactly. Now comes a little leap. The combined effect of switching $a \leftrightarrow b$ followed by switching $a \leftrightarrow c$ is equivalent to the cycle $a \rightarrow b \rightarrow c \rightarrow a$.

This explains why when we cycle the subscripts $a \rightarrow b \rightarrow c \rightarrow a$ we get back the same expression.

We have a group here. It is the group of all permutations of the letters a, b, c, and is in fact isomorphic to the group of rotations of the equilateral triangle described in §1.5.2. Those permutations which only swap two letters reverse the sign of Formula (2), all other permutations of the three letters leave the formula unchanged.

A group operating on the subscripts a, b, c is equivalent to one operating on the vertex labels A, B, C. So all the group members do is relabel the triangle with the same letters in a different order. The cycles like $A \rightarrow B \rightarrow C \rightarrow A$ merely rotate the labels. A relabelling obviously does not affect the triangle, the triangle is the same, we just relabelled the vertices, so there is no surprise that Formula (2) for the area is unchanged. This is a nice interpretation except how then do we explain that when we just swap two labels, for example, $A \leftrightarrow B$, that the area is negated? All we did was to relabel. How can that change the area?

The sign reversal on swapping two labels does change one aspect of the triangle: the original labels were alphabetically ordered clockwise around the triangle, but after swapping $A \leftrightarrow B$ they become ordered counter-clockwise. Pretend for the moment that we could flip over the entire plane of the triangle, keeping the labels glued to their vertices; this too would reverse the ordering of the labels from clockwise to counter-clockwise. This suggests that perhaps triangles (and other shapes) in the plane might have two possible orientations, face up or face down, whereby their areas are positive in one orientation but negative in the other. So, perhaps we might say that $\triangle ABC = -\triangle ACB$? All this from a little group theory.

Chapter 3.
Mathematical and Logical Notation.

3.1 **Logics and Mathematics** Although all intellectual disciplines presuppose logic, mathematics is unusual in its proximity to pure logic, and so logical notation appears frequently in mathematics. Mathematical texts generally avoid using symbols where short English words serve just as well such as "and", "or", "true", and "false". However, logical notation must be used in mathematical logic, and it must also used in some areas of mathematics.

3.2 **Why We Need Formal Logic.** One such area is analysis, a subject which in its infancy (when it was known as the calculus) achieved great success by ignoring concerns over its foundations (such as those of Bishop Berkeley) until by the 19^{th} century the mounting paradoxes caused so much confusion they could not be ignored. The resulting late 19^{th} century regime of rigor led by Weierstrasse, Dedekind, Cauchy and others gave us our concepts of continuity. It is hardly surprising that the 18^{th} century mathematicians did not have a clear idea of continuity–the definition handed down to us from Cauchy is most easily stated in a formal logic called predicate calculus. Statements expressing continuity in natural language are usually ambiguous, inaccurate, or lengthy. The definition of continuity is given in §3.7*(vi)* after I have reviewed the notations of symbolic logic.

3.3 **Logical Statements.** Let R and S be statements. Statements roughly correspond to sentences in natural language–they make meaningful, though possibly false, assertions. The notation "$R \Rightarrow S$" is a new statement constructed from statements R and S and means any and all of the following:

(i) $R \Rightarrow S$.
(ii) The statement R implies the statement S.
(iii) If R, then S.
(iv) If R is true, then so is S.
(v) R cannot be true unless S is also true.
(vi) R is a sufficient condition for S.
(vii) S is a necessary condition for R.
(viii) S is implied by R.
(ix) $S \Leftarrow R$.
(x) If S is false, then so is R.
(xi) Not-S implies not-R.
(xii) $\neg S \Rightarrow \neg R$.

In the last statement the "$\neg$" symbol stands for denial of the statement which immediately follows it, and is called the logical negation or "not" symbol. In mathematics, denial is usually indicated by crossing out the operator which stands for the verb in the sentence as in $A \notin B$, which means "A is not a member of B".

Note that "$R \Rightarrow S$" says nothing about the truth or falsity of the statement R, only that if R is true, then so is S. If we reverse the symbol $\Rightarrow$, we get the symbol for "is implied by": $S \Leftarrow R$ means $R \Rightarrow S$.

Example. Let the statement R be "n is greater than 6", and let the statement S be "n is greater than 2". It is clear that R implies S. In mathematical language this would read

$$\begin{array}{ccc} R & \Rightarrow & S \\ n > 6 & \Rightarrow & n > 2 \end{array}$$

If S is false (n is not greater than 2), then statement R must also be false (n cannot be greater than 6). Now, "..not greater than.." means the same as "..less than or equal to..", so this can be written

$$\begin{array}{ccc} \neg S & \Rightarrow & \neg R \\ n \leq 2 & \Rightarrow & n \leq 6 \end{array}$$

Lastly, note that $R \Leftarrow S$ is false: "n is greater than 2" obviously does not imply that "n is greater than 6" -- what if n was equal to 3?

The notation "$R \Leftrightarrow S$" means both $R \Rightarrow S$ and $S \Rightarrow R$. It means the same as any of the following

(i) $R \Leftrightarrow S$.
(ii) $S \Leftrightarrow R$.
(iii) R is true if and only if S is true.
(iv) R is false if and only if S is false.
(v) R and S are equivalent.
(vi) R is a necessary and sufficient condition for S.
(vii) S is a necessary and sufficient condition for R.

3.3.1 **The Propositional Calculus.** The best-known, simple, symbolic logic is the Propositional Calculus. It is one of the great triumphs of the human mind that basic classical logic can be completely formalized in the sense that any question raised in the logical system can be fully answered within the system.

The system allows the following symbols $p, q, r, s, \ldots$ to stand for propositions or statements, and has the additional symbols t and f which are constants. Informally, t means "true", and f means "false." The calculus also contains the connectives $\Rightarrow$ and $\neg$ which informally stand for respectively IMPLIES and NOT. The other connectives of logic can be defined from these:

$$\begin{aligned} p \,\&\, q &:= \neg(p \Rightarrow \neg q) \\ p \Leftrightarrow q &:= (p \Rightarrow q) \,\&\, (q \Rightarrow p) \\ p \vee q &:= \neg p \Rightarrow q \end{aligned}$$

(The symbol ":=" means "is defined to be").

In addition, there are rules of inference, which are

(i) $f \Rightarrow p$ (Law of Duns Scotus: anything follows from falsity)
(ii) If p then $(p \Rightarrow t)$
(iii) If $p \Rightarrow q$, and also p, then q (Modus Ponens)
(iv) $(p \Rightarrow q) \,\&\, (q \Rightarrow r) \Rightarrow (p \Rightarrow r)$ (Transitivity of inference)
(v) $p \,\&\, q \Rightarrow p$
(vi) $p \Rightarrow p \vee q$
(vii) $p \,\&\, q \Rightarrow q \,\&\, p$
(viii) $p \vee q \Rightarrow q \vee p$
(ix) $p \Rightarrow \neg\neg p$
(x) $\neg(p \,\&\, q) \Rightarrow \neg p \vee \neg q$ (De Morgan)
(xi) $\neg(p \,\&\, q) \Leftarrow \neg p \vee \neg q$ (De Morgan)
(xii) $p \,\&\, \neg p \Rightarrow f$ (Law of Contradiction)
(xiii) $\neg\neg p \Rightarrow p$
(xiv) $p \vee \neg p$ (Law of the Excluded Middle)

There is a simple algorithm for deciding the correctness of any statement expressed in the propositional calculus. The essential idea behind the method is to check every possible value of the propositions making up the expression and see if the expression is true under every possible truth-value of its propositions. I will illustrate the method by proving that $(p \Rightarrow q)$ is equivalent to $\neg p \vee q$.

$p \Rightarrow q$

p \ q	f	t
f	t	t
t	f	t

$\neg p \vee q$

p \ q	f	t
f	t	t
t	f	t

These tables are called **truth tables**. In both cases, the truth table shows a value of false only when p is true and q is false. Therefore, the two expressions have the same truth table, and are therefore logically equivalent. Another way to decide statements in the proposition calculus is to use the propositions of boolean algebra.

3.3.2 **Boolean Algebra.** A boolean algebra consists of a set B and a set of operators consisting of: $\wedge$ (called a "wedge"), $\vee$ (called "vee"), and $\neg$ ("negation"). The set B and the operators satisfy the axioms below for every $x, y, z \in B$:

(i)	$0 \in B,\ 1 \in B$	Special constants
(ii)	$x \vee y \in B$	Closure for the vee
(iii)	$x \wedge y \in B$	Closure for the wedge
(iv)	$\neg x \in B$	Closure for the negation
(v)	$x \vee y = y \vee x$	Commutivity of the vee
(vi)	$x \wedge y = y \wedge x$	Commutivity of the wedge
(vii)	$x \vee (y \vee z) = (x \vee y) \vee z$	Associativity of the vee
(viii)	$x \wedge (y \wedge z) = (x \wedge y) \wedge z$	Associativity of the wedge
(ix)	$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$	Distributivity of wedge into vee
(x)	$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$	Distributivity of vee into wedge
(xi)	$0 \vee x = x$	The vee identity element
(xii)	$1 \wedge x = x$	The wedge identity element
(xiii)	$\neg\neg x = x$	Double negation
(xiv)	$x \vee \neg x = 1$	
(xv)	$x \wedge \neg x = 0$	
(xvi)	$\neg(x \wedge y) = \neg x \vee \neg y$	De Morgan

George Boole invented the first algebra of this kind, calling it "laws of thought." His effort was not identical with the above axioms; they are a refinement of Boole's work by Peirce and others. Of course, Boole regarded his algebra as laws of thought because it was a calculus for deciding the truth or falsity of many types of statements in classical logic. Indeed, if we make the following substitutions in the lexicon of the propositional calculus

$$\begin{array}{rcl} \& & \rightarrow & \wedge \\ f & \rightarrow & 0 \\ t & \rightarrow & 1 \\ p \ \Rightarrow\ q & \rightarrow & \neg p \vee q \end{array}$$

we see that the propositional calculus satisfies all the axioms for a boolean algebra.

Several other simple laws can be deduced from the above axioms including:

(xvii)	$\neg(x \vee y) = \neg x \wedge \neg y$	De Morgan
(xviii)	$x \vee x = x$	Absorption for the vee
(xix)	$x \wedge x = x$	Absorption for the wedge
(xx)	$\neg 0 = 1$	
(xxi)	$\neg 1 = 0$	

The axioms of a boolean algebra strongly resemble axioms for a number system, even to the point of having two numeric constants, 0 and 1. If we transpose the boolean operators into the familiar operators of arithmetic, substituting juxtaposition for the wedge, plus for the vee operator, and an overbar for the negation, we get the following very suggestive expressions for some of the laws of a boolean algebra:

(i)	$0 \in B,\ 1 \in B$	Special constants
(ii)	$x + y \in B$	Closure for addition
(iii)	$x\,y \in B$	Closure for the product
(iv)	$\overline{x} \in B$	Closure for negation
(v)	$x + y = y + x$	Commutivity of addition
(vi)	$x\,y = y\,x$	Commutivity of the product
(vii)	$x + (y + z) = (x + y) + z$	Associativity of addition
(viii)	$x\,(y\,z) = (x\,y)\,z$	Associativity of the product
(ix)	$x\,(y + z) = x\,y + x\,z$	Distributivity of product into addition
(xi)	$0 + x = x$	The additive identity element

(xii)	$1\,x = x$	The product identity element
(xiii)	$\overline{\overline{x}} = x$	Double negation

These are laws satisfied by the familiar numbers. This is indeed a powerful analogy which is why mathematicians following Boole admired his achievement. But, before we get carried away, the analogy begins to fall apart when we apply it to the remaining laws. To begin with, we must interpret $\overline{x}$ to mean $1 - x$ not $-x$. Secondly, we must bear in mind that the variables are two-valued, taking the values of either 0 or 1. With this interpretation we can extend the analogy to the laws defining negation.

(xiv)	$x + \overline{x} = 1$	because $x + (1 - x) = 1$
(xv)	$x\,\overline{x} = 0$	true since $x = 0$ or 1
(xix)	$x\,x = x$	true since $x = 0$ or 1
(xx)	$\overline{0} = 1$	
(xxi)	$\overline{1} = 0$	

But now the analogy fails. No interpretation along the lines of numbers or analogous systems makes any sense on the remaining laws because they contradict the usual law of subtraction that we expect to hold in any number system. Below I use "$\mathcal{S} \models$" to demonstrate a contradiction with the law of subtraction and cancellation.

(x)	$x + y\,z = (x + y)\,(x + z)$	If we expand and cancel terms we get an absurdity: $\mathcal{S} \models\ x(y + z) = 0$ for all x, y, z
(xvi)	$\overline{(x\,y)} = \overline{x} + \overline{y}$	If we set $x = y$ we get $\mathcal{S} \models\ \overline{x}\overline{x} = \overline{x} + \overline{x}\ \Rightarrow\ \overline{x} = \overline{x} + \overline{x}\ \Rightarrow\ \overline{x} = 0$ (after cancelling) $\Rightarrow\ x = 1$
(xvii)	$\overline{(x + y)} = \overline{x}\,\overline{y}$	
(xviii)	$x + x = x$	$\mathcal{S} \models\ x = 0$

For this reason, we should stick to the wedge and vee notation for boolean algebras rather than promote confusion by adopting a notation suggesting laws which are not obeyed, and which does not convey laws which are obeyed.

Is there any hope left of applying the powerful techniques and results of arithmetic to the analysis of propositional logic?

3·3·3 **Binary Arithmetic To The Rescue!** We are all familiar with the fact that computers at their lowest level perform only binary calculations, the result of which is either on or off, true or false, 0 or 1. How is binary arithmetic related to the propositional calculus? An amazing fact, which only gradually dawned on people, is that there is another way to present propositional calculus as an algebra. Instead of taking the inclusive-OR as a basic connective, we adopt the exclusive-OR instead.

The inclusive-OR, $a \vee b$ is true if either a or b or both are true. The exclusive-OR on the other hand is true if exactly one of a and b is true; if both are true then the exclusive-OR is false.

We now denote the exclusive-OR by the plus sign, $+$, and we represent AND as juxtaposition, f as 0, and t as 1, and we also take $\neg x = 1 - x$ as we did before. Now, we no longer arrive at strange laws unknown in the realm of number, but instead obtain the laws of arithmetic modulo 2 ! Binary arithmetic ignores all magnitude considering only whether a number is odd or even; 0 stands for even, 1 for odd. Thus, 1+1=0 because odd+odd is even. Similarly, 0+1=1, and 0+0=0. In binary arithmetic, + is the same as − because if x is odd then so is $-x$, and if x is even so is $-x$.

(i)	$0 \in B,\ 1 \in B$	Special constants
(ii)	$x + y \in B$	Closure for addition
(iii)	$x\,y \in B$	Closure for the product
(iv)	$\overline{x} \in B$	Closure for negation
(v)	$x + y = y + x$	Commutivity of addition. Obvious from symmetry of the exclusive OR
(vi)	$x\,y = y\,x$	Commutivity of the product
(vii)	$x + (y + z) = (x + y) + z$	Associativity of addition. Tedious but easy to verify.

(viii)	$x\,(y\,z) = (x\,y)\,z$	Associativity of the product
(ix)	$x\,(y+z) = x\,y + x\,z$	Distributivity of product into addition. Tedious but easy to verify.
(xi)	$0 + x = x$	The additive identity element
(xii)	$1\,x = x$	The product identity element
(xiii)	$\overline{\overline{x}} = x$	Double negation
(xiv)	$x + \overline{x} = 1$	
(xv)	$x\,\overline{x} = 0$	
(xvi)	$\overline{(x\,y)} = \overline{x}\,\overline{y}$	
(xvii)	$\overline{(x+y)} = \overline{x} + \overline{y}$	
(xviii)	$x + x = 0$	Nilpotency of addition
(xix)	$x\,x = x$	Idempotency of the product
(xx)	$\overline{0} = 1$	
(xxi)	$\overline{1} = 0$	

The changes from Boolean logic to binary arithmetic are significant: Law *(x)* (distributivity of addition into the product) is gone–it is not obeyed by numbers. Also, laws *(xv)* through *(xviii)* have been modified. The payoff is that binary arithmetic is a commutative ring (compare with the ring axioms in §1.6.1).

3.3.8 **Logical Quantifiers.** Because logic precedes mathematics, there is no concept of number and quantity in pure logic. Nevertheless, there are two so-called logical quantifiers which have connotations of quantity. They are: the universal quantifier denoted by the symbol $\forall$ (mnemonic: "All"), and the existential quantifier denoted by the symbol $\exists$ (mnemonic: "Exists"). In formal mathematics, the quantifiers are always immediately followed by variable, then usually a comma, and then an assertion about the variable.

The $\forall$ quantifier indicates that the assertion is true for all instances of the variable. For example,

$$\forall n \in \{1, 3, 7, 11, 12\}, \quad n \text{ is less that } 20$$

In less formal mathematics, the universal quantifier often follows the assertion. This is because of the way we speak and write in common English. ("I have proved this conjecture for all triangles." "This theory applies to all stable atoms.") In this informal mathematics, the above statement would read

$$n < 20, \quad \forall n \in \{1, 3, 7, 11, 12\}$$

The $\exists$ quantifier indicates that the assertion which follows it is true for some instance. For example,

$$\exists n \in \{1, 3, 7, 11, 12\}, \quad n \text{ is divisible by } 3$$

The fact that this example has two instances for which n is divisible by 3 is irrelevant, all that is relevant is that there is at least one such instance.

Below I have listed the more common mathematical terms followed by brief translations. The table is followed by several examples which demonstrate typical usage of these terms.

3.4 Brief Glossary of Terms with Examples.

The variables appearing in the examples in the following table should be understood to be integers.

Symbol	Meaning	Brief Example
$\forall$	For all, For each, For every	$(\forall x)(x > 0 \vee x < 0 \vee x = 0)$
$\exists$	There exists... [such that]	$(\exists e)(e \circ x = x)$
	There is at least one ... [such that]	
	For at least one ...	
$\nexists$	There does not exist ... [such that]	$(\nexists x)(x + 1 = x)$
X iff Y	X is true if and only if Y is true.	$x^2 = 1$ iff $x = 1 \vee x = -1$
$X \Leftrightarrow Y$	X is false if and only if Y is false.	
	X is a necessary and sufficient condition for Y.	
X $\Rightarrow$ Y	X implies Y,	$x = 1 \Rightarrow x^2 = 1$
	X is a sufficient condition for Y.	
X $\Leftarrow$ Y	X is implied by Y,	$xy = 1 \Leftarrow y = x^{-1}$
	X is a necessary condition for Y.	
$\alpha : A \to B$	α maps the set A to the set B.	See Chapter 4
$\alpha : a \mapsto b$	α maps the element a to the element b	
w.l.o.g.	Without Loss Of Generality	
$\vdash$	All such satisfying (in set definitions)	$\mathbb{N} = \{x \in \mathbb{Z} \vdash x \geq 0\}$
$T \models$	Under the assumptions of theory T	Arithmetic $\models x + 1 > x$
s.t.	such that	Choose y s.t. $xy > 10$
$\therefore$	Therefore,	$x < 0 \therefore x \neq y^2$
	Consequently,	
	So,	
:=	...is defined to be...	$\imath := \sqrt{-1}$
$\gcd(m, n)$	Greatest Common Divisor of m and n	$\gcd(30, 12) = 6$
$\operatorname{lcm}(m, n)$	Least Common Multiple of m and n	$\operatorname{lcm}(30, 12) = 60$
$m \mid n$	m divides n with no remainder	$12 \mid 36$
QED	*Quid Erat Demonstrandum*	End of proof
$\square$	End of proof	
$\blacksquare$	End of proof	
$\sum$	Sum of all elements in a range or in a set	$\sum\{1, 22, 4\} = 27$
$\prod$	Product of all elements in a range or in a set	$\prod\{1, 22, 4\} = 88$

3.5 **Subscripts and Superscripts.** Mathematics and mathematical logic use subscripts and superscripts liberally. They almost always appear to the right of the symbol being sub- or superscripted. Subscripts are always used to denote an occurrence of one of many instances. Thus, for example, we might be given a sequence of objects x_1, x_2, x_3, x_4, x_5. We can refer to an arbitrary member of these five objects by the notation x_i, denoting the i^{th} member of the objects, where i is one of $1, 2, 3, 4,$ or 5

Superscripts on the other hand usually denote exponentiation [5]. So, x^2 usually denotes $x \times x$, assuming of course that multiplication ($\times$) is defined for the object x. For sets, the "$\times$" is defined to be the Cartesian set product. Thus, when A is a set, A^2 would normally be interpreted as the Cartesian product of A with itself. There is a special extension of this notation: 2^A denotes the set of all subsets of the set A.

3.6 **Exercise** Let $A = \{1, 2, 3, 4, 5\}$. Show that $\left|2^A\right| = 2^{|A|}$.
That is, show that the number of all subsets of A is $2^5 = 32$.

3.7 **Examples of Mathematical and Logical Notation.** I shall present some typical mathematical statements containing technical or logical terms and then follow them with the same statement in less formal language. The informal statement includes some technical terms in quotes which appear after their translation.

[5] The major exception is tensorial notation used in the Theory of Relativity.

(i) $A = \{x_i \vdash 0 \leq i < N\}$.

Translation: A equals the set ("$\{\cdots\}$") of all x_i's such that ("$x_i \vdash \ldots$") their subscripts (i) range from zero up to but excluding N ("$\ldots 0 \leq i < N$").

(ii) $A := \{b \in B \vdash b^2 \in C\}$.

Translation: A is defined to be (":=") the set of all elements of B ("$b \in B$") whose square is in a set C ("$b^2 \in C$").

(iii) $x_i = 0, \ \forall i \in A$.

Translation: The variables x_i are zero for all ("$\forall$") i that are members of the set A.

(iv) $\therefore x \leq y \ \Leftarrow \ y^3 > x^3$. $\square$

Translation: Therefore, ($\therefore$), x being less or equal ("$\leq$") to y is implied by y^3 being greater than ("$>$") x^3. End of proof ($\square$)

(v) Let $A = \{i \in \mathbb{N} \vdash 0 < i \leq 8\}$. Then, $\sum A = 36$, and $\prod A = 8! = 40320$.

Translation: Let A be the set of natural numbers ($\mathbb{N}$) greater than 0 and less than or equal to 8. Then, sum ($\sum$) of all numbers in the set A is 36, and the product ($\prod$) of the numbers in A is 8 factorial or 40320.

This example shows the economy in putting the universe of objects at the beginning of the set definition. The statement that i is in the set $\mathbb{N}$ simplifies much of the subsequent definition since we know from the outset that we are talking only of the natural numbers, not the reals, not some exotic mathematical object, merely the whole numbers.

(vi) **The Definition of Continuity.**

In this section, variables take real values.

In formal, predicate logic with mathematical symbols, the function f is said to be continuous at x if

$$(\forall\varepsilon)\Big(\varepsilon > 0 \ \Rightarrow \ (\exists\delta)\big(\delta > 0 \ \ \& \ \ (\forall y)(|x-y| < \delta \ \Rightarrow \ |f(x) - f(y)| < \varepsilon)\big)\Big)$$

It is very difficult for beginners to mathematical logic to understand a statement like this. Scanning it efficiently takes practice. For instance, the practised would guess why the author of the statement inserted the clause "$\varepsilon > 0 \ \Rightarrow \ \ldots$" etc. The reason is that the author wanted to limit ε to positive values only which the clause "$\varepsilon > 0 \ \Rightarrow \ \ldots$" effectively does. To avoid repeatedly adding technical clauses like this, mathematicians use some shortcuts, including specifying the range of values up-front as in example *(v)*. So, the definition of continuity of f at x in formal mathematics would read as:

$$\forall\varepsilon > 0, \ \exists\delta > 0 \text{ s.t. } |x-y| < \delta \ \Rightarrow \ |f(x) - f(y)| < \varepsilon$$

Translation: For all ("$\forall$") positive ε ("$\varepsilon > 0$"), there exists ("$\exists$") a positive δ ("$\delta > 0$") such that (" s.t. ") $|x-y| < \delta$ implies that (" $\Rightarrow$ ") $|f(x) - f(y)| < \varepsilon$.

In better English this would read:

> Given any positive ε, one can find a positive δ such that whenever x and y are less than δ apart then $f(x)$ and $f(y)$ are less than ε apart.

Now, having eliminated the logical symbolism leaving only English and a few mathematical symbols, the statement is reasonably clear. The formal statement contained an implicit understanding that the number δ was to be regarded as dependent on ε. This fact we are to deduce from two conventions: firstly, quantifiers are to be read left-to-right, and secondly, a variable appearing in an existential quantifier which is introduced (left-to- right order) after the introduction of another may be assumed to be "aware" of the value of the earlier variable–it may depend on it. So, in the statement of continuity, the δ may be thought of as dependent on the particular instance of ε under consideration.

Here is another example of this convention $(\exists x > 0)(\forall y > x)(\ldots$ etc.). The qualified quantifier $\forall y > x$ is allowed since x was already introduced to the left. Many authors of logic books forget to point out these conventions, causing much confusion in students.

3.8 **Commonly-Occurring Sets of Mathematics.** I list the more commonly occurring sets in mathematics with their usual symbols and a brief description. Note that both of the terms "negative" and "positive" exclude zero. Thus, if a number is non-negative, it is zero or positive.

$\mathbb{N}$ The natural numbers. All the whole numbers including zero. Considered the most basic set in mathematics. Brouwer said they were God-given. He might be explaining to God right now why the zero is God-given but the negatives are not.

$\mathbb{Z}$ The integers. The natural numbers with the negatives added. This set is an abelian group, and is also a commutative ring, the simplest such that is infinite.

$\mathbb{Q}$ The rationals. This is the mathematical term for the set of all fractions, a/b where a, b are integers, positive, negative, or in the case of a, zero. This set is a commutative ring.

$\mathbb{Q}^*$ The non-zero rationals. This set is a multiplicative group, but not a ring.

$\mathbb{R}$ The reals. All decimals, having possibly an infinity of digits following the decimal point. These correspond to all points on a line. This set is a commutative ring.

$\mathbb{C}$ The complex numbers. Complex numbers are $x + \iota y$ where x and y are real and $\iota = \sqrt{-1}$. The set of all complex numbers are usually identified with the 2-dimensional plane. This set is a commutative ring.

$\mathbb{Z}_n$, or $\mathbb{Z}/(n)$ Remainders of whole numbers after division by n. The set of remainders is usually taken to be the numbers $0, 1, 2, \ldots, n-1$.

This set is a commutative ring. It is a field iff n is a prime number.

$\mathcal{S}_n$ The symmetric group on n symbols. This is the abstract group of all permutations on n objects. Example 1.8 was $\mathcal{S}_5$.

3.9 **Statements and Proofs.**

Mathematics is very old. There is now much evidence that the recording of numbers, counts, divisions, and multiplications might have preceded the writing of ordinary sentences. Excavations in the valleys of the Euphrates and the Tigris rivers and tributaries have uncovered thousands of clay tokens of great antiquity. The best explanation for these tokens is that they were used as counters to represent produce in granaries, debts, credits, or land. Probably, these people developed techniques for performing arithmetic without any justification of the validity of these techniques. The first great revolution in mathematics, probably the greatest, was the realization by the Greeks Thales, Pythagoras, and most of all, Eudoxus and Euclid, that rigorous demonstrations were needed to establish truth in mathematics. Although they almost certainly came to this insight because of their interest in geometry, Euclid recognized, in principle, if not always in practice, the need for rigor also in arithmetic. It is remarkable that the format Euclid developed over 2,000 years ago is still the format we use today; indeed much of his terminology is still in use.

The terms "Theorem", "Proposition", "Lemma", and "Corollary" all mean the same thing: a statement which has been or is about to be proved. Their nuances are subjective and vary from author to author. Typically, "Theorem" is a major result and is often the culmination of many propositions, lemmas, and corollaries. Propositions are results which are of some interest beyond being a step toward proving a theorem. Lemmas are results which have no interest besides being required to prove other results. Corollaries are easy consequences of a theorem, proposition, another corollary, or (rarely) a lemma. However, one mathematician's lemma might be another mathematician's theorem, and vice versa.

The meaning of the term "proof" is currently controversial. I believe that a proof is an argument which if read critically without great effort but with reasonable (and not extraordinary) intelligence by a sane and alert person who is acquainted with and understands the subject matter referenced in the proof, then that person will know the statement that was to be proved is true.

The great danger in an attempted proof is that it is convincing but the statement to be proved is false.

Chapter 4.
Maps on Sets.

A **map** from one set to another is a procedure or rule for associating elements of the first set with those in the second with one restriction: given any element in the first set, the map associates exactly one element in the second.

The notation $\alpha : A \to B$ means that α is a map which assigns to each element of a set A an element of a set B. There are alternative notations; one is $A \overset{\alpha}{\to} B$, another is $\alpha(A) = B$ which is called the functional notation. The set on the left (A in this case) is called the **domain** of the map, the set on the right (B in this case) is called its **range**. The **image** of the map is all elements of B which are mapped from A. It is not assumed nor is it required that a map covers all of B with elements mapped from A. However, the image must always be a subset of (or all of) B.

An element $x \in A$ is said to be mapped to y if α assigns to x the element y. Again the mapping of an element can be denoted in a few ways: by $\alpha : x \mapsto y$ (map notation), by $x \overset{\alpha}{\mapsto} y$ (also called map notation), and by $\alpha(x) = y$ or $y = \alpha(x)$ (functional notation).

4.1.1 Example of a map

(i) Define α to be a map from the days of the week to the set $\mathbb{N}$, the natural numbers $(0, 1, 2, 3, \ldots)$. We define α as follows:

Domain	$\overset{\alpha}{\longrightarrow}$	Range
Saturday	$\longrightarrow$	0
Sunday	$\longrightarrow$	0
Monday	$\longrightarrow$	1
Tuesday	$\longrightarrow$	2
Wednesday	$\longrightarrow$	3
Thursday	$\longrightarrow$	4
Friday	$\longrightarrow$	5

These associations show several features of maps:

(i) Each day is associated with exactly one integer. This is an essential feature of a map.
(ii) Every day in the week is associated with some integer. This too is an essential feature.
(iii) The numbers greater than 5 are not associated to any day. Maps are not required to map to the whole of the range; they are allowed to map into subsets of the range.
(iv) Two days are mapped to the same integer (0). Maps are allowed to map several in the domain to the same value in the range.

4.1.2 Example

Pick any one of the six transformations of the triangle described in §1.5.2, let us say we pick the rotation followed by a twist, rt. This transformation moves a triangle in one orientation to another orientation depending only on the starting orientation. So this is another example of a map.

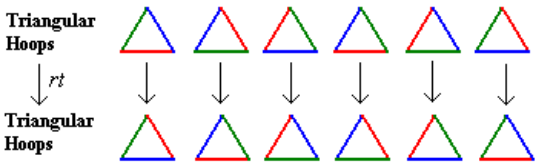


Diagram 4.1.2. A Map of Triangular Hoops

This is an example of a map whose range, domain, and image are all the same set. Note that you can reverse all the arrows and you get another valid map.

4.1.3 Functions

Maps are a generalization of mathematical functions. Consider the mathematical function $f(x) = x^2$. It assigns to each number, x, the unique value, x^2, and so is a map. We can think of this map as $f : \mathbb{R} \to \mathbb{R}$, a map from the real numbers into the real numbers. Because the product of any

number is always non-negative, the negative numbers are not in the image of this map. Also, every positive number has two real numbers mapped to it. $f : -x \mapsto x^2$ and $f : x \mapsto x^2$.

4.1.4 **Example** A car travelling at a speed of v miles per hour under good road conditions requires $S(v)$ seconds to stop from the moment the driver slams on the brakes.

The function S converts speeds in miles per hour which can vary from -20 to 100 miles per hour say, and returns values in the range 0 to infinity feet.

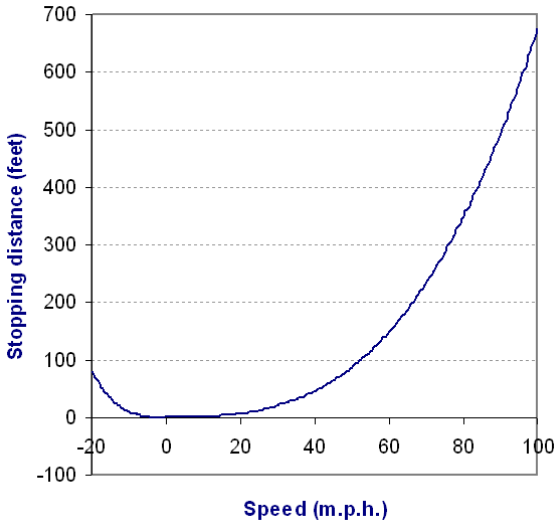


Diagram 4.1.4. A Function Representing Car Stopping Distance

The essential feature of the function is that for a given velocity there is only one stopping distance. The opposite is not true. Given a stopping distance, one roughly in the range 0-80 feet, there can be two velocities which can result in that stopping distance, one positive (car moving forward) and the other negative (the car in reverse).

4.1.5 **Example** Define $\mu(n)$ to be the number of months having n days in any four-year period. Thus, $\mu : \mathbb{N} \to \mathbb{N}$. (Recall that $\mathbb{N}$ denotes the natural numbers.) Here is the map.

$$\begin{array}{ccccccccccccc} \mathbb{N} & 0 & 1 & 2 & \cdots & 27 & 28 & 29 & 30 & 31 & 32 & \cdots \\ \mu\downarrow & \downarrow & \downarrow & \downarrow & \cdots & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \cdots \\ \mathbb{N} & 0 & 0 & 0 & \cdots & 0 & 3 & 1 & 16 & 28 & 0 & \cdots \end{array} \tag{1}$$

The domain and range of μ are both $\mathbb{N}$. The image of μ is the set $\{0, 1, 3, 16, 28\}$. Almost all the numbers are mapped to zero. The set of those numbers not mapped to zero is called the **support** of the function. The support of μ is $\{28, 29, 30, 31\}$.

This is a slightly interesting example because μ as originally defined is actually not a function because some 4-year periods during which the century ends and another starts can be missing a leap day in which case $\mu(29) = 0$. In other words, $\mu(n)$ has two possible values at $n = 29$, and so it is not a function. However, if we take Formula (1) as the definition of μ then of course it is a function. Nevertheless, this example shows how a mathematician can easily make a mistake and define what appears to be a function but actually is not.

4.2 Ranges and Images of Maps.

Why bother with a range when the image is all that is relevant to a map? There are many reasons, one is that the precise image might not be known with certainty: in Example 4.1.4 we knew only that the image was from 0 to the stopping time at 100 mph, whatever that was. Another reason is that we might have several maps from A to B whose images are not necessarily the same. By allowing the range to be any set enclosing all possible images, we can sensibly talk of maps (plural) from A to B.

4.3 **Restrictions of Maps.** Again, let $\alpha : A \to B$, and suppose X is a subset of A, $X \subset A$. Since α is defined on every element of A it is in particular defined on every element of X, and so α maps the subset X into B. The image of X under α is denoted by $\alpha(X)$ and is the set $\{b \in B \vdash \exists x \in X,\ \alpha(x) = b\}$.

The map α regarded as a map on X is call the **restriction** of α to X. It is denoted by $\alpha|X$ or $\alpha|_X : X \to B$. Even though $\alpha|X$ has the same values as α, it is nevertheless formally distinct from α. However, if the distinction is irrelevant, then α is used regardless.

Recall from Chapter 2 that the empty set, $\emptyset$, is a subset of every set. So, we can set $X = \emptyset$ in the above and deduce that α maps the empty set to itself; that is, $\alpha(\emptyset) = \emptyset$.

4.3.1 **Example** Let $A = \{0, 1, 2, 3, 4, 5, 6\}$, let $X = \{1, 4\}$, let B be the set of integers divisible by 3. This set is usually denoted by $3\mathbb{Z}$. let $\alpha : A \to 3\mathbb{Z}$ be the map defined as

$$\alpha : \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 0 & 9 & 18 & 0 & 9 & 18 & 0 \end{array}$$

Then,

$$\alpha|_X : \begin{array}{cc} 1 & 4 \\ \downarrow & \downarrow \\ 9 & 9 \end{array}$$

The image of α is $\{0, 9, 18\}$, and the image of $\alpha|_X$ is $\{9\}$. As it happens, the image of α is actually in the smaller set $9\mathbb{Z}$ (the integers which are multiples of 9). We can equally well regard α as mapping A into $9\mathbb{Z}$; mathematicians would not normally draw a distinction between α mapping into $3\mathbb{Z}$ or into $9\mathbb{Z}$.

4.4 **Inverse Maps.**

Let $\alpha : A \to B$ and let $Y \subset B$. The set of elements mapped into Y by α is denoted by $\alpha^{-1}(Y)$. and is called the **inverse image** of Y under α. This inverse map, α^{-1}, is defined only on the image, $\alpha(A)$. Beyond the image it is undefined. Hence, α^{-1} is a map on all of B only if the image and the range are the same. Furthermore, the inverse is not a map to A, but subsets of A.

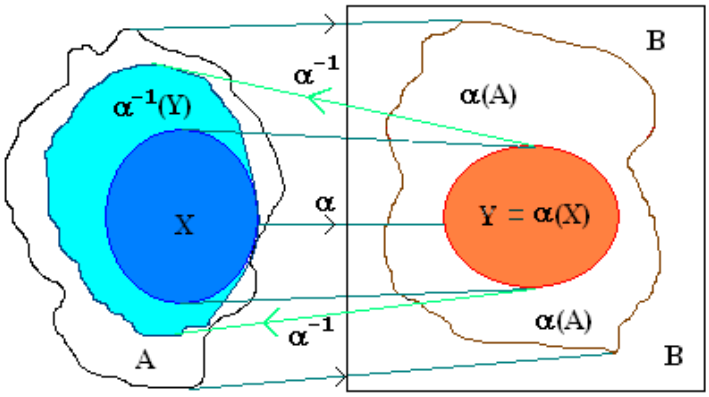


Diagram 4.4. The Inverse Map.

Diagram 4.4 shows a typical map and its inverse. The map is $\alpha : A \to B$. The big, irregular area on the left is the domain, A. The range is B, and the diagram illustrates a case where the image is not whole of B; the image is the irregular area bounded in red within B. The diagram also illustrates a case where a subset X of A is mapped to a subset of B whose inverse image is not X. In fact $\alpha(X) = Y \subset B$, and $\alpha^{-1}(Y) \neq X$, although necessarily $X \subset \alpha^{-1}(Y)$.

Useful though Diagram 4.4 is it obfuscates a small complication by its very simplicity. To explain, I shall take Example 4.3.1 above as a concrete example. The inverse map in this example is $\alpha^{-1} : \{0, 9, 18\} \to$

subsets of A given by

$$\alpha^{-1}: \quad \begin{array}{ccc} 0 & 9 & 18 \\ \downarrow & \downarrow & \downarrow \\ \{0,3,6\} & \{1,4\} & \{2,5\} \end{array}$$

Hence the image of α^{-1} is $\{\{0,3,6\},\{1,4\},\{2,5\}\}$. Straightaway, we see that α^{-1} does not map to A, but to subsets of A. So what can we take as the range of α^{-1}? The simplest range, and the one most typically used, is so-called power set of A. The **power set** of a set is set of all its subsets. It is written 2^A (for reasons that will become clear in §4.4.6). In the case of our example 4.3.1, the power set contains 2^7 elements which are all the subsets of $A = \{0,1,2,3,4,5,6\}$:

$$2^A \;=\; \{\emptyset,\{0\},\{1\},\ldots,\{0,1\},\{0,2\},\ldots,\;\ldots,\{0,1,2,3,4,5\},\{0,1,2,3,4,6\},\ldots,\{0,1,2,3,4,5,6\}\}$$

Notice that I did not forget to include the empty set (the first listed) nor did I forget A itself (every set is a subset of itself).

The reason to pick the power set as the range is simple: the inverse of every conceivable map from A to anything else is a map into 2^A. So knowing nothing about a map besides that it maps A into some other set, we at least know that its inverse is a map into 2^A. To summarize

$$\alpha^{-1}: \alpha(A) \;\to\; 2^A$$

This should make it pretty obvious that if $x \in A$, we should not expect $\alpha^{-1}(\alpha(x))$ to be equal to x. Yes, $\alpha^{-1}(\alpha(x))$ could conceivably be equal to $\{x\}$, but it certainly could not equal x. But, alas, this a frequent assumption made by mathematicians. The reason is this. The conjunction of α^{-1} with α to a mathematician looks just like the conjunction of $1/y$ with y to a high school student. Just as there is an urge to cancel, to say that $(1/y)y = 1$, there is a compulsion to think of $\alpha^{-1}\alpha$ as the identity map.

The idea that $\alpha^{-1}\alpha(x)$ could be identified with x is just too good a prospect for mathematicians, and the truth is that the situations where this identification becomes possible in a consistent and well-defined way are too important for us to pass up the opportunity thus afforded.

As a first step toward this ideal, we re-examine Diagram 4.4. As I said this diagram obfuscated a complication; the complication is that, as it stands, $\alpha^{-1}(Y)$ is not a subset of A, rather it is a subset of 2^A. So how can we say that $X \subset \alpha^{-1}(Y)$? Well, strictly we cannot, but we would like to. So this is what we do. We say α^{-1} maps subsets of the image to subsets of A. Specifically, for all subsets $Y \subset B$, we redefine

$$\alpha^{-1}(Y) \;:=\; \bigcup\{\alpha^{-1}(y) \vdash\; y \in Y\}$$

That big "U" means union of all the sets inside the curly braces. What this does is to set $\alpha^{-1}(Y)$ to be not the collection of subsets in the inverse image but instead the union of these subsets. To jump ahead a little bit, all those subsets $\alpha^{-1}(y)$ are not only distinct they are also non-overlapping, and their union is simply

$$\alpha^{-1}(Y) \;=\; \{a \in A \vdash\; \alpha(a) \in Y\}$$

We now move to the second step toward achieving the ideal.

4.4.4 Onto Maps, 1-1 Maps, and Bijections.

Let $\alpha : A \to B$.

If $\alpha(A) = B$ (or equivalently, the image is the entire range) then the map α is called **onto**, and it is said to be onto its range.

If α satisfies the condition $\alpha(a) = \alpha(a') \;\Rightarrow\; a = a'$ then α is said to be **1-1** (pronounced "one-to-one"). What this means is that α does not map two distinct points to the same point in B. That is, given any point in the image there is one and only one point mapped to it.

If α is both onto and 1-1, then it is called a **bijection**, and is said to be **bijective**.

There is a simple theorem which states that a bijection has a unique map which is both a left and right inverse to it. That is, if α is bijective, then there exists $\beta : B \to A$ such that $\alpha(\beta(b)) = b, \forall b \in B$, and

$\beta(\alpha(a)) = a, \forall a \in A$. This β is denoted by (you guessed it) α^{-1}, and is called the "inverse map," except this time the inverse map really is a map.

A bijection implies that the domain and the range of the map are equivalent in some sense. If the map is nothing more than a map then the equivalence is merely that every element in one is associated with exactly one element in the other. But usually, a bijection is something more.

This is a bit of a digression, but an important one. Let us take as an example a map between two groups, $\alpha : G \to H$. Such a map would normally only be of interest if it respected the group operations in the two groups. What does this mean? Suppose the group product in G is "$\circ$", and the product in H is "$\cdot$". Then, the map respects the product in the groups if $\alpha(g_1 \circ g_2) = \alpha(g_1) \cdot \alpha(g_2)$ for every $g_1, g_2 \in G$. The map should also respect the identity: it should map the identity of G into that of H. With these properties, a bijection α is called an **isomorphism** of groups. Two groups connected by an isomorphism are called **isomorphic**. Isomorphic groups are equivalent in a very strong sense. To all intents and purposes they are the same group. I discussed this in §1.5.4 when I defined the abstract group—isomorphic groups are all manifestations of the same abstract group.

This example of maps connecting groups extends to other types of sets. Another example is rings. A bijective map between two rings is called an isomorphism if it is an isomorphism on the additive group of the ring and also respects the multiplication of the ring. Let the map be $\alpha : R \to S$. The map should obey $\alpha(r_1 + r_2) = \alpha(r_1) + \alpha(r_2)$, $\alpha(0) = 0$, and $\alpha(r_1 r_2) = \alpha(r_1)\alpha(r_2)$. As in groups, isomorphic rings are in the same sense all manifestations of the same abstract ring—there is no ring-theoretic difference between them.

4.4.5 Exercises

(i) Show that if $\alpha : A \to B$ is 1-1 then $\alpha : A \to \alpha(A)$ is bijective. For this reason, 1-1 maps are sometimes called embeddings.

(ii) Show that if A is finite and $\alpha : A \to A$, then α is onto iff α is 1-1 iff α is bijective.

4.4.6 The Power Set and Partitions.[6]

Recall from §4.3 that the set of subsets of a set A is denoted by 2^A and is called the **power set** of A. The inverse map α^{-1} acting on subsets of B is a map $\alpha^{-1} : 2^B \to 2^A$ whereas α^{-1} acting on single elements of B is a map $\alpha^{-1} : B \to 2^A$. As mentioned above, these two meanings of α^{-1} should strictly have different symbols, but in practice the meaning of α^{-1} is clear from the context.

A **partition** of a set A is any set of non-intersecting subsets of A whose union is A. A partition of A is said to be finite if it contains only a finite number of subsets of A. There is a simple criterion for P being a partition for A when A is a finite set. $P = \{A_1, A_2, \ldots, A_n\}$ is a partition of A iff each A_i is a subset of A and $|A_1| + |A_2| + \cdots + |A_n| = |A|$.

Given two partitions P_1 and P_2 of A, P_1 is said to be **finer** than P_2 if given any $S \in P_2$ there is a subset of P_1 which is a partition for S. That is, every set in the P_2 partition is partitioned by a collection of sets from P_1. In symbols, this is written $P_1 \prec P_2$.

Let $\alpha^{-1} : B \to 2^A$. This map defines a partitioning of A given by $P_\alpha = \{\alpha^{-1}(b) \vdash b \in B\}$. Hence, via the inverse α^{-1}, every map defines a partition of its domain. One can see from the definition of P_α that α^{-1} assigns an element $b \in B$ its partition set $\alpha^{-1}(b)$. Therefore, $\alpha^{-1} : B \to P_\alpha$.

Example. Take the example of section 4.3 again. (See Diagram 4.2.) The partition P_α of A is

$$P_\alpha \;=\; \{\{0,3,6\},\ \{1,4\},\ \{2,5\},\ \emptyset\}$$

$$\alpha^{-1}(0) = \{0,3,6\}, \quad \alpha^{-1}(9) = \{1,4\}, \quad \alpha^{-1}(18) = \{2,5\}$$

Note that, as required, none of the sets in P_α intersect, and their union equals $\{0,3,6\} \cup \{1,4\} \cup \{2,5\} = \{0,1,2,3,4,5,6\} = A$. The empty set, $\emptyset$, is in the partition because it is the inverse image of all those elements of the range not in the image, that is, all those elements of B not equal to 0, 9, or 18.

[6] This section can be skipped.

We have seen that if $\alpha : A \to B$ then there is a map $\alpha^{-1} : B \to P_\alpha$ where P_α is the partition of A defined by α. This α^{-1} map is onto and if α is onto then α^{-1} is also 1-1, and hence bijective. (See Exercise 4.5*(iii).*) Let its inverse be $\bar{\alpha} : P_\alpha \to B$. It is given by: $\bar{\alpha}(S)$ equals the one and only element in $\alpha(S)$. There is also a map $\nu_\alpha : A \to P_\alpha$ given by $\nu_\alpha(a) = \alpha^{-1}\alpha(a)$. This is called the **natural map** from A to P_α and exists whether or not α is onto. The action of ν_α is simple; it assigns each element of A the partition set it belongs to. We therefore have a bijection $\bar{\alpha} : \nu_\alpha(A) \to B$ whenever α is onto.

4.4.7 **Exercise.** Why must α be onto for $\alpha^{-1} : B \to P_\alpha$ to be bijective? Can the requirement be relaxed?

4.5 **Equivalence Relationships.**[7] Equivalence relationships are intimately connected with partitions. Take any set X, and let $S(a, b)$ be a (true of false) statement about any members a, b of X. For example, X might be the cities of North America and $S(a, b)$ might be the statement: "There is a direct rail link from a to b." The statement need not relate a and b. For instance, S might be the statement "a is over 200 feet and b is over 400 feet above sea-level."

Of all the possible statements involving two members of a set, there is one type which is particularly important, the equivalence relationship. An equivalence relationship satisfies the following axioms:

(i) $S(x, x)$ Reflexivity

(ii) $S(x, y) \Rightarrow S(y, x)$ Symmetry

(iii) $S(x, y)$ & $S(y, z) \Rightarrow S(x, z)$ Transitivity

How are equivalence relationships connected to partitions? One of the most elementary theorems of mathematics is that an equivalence relationship defines a partition and *vice versa*. To see this, suppose we are first given the equivalence relationship, S. Then, a typical set in the partition is all those elements of X which satisfy $S(x, y)$ for some $x \in X$. Conversely, if we are given a partition $\mathcal{P}$, we define $S(x, y)$ to be the statement x and y are members of the same set in the partition $\mathcal{P}$.

4.6 **Maps on Combinations of Sets.** The following statements are exercises. Their proof requires only simple but clear logic. Let $\alpha : A \to B$, and let X and Y be subsets of A, and U and V subsets of B.

(i) $\alpha(X \cup Y) = \alpha(X) \cup \alpha(Y)$

(ii) $\alpha(X \cap Y) \subset \alpha(X) \cap \alpha(Y)$

(iii) $\alpha^{-1}(U \cup V) = \alpha^{-1}(U) \cup \alpha^{-1}(V)$

(iv) $\alpha^{-1}(U \cap V) = \alpha^{-1}(U) \cap \alpha^{-1}(V)$

The exception here is *(ii)*. Consider $X = \{1, 2\}$, $Y = \{3, 4\}$, and let $\alpha(1) = \alpha(2) = \alpha(3) = \alpha(4) = 1$. Then, $\alpha(X \cap Y) = \alpha(\emptyset) = \emptyset$, whereas $\alpha(X) \cap \alpha(Y) = \{1\}$. That is, $\alpha(X \cap Y) \neq \alpha(X) \cap \alpha(Y)$.

4.7 **Reflexive Maps. Composition of Maps.** An important case is where a set is mapped into itself, $\alpha : A \to A$. In this case, powers of α can be defined to mean successive applications of α. Thus, $\alpha^3(a) = \alpha(\alpha(\alpha(a)))$. If $\alpha : A \to A$ is bijective, then powers of α are defined for positive and negative integers. Thus, $\alpha^{-3}(a) = \alpha^{-1}(\alpha^{-1}(\alpha^{-1}(a)))$. The **identity** map on A is the map which assigns each element to itself, which is to say that $\alpha(a) = a$ for every $a \in A$. The identity map is sometimes denoted by id: $A \to A$, sometimes by $\imath : A \to A$, and often by $1 : A \to A$, or $1_A : A \to A$ if it could be mistaken for an identity on

[7] This section can be skipped.

another set. It is the unique map which satisfies $\imath\alpha = \alpha\imath = \alpha$ for all maps $\alpha : A \to A$. Conventionally, the zeroth power of an arbitrary map is defined to be the identity map.

Now suppose that $\alpha : R \to S$ and $\beta : S \to T$. Then the map $\beta\alpha : R \to T$ is defined by $\beta\alpha(x) = \beta(\alpha(x))$. This example constructs a map from two others by defining a product between maps. This is the normal product law for maps, and is called the **composition** of the maps. The product is usually represented positionally as in $\beta\alpha$, but sometimes (often to distinguish it from multiplication of values) a circle is interposed to emphasize that composition of the two maps is intended: $\beta \circ \alpha$. However, one can always be unambiguous by merely writing what the composition means, namely $\beta(\alpha(x))$. The drawback of this last notation is that it requires a possibly irrelevant variable such as x. In all three notations, the rightmost map acts first and the leftmost acts last. This convention is to keep the order of maps in the product notation the same as that in the functional notation; clearly, $\beta(\alpha(x))$ means β applied to the result of $\alpha(x)$. That is, α is applied first then β. Since this is contrary to the European method of reading left to right, you should always take care to read compositions of maps backwards.

Two maps α and β are said to be equal, written $\alpha = \beta$, iff they have the same domains, and $\alpha(x) = \beta(x)$ for every x in the (common) domain.

Examples

(i) Let α map A to itself. Then, $\alpha^i\alpha^j = \alpha^{i+j}$, $\forall i, j \geq 0$, and if α is bijective then this holds for all integers i, j.

(ii) Let $\alpha : A \to B$, and $\beta : B \to C$. Then, $\beta|_{\alpha(A)} \circ \alpha = \beta \circ \alpha$.

(iii) Let $\alpha : A \to B$ and $\beta : B \to C$ be bijections. Then $(\beta\alpha)^{-1} = \alpha^{-1}\beta^{-1}$. That is, inversion reverses the order of the maps.

The composition of maps is associative. This is so important and so simple to prove that I prove it.

4.8 **Proposition** Let $\alpha : A \to B$, $\beta : B \to C$, and $\gamma : C \to D$ be maps. Then, $(\gamma\beta)\alpha = \gamma(\beta\alpha)$.

Proof. First I shall show that $\gamma(\beta\alpha)$ and $(\gamma\beta)\alpha$ have the same domains.

For all $a \in A$, let $b = \alpha(a)$. The map β is defined for the whole of B, so let $c = \beta(b)$. The map γ is defined for all of C, so let $d = \gamma(c)$. Hence $d = \gamma(\beta(\alpha(a)))$ exists for all $a \in A$. Therefore, the domain of the map $\gamma(\beta\alpha)$ is A.

For all $b \in B$, let $c = \beta(b)$. Again, γ is defined at c, so let $d = \gamma(c)$. Then, $d = \gamma\beta(b)$. Therefore, $\gamma\beta : B \to D$. Let $a \in A$, then $\alpha(a) \in B$. Hence, $(\gamma\beta)(\alpha(a)) \in D$. So, $(\gamma\beta)\alpha : A \to D$; that is, the domain of $(\gamma\beta)\alpha$ is also A. QED (Equal domains).

Now I shall show that $\gamma(\beta\alpha)(a)$ and $(\gamma\beta)\alpha(a)$ are equal for every $a \in A$. By definition of composition, $\gamma(\beta\alpha)(a) = \gamma(\beta(\alpha(a))$. Now, let us evaluate $(\gamma\beta)\alpha(a)$. By definition of composition, $(\gamma\beta)\alpha(a) = (\gamma\beta)(\alpha(a))$. Let $\alpha(a) = b$. Then, $(\gamma\beta)(\alpha(a)) = (\gamma\beta)(b) = \gamma(\beta(b))$, by definition of composition. But $\gamma(\beta(b)) = \gamma(\beta(\alpha(a)))$. □

4.9 **Map Diagrams and Commutivity of Diagrams.** Compositions of maps can be depicted graphically with map diagrams. Consider the diagram below.

$$\begin{array}{ccc} & A & \\ {\scriptstyle\alpha}\swarrow & & \searrow{\scriptstyle\gamma} \\ B & \xrightarrow[\beta]{} & C \end{array}$$

Diagram 4.8.

This diagram declares the existence of maps $\alpha : A \to B$, $\beta : B \to C$, and $\gamma : A \to C$. The diagram is said to be **commutative** if $\beta\alpha = \gamma$; sometimes, the map β is said to **agree** with α and γ, or α is said to agree with β and γ.

Map diagrams can get quite complex, especially in certain subjects such as algebraic topology. Commutivity of more complex diagrams means that no matter which route you take in the diagram by following

arrows, the result is the same. Sometimes, the arrows allow for more than one route starting from several points. A diagram may be declared commutative at a particular point which means that all routes starting at that point to any given point give the same result.

Quite often we are given the maps $\alpha : A \to B$ and $\gamma : A \to C$ as in Diagram 4.8, but not the map β, although we might suspect that a map exists from B to C which agrees with α and γ. Such a map does exist iff the partition of A due to α is finer than the partition due to γ, that is, $P_\alpha \prec P_\gamma$. (See §3.3.6.) When this is the case, the map is given by $\beta(b) = \gamma\alpha^{-1}(b)$. The map γ is here regarded as mapping the subset $\alpha^{-1}(b)$ to the single element in its image.

4.10 Conjugation and Similarity.[8]

We are given the two maps $\rho : A \to B$ and $\alpha : A \to A$. We would like to deduce the existence of β mapping B to itself which agrees with α. (See diagram 8 below.)

$$\begin{array}{ccc} & \alpha & \\ A & \longrightarrow & A \\ \rho\big\downarrow & & \big\downarrow\rho \\ B & \longrightarrow & B \\ & \beta? & \end{array}$$

Diagram 4.9.

In this situation agreement means that $\beta\rho = \rho\alpha$. I shall derive a sufficient condition for the existence of β.

4.10.1 Proposition[9]

A sufficient condition that β exists such that Diagram 4.9 is commutative is that $P_\rho \prec P_\alpha$.

Proof. If we set $\gamma = \rho\alpha$, we have the same situation as in diagram 7 with C set to B and α set to ρ. Hence, β can be defined such that Diagram 4.9 commutes iff the partition defined by ρ is finer than that defined by γ. To see what is the partition due to γ trace an element $b \in B$ backwards through the maps ρ and α. The element b is in the image of the set $\rho^{-1}(b) \subset P_\rho$, and this set is the image of the union of the sets $\alpha^{-1}(a) \subset P_\alpha$ for every $a \in \rho^{-1}(b)$. Hence, the inverse image of b in P_γ is the set

$$\bigcup_{a\in\rho^{-1}(b)} \alpha^{-1}(a) \;=\; \alpha^{-1}\left(\bigcup_{a\in\rho^{-1}(b)} \{a\}\right) \;=\; \alpha^{-1}\rho^{-1}(b)$$

I have two points to make about the above derivation. Firstly, the map α^{-1} on the left is the map $\alpha^{-1} : A \to P_\alpha$. Thereafter it is the inverse image map on subsets of A. Secondly, I used a generalization of statement *(iii)* in 4.6. There it is stated that an inverse map preserves the union of two sets, and hence of any finite number of unions. In fact, it preserves the union of any number of sets, including an infinity of them.

So a typical set in the partition P_γ is $\alpha^{-1}\rho^{-1}(b)$ where $b \in B$. We can deduce that $P_\rho \prec P_\gamma$ and β can be defined if and only if for every $y \in B$, there is a b also in B such that $\rho^{-1}(y) \subset \alpha^{-1}\rho^{-1}(b)$. This is so if and only if $\alpha(\rho^{-1}(y)) \subset \rho^{-1}(b)$. Hence, α must map sets in the partition P_ρ into sets in P_ρ.

Now suppose $P_\rho \prec P_\alpha$ which is the condition in the statement of the proposition. Then α maps sets of P_ρ to single points, and single points must be in partition sets no matter what partition it is. In particular, α maps sets of P_ρ into sets of P_ρ as required. □

Although $P_\rho \prec P_\alpha$ is only a sufficient condition for the existence of β, it is simple and it is a condition that is frequently satisfied. For instance, when the map ρ is a bijection, the partition P_ρ consists of single points, so will automatically be finer that any other partition, and β will exist for every map α. Indeed, $\beta = \rho\alpha\rho^{-1}$.

---

[8] This section can be skipped.

[9] §4.4.6 is a prerequisite for this proposition.

4.10.2 **Conjugation.**
If the vertical arrows are reversed we get diagram 9.

$$\begin{array}{ccc} & \alpha & \\ A & \longrightarrow & A \\ \tau\uparrow & & \uparrow\tau \\ B & \longrightarrow & B \\ & \beta? & \end{array}$$

Diagram 4.9.2.

In this situation β can always be defined to agree with α. This is how.

For each $b \in B$, pick any member y of the set $\tau^{-1}\alpha\tau(b)$, and define $\beta(b) := y$. I leave it as an exercise to verify that β is a properly defined map and agrees with α. But one needs to be aware that β is not uniquely defined–it depends on the choice of y for each $b \in B$.

Assume additionally that τ is bijective, then map $\tau^{-1}\alpha\tau$ is uniquely defined, and is called the **conjugation** of α by τ, and is often written α^τ. When the maps α and τ are matrix transformations, α^τ is called instead the **similarity** transformation of α by τ. When we are given the map α but not β, we say that β is **induced** by α (under τ). Since τ is bijective, we can also say that α is induced by β when it is β that is given and not α.

4.10.3 **Exercise.** Suppose we are given maps $\alpha, \beta, \rho, \tau : A \to A$. Since the domains and ranges are all the same set we can define arbitrary compositions of these maps. If ρ and τ are bijections show that $(\alpha\beta)^\rho = \alpha^\rho\beta^\rho$, and that $(\alpha^\rho)^\tau = \alpha^{\rho\tau}$. In other words, conjugation satisfies the usual law of exponents.

4.11 **Axiom of Choice.** I have a confession. Where I argued above that $\beta(b)$ could always be defined by $\beta(b) = y$ where y is any element of $\tau^{-1}\alpha\tau(b)$, I made an unwarranted assumption. However, the assumption is subtle, so subtle that mathematicians unconsciously took it for granted for centuries before it was seen for what it was, an unsupported assumption. The assumption occurs in the phrase "where y is any element of $\tau^{-1}\alpha\tau(b)$". It is assumed that we can construct a set, Y, consisting of one element each chosen from the set of sets $\{\tau^{-1}\alpha\tau(b) \vdash b \in B\}$. But, we are not given a rule for constructing such a set. Furthermore, without further information about the maps τ and α, we cannot devise such a rule. This assumption is equivalent to what is now called the Axiom of Choice:

Statement of the Axiom of Choice. "Given any set P whose elements are all non-empty sets, there exists a set consisting of one element from each element of P."

For example, let $P = \{\{1,2,3,6\},\ \{4,9\},\ \{5,7\}\}$ then according to the Axiom of Choice a set exists which contains one element from each of the sets $\{1,2,3,6\}$, $\{4,9\}$, and $\{5,7\}$. Well, obviously, such a set exists, $\{1,9,5\}$ to give one example.

When all sets are finite as in this example, that such a set exists is so obvious that most people would say that the Axiom of Choice is self-evident. The problem really pertains to infinite sets of sets. The Axiom of Choice says that we can make an infinity of choices, and this is clearly impossible even in principle since we are all mortal. Nevertheless, we can specify any particular choice, and there is no choice that we cannot make. Furthermore, in many cases, we can specify an infinity of choices by laying down a rule which uniquely determines which element to pick from each element of P. For example, if P were a set of sets of numbers, we could specify the rule: "Pick the smallest number from each set in P." For these and other reasons, most mathematicians do assume the Axiom of Choice, as shall I.

There is an interesting addendum to the Axiom of Choice. Not so long ago (in living memory), another common and unwarranted assumption was uncovered by a fellow by the name Max Zorn (1906-1993). It is now called Zorn's Lemma. To explain the lemma, we need some definitions.

A set S is said to be **partially ordered** if a relationship "$a \preceq b$" exists between some pairs of elements (a, b) of S. You can think of this relationship $a \preceq b$ as meaning "a is either below or equal to b."

This relationship obeys the following rules:

(i) $a \preceq a, \quad \forall a \in S$ (Reflexivity: a is not below itself)
(ii) $a \preceq b \ \& \ b \preceq c \ \Rightarrow \ a \preceq c$ (Transitivity: if a is below b and b is below c then a is below c)
(iii) $a \preceq b \ \& \ b \preceq a \ \Rightarrow \ a = b$ (No cycles)

A **chain** in S is a sequence $a \preceq b \preceq c \preceq \cdots$ possibly infinite in length. So a chain is a sequence which at each link or step it either pauses or ascends.

Example Let S be the set of all integer divisors of 36, and define $a \preceq b$ to be "a divides b". Then,

$$S \ = \ \{1, 2, 3, 4, 6, 9, 12, 18, 36\}$$

Here are some of the chains in S:

$$1 \preceq 2 \preceq 4 \preceq 12 \preceq 36$$
$$3 \preceq 9 \preceq 36$$
$$1 \preceq 9 \preceq 9 \preceq \cdots \preceq 9 \preceq 18$$

Suppose we have a partially-ordered set A and let $C \subset A$ be a chain. If there exists $x \in A$ such that $C \preceq x$ (that is $a \preceq x$ for every $a \in C$), then x is said to be an **upper bound** on C.

Let $y \in A$. If there are no elements above y in A, then y is called a **maximal** element of A.

Statement Z. A partially ordered set in which every chain has an upper bound has a maximal element.

Well, the stinger is that Zorn proved that Statement Z is logically equivalent to the Axiom of Choice!

Max Zorn was a German mathematician, and an exceptional man. He discerned the true nature of the Nazis very early, and though not Jewish, he had to flee his homeland in 1933 because of his opposition to them. He settled in the US, and ended up as professor at the University of Indiana. His life there seems to have been quite settled, an anodyne for his earlier existence at the mercy of the volatile politics of Hindenburg Germany.

Chapter 5.
Sets Gone Wild!

Georg Cantor (1845-1918) was a German mathematician who made important contributions to a branch of mathematics called Fourier Analysis. Fourier Analysis has great practical value; it is used in almost all spheres of science and technology including electric power transmission, orbital dynamics, acoustics, predator-prey population cycles, CAT scans, and analyzing the cosmic background radiation. Through his research into Fourier Analysis, Cantor was drawn into the question of what is the number of elements in certain, infinite sets. Well, the obvious answer is, of course, infinity. But, Cantor wanted to know if there was more to it. Were there bigger and smaller infinities? How could one know if one infinity was equal to another?

The difficulty facing Cantor was that he needed to define equality of number without counting; after all, a set having an infinity of elements cannot be counted. Cantor's first achievement was to reduce the act of comparing numbers to its essence: he saw that two sets, A and B, have the same number of elements if the elements in A can be placed in a well-defined correspondence to the elements in B, and *vice versa*. This can be more succinctly stated in the language of Chapter 4. According to Cantor:

Two sets have the same number of elements if one can find a bijective mapping from one to the other.

Let us check that this definition makes sense when applied to the familiar situation of two finite sets. Let us take the two sets to be A and B where

$$A := \{1, 2, 3, 4\}, \quad B := \{\text{pony}, \text{owl}, \text{cat}, \text{dog}\}$$

We define a map $\alpha : A \to B$ by

$$\alpha : 1 \mapsto \text{cat}$$
$$\alpha : 2 \mapsto \text{dog}$$
$$\alpha : 3 \mapsto \text{pony}$$
$$\alpha : 4 \mapsto \text{owl}$$

the inverse map α^{-1} is the same diagram but with the arrows reversed. The map α is a bijection. Therefore, by Cantor's criterion the two sets A and B have the same number. Of course we know this because we can see by counting that they have four elements each.

What happens when the two sets do not have the same number of elements? Suppose B contains the additional element 'fox' so that $B = \{$ cat, pony, owl, dog, fox $\}$. Now, there is no bijective map $\alpha : A \to B$ because any map $A \to B$ would necessarily omit at least one member of B. We intuitively know this, of course, because we can count, enabling us to see that there are more elements in B than A, and so one element of B must be unmapped no matter what bijection we choose from A to B.

Cantor's genius was to use bijections as a method of comparing numbers in sets which did not require counting the elements in the sets.

An important aspect of Cantor's definition is the fact that bijection between two sets is an equivalence relationship on sets. This means that

(i) there is a bijection between A and itself;

(ii) if A and B are bijective, then so are B and A; and

(iii) if A and B are bijective, and B and C are bijective, then so are A and C.

This implies (see §4.5) that sets can be partitioned into classes such that each class contains all sets which are mutually bijective. (Informally, we would say that a class consists of all sets which have the same number of elements.)

5.1 Cardinality

We define the **cardinality** of a set to be the bijective class that it belongs to. Informally, we identify cardinality of a set with the number of elements in the set, but since we cannot count these elements in infinite sets, we must replace the concrete specification of the number of elements in a set A with the more

abstract specification of the collection of sets to which A is bijective. The identification of a number n with all those sets which have n elements goes back to Russell and Whitehead's Principia Mathematica. Even though this work failed to achieve its primary goal, it nevertheless brought forth important new ideas.

We shall denote the cardinality of a set A by $|A|$. (Some authors use $\#A$, and others use card$A$.) In the case of the finite sets, we can safely identify the cardinality with the number of elements in the set, and we say that the cardinality is finite. For example, if $A = \{1, 4, 3\}$, then $|A| = 3$. When a set is not finite, we say its cardinality is infinite.

It is important to bear in mind that the definition of cardinality is merely a means of extending our commonplace idea of number to infinite sets.

We define an order on sets by insisting that if $A \subset B$, then $|A| \leq |B|$. If there is a 1-1 map from A to B, but no bijection between A and B, then we say $|A| < |B|$. In our commonplace language of numbers we would say "A has fewer members than B." This inequality introduces an ordering to cardinalities:

$$0 < 1 < 2 < \cdots < 1904787145 < 1904787146 < \cdots < \infty < ??$$

As the question marks in the above sequence remind us, the question now is whether there is only one infinite cardinality or many. We know that the natural numbers, $\mathbb{N}$, are infinite, so the question is: Are there infinite sets which are strictly smaller or bigger than $\mathbb{N}$?

If we assume that any set can be counted (possibly without end), then it is easy to eliminate infinite sets smaller than $\mathbb{N}$ – by counting such a set, we are implicitly assuming $\mathbb{N}$ can be mapped 1-1 into it, and so it is at least as numerous as $\mathbb{N}$. Because $\mathbb{N}$ is therefore the smallest possible infinity, we give its cardinality a special symbol, $\aleph_0$, which is the Hebrew letter "aleph" with subscript 0. Any set of cardinality $\aleph_0$ or less, that is any set which is finite or bijective to the natural numbers, is called **countable**. All other sets, if they exist, are called **uncountable**.

What about the integers, $\mathbb{Z}$? These include not only the natural numbers but also the negative integers. So, we might be inclined to say that $|\mathbb{N}| < |\mathbb{Z}|$. But this is false. In comparing infinite sets, we cannot rely on our intuitions which work so well in our finite world, we must stick to Cantor's criterion. In fact, one can easily construct a bijection between $\mathbb{N}$ and $\mathbb{Z}$. Define $\alpha : \mathbb{Z} \to \mathbb{N}$ by

$$\alpha(x) \;:=\; \begin{cases} 2x & \text{if } x \geq 0 \\ -2x + 1 & \text{if } x < 0 \end{cases}$$

The map α does the following assignments:

(i) the non-negative integers are mapped 1-1 onto the even natural numbers,

$$0 \mapsto 0, \quad 1 \mapsto 2, \quad 2 \mapsto 4, \quad 3 \mapsto 6, \quad \ldots, \quad \text{etc., and}$$

(ii) the negative integers are mapped 1-1 onto to the odd natural numbers,

$$-1 \mapsto 1, \quad -2 \mapsto 3, \quad -3 \mapsto 5, \quad -4 \mapsto 7, \quad \ldots, \quad \text{etc.}$$

By Cantor's criterion, we conclude that

$$|\mathbb{Z}| \;=\; |\mathbb{N}| \;=\; \aleph_0$$

But wait, how can this be? The natural numbers, $\mathbb{N} = \{0, 1, 2, \ldots\}$ account for only "half" of the integers, $\mathbb{Z}$, since $\mathbb{Z}$ also includes all the negative numbers; yet we are saying that the two have the same number of elements? Counterintuitive though this is, it is in fact a characteristic of all infinite sets. It serves as a warning that we must very careful in our definitions and assumptions when dealing with infinities.

5.2 **The Rationals, $\mathbb{Q}$.** Recall that the set of all fractions (all m/n where $m, n \in \mathbb{Z}$, $n \neq 0$) is called the rationals, and is denoted by $\mathbb{Q}$. Having shown that the integers had the same cardinality as the natural numbers, Cantor then asked what might be the cardinality of the rationals. People (including Cantor) believed that the rationals must be much greater in number than the set of integers. After all, if we mark off the integers on a line, between any two marks, there will be an infinitude of fractions. In fact, between any two fractions there are infinitude of fractions! For suppose we have two distinct, positive fractions a/b, and c/d, then $a+b/c+d$ is strictly between them. We can now repeat the argument taking as our starting point the two fractions a/b and $a+b/c+d$; we again see that they too have a fraction strictly between them. The same argument applies to $a+b/c+d$ and c/d. We can proceed thus indefinitely which shows that there must be an infinity of fractions between any two. (See Diagram 5.2.)

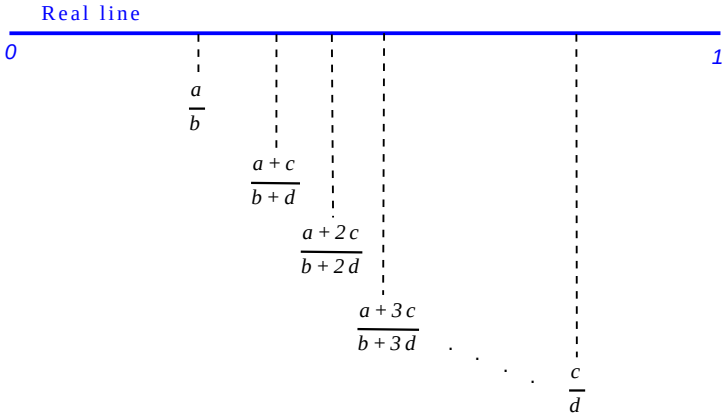


Diagram 5.2. Divisions of the line from 0 to 1.

Exercise Given positive integers a, b, c, d with $a/c < b/d$.

(i) Show that

$$\frac{a}{b} < \frac{ia+jc}{ib+jd} < \frac{c}{d}$$

where i and j are positive, but otherwise arbitrary, integers.

(ii) Let F_n be the set of fractions of the form i/j where $1 \leq i < j \leq n$. Show that by taking n large enough we can make the greatest difference between neighbors in F_n as small as we please.

For example, here is F_5 ordered by magnitude:

$$\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}$$

The above exercise explicitly constructs as many fractions as we want between any two fractions x, y lying between 0 and 1–there are no gaps in the fractions. Topologists would describe this situation by saying that the rationals are dense in the line.

So you can imagine everyone's surprise when Cantor discovered in 1873 that the rationals are countable! This is worth demonstrating.

We arrange the positive rationals in an infinite table with the denominators listed along the top, and the numerators listed down the left as shown below; the second number in bold type in each entry is the count. Notice that the count proceeds along diagonals of the table.

	1	2	3	4	...
1	$1/1$ **1**	$1/2$ **2**	$1/3$ **4**	$1/4$ **7**	...
2	$2/1$ **3**	$2/2$ **5**	$2/3$ **8**	$2/4$ **12**	...
3	$3/1$ **6**	$3/2$ **9**	$3/3$ **13**	$3/4$ **18**	...
4	$4/1$ **10**	$4/2$ **14**	$4/3$ **19**	$4/4$ **25**	...
⋮	⋮	⋮	⋮	⋮	⋱

This method of counting the positive rationals actually counts each fraction more than once. For example $1/2$ is counted many times, $1/2 = 2/4 = 3/6 = \cdots$. Yet, the natural numbers are numerous enough to count all the entries in the table despite these many repetitions. Therefore, the positive rationals are countable. Given this, it is an easy matter to show that the rationals (the positive and negative) rationals are countable.

People now began to wonder if every set might be countable. The next big test would come with the algebraic numbers.

5·3 Algebraic and Transcendental Numbers

At around this time (in the mid 1870's) another line of inquiry was almost stymied for lack of progress. To explain the problem we need to briefly visit ancient Greek mathematics. Pythagoras had found that $\sqrt{2}$ was not rational (such numbers are called "irrational"; the term is purely technical and implies nothing about a lack of logic in such numbers). In modern notation, Pythagoras had discovered that $\sqrt{2} \notin \mathbb{Q}$. So mathematicians in Cantor's time were well aware that the real numbers consisted of more than just the rationals. Mathematicians eventually generalized numbers like the rationals and $\sqrt{2}$ into something called the algebraic numbers.

A number, u, is said to be **algebraic** if $f(u) = 0$ for some polynomial f having integer coefficients. For example, if the polynomial is $f(x) = ax + b$ where $a, b \in \mathbb{Z}$ (the integers), and u is a root of this polynomial, $f(u) = 0$, then $u = -b/a$ is algebraic, and in fact, it is rational. So all rationals are also algebraic. But the converse is false. For example, consider $f(x) = x^2 - 2$. One root is $u = \sqrt{2}$. Hence, although $\sqrt{2}$, as Pythogoras discovered, is irrational, it is nevertheless algebraic.

A real number which is not algebraic is called **transcendental**, if any such numbers exist. At the time no transendentals were known, though there were a few numbers suspected of being transcendental like $\pi = 3.14159\cdots$. So the burning question was: Do transcendental numbers exist?

Regardless, in 1874, Cantor discovered that the algebraic numbers too were countable. Below is a proof. The proof counts the roots of all polynomials with integer coefficients.

It is helpful at this point to introduce the standard notation for the set of polynomials with integer coefficients; it is $\mathbb{Z}[x]$[10].

5·4 Theorem The Real Algebraic Numbers are Countable.

Proof. The set of real algebraic numbers is the set of all real roots to all polynomials in $\mathbb{Z}[x]$ (polynomials having integer coefficients). We shall start by counting $\mathbb{Z}[x]$.

To fix your mind on what we mean, below are two examples which can be referred to for illustration in the sequel; call them $g(x)$ and $h(x)$.

$$g(x) \;=\; 3 - 8x + x^3 \tag{5.4a}$$

$$h(x) \;=\; 48\,104\,547 \;+\; 38\,430\,728\,588\,x \;-\; 4\,900\,944\,499\,x^3 \;-\; 498\,832\,445\,923\,x^7 \tag{5.4b}$$

[10] More generally, $R[x]$ denotes the set of polynomials with coefficients in a ring, R.

The polynomial g is of degree 3 (usually called a "cubic"), and $h(x)$ is of degree 7. The degree is the highest occurring power of the unknown in the polynomial. The roots of g are

$$\begin{aligned} r_1 &= -3 \\ r_2 &= \tfrac{1}{2}\left(3+\sqrt{5}\right) = 2.6180339887\cdots \\ r_3 &= \tfrac{1}{2}\left(3-\sqrt{5}\right) = 0.3819660113\cdots \end{aligned}$$

Note that the number of roots is the same as the degree of the polynomial.

There are no expressions for the roots of h in terms of integers and surds of integers, but I calculated its roots with the help of a computer to 10 decimal places. Here they are

$$\begin{array}{ll} -0.0012517212, & \\ -0.6461676567, & \\ \ \ 0.6466002648, & \\ \ \ 0.3322459196 + 0.5649470238\imath, & \ \ 0.3322459196 - 0.5649470238\imath, \\ -0.3318363630 + 0.5649341892\imath, & -0.3318363630 - 0.5649341892\imath. \end{array}$$

The first three roots of h are real, and the last four are complex (they have imaginary parts indicated by $\imath$, the Greek *iota* which conventionally stands for $\sqrt{-1}$). Again, the number of roots is seven which is also the degree of the polynomial. This is not a coincidence. It was a fact well-known to Cantor that a polynomial of degree n has n roots, some of which may be equal to others. So the number of distinct roots cannot exceed the degree of the polynomial. As you can see in the example, the roots can be complex numbers. Just the same, we can still say that a polynomial of degree n has at most n real roots.

We shall actually show that $\mathbb{Z}[x]$ can be mapped 1-1 into a subset of the rationals. This is sufficient to show that $\mathbb{Z}[x]$ is countable since we already know by the diagonal argument that the rationals are countable.

Take any $p(x) \in \mathbb{Z}[x]$. Let us say that $p(x)$ is of degre n and

$$p(x) \quad = \quad a_0 \ + \ a_1 x \ + \ a_2 x^2 \ + \ a_3 x^3 \ + \ \cdots \ + \ a_i x^i \ + \ \cdots \ + \ a_n x^n$$

and $a_0, a_1, \ldots, a_n$ are integers with $a_n \neq 0$.

Notice that the polynomial is completely specified by the sequence of its coefficients, $a_0, a_1, \ldots, a_n$. In the example of $g(x)$, the sequence is $3, -8, 1$, and in the case of $h(x)$ the sequence would be 48104547, 38430728588, 0, -4900944499, 0, 0, 0, -498832445923.

We define a map α from $\mathbb{Z}[x]$ to $\mathbb{Q}$ by

$$\alpha(p) \ = \ 3^{a_0}\, 5^{a_1}\, 7^{a_2}\, 11^{a_3} \ \cdots \ q_{i+2}^{a_i} \ \cdots \ q_{n+2}^{a_n}$$

where q_i stands for the i^{th} prime number. So the product is formed by raising the $(i+2)^{\text{th}}$ prime to the power of the i^{th} polynomial coefficient. We have omitted the prime number 2 for a reason that will later become clear.

Applied to the two polynomials in example 5.4, the α map gives

$$\begin{aligned} \alpha(g) &= \frac{11 \cdot 3^3}{5^8} = \frac{297}{390\,625} = 0.00076032 \\ \alpha(h) &= \frac{3^{48104547}\ 5^{38430728588}}{11^{4900944499}\ 23^{498832445923}} = 1.1028823\cdots \end{aligned}$$

In $\alpha(g)$, 5^8 is in the denominator since the coefficient of x in $g(x)$ is negative 8, and $5^{-8} = 1/5^8$. The final decimal value for $\alpha(g)$ is exact. In the fraction for h, note that 23 is the 9^{th} prime. The coefficients of x^2 and x^7 are negative so their primes raised to these values appear in the denominator of the fraction[11].

[11] The fact that $\alpha(h)$ is neither large nor small is very unusual for a polynomial of such high degree. More typically $\alpha(h)$ would be either extremely small or extremely large.

We need to show that α cannot map two different polynomials to the same rational. Suppose we are told the value of $\alpha(p)$ is a/b say, where a, b are positive integers. Can we reconstruct the original polynomial uniquely? Yes, we can. We can assume that the fraction a/b has been reduced (any common factors having been cancelled out). We then express a as a product of prime powers, which we can always do uniquely by the Fundamental Theorem of Arithmetic.

In case you have forgotten or maybe never heard of the Fundamental Theorem of Arithmetic I shall review it briefly. The prime numbers are those integers greater than 1 which are divisible only by 1 and themselves. The primes under 100 are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97. The list goes on, in fact, for ever. Every integer greater than 1 can be factored into a product of one or more prime numbers. For example, $10 = 2 \times 5$, $13 = 13$, $12 = 2 \times 2 \times 3$. The important thing is that this product is unique so long as we sequence the primes in increasing order. This is the Fundamental Theorem of Arithmetic.

So, coming back to the fraction a/b, there is one and only one sequence of primes whose product equals a, and likewise there is one and only one sequence of primes whose product equals b. So given the fraction a/b we can deduce the coefficients of the original polynomial. The powers of the primes appearing in a give us the positive coefficients, and those in b the negative coefficients. In other words, α is 1-1.

We have proved that $\mathbb{Z}[x]$, the set of polynomials with integer coefficients, are no more numerous than the rationals, but we already know that the rationals are countable; therefore $\mathbb{Z}[x]$ must also be countable.

But we're not done yet. What we actually want to show is that the set of roots of $\mathbb{Z}[x]$ are countable.

Well, one prime, 2, is missing from the construction of the map α. We shall use this prime to pick out the root of the polynomial in $\mathbb{Z}[x]$.

The polynomial p of degree n can have up to n distinct real roots. We order them by magnitude $r_1 < r_2 < \cdots < r_m$ where m is the actual number of distinct roots (and $m \leq n$). We have a map for the polynomials, now we need to incorporate a map for their roots. We define the final map $\beta : \mathbb{N} \times \mathbb{Z}[x] \to \mathbb{Q}$[12].

$$\beta(i, p) \;=\; 2^i \alpha(p)$$

$p \in \mathbb{Z}[x]$, and $i \in \mathbb{N}$ is the ordinal number of a real root of p assuming the real roots are ordered by magnitude.

Since α is rational, so is β. The prime 2 does not appear in $\alpha(p)$, so if given the value of $\beta(r_i)$, then α is that part of β which does not contain a power of 2. From α we recover the polynomial p, and from the power of 2 we find i, the position of the real root among all the real roots of p. Thus, given the rational value of $\beta(r_i)$, we have specified the value of the arbitrary root, r_i, of an arbitrary polynomial with integer coordinates. □

Whew! That was a big proof. I promise to inflict none other so lengthy.

Anyway, it now seemed to Cantor's contempories that the accumulating evidence was now indicating that all sets were indeed countable.

Returning to our story of the transendentals, by the mid-1870's mathematicians after much work and ingenious arguments had succeeded in finding a handful of unrelated transcendental numbers[13], but still had no clue as to how many more there might be still to be discovered. Could it be that there were just a few genuinely different transcendentals, perhaps a few dozen? If that were so, then the countability of all sets would be clinched.

In his most celebrated discovery, one of the greatest theorems in all mathematics, Cantor found that the real numbers were uncountable.

At one fell swoop, Cantor had shown that the transcendental numbers were not merely numerous but infinite beyond all reckoning.

The proof is one of the best. Here it is.

---

[12] Recall from §2.5 that $\mathbb{N} \times \mathbb{Z}[x]$ is the Cartesian product of the sets $\mathbb{N}$ and $\mathbb{Z}[x]$. It is the set of all pairs (n, p) where $n \in \mathbb{N}$ and $p \in \mathbb{Z}[x]$.

[13] I do not count transcendentals which could be almost trivially created from the few found. For example, given that π is transcendental (which it is), then so is $f(\pi)$ for every non-constant $f \in \mathbb{Z}[x]$.

5.5 **Theorem** The real numbers are uncountable. I.e. $|\mathbb{R}| > \aleph_0$

Proof. Let I be the set of real numbers in the interval 0 to 1 inclusive. A typical number in I can be written in decimal notation as

$$r \;=\; 0.d_1d_2d_3d_4\cdots$$

where d_1 is the first decimal, d_2 the second decimal, and so on.

Let us assume that I is countable. Wait a minute! Doesn't this directly contradict what we intend to prove? Yes, it does, but our intention is to deduce a contradiction, thereby demonstrating that this assumption is false. This type of argument is called a *reductio ad absurdum*, and is routine[14] in mathematics.

Since I is countable by assumption, we can enumerate the members of I with the natural numbers. Let this enumeration be $r_1, r_2, \ldots$. Let the decimal expansion of r_i be

$$r_i \;=\; 0\,.\,d_{i,1}\;d_{i,2}\;d_{i,3}\;d_{i,4}\;\cdots$$

where each $d_{i,j}$ is a decimal digit, $0 \le d_{i,j} \le 9$.

The enumeration is below; it starts with the first, r_1, followed by the second, r_2, the third, r_3, and so on.

$$\begin{aligned}
r_1 &= 0\,.\,d_{1,1}\;d_{1,2}\;d_{1,3}\;d_{1,4}\;\cdots\\
r_2 &= 0\,.\,d_{2,1}\;d_{2,2}\;d_{2,3}\;d_{2,4}\;\cdots\\
r_3 &= 0\,.\,d_{3,1}\;d_{3,2}\;d_{3,3}\;d_{3,4}\;\cdots\\
\vdots &\quad \vdots\\
r_i &= 0\,.\,d_{i,1}\;d_{i,2}\;d_{i,3}\;d_{i,4}\;\cdots\\
\vdots &\quad \vdots
\end{aligned}$$

Let us now construct a number $t \in I$ using this enumeration. We define t by its decimal expansion as

$$t \;:=\; 0\,.\,\bar{d}_{1,1}\;\bar{d}_{2,2}\;\bar{d}_{3,3}\;\bar{d}_{4,4}\;\cdots$$

where a bar over a digit indicates that the digit is incremented modulo 10. That is, $\bar{x}$ means $x+1 \bmod 10$. Thus, $\bar{0} = 1$, $\bar{1} = 2$, …, $\bar{9} = 0$. By this construction, the i^{th} digit of t is guaranteed not to equal the i^{th} digit of r_i.

Because t has a decimal expansion starting with $0.\cdots$ it must be in the interval 0 to 1 inclusive. Therefore, $t \in I$. Therefore, $t = r_j$ for some j. But this is impossible because t differs from r_j at the j^{th} decimal position. Contradiction. □

5.6 **The Continuum Hypothesis.**

Recall the definition of the power set, 2^A, of a set A from §4.4.6: it is the set of all subsets of A. One can show in complete generality (that is for finite as well as infinite cardinals) that $|A| < \left|2^A\right|$. We have just shown that $|\mathbb{N}| < |\mathbb{R}|$. Where does $2^{\mathbb{N}}$ fit in? Do we have $|\mathbb{N}| < \left|2^{\mathbb{N}}\right| < |\mathbb{R}|$, or do we have $|\mathbb{N}| < |\mathbb{R}| < \left|2^{\mathbb{N}}\right|$? In fact, it is not difficult to show that $\left|2^{\mathbb{N}}\right| = |\mathbb{R}|$ which is more commonly written as the equation

$$2^{\aleph_0} \;=\; |\mathbb{R}|$$

Does there exist a set with cardinality strictly between $\aleph_0$ and $2^{\aleph_0}$? That is,

$$\exists A?, \quad \aleph_0 \;<\; |A| \;<\; 2^{\aleph_0}$$

[14] There is a school of mathematicians, called the Intuitionists, who object to the use of *reductio ad absurdum* because such a proof is not constructive, but there is not enough room here to discuss their very interesting but controversial stance.

The surprising answer is that we do not know, and even more unsettling, that we can never know. There are a lot of technicalities behind this statement. But, in essence, the existence (or non-existence) of such a cardinality is independent of the usual axioms of arithmetic. We are therefore free to choose whatever answer pleases us the most. The usual assumption, called the Continuum Hypothesis, is that there is no cardinality between $\aleph_0$ and $2^{\aleph_0}$. At the time that this question first arose, this seemed the most natural assumption to mathematicians, a sort of Occam's Razor–why unnecessarily postulate the existence of objects? Hence, the Continuum Hypothesis became part of the Standard Model of Arithmetic.

I shall digress briefly into a personal experience. When I was supposed to be studying important matters such as mathematics, or failing that, logic, or failing even that, theology, I was in fact spending my time trying to build a cantilever made of playing cards that would stretch as far as possible away from the edge of a dining table. I discovered that, at least theoretically, for the cantilever to stretch x units of distance from the edge of the table would require roughly 2^x playing cards. But, being a student, I could afford to buy only a few decks of playing cards. Chagrined, I went to bed. That night, I dreamt that I had become a count, and was fabulously rich, and could buy as many playing cards as I wanted. I went straight to work to find out how far I could grow the cantilever. I wanted it to reach for infinity (counts are immodest). Then, I found out, being a count, that my money was countable, but that the number of cards required was $2^{\aleph_0}$ which was uncountable!

When I woke from the dream, I realized that there was indeed a conundrum: At what length exactly does the number of cards in the cantilever become infinite? Surely, the number of cards in the cantilever will first become $\aleph_0$ before becoming $2^{\aleph_0}$. What will be the length of the cantilever when it consists of $\aleph_0$ cards? Not $\aleph_0$ since such a length would require $2^{\aleph_0}$ cards. Surely then, there must be cardinalities other than $\aleph_0$ below $|\mathbb{R}|$. You might now be feeling an urge to check my argument that there are no infinite cardinalities less than $\aleph_0$.

Anyway, the point of this digression is to show you that the choice of the Continuum Hypothesis is not necessarily the most intuitive, at least to amateur civil engineers, although the hypothesis is certainly the simplest.

Chapter 6.
Point Set Topology.

This chapter is still being written.

APPENDIX A. The Greek Alphabet

Mathematics has long used the Greek alphabet and continues to do so liberally. Many people find that knowing how a Greek letter is pronounced is conducive to reading mathematical formulæ. Below I have provided a transliteration from Greek to English.

Greek lower−case	Greek upper−case	Transliteration	English Pronunciation	Greek lower−case	Greek upper−case	Transliteration	English Pronunciation
α	A	alpha	al'fa	ν	N	nu	as in numeric
β	B	beta	be'ta (U.S.), or	ξ	Ξ	xi	ksī
			bee'ta (U.K.)	o	O	omicron	om'ikron
γ	Γ	gamma	gam'ma	π	Π	pi	pī as in pie
δ	Δ	delta	del'ta	ρ	P	rho	rho
ϵ, ε	E	epsilon	epsi'lon	σ, ς	Σ	sigma	sig'ma
ζ	Z	zeta	zeet'a	τ	T	tau	vowel as in loud
η	H	eta	et'a or eet'a	υ	Υ	upsilon	ipsi'lon
θ	Θ	theta	the'ta or thee'ta	ϕ	Φ	phi	fī
ι	I	iota	īo'ta, Eye ota	χ	X	chi	khī
κ	K	kappa	kap'pa	ψ	Ψ	psi	psī
λ	Λ	lambda	lam'da	ω	Ω	omega	omeg'a (U.S.)
μ	M	mu	mu as in mute				o'miga (U.K.)

ī is pronounced as in the English words "tie" and "pie",
the single letter 'e' as in "red" (but lengthened as in "rare"),
the double 'ee' as in "meet" and "seat",
the letter 'a' as in "cat", and
the letters "kh" is pronounced as in the Scottish "loch" or the Welsh "chwarel."

Appendix B. Glossary of Terms.

Abelian	§1.5.1	A group which satisfies the commutivity axiom: $x \circ y = y \circ x$. Also called "additive."
Abstract group	§1.5.3	A group whose elements are divorced from any concrete realization. The set of all groups which are isomorphic.
Affine Group	§1.5.4	A group of translations and rotations of an object in space.
Algebraic number	§5.3	u is algebraic if $f(u) = 0$ and f is a polynomial with integer coefficients
Associativity	§1.3, §1.4.3	An operation is associative if parentheses around groups of terms can be removed
Axiom of Choice	§4.12	
Axioms	§1.3	Axioms for a group
Axioms	§1.6.1	Axioms for a ring
Axioms	§3.3.2	Axioms for a boolean algebra
Axioms	§3.3.3	Axioms for binary arithmetic
Bijection	§4.3.4	A map which is invertible. A 1-1 and onto map
Bijective map	§4.3.4	A bijection
Closure	§1.3	A set S satisfies closure under "$\circ$" if $a, b \in S \ \Rightarrow \ a \circ b \in S$
Commutative	§1.3, §1.5	x, y commutative means $x \circ y = y \circ x$
Commutative	§4.8	A diagram is commutative if the result of maps does not depend on the choice of path through the diagram
Conjugate	§4.9	In groups, x is conjugate to y if y is a conjugation of x by g for some g in the group
Conjugation	§4.9	In groups, conjugation of x by g is written x^g and means $g^{-1}xg$
Continuum Hypothesis	§5.6	There is no cardinality between $\aleph_0$ and $2^{\aleph_0}$
Domain	§4.1	If $f : X \to Y$, then X is the domain of f.
Group	§1.3	A set which is closed under a product operation and obeys certain axioms
Identity	§1.3, §4.6	In a set endowed with a binary operation, $\circ$, say, an identity if it exists, is a special element, usually denoted by "1", which satisfies $1 \circ x = x \circ 1 = x$
Image	§4.1	If $f : X \to Y$, then $f(X)$ is the image of f.
Integers	§1.5.1, §1.5.2 §3.8	The whole numbers including negative numbers, denoted by $\mathbb{Z}$. $\mathbb{Z} \ = \ \{\ldots -3, -2, -1, 0, 1, 2, 3, \ldots\}$
Integral Domain	§1.6.2	A ring where the cancellation law holds: if $x \neq 0$ and $xy = xz$ then $y = z$
Invariance	§1.5.5	If G is a group of actions on X, then a map f defined on X is invariant with respect to G if $f(gx) = f(x)$ for all $g \in G$. I.e. G has no effect on f.
Inverse	§1.3	The inverse of x, call it $\bar{x}$, satisfies $x \circ \bar{x} = \bar{x} \circ x = 1$. where 1 is the identity.
Map	§4	
Permutation group	§1.5.3	A group consisting of rearrangements, re-orderings, shuffles, or permutations of some objects.
Power set	§4.3.6	The set of subsets of a set
Proof	§3.9	
Range	§4.1	If $f : X \to Y$, then Y is the range of f.
Ring	§1.6	A set with two operations, under one of which,"+", the set is a commutative group.
Similarity	§4.9	Two matrixes A and B are similar if there is a third matrix, S, which is invertible, and such that $A = S^{-1}BS$ (see "Conjugate")
Transcendental	§5.3	A real number that is not algebraic.

Appendix C. Glossary of Symbols.

$\aleph_0$	§5	The cardinality of the natural numbers, $\mathbb{N}$
2^A	§4.3.6	If A is a set, the power set of A
$\forall x$	§3.3.8	Logical universal quantifier: for all x
$\exists x$	§3.3.8	Logical existential quantifier: there is an x
$f\|B$	§4.2	The map f restricted to the subset B of the domain of f.
1-1 map	§4.3	A map $f : X \to Y$ is 1-1 if $\left\|f(^{-1}(y)\right\| = 0$ or 1 for every $y \in Y$.
$\|x\|$		If x is a number, the absolute value or magnitude of x: $\|x\| := \begin{cases} -x & \text{if } x < 0 \\ x & \text{otherwise} \end{cases}$
$\|A\|$	§5.1	If A is a set, the cardinality of A, or the number of elements in A
$\emptyset$	§2.1	The empty set
$A \cap B$	§2.1	Set intersection, all elements which A and B have in common
$A \cup B$	§2.1	Set union, all elements which are either in A or in B
A^c	§2.1	Set complement, all elements not in A.
$A - B$	§2.1	Set difference. All elements of A not in B.
$A \subset B$	§2.1	A is a subset of B, and possibly equal
$A \subsetneq B$	§2.1	A is a subset of B, and not equal
$A \not\subset B$	§2.1	A is neither a subset of nor equal to B
$S \Rightarrow T$	§3	Logical IMPLIES: S implies T
$S \Leftarrow T$	§3	Logical IMPLIED: T implies S
$S \Leftrightarrow T$	§3	Logical equivalence: S if and only if T
$\neg S$	§3	Logical Negation: not S.
$S \vee T$	§3	Logical OR: S or T, also Boolean vee.
$S \,\&\, T$	§3	Logical AND: S and T
$S \wedge T$	§3	Boolean wedge.
$\sum_i x_i$	§3.4	Summation: the sum of all x_i's
$\prod_i x_i$	§3.4	Product: the product of all x_i's
Onto map	§4.3	A map $f : X \to Y$ is onto if $f(X) = Y$. I.e. if the image equals the range.
$\square$	§3.4	End of proof. QED
QED	§3.4	End of proof.
$\mathbb{Q}$	§3.8	The rationals. The set of all fractions including negative fractions.
$\mathbb{R}$	§3.8	The real numbers. All distances along a straight line. Infinitely long decimals.
$\mathcal{S}_n$	§3.8	The Symmetric Group of Degree n: The set of all permutations on n ordered objects.
$\mathbb{Z}$	§3.8	The integers, from the German word *Zahlen*.
$\mathbb{Z}_n$	§1.6.2	Remainders modulo n, usually taken to be the set $\{0, 1, 2, \ldots, n-1\}$.

[BM] Garrett Birkhoff and Saunders MacLane, "A Survey of Modern Algebra", AK Peters, Ltd., 1997

[Rot] J.J. Rotman, "The Theory of Groups", Allyn and Bacon, Boston, 1965.

[RW] Bertrand Russell & Alfred North Whitehead, "Principia Mathematica", Cambridge University Press, Cambridge, UK, 1962.

[WG] Wikipedia article on Évariste Galois. http://en.wikipedia.org/wiki/and an expanded version in French at http://fr.wikipedia.org/wiki/Evariste_Galois