CLASS NOTES FOR DISCRETE MATHEMATICS

NOTE ADDED 14 June 2008

These class notes were used for fifteen years in a discrete math class taught at Case Western Reserve University until I retired in 1999. I am making them available as a resource to anyone who wishes to use them. They may be copied and distributed for educational use, provided that the recipients are charged only the copying costs.

If I were revising these notes today I would make some sizeable changes. The most important would be to reformulate the definition of division on page 4 to require that the divisor be nonzero. The result would change the statement "0 divides 0" from true to false, and would affect the answers to a number of exercises.

I will be glad to receive comments and suggestions at charles@abstractmath.org. Interested readers may wish to look at my other books and websites concerned with teaching:

The Abstract Math website
Astounding Math Stories
The Handbook of Mathematical Discourse

Charles Wells
charles (at) abstractmath.org

DISCRETE MATHEMATICS

Charles Wells
June 22, 1999

Supported in part by the Fund for the Improvement of Post-Secondary Education (Grant GCO8730463)

Charles Wells
Department of Mathematics
Case Western Reserve University
10900 Euclid Avenue
Cleveland, OH 44106-7058, USA
Email: `charles@freude.com`
Home Page: `http://www.cwru.edu/artsci/math/wells/home.html`

Contents

About these notes

These class notes are for MATH 304, Fall semester, 1999. Previous versions are not usable because the text has been rewritten.

It would be a good idea to leaf through this copy to see that all the pages are there and correctly printed.

Labeled paragraphs This text is written in an innovative style intended to make the logical status of each part of the text as clear as possible. Each part is marked with labels such as "**Theorem**", "**Remark**", "**Example**", and so on that describe the intent of that part of the text. These descriptions are discussed in more detail in Chapter 1.

Exercises *The key to learning the mathematics presented in these notes is in doing all the exercises.* Many of them are answered in the back; when that is so, the text gives you the page the answer is on. You should certainly attempt every exercise that has an answer and as many of the others that you have time for.

Exercises marked "(discussion)" may be open-ended or there may be disagreement as to the answer. Exercises marked "(Mathematica)" either require Mathematica or will be much easier to do using Mathematica. A few problems that require knowledge of first-year calculus are marked "(calculus)".

Indexes On each page there is a computer-generated index of the words that occur on that page that are defined or discussed somewhere in the text. In addition, there is a complete computer-generated index on page 254. In some cases the complete index has entries for later pages where significant additional information is given for the word.

There is also an index of symbols (page 260).

Bibliography The bibliography is on page 253. References to books in the bibliography are written like this: [Hofstadter, 1979]. Suggestions for other books to include would be welcome.

Acknowledgments A grant from the Fund for the Improvement of Post-Secondary Education supported the development of these class notes. A grant from the Consolidated Natural Gas Corporation supported the development of the Mathematica package `dmfuncs.m` and the concomitant revisions to these notes.

I would like to thank Michael Barr, Richard Charnigo, Otomar Hájek, Ernest Leach, Marshall Leitman and Arthur Obrock for finding mistakes and making many helpful suggestions.

I would appreciate being notified of any errors or ambiguities. You may contact me at `charles@freude.com`.

Charles Wells

1. How to read these notes

This text introduces you to the subject matter of discrete mathematics; it includes a substantial portion of the basic language of mathematics used by all mathematicians, as well as many topics that have turned out to be useful in computer science.

In addition, this text constitutes a brief introduction to *mathematical reasoning.* This may very well be the first mathematics course in which you are expected to produce a substantial amount of correct mathematical reasoning as well as to compute answers to problems.

Most important concepts can be visualized in more than one way, and it is vital to be able to conceive of these ideas in some of the ways that mathematicians and computer scientists conceive of them. There is discussion in the text about most of the concepts to help you in doing this. The problem is that this type of discussion in general *cannot be cited in proofs*; the steps of a proof are allowed to depend only on definitions, and previously proved theorems. That is why the text has labels that distinguish the logical status of each part.

What follows is a brief glossary that describes many of the types of prose that occur in this book.

1.1 Glossary

Corollary A corollary to a theorem P is another theorem that follows easily from P.

Definition Provides a definition of one or more concepts. *Every statement to be proved should be rewritten to eliminate terms that have definitions.* This is discussed in detail in Chapter 3.

Not all concepts are defined in this text. Basic ideas such as integers and real numbers are described but not defined; we depend on your familiarity with them from earlier courses. We give a specification for some of these.

Example An example of a concept is a mathematical object that fits the definition of the concept. Thus in Definition 4.1, we define “divides” for integers, and then Example 4.1.1 we observe that 3 and 6 form an example of “divides” (3 divides 6).

For study purposes it is worthwhile to *verify that each example does fit the definition.* This is usually easy.

A few examples are actually non-examples: mathematical objects that you might think are examples of the concept but in fact are not.

Fact A fact is a precise statement about mathematics that is correct. A fact is a theorem, but one that is easy to verify and not necessarily very important. The statements marked “fact” in this text are usually immediately obvious from the definitions.

This usage is peculiar to these notes. Many texts would mark what we call facts as “propositions”, but here the word “proposition” is used in a slightly different way.

corollary 1
fact 1
lemma 2
proof 4
theorem 2
usage 2
warning 2

Lemma A lemma is a theorem that is regarded as a tool to be used in proving other theorems rather than as interesting in its own right. In fact, some theorems are traditionally called lemmas that in fact are now perceived as quite important.

Method A paragraph marked “Method” provides a method for calculating some object or for determining the truth of a certain type of statement.

Proof A mathematical proof of a statement is a sequence of closely reasoned claims about mathematical objects (numbers, sets, functions and so on) with each claim depending on the given assumptions of the statement to be proved, on known definitions and previously proved theorems (including lemmas, corollaries and facts), and on the previous statements in the proof.

Proofs are discussed in more detail in Chapters 3, 5, and in a sequence of chapters beginning with Chapter 80. Particular proof techniques are described in smaller sections throughout the text.

“Show” is another word for “prove”. (Not all math texts use the word “show” in this way.)

Remark A remark is a statement that provides some additional information about a concept. It may describe how to think about the concept, point out some aspects that follow (or don’t follow!) from the definition that the reader on first reading might miss, or give further information about the concept.

Note: As of this revision (June 22, 1999) there are some statements called “remark” that perhaps should be called “fact”, “usage” or “warning”. The author would appreciate being told of any mislabeled statement.

Specification A specification of a mathematical concept describes some basic properties of the concept but does not pin down the concept in terms of other concepts the way a definition does.

Theorem A theorem is a precise statement about mathematics that has been proved (proved somewhere — not always in this text). Theorems may be quoted as reasons in a proof, unless of course the statement to be proved is the theorem being quoted!

Corollaries, lemmas and facts are all theorems. Statements marked “Theorem” are so marked because they are important. Particularly important theorems are enclosed in a box.

Usage A paragraph marked “Usage” describes the way some terminology or symbolism is used in mathematical practice. Sometimes usage varies from text to text (example: Section 2.2.1) and in many cases, the usage of a term or symbol in mathematical texts is different, often in subtle ways, from its usage in other texts (example: Section 14.1.2).

Warning A paragraph marked “Warning” tells you about a situation that has often (in my experience) misled students.

2. Integers

2.1 Specification: integer
An **integer** is any whole number. An integer can be zero, greater than zero or less than zero.

2.1.1 Remark Note that this is not a formal definition; it is assumed that you are familiar with the integers and their basic properties.

2.1.2 Example -3, 0, 55 and one million are integers.

2.2 Definition: Properties of integers
For any integer n:

a) n is **positive** if $n > 0$.
b) n is **negative** if $n < 0$.
c) n is **nonnegative** if $n \geq 0$.
d) An integer n is a **natural number** if n is nonnegative.

2.2.1 Usage

a) A few authors define zero to be both positive and negative, but that is not common mathematical practice in the USA.
b) In pure mathematics the phrase "natural number" historically meant *positive* integer, but the meaning "nonnegative integer" used in this book has become more common in recent years.

The following theorem records some familiar facts.

2.3 Theorem
If m and n are integers, then so are $m+n$, $m-n$ and mn. If m and n are not both zero and n is nonnegative, then m^n is also an integer.

2.3.1 Remarks

a) In this text, 0^0 is undefined.
b) Observe that m^n may not be an integer if n is negative.

2.3.2 Exercise Describe precisely all integers m and n for which m^n is an integer. Note that Theorem 2.3 does not quite answer this question!

3. Definitions and proofs in mathematics

Each Definition in this text gives the word or phrase being defined in **boldface**. Each definition gives a precise description of what is required for an object to fit that definition. The only way one can verify for sure that a statement about a defined object is correct is to give a proof that it is correct *based on the definition* or on previous facts proved using the definition.

Definition 2.2 gives a precise meaning to the words "positive", "negative", "nonnegative" and "natural number". Any question about whether a given integer is positive or negative or is a natural number must be answered by checking this definition.

Referring to the definition in trying to understand a concept is the first of many methods which are used throughout the book. We will give such methods formal status, like this:

> **3.1.1 Method**
> To prove that a statement involving a concept is true, begin by using the definition of the concept to rewrite the statement.

3.1.2 Example The statement "0 is positive" is false. This claim can be justified by rewriting the statement using Definition 2.2: "$0 > 0$". Since this last statement is false, 0 is not positive.

3.1.3 Remark The preceding example illustrates the use of Method 3.1.1: I justified the claim that "0 is positive" is false by using the definition of "positive".

3.1.4 Example It also follows from Definition 2.2 that 0 is not negative (because the statement $0 < 0$ is false), but it *is* nonnegative (because the statement $0 \geq 0$ is true).

3.1.5 Exercise Is $-(-3)$ positive? (Answer on page 243.)

4. Division

> **4.1 Definition: division**
> An integer n **divides** an integer m if there is an integer q for which $m = qn$. The symbol for "divides" is a vertical line: $n \mid m$ means n divides m.

4.1.1 Example Because $6 = 2 \times 3$, it is true that $3 \mid 6$. It is also true that $-3 \mid 6$, since $6 = (-2) \times (-3)$, but it is not true that $4 \mid 14$ since there is no *integer* q for which $14 = 4q$. There is of course a *fraction* $q = 14/4$ for which $14 = 4q$, but $14/4$ is not an integer.

4.1.2 Exercise Does $13 \mid 52$? (Answer on page 243.)

4.1.3 Exercise Does $-37 \mid 111$?

4.1.4 Usage If n divides m, one also says that n is a **factor** of m or that n is a **divisor** of m.

4.1.5 Worked Exercise Find all the factors of 0, 1, 10 and 30.

Answer

Number	Factors
0	every integer
1	-1, 1
10	-1, -2, -5, -10, 1, 2, 5, 10
30	-1, -2, -3, -5, -6, -10, -15, -30, 1, 2, 3, 5, 6, 10, 15, 30

4.1.6 Exercise Find all the factors of 7, 24, 26 and 111.

4.1.7 Remarks

a) **Warning:** Don't confuse the vertical line "|", a *verb* meaning "divides", with the slanting line "/" used in fractions. The expression "$3 \mid 6$" is a sentence, but the expression "$6/3$" is the name of a number, and does not form a complete sentence in itself.

b) **Warning:** Definition 4.1 of "divides" requires that the numbers involved be integers. So it doesn't make sense in general to talk about one real number dividing another. It is tempting, for example, to say that 2 divides 2π, but according to the definition given here, that statement is meaningless.

c) Definition 4.1 does not say that there is only one integer q for which $m = qn$. However, it is true that if n is nonzero then there is only one such q, because then $q = m/n$. On the other hand, for example $0 = 5 \cdot 0 = 42 \cdot 0$ so $0 \mid 0$ and there is more than one q proving that fact.

d) Definition 4.1 says that $m \mid n$ if an integer q exists that satisfies a certain property. A statement that asserts the existence of an object with a property is called an **existential statement**. Such statements are discussed in more detail on page 113.

4.1.8 Example According to the definition, 0 divides itself, since $0 = 0 \times 0$. On the other hand, 0 divides no other integer, since if $m \neq 0$, then there is no integer q for which $m = q \times 0$.

4.1.9 Usage Many authors add the requirement that $n \neq 0$ to Definition 4.1, which has the effect of making the statement $0 \mid 0$ meaningless.

4.1.10 Exercise Find all the integers m for which $m \mid 2$. (Answer on page 243.)

4.2 Definition: even and odd

An integer n is **even** if $2 \mid n$. An **odd** integer is an integer that is not even.

4.2.1 Example -12 is even, because $-12 = (-6) \times 2$, and so $2 \mid -12$.

definition 4
divide 4
division 4
integer 3
proof 4
theorem 2

5. More about proofs

We will state and prove some simple theorems about division as an illustration of some techniques of proof (Methods 5.1.2 and 5.3.3 below.)

5.1 Theorem

Every integer divides itself.

Proof Let m be any integer. We must prove that $m \mid m$. By Definition 4.1, that means we must find an integer q for which $m = qm$. By first grade arithmetic, we can use $q = 1$.

5.1.1 How to write a proof (1) In the preceding proof, we *start with what is given* (an arbitrary integer m), we write down what must be proved (that $m \mid m$), we *apply the definition* (so we must find an integer q for which $m = qm$), and we then write down how to accomplish our goal (which is one step in this simple proof – let $q = 1$).

We will continue this discussion in Section 5.3.7.

The proof of Theorem 5.1 also illustrates a method:

> **5.1.2 Method: Universal Generalization**
> To prove a statement of the form "Every x with property P has property Q", begin by assuming you have an x with property P and prove *without assuming anything special about* x (other than its given properties) that it has property Q.

5.1.3 Example Theorem 5.1 asked us to prove that every integer divides itself. Property P is that of being an integer and property Q is that of dividing itself. So we began the proof by assuming m is an integer. (Note that we chose a name, m, for the integer. Sometimes the theorem to be proved gives you a name; see for example Theorem 5.4 on page 8.) The proof then proceeds without assuming anything special about m. It would have been wrong, for example, to say something like "Assume $m = 5$" because then you would have proved the theorem only for 5.

5.2 Theorem

Every integer divides 0.

Proof Let m be an integer (Method 5.1.2!). By Definition 4.1, we must find an integer q for which $0 = qm$. By first grade arithmetic, we can use $q = 0$.

5.2.1 Remark Theorem 5.2 may have surprised you. You can even find texts in which the integer q in the definition of division is required to be unique. For those texts, it is false that every integer divides 0.

This illustrates two important points:

a) The definition of a mathematical concept determines the truth of every statement about that concept. Your intuition and experience don't count in determining the mathematical truth of a statement. Of course they *do* count in being able to do mathematics effectively!

b) There is no agency that standardizes mathematical terminology. (There are such agencies for physics and chemistry.)

5.3 Theorem

1 *divides every integer.*

Proof Let m be any integer. By Definition 4.1, we must find an integer q for which $m = q \cdot 1$. By first grade arithmetic, we can use $q = m$.

5.3.1 Exercise Prove that if $m \mid n$ and a and b are nonnegative integers such that $a \leq b$, then $m^a \mid n^b$.

5.3.2 Worked Exercise Prove that 42 is a factor of itself.
Proof Theorem 5.1 says that every integer is a factor of itself. Since 42 is an integer, it is a factor of itself.

This worked exercise uses another proof method:

5.3.3 Method: Universal Instantiation
If a theorem says that a certain statement is true of every object of a certain type, and c is an object of that type, then the statement is true of c.

5.3.4 Example In Example 5.3.2, the theorem was Theorem 5.1, the type of object was "integer", and c was 42.

5.3.5 Remark Make sure you understand the difference between Method 5.1.2 and Method 5.3.3.

5.3.6 Worked Exercise Prove that 0 is even.
Answer Bu definition of even, we must show that $2 \mid 0$. By Theorem 5.15.2, every integer divides 0. Hence 2 divides 0 (Method 5.3.3).

5.3.7 How to write a proof (2) Worked Exercise 5.3.8 below illustrates a more complicated proof. In writing a proof you should normally include all these steps:

- PS.1 Write down *what is given*, and translate it according to the definitions of the terms involved in the statement of what is given. This translation may involve naming some of the mathematical objects mentioned in the statement to be proved.
- PS.2 Write down *what is to be proved*, and translate it according to the definitions of the terms involved.
- PS.3 Carry out some reasoning that, *beginning with what is given, deduces what is to be proved.*

The third step can be quite long. In some very simple proofs, steps PS-1 and PS-2 may be trivial. For example, Theorem 5.3 is a statement about every integer. So for step PS-1, one merely names an arbitrary integer: "Let m be any integer." Even, here, however, we have *named what we will be talking about.*

Another very important aspect of proofs is that *the logical status of every statement* should be clear. Each statement is either:

a) Given by the hypothesis of the theorem.
b) A statement of what one would like to prove (a goal). Complicated proofs will have intermediate goals on the way to the final goal.
c) A statement that has been deduced from preceding known statements. For each of these, a reason must be given, for example “Universal Instantiation” or “high school algebra”.

5.3.8 Worked Exercise Prove that any two nonnegative integers which divide each other are the same.

Answer First, we follow PS-1 and write down what we are given and translate it according to the definition of the words involved (“divides” in this case): Assume we are given integers m and n. Suppose $m \mid n$ and $n \mid m$. By Definition 4.1, the first statement means that for some q, $n = qm$. The second statement means that for some q', $m = q'n$. Now we have written and translated what we are *given.*

PS-2: We must prove that $m = n$. (This translates the phrase “are the same” using the names we have given the integers.)

PS-3: We put these statements that we have assumed together by simple algebra: $m = q'n = q'qm$. Now we have two cases: either $m = 0$ or $m \neq 0$.

a) If $m = 0$, then $n = qm = q \times 0 = 0$, so $m = n$.
b) If $m \neq 0$, then also $n \neq 0$, since $m = q'n$. Then the fact that $m = q'n = q'qm$ means that we can cancel the m (because it is nonzero!) to get $qq' = 1$. This means either $q = q' = 1$, so $m = n$, or $q = q' = -1$, so $m = -n$. But the latter case is impossible since m and n are both positive. So the only possibility that is left is that $m = n$.

We give another illustration of writing a proof by rewriting what is given and what is to be proved using the definitions by proving this proposition:

5.4 Theorem

For all integers k, m and n, if $k \mid m$ and $k \mid n$ then $k \mid m + n$.

Proof What we are *given* is that $k \mid m$ and $k \mid n$. If we rewrite these statements using Definition 4.1, we get that there are integers q and q' for which $m = qk$ and $n = q'k$. What we *want to show*, rewritten using the definition, is that there is an integer q'' for which $m + n = q''k$. Putting the hypotheses together gives

$$m + n = qk + q'k = (q + q')k$$

so we can set $q'' = q + q'$ to prove the theorem.

5.4.1 Usage In the preceding paragraph, I follow common mathematical practice in putting primes on a variable like q or r in order to indicate another variable q' of the same type. This prime has nothing to do with the concept of derivative used in the calculus.

5.4.2 Existential Bigamy In the proof of Theorem 5.4, we were given that $k \mid m$ and $k \mid n$. By using the definition of division, we concluded that there are integers q and q' for which $m = qk$ and $n = q'k$. It is a common mistake called **existential bigamy** to conclude that there is *one* integer q for which $m = qk$ and $n = qk$.

Consider that the phrase "Thurza is married" by definition means that there is a person P to whom Thurza is married. If you made the mistake just described you would assume that if Amy and Thurza were both married, then they would be *married to the same person.* That is why it is called "existential bigamy".

Mrs. Thurza Golightly White was the author's great great grandmother, and Mrs. Amy Golightly Walker was her sister. They were very definitely married to different people.

5.5 Exercise set

In problems 5.5.1 through 5.5.5, you are asked to prove certain statements about integers and division. Your proofs should involve only integers — no fractions should appear. This will help insure that your proof is based on the definition of division and not on facts about division you learned in high school. As I mentioned before, you may use algebraic facts you learned in high school, such as that fact that for any integers, $a(b+c) = ab+ac$.

5.5.1 Exercise Prove that $37 \mid 333$. (Answer on page 243.)

5.5.2 Exercise Prove that if $n > 0$, then any nonnegative integer less than n which is divisible by n must be 0. (Answer on page 243.)

5.5.3 Exercise Prove that if k is an integer which every integer divides, then $k = 0$.

5.5.4 Exercise Prove that if k is an integer which divides every integer, then $k = 1$ or $k = -1$.

5.5.5 Exercise Prove that if $k \mid m$ and $m \mid n$ then $k \mid n$.

5.6 Factors in Mathematica

The `DmFuncs` package contains the function `DividesQ[k,n]`. It returns `True` if $k \mid n$ and `False` otherwise. For example, `DividesQ[3,12]` returns `True` but `DividesQ[5,12]` returns `False`.

You can get a list of all the positive factors of n by typing `AllFactors[n]`. Thus `AllFactors[12]` returns `{1,2,3,4,6,12}`. As always, lists in Mathematica are enclosed in braces.

5.6.1 Remark `AllFactors` returns only the positive factors of an integer. In this text, however, the phrase "all factors" includes all the positive and all the negative factors.

6. Primes

Prime numbers are those, roughly speaking, which don't have nontrivial factors. Here is the formal definition:

6.1 Definition: prime number

A positive integer n is a **prime** if and only if it is greater than 1 and its only positive factors are 1 and n. Numbers bigger than 1 which are not primes are called **composite** numbers.

6.1.1 Example The first few primes are $2,3,5,7,11,13,17,\ldots$.

6.1.2 Example 0 and 1 are not primes.

6.1.3 Worked Exercise Let k be a positive integer. Prove that $4k+2$ is not a prime.

Answer $4k+2=2(2k+1)$ Thus it has factors 1, 2, $2k+1$ and $4k+2$. We know that $2 \neq 4k+2$ because k is positive. Therefore $4k+2$ has other positive factors besides 1 and $4k+2$, so $4k+2$ is not prime.

6.1.4 Exercise Prove that any even number bigger than 2 is composite.

6.1.5 Exercise Which of these integers are prime and which are composite? Factor the composite ones: 91, 98, 108, 111. (Answer on page 243.)

6.1.6 Exercise Which of these integers are prime and which are composite? Factor the composite ones: 1111, 5567, 5569.

6.1.7 Exercise Prove that the sum of two odd primes cannot be a prime.

6.2 Primes in Mathematica

The command `PrimeQ` determines if an integer is prime (it is guaranteed to work for $n < 2.5 \times 10^{10}$). Thus `PrimeQ[41]` will return `True` and `PrimeQ[111]` will return `False`.

The command `Prime[n]` gives the nth prime in order. For example, `Prime[1]` gives `2`, `Prime[2]` gives `3`, and `Prime[100]` gives `541`.

6.2.1 Exercise (Mathematica) Find all the factors of your student number.

7. Rational numbers

7.1 Definition: rational number

A **rational number** is a number representable as a fraction m/n, where m and n are integers and $n \neq 0$.

7.1.1 Example The numbers 3/4 and $-11/5$ are rational. 6 is rational because $6 = 6/1$. And .33 is rational because $.33 = 33/100$.

7.2 Theorem

Any integer is rational.

Proof The integer n is the same as the fraction $n/1$.

7.2.1 Remark The representation of a rational number as a fraction is not unique. For example,

$$\frac{3}{4} = \frac{6}{8} = \frac{-9}{-12}$$

7.2.2 Fact Two representations m/n and r/s give the same rational number if and only if $ms = nr$.

7.3 Definition: lowest terms

Let m/n be the representation of a rational number with $m \neq 0$ and $n > 0$. The representation is in **lowest terms** if there is no integer $d > 1$ for which $d \mid m$ and $d \mid n$.

7.3.1 Example 3/4 is in lowest terms but 6/8 is not, because 6 and 8 have 2 as a common divisor.

7.3.2 Exercise Is $\frac{37}{111}$ in lowest terms?

7.4 Theorem

The representation in lowest terms described in Definition 7.3 exists for every rational number and is unique.

Proof Left for you to do (Problems 64.2.5 and 63.4.1).

7.4.1 Warning You can't ask if a rational number is in lowest terms, only if its *representation* as a fraction of integers is in lowest terms.

7.5 Operations on rational numbers

Rational numbers are added, multiplied, and divided according to the familiar rules for operating with fractions. Thus for rational numbers a/b and c/d, we have

$$\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd} \quad \text{and} \quad \frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd} \tag{7.1}$$

7.5.1 Exercise If a/b and c/d are representations of rational numbers in lowest terms, must their sum $(ad+bc)/bd$ and their product ac/bd be in lowest terms? (Answer on page 243.)

8. Real numbers

8.1 Specification: real number
A **real number** is a number which can be represented as a directed distance on a straight line. A real number r is **positive** if $r > 0$ and **negative** if $r < 0$.

8.1.1 Remark Specification 8.1 is informal, but it's all you are going to get, since a formal definition is quite involved.

8.1.2 Example Any integer or rational number is a real number, and so are numbers such as π and $\sqrt{2}$. We will see a proof in Section 86 that $\sqrt{2}$ is not rational, which shows that there are real numbers that are not rational.

8.1.3 Usage The symbol $\sqrt{4}$ denotes 2. It does not denote -2. In general, for a positive real number x, the notation $\sqrt{x}$ denotes the positive square root of x, which is precisely the unique positive real number r with the property that $r^2 = x$. The unique negative number s such that $s^2 = x$ is denoted by $-\sqrt{x}$.

This usage may conflict with usage you saw in high school, but it is standard in college-level and higher mathematics.

8.1.4 Exercise For what real numbers x is it true that $\sqrt{(-x)^2} = x$?

8.2 Infinity

In calculus you may have used the symbols ∞ and $-\infty$ in connection with limits. By convention, ∞ is bigger than any real number and $-\infty$ is less than any real number. However, *they are not themselves real numbers.* There is no largest real number and there is no smallest real number.

9. Decimal representation of real numbers

A real number always has a decimal representation, possibly with an unending sequence of digits in the representation. For example, as you know, the first few decimal places of π are $3.14159\ldots$. As a general rule, you don't expect to know the exact value of a real number, but only an approximation to it by knowing its first few decimal places. Note that $22/7$ is *not* π, although it is close to it.

9.1.1 Usage The decimal representation is also called the **decimal expansion**.

9.1.2 Approximations Mathematicians on the one hand and scientists and engineers on the other tend to treat expressions such as "3.14159" in two different ways. The mathematician will think of it as a precisely given number, namely $\frac{314159}{100000}$, so in particular it represents a rational number. The scientist or engineer will treat it as the known part of the decimal representation of a real number. From their point of view, one knows 3.14159 to six significant figures. *This book always takes the mathematician's point of view.*

Mathematicians referring to an approximation may use an ellipsis (three dots), as in "π is approximately $3.14159\ldots$".

The decimal representations of two different real numbers must be different. However, two different decimal representations can, in certain circumstances, represent the same real number. This is specified precisely by the following rule:

9.2 Theorem
If $m = d_0.d_1d_2d_3\ldots$ and $n = e_0.e_1e_2e_3\ldots$, where all the d_i and e_i are decimal digits, and for some integer $k \geq 0$ the following four statements are all correct, then $m = n$:
DR.1 $d_i = e_i$ *for* $0 \leq i < k$*;*
DR.2 $d_k = e_k + 1$*;*
DR.3 $d_i = 0$ *for all* $i > k$*; and*
DR.4 $e_i = 9$ *for all* $i > k$*.*
Moreover, if the decimal representations of m and n are not identical but do not follow this pattern for some k, then $m \neq n$.

9.2.1 Usage We use a line over a string of digits to indicate that they are repeated infinitely often.

9.2.2 Example $4.\overline{9} = 5$ (here $k = 0$ in Theorem 9.2) and $1.45\overline{9} = 1.46$ (here $k = 2$).

9.2.3 Remarks
a) As it stands, Theorem 9.2 applies only to real numbers between 0 and 10, but that was only to avoid cumbersome notation. By multiplying or dividing by the appropriate power of 10, you can apply it to any real number. For example, $499.\overline{9} = 500$, since Theorem 9.2 applies to those numbers divided by 100.
b) The proofs of Theorems 9.2 and 10.1 (below) are based on the theory of geometric series (and are easy if you are familiar with that subject) but that belongs to continuous mathematics rather than discrete mathematics and will not be pursued here.

9.2.4 Exercise Which of these pairs of real numbers are equal?
a) 1.414, $\sqrt{2}$.
b) $473, 472.999$.
c) $4.0\overline{9}$, 4.1.
(Answer on page 243.)

9.2.5 Exercise Which of these pairs of real numbers are equal?
a) $53.\overline{9}$, 53.0.
b) $39/13$, $2.\overline{9}$.
c) $5698/11259$ and $.506084$.

9.2.6 Exercise If possible, give two different decimal representations of each number. If not possible, explain why not.

a) $\frac{25}{3}$.
b) $\frac{25}{4}$.
c) 105.3.

10. Decimal representation of rational numbers

The decimal representation of a rational number m/n is obtainable by dividing n into m using long division. Thus $9/5 = 1.8$ and $1/3 = 0.333\ldots$

A decimal representation which is all 0's after a certain point has to be the decimal representation of a rational number. For example, 1.853 is the rational number $1853/10^3$. On the other hand, the example of $1/3$ shows that the decimal representation of a rational number can go on forever.

The following fact is useful: If the decimal representation of a number n starts repeating in blocks after a certain point, then n is rational. For example, $1/7 = 0.\overline{142857}$ with the block 142857 repeated forever.

The following theorem says exactly which rational number is represented by a decimal representation with a repeating block of consecutive digits:

10.1 Theorem
If $n = 0.bbb\ldots$, where b is a block of k consecutive digits, then $n = b/(10^k - 1)$.

10.1.1 Example $0.\overline{13}$ is $13/99$. As another example, the theorem says that $0.\overline{3}$ is $3/9$, which of course is correct.

10.1.2 Exercise Give the exact rational value in lowest terms of $5.\overline{1}$, $4.\overline{36}$, and $4.1\overline{36}$. (Answer on page 243.)

10.1.3 Remark Theorem 10.1 says that if the decimal representation of a real number repeats in blocks then the number is rational, and moreover it tells you how to calculate it. Actually, the reverse is true, too: the decimal representation of a rational number must repeat in blocks after a certain point.

You can see why this is true by thinking about the process of long division: Suppose you have gone far enough that you have used up all the digits in the dividend (so all further digits are zero). Then, if you get a certain remainder in the quotient twice, the process necessarily repeats the second time what it did the first time.

10.2 Representations in general

It is important to distinguish between a mathematical object such as a number and its representation, for example its decimal representation or (in the case of a rational number) its representation as a fraction of integers. Thus $9/5$, $27/15$ and 1.8 all represent the *same number* which is in fact a rational number. We will return to this idea several times, for example in Section 17.1.3 and in Section 66.8.

10.3 Types of numbers in Mathematica

Mathematica knows about integers, rational numbers and real numbers. It treats a number with no decimal point as an integer, and an explicit fraction, for example `6/14`, as a rational number. If the number has a decimal point, it is always regarded as real number.

`IntegerQ[n]` returns `True` if n is represented as an integer in the sense just described. Thus `IntegerQ[3]` returns `True`, but `IntegerQ[3.0]` returns `False`.

Mathematica will store a number given as the fraction of two integers as a rational number in lowest terms. For example, if you type `6/14`, you will get `3/7` as the answer. It will return the sum, product, difference and quotient of rational numbers as rational numbers, too. Try typing `3/7+5/6` or `(3/7)/(5/6)`, for example.

The function that gives you the decimal representation of a number is `N`. For example, `N[3/7]` gives `0.4285714285714286`. You may give a second input to `N` that gives the number of decimal digits that you want. Thus `N[3/7,20]` gives

```
0.42857142857142857143
```

You can invoke `N` by typing `//N` *after* an expression, too. For example, instead of typing `N[3/7+5/4]`, you can type `3/7 + 5/4 //N`.

11. Propositions

Sentences in English can express emotion, state facts, ask questions, and so on. A sentence in a computer language may state a fact or give a command. In this section we are concerned with sentences that are either true or false.

> **11.1 Specification: proposition**
> A **proposition** is a statement which is either true or false.

11.1.1 Example Let P be the proposition "$4 \geq 2$", and Q the proposition "$25 \leq -2$". Both statements are meaningful; P is true and Q is false.

11.1.2 Example In Example 3.1.2, page 4, we showed that 0 is not positive by using the definition of positive to see that 0 is positive if the proposition $0 > 0$ is true. Since it is not true, 0 is not positive.

11.1.3 Example The statement $x > 4$ is *not* a proposition, since we don't know what x is. It is an example of a predicate.

11.1.4 Usage In many textbooks on logic a proposition is called a **sentence**.

11.1.5 Remark Textbooks on logic *define* propositions (and predicates, the subject of the next chapter) rather than merely specifying them as we have done. The definition is usually by an recursive process and can be fairly complicated. In order to prove theorems about logic, it is necessary to do this. This text explains some of the basic ideas about logic but does not prove theorems in logic.

11.2 Propositions in Mathematica

A statement such as `2 < 3` is a proposition in Mathematica; if you type it in, it will return `True`. The symbol for equals is `==` rather than "=", so for example `2 == 3` returns `False`.

12. Predicates

> **12.1 Specification: predicate**
> A **predicate** is a meaningful statement containing variables that becomes true or false when appropriate values are substituted for the variables. The proposition obtained by substituting values for each of the variables in a predicate is called an **instance** of the predicate.

12.1.1 Usage In other texts, a predicate may be called a "formula" or an "open sentence".

12.1.2 Example If x is a variable of type integer, the statement "$25 \leq x$" is a predicate. If you substitute an integer for x, the statement becomes true or false depending on the integer. If you substitute 44 for x you get the proposition "$25 \leq 44$", which is true; if you substitute 5 for x, you get the proposition "$25 \leq 5$", which is false.

12.1.3 Usage We will regard a proposition as a predicate with no variables. In other words, every proposition is a predicate.

12.1.4 Algebraic expressions and predicates An **algebraic expression** is an arrangement of symbols such as

$$x^2 - \frac{6}{x} + 4y \tag{12.1}$$

It consists of variables (x and y in this case) and operation symbols. The expression must be correctly formed according to the rules of algebra.

A predicate is analogous to an algebraic expression, except that it also contains symbols such as "$<$" and "$=$" (called **relational symbols**) that make the expression denote a statement instead of a number.

12.1.5 Example The expression

$$x^2 - \frac{6}{x} + 4y > x + y \tag{12.2}$$

is a predicate.

12.2 Substitution

When numbers are substituted for the variables in an algebraic expression, the result is a number.

12.2.1 Example Setting $x = 2$ and $y = 3$ in the expression (12.1) gives the number 13.

On the other hand, if data of the correct type are substituted into a predicate the result is not a number but *a statement which is true or false*, in other words a proposition.

12.2.2 Example If you substitute $x = 3$ into the predicate $x^2 < 4$ you get the proposition $9 < 4$, which is false. The substitution $x = 1$ gives $1 < 4$, which is true.

12.2.3 Example Substituting $x = 2$ and $y = 3$ into the expression (12.2) gives the proposition $13 > 5$, which is true.

12.2.4 Exercise Find a pair of numbers x and y that when substituted in 12.2 give a false statement.

12.2.5 Example Expressions can be substituted into other expressions as well. For example one can substitute xy for x in the expression (12.2) to get

$$x^2y^2 - \frac{6}{xy} + 4y > xy + y$$

In doing such substitution you must take into account the rules concerning how algebra is written; for example to substitute $x + y$ for x and $y + z$ for y in (12.1) you must judiciously add parentheses:

$$(x+y)^2 - \frac{6}{x+y} + 4(y+z) > x + y + y + z$$

And the laws of algebra sometimes disallow a substitution; for example you cannot substitute 0 for x in 12.2.

12.2.6 Exercise Write the result of substituting x for both x and for y in 12.2. (Answer on page 243.)

12.3 Types

In this book, variables are normally assumed to be of a particular type; for example the variable x mentioned in Example 12.1.2 is of type integer. We do not always specify the type of variables; in that case, you can assume that the variable can be replaced by any data that makes the predicate make sense. For example, in the predicate $x \leq 25$, x can be any number for which "$\leq$" makes sense — thus any real number number, but not a complex number. This informal practice would have to be tightened up for a correct formal treatment of predicates; the intent here is to provide an informal introduction to the subject in which predicates are used the way they are normally used in common mathematical practice.

12.3.1 Usage A **real variable** is a variable of type real. An **integer variable** is a variable of type integer. Don't forget that both integer variables and real variables are allowed to have negative values.

12.3.2 Worked Exercise Let x be a variable of type real. Find a value of x that makes the statement "$x > 1$ and $x < 2$" true, and another that makes it false. Do the same for the case that x is an integer variable.

Answer Any real number between 1 and 2 makes "$x > 1$ and $x < 2$" true, for example $x = \frac{1}{2}$ or $x = \sqrt{2}$. The values $x = 0$, $x = 1$, $x = -1$, and $x = 42$ all make it false.

No integer value of x makes the statement true; it is false for every integer.

12.4 Exercise set

Let m be an integer variable. For each predicate in problems 12.4.1 through 12.4.5, give (if it is possible) a value of m for which it is true and another value for which it is false.

12.4.1 $m \mid 4$. (Answer on page 243.)

12.4.2 $m = m$. (Answer on page 243.)

12.4.3 $m = m + 1$.

12.4.4 $m = 2m$.

12.4.5 $m^2 = m$.

12.5 Naming predicates

We will name predicates with letters in much the same way that we use letters to denote numbers in algebra. It is allowed, *but not required*, to show the variable(s) in parentheses. For example, we can say: let $P(x)$ denote the predicate "$25 \leq x$". Then $P(42)$ would denote the proposition "$25 \leq 42$", which is true; but $P(-2)$ would be false. $P(42)$ is obtained from $P(x)$ by substitution.

We can also say, "Let P denote the predicate $25 \leq x$" without the x being exhibited. This is useful when we want to refer to an arbitrary predicate without specifying how many variables it has.

Predicates can have more than one variable. For example, let $Q(x, y)$ be "$x \leq y$". Then $Q(25, 42)$ denotes the proposition obtained by substituting 25 for x and 42 for y. $Q(25, 42)$ is true; on the other hand, $Q(25, -2)$ is false, and $Q(25, y)$ is a predicate, neither true nor false.

12.5.1 Worked Exercise Let m and n be integer variables. Let $P(n)$ denote the predicate $n < 42$ and $Q(m, n)$ the predicate $n \mid (m + n)$. Which of these predicates is true when 42 is substituted for m and 4 is substituted for n?

Answer $P(4)$ is $4 < 42$, which is true, and $Q(42, 4)$ is $4 \mid 46$, which is false.

12.5.2 Exercise If $Q(x)$ is the predicate $x^2 < 4$, what are $Q(-1)$ and $Q(x-1)$? (Answer on page 243.)

12.5.3 Exercise Let $P(x,y,z)$ be the predicate $xy < x+z+1$. Write out each of these predicates.

a) $P(1,2,3)$.
b) $P(1,3,2)$.
c) $P(x,x,y)$
d) $P(x,x+y,y+z)$.

(Answer on page 243.)

12.5.4 Exercise Let P be the predicate of Exercise 12.5.3. Write out $P(x,x,x)$ and $P(x,x-1,x+1)$ and for each predicate give a value of x for which it is true and another value for which it is false.

12.5.5 Warning You may have seen notation such as "$f(x)$" to denote a function. Thus if $f(x)$ is the function whose value at x is $2x+5$, then $f(3) = 11$. We will consider functions formally in Chapter 39. Here we only want to call your attention to a difference between that notation and the notation for predicates: If $f(x) = 2x+5$, then "$f(x)$" is an *expression.* It is the name of something. On the other hand, if $P(x)$ denotes the predicate "$25 \leq x$", then $P(x)$ is a *statement* – a complete sentence with a subject and a verb. It makes sense to say, "If $a = 42$, then $P(a)$", for that is equivalent to saying, "If $a = 42$, then $25 \leq a$". It does not make sense to say, "If $a = 42$, then $f(a)$", which would be "If $a = 42$, then $2a+5$". Of course, it *is* meaningful to say "If $a = 42$, then $f(a) = 89$".

12.6 Predicates in Mathematica

A statement such as `2 < x` is a predicate. If `x` has not been given a value, if you type `2 < x` you will merely get `2 < x` back, since Mathematica doesn't know whether it is true or false.

13. Universally true

13.1 Definition: universally true predicate

A predicate containing a variable of some type that is true for *any* value of that type is called **universally true**.

13.1.1 Example If x is a real number variable, the predicate "$x^2 - 1 = (x+1)(x-1)$" is true for any real number x. In this example the variable of the definition is x, its type is "real", and so any value of that type means any real number. In particular, 42 is a real number so we know that $42^2 - 1 = (42+1)(42-1)$

13.1.2 Usage In some contexts, a universally true predicate is called a **law**. When a universally true predicate involves equality, it is called an **identity**.

13.1.3 Example The predicate "$x^2-1=(x+1)(x-1)$" is an identity. An example of a universally true predicate which is not an identity is "$x+3 \geq x$" (again, x is real number).

13.1.4 Remark If $P(x)$ is a predicate and c is some particular value for x for which $P(c)$ is false, then $P(x)$ is not universally true. For example, $x>4$ is not universally true because $3>4$ is false (in this case, $c=3$). This is discussed further in Chapter 75.

13.2 Definition: $\forall$
We will use the notation $(\forall x)$ to denote that the predicate following it is true of all x of a given type.

13.2.1 Example $(\forall x)(x+3 \geq x)$ means that for every x, $x+3 \geq x$.

13.2.2 Worked Exercise Let x be a real variable. Which is true? (a) $(\forall x)(x > x)$. (b) $(\forall x)(x \geq x)$. (c) $(\forall x)(x \neq 0)$.
Answer (a) is false, (b) is true and (c) is false.

13.2.3 Remark In Exercise 13.2.2, it would be wrong to say that the answer to (c) is "almost always true" or to put any other qualification on it. *Any universal statement is either true or false, period.*

13.2.4 Example The statement "$x \neq 0$" is true for $x=3$ and false for $x=0$, but the statement $(\forall x)(x \neq 0)$ is just plain false.

13.2.5 Exercise Let x be a real variable. Which is true? (a) $(\forall x)(x \neq x)$. (b) $(\forall x)((\forall y)(x \neq y))$. (c) $(\forall x)((\forall y)(x \geq y))$.

13.2.6 Usage The symbol "$\forall$" is called a **quantifier** We take a more detailed look at quantifiers in Chapter 75.

13.2.7 Exercise Which of these statements are true? n is an integer and x a real number.

a) $(\forall n)(n+3 \geq n)$.
b) $(\forall x)(x+3 \geq x)$.
c) $(\forall n)(3n > n)$.
d) $(\forall n)(3n+1 > n)$.
e) $(\forall x)(3x > x)$.

(Answer on page 243.)

14. Logical Connectives

Predicates can be combined into compound predicates using combining words called **logical connectives**. In this section, we consider "and", "or" and "not".

14.1 Definition: "and"
If P and Q are predicates, then $P \wedge Q$ ("P and Q") is also a predicate, and it is true precisely when both P and Q are true.

14.1.1 Worked Exercise Let n be an integer variable and let $P(n)$ be the predicate ($n > 3$ and n is even). State whether $P(2)$, $P(6)$ and $P(7)$ are true.
Answer $P(2)$ is false, $P(6)$ is true and $P(7)$ is false.

14.1.2 Usage
a) A predicate of the form "$P \wedge Q$" is called a **conjunction**.
b) Another notation for $P \wedge Q$ is "PQ". In Mathematica, "$P \wedge Q$" is written `P && Q`.

14.2 Definition: "or"
$P \vee Q$ ("P or Q") is a predicate which is true when at least one of P and Q is true.

14.2.1 Usage
a) A compound predicate of the form $P \vee Q$ is called a **disjunction**.
b) Often "$P + Q$" is used for "$P \vee Q$". In Mathematica, it is written `P || Q`.

14.2.2 Example If P is "$4 \geq 2$" and Q is "$25 \leq -2$", then "$P \wedge Q$" is false but "$P \vee Q$" is true.

14.2.3 Exercise For each predicate $P(n)$ given, state whether these propositions are true: $P(2)$, $P(6)$, $P(7)$.
a) ($n > 3$ or n is even)
b) ($n \mid 6$ or $6 \mid n$)
c) n is prime or ($n \mid 6$)
(Answer on page 243.)

14.2.4 Exercise For each predicate give (if possible) an integer n for which the predicate is true and another integer for which it is false.
a) $(n+1 = n) \vee (n = 5)$.
b) $(n > 7) \vee (n < 4)$.
c) $(n > 7) \wedge (n < 4)$.
d) $(n < 7) \vee (n > 4)$.
(Answer on page 243.)

14.2.5 Exercise Which of the predicates in Problem 14.2.4 are universally true for integers? (Answer on page 243.)

14.3 Truth tables

The definitions of the symbols '$\wedge$' and '$\vee$' can be summarized in **truth tables**:

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

14.3.1 Remark As the table shows, the definition of '$\vee$' requires that $P \vee Q$ be true if *either or both* of P and Q are true; in other words, this is "or" in the sense of "and/or". This meaning of "or" is called "inclusive or".

14.3.2 Usage In computer science, "1" is often used for "true" and "0" for "false".

14.4 Definition: "xor"

If P and Q are predicates, the compound predicate $P\,\mathrm{XOR}\,Q$ is true if exactly one of P and Q is true.

14.4.1 Fact The truth table of XOR is

P	Q	P XOR Q
T	T	F
T	F	T
F	T	T
F	F	F

14.4.2 Usage

a) XOR in Mathematica is `Xor`. $P\,\mathrm{XOR}\,Q$ may be written either `P ~Xor~ Q` or `Xor[P,Q]`.
b) In mathematical writing, "or" normally denotes the inclusive or, so that a statement like, "Either a number is bigger than 2 or it is smaller than 4" is considered correct. The writer might take pity on the reader and add the phrase, "or both", but she is not obliged to.

14.4.3 Worked Exercise Which of the following sentences say the same thing? In each sentence, n is an integer.

a) Either n is even or it is positive.
b) n is even or positive or both.
c) n is both even and positive.

Answer (a) and (b) say the same thing. (c) is not true of 7, for example, but (a) and (b) are true of 7.

14.5 Definition: "not"

The symbol '$\neg P$' denotes the **negation** of the predicate P.

14.5.1 Example For real numbers x and y, $\neg(x < y)$ means the same thing as $x \geq y$.

14.5.2 Fact Negation has the very simple truth table

P	$\neg P$
T	F
F	T

14.5.3 Usage

a) Other notations for $\neg P$ are $\overline{P}$ and $\sim P$.
b) The symbol in Mathematica for "not" is `!`, the exclamation point. $\neg P$ is written `!P`.
c) The symbol '$\neg$' always applies to the first predicate after it only. Thus in the expression $\neg P \vee Q$, only P is negated. To negate the whole expression $P \vee Q$ you have to write "$\neg(P \vee Q)$".

14.5.4 Warning Negating a predicate is not (usually) the same thing as stating its opposite. If P is the statement "$3 > 2$", then $\neg P$ is "3 is not greater than 2", rather than "$3 < 2$". Of course, $\neg P$ can be *reworded* as "$3 \leq 2$".

14.5.5 Example Writing the negation of a statement in English can be surprisingly subtle. For example, consider the (false) statement that 2 divides every integer. The negation of this statement is true; one way of wording it is that there is *some* integer which is not divisible by 2. In particular, the statement, "All integers are not divisible by 2" is *not* the negation of the statement that 2 divides every integer.

We will look at this sort of problem more closely in Section 77.

14.6 Truth Tables in Mathematica

The `dmfuncs.m` package has a command `TruthTable` that produces the truth table of a given Mathematica logical expression. For example, if you define the expression

```
e = a && (b || !c)
```

then `TruthTable[e]` produces

a	b	c	a && (b || !c)
T	T	T	T
T	T	F	T
T	F	T	F
T	F	F	T
F	T	T	F
F	T	F	F
F	F	T	F
F	F	F	F

15. Rules of Inference

15.1 Definition: rule of inference
Let P_1, P_2, $\ldots P_n$ and Q be predicates. An expression of the form

$$P_1, \ldots, P_n \vdash Q$$

is a **rule of inference**. Such a rule of inference is **valid** if whenever P_1, $P_2 \ldots$ and P_n are all true then Q must be true as well.

15.1.1 Example If you are in the middle of proving something and you discover that $P \wedge Q$ is true, then you are entitled to conclude that (for example) P is true, if that will help you proceed with your proof. Hence

$$P \wedge Q \vdash P \tag{15.1}$$

is a valid rule of inference.

That is not true for '$\vee$', for example: If $P \vee Q$ is true, you know that at least one of P and Q are true, but you don't know which one. Thus the purported rule of inference "$P \vee Q \vdash P$" is *invalid.*

15.1.2 Usage The symbol '$\vdash$' is called the "turnstile". In this context, it can be read "yields".

15.1.3 Example The basic rules of inference for "or" are

$$P \vdash P \vee Q \quad \text{and} \quad Q \vdash P \vee Q \tag{15.2}$$

These say that if you know P, you know $P \vee Q$, and if you know Q, you know $P \vee Q$.

15.1.4 Example Another rule of inference for "and" is

$$P, Q \vdash P \wedge Q \tag{15.3}$$

15.1.5 Exercise Give at least two nontrivial rules of inference for XOR. The rules should involve only propositional variables and XOR and other logical connectives.

15.1.6 Exercise Same instructions as for Exercise 15.1.5 for each of the connectives defined by these truth tables:

P	Q	$P * Q$
T	T	F
T	F	F
F	T	T
F	F	F

(a)

P	Q	P NAND Q
T	T	F
T	F	T
F	T	T
F	F	T

(b)

P	Q	P NOR Q
T	T	F
T	F	F
F	T	F
F	F	T

(c)

15.2 Definitions and Theorems give rules of inference

What Method 3.1.1 (page 4) says informally can be stated more formally this way: *Every definition gives a rule of inference.*

Similarly, any Theorem gives a rule of inference.

15.2.1 Example The rule of inference corresponding to Definition 4.1, page 4, is that for m, n and q integers,

$$m = qn \vdash n \mid m$$

One point which is important in this example is that it must be clear in the rule of inference what the types of the variables are. In this case, we required that the variables be of type integer. Although $14 = (7/2) \times 4$, you cannot conclude that $4 \mid 14$, because $7/2$ is not an integer.

15.2.2 Worked Exercise State Theorem 5.4, page 8, as a rule of inference.
Answer $k \mid m, k \mid n \vdash k \mid m + n$.

15.2.3 Exercise (discussion) What is the truth table for the English word "but"?

16. Sets

The concept of set, introduced in the late nineteenth century by Georg Cantor, has had such clarifying power that it occurs everywhere in mathematics. Informally, a set is a collection of items. An example is the set of all integers, which is traditionally denoted Z.

We give a formal specification for sets in 21.1.

16.1.1 Example Any data type determines a set — the set of all data of that type. Thus there is a set of integers, a set of natural numbers, a set of letters of the English alphabet, and so on.

16.1.2 Usage The items which constitute a set are called the **elements** or **members** of the set.

16.2 Standard notations

The following notation for sets of numbers will be used throughout the book.

a) N is the set of all nonnegative integers
b) N^+ is the set of all positive integers.
c) Z is the set of all integers.
d) Q is the set of all rational numbers.
e) R is the set of all real numbers.
f) R^+ is the set of all nonnegative real numbers.
g) R^{++} is the set of all positive real numbers.

16.2.1 Usage Most authors adhere to the notation of the preceding table, but some use N for N^+ or I for Z.

16.3 Definition: "$\in$"

If x is a member of the set A, one writes "$x \in A$"; if it is not a member of A, "$x \notin A$".

16.3.1 Example $4 \in \mathrm{Z}$, $-5 \in \mathrm{Z}$, but $4/3 \notin \mathrm{Z}$.

16.4 Sets, types and quantifiers

When using the symbol $\forall$, as in Section 13.1, the type of the variable can be exhibited explicitly with a colon followed by the name of a set, as is done in Pascal and other computer languages. Thus to make it clear that x is an integer, one could write $(\forall x{:}\mathrm{Z})P(x)$.

16.4.1 Worked Exercise Which of these statements is true?

a) $(\forall x{:}\mathrm{Z})x \geq 0$
b) $(\forall x{:}\mathrm{N})x \geq 0$

Answer Part (a) says that every integer is nonnegative. That is false; for example, -3 is negative. On the other hand, part (b) is true.

17. List notation for sets

There are two common methods for defining sets: list notation, discussed here, and setbuilder notation, discussed in the next chapter.

17.1 Definition: list notation

A set with a small number of members may be denoted by listing them inside curly brackets.

17.1.1 Example The set $\{2,5,6\}$ contains the numbers 2, 5 and 6 as elements, and no others. So $2 \in \{2,5,6\}$ but $7 \notin \{2,5,6\}$.

17.1.2 Remark

a) In list notation, the order in which the elements are given is irrelevant: $\{2,5,6\}$ and $\{5,2,6\}$ are the *same set*.
b) Repetitions don't matter, either: $\{2,5,6\}$, $\{2,2,5,6\}$ and $\{2,5,5,5,6,6\}$ are all the same set. Note that $\{2,5,5,6,6\}$ has *three* elements.

17.1.3 Remark The preceding remarks indicate that the symbols $\{2,5,6\}$ and $\{2,2,5,6\}$ are different representations of the same set. We discussed different representations of numbers in Section 10.2. Many mathematical objects have more than one representation.

17.1.4 Exercise How many elements does the set $\{1,1,2,2,3,1\}$ have? (Answer on page 243.)

17.2 Sets in Mathematica

In Mathematica, an expression such as

```
{2,2,5,6}
```

denotes a *list* rather than a set. (Lists are treated in detail in Chapter 109.) Both order and repetition matter. In particular, `{2,2,5,6}` is not the same as `{2,5,6}` and neither are the same as `{2,6,5}` .

A convenient way to list the first n integers is `Table[k,{k,1,n}]`. For example, `Table[k,{k,1,10}]` returns `{1,2,3,4,5,6,7,8,9,10}`.

17.3 Sets as elements of sets

A consequence of Specification 21.1 is that a set, being a "single entity", can be an element of another set. Furthermore, if it is, its elements are not necessarily elements of that other set.

17.3.1 Example Let $A = \{\{1,2\},\{3\},2,6\}$. It has four elements, two of which are sets.

Observe that $1 \in \{1,2\}$ and $\{1,2\} \in A$, but the number 1 is *not* an element of A. The set $\{1,2\}$ is *distinct from its elements*, so that even though one of its elements is 1, the set $\{1,2\}$ itself is *not* 1. On the other hand, 2 *is* an element of A because it is explicitly listed as such.

17.3.2 Exercise Give an example of a set that has $\{1,2\}$ as an element and 2 as an element but which does not have 1 as an element.

18. Setbuilder notation

18.1 Definition: setbuilder notation

A set may be denoted by the expression $\{x \mid P(x)\}$, where P is a predicate. This denotes the set of *all* elements of the type x for which the predicate $P(x)$ is true. Such notation is called **setbuilder notation**. The predicate P is called the **defining condition** for the set, and the set $\{x \mid P(x)\}$ is called the **extension** of the predicate P.

18.1.1 Usage

a) Sometimes a colon is used instead of '|' in the setbuilder notation.
b) The fact that one can define sets using setbuilder notation is called **comprehension**. See 18.1.11.

18.1.2 Example The set $\{n \mid n \text{ is an integer and } 1 < n < 6\}$ denotes the set $\{2,3,4,5\}$.

18.1.3 Example The set $S = \{n \mid n \text{ is an integer and } n \text{ is prime}\}$ is the set of all primes.

18.1.4 Worked Exercise List the elements of these sets, where n is of type integer.

a) $\{n \mid n^2 = 1\}$.
b) $\{n \mid n \text{ divides } 12\}$.
c) $\{n \mid 1 < n < 3\}$.

Answer a) $\{-1, 1\}$. b) $\{1, 2, 3, 4, 6, 12, -1, -2, -3, -4, -6, -12\}$. c) $\{2\}$.

18.1.5 Exercise How many elements do each of the following sets have? In each case, x is real.

a)	$\{2, 1, 1, 1\}$	c)	$\{x \mid x^2 - 1 = 0\}$
b)	$\{1, 2, -1, \sqrt{4}, \lvert -1 \rvert\}$	d)	$\{x \mid x^2 + 1 = 0\}$

(Answer on page 243.)

18.1.6 Example The extension of the predicate

$$(x \in \mathrm{Z}) \wedge (x < 5) \wedge (x > 2)$$

is the set $\{3, 4\}$.

18.1.7 Example The extension of a predicate whose main verb is "equals" is what one would normally call the solution set of the equation. Thus the extension of the predicate $x^2 = 4$ is $\{-2, 2\}$.

18.1.8 Exercise Write predicates whose extensions are the sets in exercise 18.1.5 (a) and (b). Use a real variable x.

18.1.9 Exercise Give these sets in list notation, where n is of type integer.

a) $\{n \mid n > 1 \text{ and } n < 4\}$.
b) $\{n \mid n \text{ is a factor of } 3\}$.

18.1.10 Usage In some texts, a predicate is *defined* to be what we have called its extension here: in those texts, a predicate $P(x)$ is a subset (see Chapter 31) of the set of elements of type x. In such texts, "$(x = 2) \vee (x = -2)$" would be regarded as the *same predicate* as "$x^2 = 4$".

18.1.11 Method: Comprehension
Let $P(x)$ be a predicate and let $A = \{x \mid P(x)\}$. Then if you know that $a \in A$, it is correct to conclude that $P(a)$. Moreover, if $P(a)$, then you know that $a \in A$.

18.1.12 Remark The Method of Comprehension means that the elements of $\{x \mid P(x)\}$ are exactly all those x that make $P(x)$ true. If $A = \{x \mid P(x)\}$, then *every* x for which $P(x)$ is an element of A, and nothing else is.

This means that in the answer to Worked Exercise 18.1.4, the *only correct answer* to part (b) is $\{1, 2, 3, 4, 6, 12, -1, -2, -3, -4, -6, -12\}$. For example, the set $\{1, 2, 3, 4, 6, -3, -4, -6, -12\}$ would not be a correct answer because it does not include *every* integer that makes the statement "n divides 12" true (it does not contain -2, for example).

18.1.13 Rules of inference for sets It follows that we have two rules of inference: If $P(x)$ is a predicate, then for any item a of the same type as x,

$$P(a) \vdash a \in \{x \mid P(x)\} \tag{18.1}$$

and

$$a \in \{x \mid P(x)\} \vdash P(a) \tag{18.2}$$

18.1.14 Example The set

$$\mathrm{I} = \{x \mid x \text{ is real and } 0 \le x \le 1\} \tag{18.3}$$

which has among its elements 0, $1/4$, $\pi/4$, 1, and an infinite number of other numbers. I is fairly standard notation for this set — it is called the **unit interval**.

18.1.15 Usage Notation such as "$a \le x \le b$" means $a \le x$ *and* $x \le b$. So the statement "$0 \le x \le 1$" in the preceding example means "$0 \le x$" and "$x \le 1$". Note that it follows from this that $5 \le x \le 3$ means $(5 \le x) \wedge (x \le 3)$ — there are no numbers x satisfying that predicate. It does *not* means "$(5 \le x) \vee (x \le 3)$"!

18.1.16 Exercise What is required to show that $a \notin \{x \mid P(x)\}$? (Answer on page 243.)

19. Variations on setbuilder notation

Frequently an expression is used left of the vertical line in setbuilder notation, instead of a single variable.

19.1 Typing the variable

One can use an expression on the left side of setbuilder notation to indicate the type of the variable.

19.1.1 Example The unit interval I could be defined as

$$\mathrm{I} = \{x \in \mathrm{R} \mid 0 \leq x \leq 1\}$$

making it clear that it is a set of real numbers rather than, say rational numbers.

19.2 Other expressions on the left side

Other kinds of expressions occur before the vertical line in setbuilder notation as well.

19.2.1 Example The set $\{n^2 \mid n \in \mathrm{Z}\}$ consists of all the squares of integers; in other words its elements are $0, 1, 4, 9, 16, \ldots$.

19.2.2 Example Let $A = \{1,3,6\}$. Then

$$\{n-2 \mid n \in A\} = \{-1, 1, 4\}$$

19.2.3 Remark The notation introduced in the preceding examples is another way of putting an additional condition on elements of the set. Most such definitions can be reworded by introducing an extra variable. For example, the set in Example 19.2.1 could be rewritten as

$$\{n^2 \mid n \in \mathrm{Z}\} = \{k \mid (k = n^2) \wedge (n \in \mathrm{Z})\}$$

and the set in Example 19.2.2 as

$$\{n-2 \mid n \in A\} = \{m \mid (m = n-2) \wedge (n \in A)\}$$

19.2.4 Warning Care must be taken in reading such expressions: for example, the integer 9 *is* an element of the set $\{n^2 \mid n \in \mathrm{Z} \wedge n \neq 3\}$, because although $9 = 3^2$, it is also true that $9 = (-3)^2$, and -3 is an integer not ruled out by the predicate on the right side of the definition.

19.2.5 Exercise Which of these equations are true?
a) $\mathrm{R}^+ = \{x^2 \mid x \in \mathrm{R}\}$
b) $\mathrm{N} = \{x^2 \mid x \in \mathrm{N}\}$
c) $\mathrm{R} = \{x^3 \mid x \in \mathrm{R}\}$
(Answer on page 243.)

19.2.6 Exercise List the elements of these sets.
a) $\{n-1 \in \mathrm{Z} \mid n \text{ divides } 12\}$
b) $\{n^2 \in \mathrm{N} \mid n \text{ divides } 12\}$
c) $\{n^2 \in \mathrm{Z} \mid n \text{ divides } 12\}$
(Answer on page 243.)

19.2.7 Exercise List the elements of these sets, where x and y oare of type real:
a) $\{x + y \mid y = 1 - x\}$.
b) $\{3x \mid x^2 = 1\}$.

19.2.8 Exercise How many elements does the set

$$\{\frac{1}{x^2} \mid x = -\frac{1}{2}, \frac{1}{2}, -2, 2\}$$

have?

19.3 More about sets in Mathematica

The `Table` notation described in 17.2 can use the variations described in 19. For example, `Table[k^2,{k,1,5}]` returns `{1,4,9,16,25}`.

Defining a set by setbuilder notation in Mathematica is accomplished using the command `Select`. `Select[list,criterion]` lists all the elements of the list that meet the criterion. For example, `Select[{2,5,6,7,8},PrimeQ]` returns `{2,5,7}`. The criterion must be a Mathematica command that returns `True` or `False` for each element of the list. The criterion can be such a command you defined yourself; it does not have to be built in.

19.3.1 Exercise (Mathematica) Explain the result you get when you type

```
Select[{2,4,Pi,5.0,6.0},IntegerQ]
```

in Mathematica.

20. Sets of real numbers

Now we use the setbuilder notation to define a notation for intervals of real numbers.

20.1 Definition: interval

An **open interval**

$$(a\,..\,b) = \{x \in \mathrm{R} \mid a < x < b\} \tag{20.1}$$

for any specific real numbers a and b. A **closed interval** includes its endpoints, so is of the form

$$[a\,..\,b] = \{x \in \mathrm{R} \mid a \le x \le b\} \tag{20.2}$$

20.1.1 Example The interval I defined in (18.3), page 29, is $[0\,..\,1]$.

20.1.2 Usage The more common notation for these sets uses a comma instead of two dots, but that causes confusion with the notation for ordered pair which will be introduced later.

20.1.3 Exercise Which of these are the same set? x is real.

a)	$\{0,1,-1\}$	d)	$\{x \mid x^3 = -x\}$
b)	$\{x \mid x = -x\}$	e)	$[-1\,..\,1]$
c)	$\{x \mid x^3 = x\}$	f)	$(-1\,..\,1)$

(Answer on page 243.)

20.2 Bound and free variables

The variable in setbuilder notation, such as the x in Equation (18.3), is **bound**, in the sense that you cannot substitute anything for it. The "dummy variable" x in an integral such as $\int_a^b f(x)\,dx$ is bound in the same sense. On the other hand, the a and b in Equation (20.2) are **free variables**: by substituting real numbers for a and b you get specific sets such as $[0\,..\,2]$ or $[-5\,..\,3]$. Free variables which occur in a definition in this way are also called **parameters** of the definition.

21. A specification for sets

We said that Method 18.1.11 "determines the set $\{x \mid P(x)\}$ precisely." Actually, what the method does is explain how the notation determines the *elements* of the set precisely. But that is the basic fact about sets: *a set is determined by its elements.*

Indeed, the following specification contains everything about what a set is that you need to know (for the purposes of reading this book!).

21.1 Specification: set
A set is a single entity distinct from, but completely determined by, its elements (if there are any).

21.1.1 Remarks

a) This is a specification, rather than a definition. It tells you the *operative properties* of a set rather than giving a definition in terms of previously known objects.

Thus a set is a single abstract thing (entity) like a number or a point, even though it may have many elements. It is *not* the same thing as its elements, although it is determined by them.

b) In most circumstances which arise in mathematics or computer science, a kind of converse to Specification 21.1 holds: any collection of elements forms a set. However, this is not true universally. (See Section 24.)

21.2 Consequences of the specification for sets

A consequence of Specification 21.1 is the observation in Section 17.1 that, in using the list notation, the order in which you list the elements of a set is irrelevant. Another consequence is the following method.

21.2.1 Method
For any sets A and B, $A = B$ means that

a) Every element of A is an element of B and
b) Every element of B is an element of A.

21.2.2 Example For x real,

$$\{x \mid x^2 = 1\} = \{x \mid (x = 1) \vee (x = -1)\}$$

We will prove this using Method 21.2.1. Let

$$A = \{x \mid x^2 = 1\} \text{ and } B = \{x \mid (x = 1) \vee (x = -1)\}$$

Suppose $x \in A$. Then $x^2 = 1$ by 18.2. Then $x^2 - 1 = 0$, so $(x - 1)(x + 1) = 0$, so $x = 1$ or $x = -1$. Hence $x \in B$ by 18.1. On the other hand, if $x \in B$, then $x = 1$ or $x = -1$, so $x^2 = 1$, so $x \in A$.

21.2.3 Remark The two statements, "$x^2 = 1$" and "$(x = 1) \vee (x = -1)$" are different statements which nevertheless say the same thing. On the other hand, the descriptions $\{x \mid x^2 = 1\}$ and $\{x \mid (x = 1) \vee (x = -1)\}$ *denote the same set;* in other words, the predicates "$x^2 = 1$" and "$(x = 1) \vee (x = -1)$" have the same extension. This illustrates that the defining property for a particular set can be stated in various equivalent ways, but *what the set is is determined precisely by its elements.*

22. The empty set

22.1 Definition: empty set

The **empty set** is the unique set with no elements at all. It is denoted $\{\}$ or (more commonly) $\emptyset$.

22.1.1 Remark The existence and uniqueness of the empty set follows directly from Specification 21.1.

22.1.2 Example $\{x \in \mathrm{R} \mid x^2 < 0\} = \emptyset$.

22.1.3 Example The interval notation "$[a\,..\,b]$" introduced in 20.1 defines the empty set if $a > b$. For example, $[3\,..\,2] = \emptyset$.

22.1.4 Example Since the empty set is a set, it can be an element of another set. Consider this: although "$\emptyset$" and "$\{\}$" both denote the empty set, $\{\emptyset\}$ is *not* the empty set; it is a set whose only element is the empty set.

22.1.5 Usage This symbol "$\emptyset$" should not be confused with the Greek letter phi, written ϕ, nor with the way the number zero is sometimes written by older printing terminals for computers.

22.1.6 Exercise Which of these sets is the empty set?
a) $\{0\}$.
b) $\{\emptyset, \emptyset\}$.
c) $\{x \in \mathrm{Z} \mid x^2 \leq 0\}$.
d) $\{x \in \mathrm{Z} \mid x^2 = 2\}$.
(Answer on page 243.)

23. Singleton sets

23.1 Definition: singleton
A set containing exactly one element is called a **singleton set**.

23.1.1 Example $\{3\}$ is the set whose only element is 3.

23.1.2 Example $\{\emptyset\}$ is the set whose only element is the empty set.

23.1.3 Remark Because a set is distinct from its elements, a set with exactly one element is not the same thing as the element. Thus $\{3\}$ is a set, not a number, whereas 3 is a number, not a set. Similarly, the President is not the same as the Presidency, although the President is the only holder of that office.

23.1.4 Example $[3\,..\,3]$ is a singleton set, but $(3\,..\,3)$ is the empty set.

23.1.5 Exercise Which of these describe (i) the empty set (ii) a singleton?

a)	$\{1, -1\}$	e)	$\{x \in \mathrm{R}^+ \mid x < 1\}$
b)	$\{x \in \mathrm{N} \mid x < 1\}$	f)	$\{x \in \mathrm{R} \mid x^2 - 1 = 0\}$
c)	$\{x \in \mathrm{R} \mid x^2 = 0\}$	g)	$\{x \in \mathrm{R} \mid x^3 + x = 0\}$
d)	$\{x \in \mathrm{R} \mid x^2 < 0\}$		

(Answer on page 243.)

23.1.6 Exercise For each positive integer n, let D_n be the set of positive divisors of n.
a) For which integers n is D_n a singleton?
b) Which integers k are elements of D_n for every positive integer n?
(Answer on page 243.)

23.1.7 Exercise Simplify these descriptions of sets as much as possible, where n is of type integer.
a) $\{n \mid 1 < n < 2\}$.
b) $\{n \mid |n| < 2\}$.
c) $\{n \mid \text{for all integers } m, n < m\}$.

24. Russell's Paradox

The setbuilder notation has a bug: for some predicates $P(x)$, the notation $\{x \mid P(x)\}$ does not define a set. An example is the predicate "x is a set". In that case, if $\{x \mid x \text{ is a set}\}$ were a set, it would be the set of all sets. However, there is no such thing as the set of all sets. This can be proved using the theory of infinite cardinals, but will not be done here.

We now give another example of a definition $\{x \mid P(x)\}$ which does not give a set, and we will prove that it does not give a set. It is historically the first such example and is due to Bertrand Russell. He took $P(x)$ to be "x is a set and x is not an element of itself." This gives the expression "$\{x \mid x \notin x\}$".

We now prove that that expression does not denote a set. Suppose $S = \{x \mid x \notin x\}$ is a set. There are two possibilities: (i) $S \in S$. Then *by definition of* S, S *is not an element of itself*, i.e., $S \notin S$. (This follows from the rule of inference (18.1) on page 29.) (ii) $S \notin S$. In this case, since S is not an element of S and S is the set of *all* sets which are not elements of themselves, it follows from Rule (18.1) that $S \in S$. Both cases are impossible, so there is no such set as S. This is an example of a proof by contradiction, which we will study in detail in Section 86, page 125.

As a result of the phenomenon that the setbuilder notation can't be depended on to give a set, set theory as a mathematical science (as opposed to a useful language) had to be developed on more abstract grounds instead of in the naive way described in this book. The most widely-accepted approach is via Zermelo-Frankel set theory, which unfortunately is complicated and not very natural in comparison with the way mathematicians actually use sets.

Luckily, for most practitioners of mathematics or computer science, this difficulty with the setbuilder notation does not usually arise. In most applications, the notation "$\{x \mid P(x)\}$" has x varying over a specific type whose instances (unlike the type "set") are already known to constitute a set (e.g., x is real — the real numbers form a set). In that case, any meaningful predicate defines a set $\{x \mid P(x)\}$ of elements of that type.

For more about Russell's Paradox, see [Wilder, 1965], starting on page 57.

24.0.8 Exercise (discussion) In considering Russell's Paradox, perhaps you tried unsuccessfully to think of a set which is an element of itself. In fact, most axiomatizations of set theory rule out the possibility of a set being an element of itself. Does doing this destroy Russell's example? What does it say about the collection of all sets?

25. Implication

In Chapter 14, we described certain operations such as "and" and "or" which combine predicates to form compound predicates. There is another logical connective which denotes the relationship between two predicates in a sentence of the form "If P, then Q", or "P implies Q". Such a statement is called an **implication**.

Implications are at the very heart of mathematical reasoning. Mathematical proofs typically consist of chains of implications.

25.1 Definition: implication

For predicates P and Q, the **implication** $P \Rightarrow Q$ is a predicate defined by the truth table

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

In the implication $P \Rightarrow Q$, P is the **hypothesis** or **antecedent** and Q is the **conclusion** or **consequent**.

25.1.1 Example Implication is the logical connective used in translating statements such as "If $m > 5$ and $5 > n$, then $m > n$" into logical notation. This statement could be reworded as, "$m > 5$ and $5 > n$ implies that $m > n$." If we take $P(m,n)$ to be "$(m > 5) \wedge (5 > n)$" and $Q(m,n)$ to be "$m > n$", then the statement "If $m > 5$ and $5 > n$, then $m > n$" is "$P(m,n) \Rightarrow Q(m,n)$".

25.1.2 Usage The implication connective is also called the **material conditional**, and $P \Rightarrow Q$ is also written $P \supset Q$. An implication, that is, a sentence of the form $P \Rightarrow Q$, is also called a **conditional sentence**.

25.1.3 Remarks

a) Definition 25.1 gives a technical meaning to the word "implication". It also has a meaning in ordinary English. Don't confuse the two. The technical meaning makes the word "implication" the *name of a type of statement.*

b) **Warning:** The truth table for implication has surprising consequences which can cause difficulties in reading technical articles. The first line of the truth table says that if P and Q are both true then $P \Rightarrow Q$ is true. In Example 25.1.1, we have "$7 > 5$ and $5 > 3$ implies $7 > 3$" which you would surely agree is true.

   However, the first line of the truth table also means that other statements such as "If $2 > 1$ then $3 \times 5 = 15$" are true. You may find this odd, since the fact that $3 \times 5 = 15$ doesn't seem to have anything to do with the fact that $2 > 1$. Still, it fits with the truth table. Certainly you wouldn't want the fact that P and Q are both true to be grounds for $P \Rightarrow Q$ being *false.*

25.1.4 Exercise Which of these statements are true for all integers m?

a) $m > 7 \Rightarrow m > 5$.
b) $m > 5 \Rightarrow m > 7$.
c) $m^2 = 4 \Rightarrow m = 2$.

(Answer on page 243.)

26. Vacuous truth

The last two lines of the truth table for implication mean that if the hypothesis of an implication is false, the implication is automatically true.

> **26.1 Definition: vacuously true**
> In the case that $P \Rightarrow Q$ is true because P is false, the implication $P \Rightarrow Q$ is said to be **vacuously true**.

26.1.1 Remark The word "vacuous" refers to the fact that in that case the implication says nothing interesting about either the hypothesis or the conclusion. In particular, *the implication may be true, yet the conclusion may be false* (because of the last line of the truth table).

26.1.2 Example Both these statements are vacuously true:
a) If 4 is odd then $3 = 3$.
b) If 4 is odd then $3 \neq 3$.

26.1.3 Remarks Although this situation may be disturbing when you first see it, making either statement in Example 26.1.2 false would result in even more peculiar situations. For example, if you made $P \Rightarrow Q$ false when P and Q are both false, you would then have to say that the statement discussed previously,

"For any integers m and n, if $m > 5$ and $5 > n$ then $m > n$,"

is not always true (substitute 3 for m and 4 for n and you get both P and Q false). This would surely be an unsatisfactory state of affairs.

Most of the time in mathematical writing the implications which are actually stated involve predicates containing variables, and the assertion is typically that the implication is true for all instances of the predicates. Implications involving propositions occur only implicitly in the process of checking instances of the predicates. That is why a statement such as, "If $3 > 5$ and $5 > 4$, then $3 > 4$" seems awkward and unfamiliar.

26.1.4 Example Vacuous truth can cause surprises in connection with certain concepts which are defined by using implication. Let's look at a made-up example here: to say that a natural number n is **fourtunate** (the spelling is intentional) means that if 2 divides n then 4 divides n. Thus clearly 4, 8, 12 are all fourtunate. But *so are* 3 *and* 5. They are *vacuously* fourtunate!

26.1.5 Exercise For each implication, give (if possible) an integer n for which it is true and another for which it is false.

a)	$(n > 7) \Rightarrow (n < 4)$	d)	$(n = 1 \vee n = 3) \Rightarrow (n \text{ is odd})$
b)	$(n > 7) \Rightarrow (n > 4)$	e)	$(n = 1 \wedge n = 3) \Rightarrow (n \text{ is odd})$
c)	$(n > 7) \Rightarrow (n > 9)$	f)	$(n = 1 \vee n = 3) \Rightarrow n = 3$

(Answer on page 243.)

26.1.6 Exercise If possible, give examples of predicates P and Q for which each of these is (i) true and (ii) false.

a) $P \Rightarrow (P \Rightarrow Q)$
b) $Q \Rightarrow (P \Rightarrow Q)$
c) $(P \Rightarrow Q) \Rightarrow P$
d) $(P \Rightarrow Q) \Rightarrow Q$

27. How implications are worded

Implication causes more trouble in reading mathematical prose than all the other logical connectives put together. An implication may be worded in various ways; it takes some practice to get used to understanding all of them as implications.

The five most common ways of wording $P \Rightarrow Q$ are

WI.1 If P, then Q.
WI.2 P only if Q.
WI.3 P implies Q.
WI.4 P is a sufficient condition for Q.
WI.5 Q is a necessary condition for P.

27.1.1 Example For all $x \in \mathrm{Z}$,

a) If $x > 3$, then $x > 2$.
b) $x > 3$ only if $x > 2$.
c) $x > 3$ implies $x > 2$.
d) That $x > 3$ is sufficient for $x > 2$.
e) That $x > 2$ is necessary for $x > 3$.

all mean the *same thing*.

27.1.2 Remarks

a) Watch out particularly for Example 27.1.1(b): it is easy to read this statement backward when it occurs in the middle of a mathematical argument. Perhaps the meaning of (b) can be clarified by expanding the wording to read: "x can be greater than 3 only if $x > 2$."

   Note that sentences of the form "P only if Q" about ordinary everyday things generally do *not* mean the same thing as "If P then Q"; that is because in such situations there are considerations of time and causation that do not come up with mathematical objects. Consider "If it rains, I will carry an umbrella" and "It will rain only if I carry an umbrella".

b) Grammatically, Example 27.1.1(c) is quite different from the first two. For example, (a) is a statement about x, whereas (c) is a statement about statements about x. However, the information they communicate is the same. See 27.3 below.

27.1.3 Exercise You have been given four cards each with an integer on one side and a colored dot on the other. The cards are laid out on a table in such a way that a 3, a 4, a red dot and a blue dot are showing. You are told that, if any of the cards has an even integer on one side, it has a red dot on the other. What is the smallest number of cards you must turn over to verify this claim? Which ones should be turned over? Explain your answer.

27.2 Universally true implications

Implications which are universally true are sometimes stated using the word "every" or "all". For example, the implication, "If $x > 3$, then $x > 2$", could be stated this way: "Every integer greater than 3 is greater than 2" or "All integers greater than 3 are greater than 2". You can recognize such a statement as an implication if what comes after the word modified by "every" or "all" can be reworded as a predicate ("greater than 3" in this case).

27.2.1 Exercise Which of the following sentences say the same thing?

a) If a real number is positive, it has a square root.
b) If a real number has a square root, it is positive.
c) A real number is positive only if it has a square root.
d) Every positive real number has a square root.
e) For a real number to be positive, it is necessary that it have a square root.
f) For a real number to be positive, it is sufficient that it have a square root.

(Answer on page 243.)

27.2.2 Exercise Suppose you have been told that the statement $P \Rightarrow Q$ is false. What do you know about P? About Q?

27.3 Implications and rules of inference

Suppose P and Q are any predicates. If $P \Rightarrow Q$, then the rule of inference $P \vdash Q$ is valid, and conversely if $P \vdash Q$ is valid, then $P \Rightarrow Q$ must be true. This is stated formally as a theorem in texts on logic, but that requires that one give a formal definition of what propositions and predicates are. We will take it as known here.

27.3.1 Example It is a familiar fact about real numbers that for all x and y, $(x > y) \Rightarrow (x > y - 1)$. This can be stated as the rule of inference $x > y \vdash x > y - 1$.

28. Modus Ponens

The truth table for implication may be summed up by saying:

An implication is true unless the hypothesis is true and the conclusion is false.

This fits with the major use of implications in reasoning: if you know that the implication is true and you know that its hypothesis is true, then you know its conclusion is true. This fact is called "modus ponens", and is the most important rule of inference of all:

28.1 Definition: modus ponens

Modus ponens is the rule of inference

$$(P,\, P \Rightarrow Q) \vdash Q \tag{28.1}$$

which is valid for all predicates P and Q.

28.1.1 Remark That modus ponens is valid is a consequence of the truth table for implication (Definition 25.1). If P is true that means that one of the first two lines of the truth table holds. If $P \Rightarrow Q$ is true, one of lines 1, 3 or 4 must hold. The only possibility, then, is line 1, which says that Q is true.

28.2 Uses of modus ponens

A theorem (call it Theorem T) in a mathematical text generally takes the form of an implication: "If [hypotheses $H_1, \ldots, H_n$] are true, then [conclusion]." It will then typically be applied in the proof of some subsequent theorem using modus ponens. In the application, the author will verify that the hypotheses $H_1, \ldots, H_n$ of Theorem T are true, and then will be able to assert that the conclusion is true.

28.2.1 Example As a baby example of this, we prove that $3 \mid 6$ using Theorem 5.1 and Theorem 5.4. By Theorem 5.1, $3 \mid 3$. The *hypotheses* of Theorem 5.4 are that $k \mid m$ and $k \mid n$. Using $k = m = n = 3$ this becomes $3 \mid 3$ and $3 \mid 3$, which is true. Therefore the *conclusion* $3 \mid 3 + 3$ must be true by Theorem 5.4. Since $3 + 3 = 6$ we have that $3 \mid 6$.

29. Equivalence

29.1 Definition: equivalence

Two predicates P and Q are **equivalent**, written $P \Leftrightarrow Q$, if for any instance, both P and Q are true or else both P and Q are false. The statement $P \Leftrightarrow Q$ is called an **equivalence** or a **biconditional**.

29.1.1 Fact The truth table for equivalence is

P	Q	$P \Leftrightarrow Q$
T	T	T
T	F	F
F	T	F
F	F	T

This is the same as $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$.

29.1.2 Usage The usual way of saying that $P \Leftrightarrow Q$ is, "P if and only if Q", or "P is equivalent to Q." The notation "iff" is sometimes used as an abbreviation for "if and only if".

29.1.3 Example $x > 3$ if and only if both $x \geq 3$ and $x \neq 3$.

29.1.4 Warning The statement "$P \Leftrightarrow Q$" *does not say that P is true.*

29.2 Theorem

Two expressions involving predicates and logical connectives are equivalent if they have the same truth table.

29.2.1 Example $P \Rightarrow Q$ is equivalent to $\neg P \vee Q$, as you can see by constructing the truth tables. This can be understood as saying that $P \Rightarrow Q$ is true if and only if either P is *false* or Q is *true.*

29.2.2 Worked Exercise Construct a truth table that shows that $(P \vee Q) \wedge R$ is equivalent to $(P \wedge R) \vee (Q \wedge R)$.
Answer

P	Q	R	$P \vee Q$	$(P \vee Q) \wedge R$	$P \wedge R$	$Q \wedge R$	$(P \wedge R) \vee (Q \wedge R)$
T	T	T	T	T	T	T	T
T	T	F	T	F	F	F	F
T	F	T	T	T	T	F	T
T	F	F	T	F	F	F	F
F	T	T	T	T	F	T	T
F	T	F	T	F	F	F	F
F	F	T	F	F	F	F	F
F	F	F	F	F	F	F	F

29.2.3 Exercise Construct truth tables showing that the following three statements are equivalent:
a) $P \Rightarrow Q$
b) $\neg P \vee Q$
c) $\neg(P \wedge \neg Q)$

29.2.4 Exercise Write English sentences in the form of the three sentences in Exercise 29.2.3 that are equivalent to

$$(x > 2) \Rightarrow (x \geq 2)$$

30. Statements related to an implication

30.1 Definition: converse
The **converse** of an implication $P \Rightarrow Q$ is $Q \Rightarrow P$.

30.1.1 Example The converse of

$$\text{If } x > 3\text{, then } x > 2$$

is

$$\text{If } x > 2\text{, then } x > 3$$

The first is true for all real numbers x, whereas there are real numbers for which the seconxd one is false: *An implication does not say the same thing as its converse.* (If it's a cow, it eats grass, but if it eats grass, it need not be a cow.)

30.1.2 Example In Chapter 10, we pointed out that if the decimal expansion of a real number r is all 0's after a certain point, then r is rational. The converse of this statement is that if a real number r is rational, then its decimal expansion is all 0's after a certain point. *This is false*, as the decimal expansion of $r = 1/3$ shows.

The following Theorem says more about an implication and its converse:

30.2 Theorem
If $P \Rightarrow Q$ and its converse are both true, then $P \Leftrightarrow Q$.

30.2.1 Exercise Prove Theorem 30.2 using truth tables and Theorem 29.2.

30.3 Definition: contrapositive
The **contrapositive** of an implication $P \Rightarrow Q$ is the implication $\neg Q \Rightarrow \neg P$. (Note the reversal.)

30.3.1 Example The contrapositive of

$$\text{If } x > 3\text{, then } x > 2$$

is (after a little translation)

$$\text{If } x \leq 2\text{, then } x \leq 3$$

These two statements are equivalent. This is an instance of a general rule:

30.4 Theorem
An implication and its contrapositive are equivalent.

30.4.1 Exercise Prove Theorem 30.4 using truth tables.

30.4.2 Remark To say, "If it's a cow, it eats grass," is logically the same as saying, "If it doesn't eat grass, it isn't a cow." Of course, the emphasis is different, but the two statements communicate the same facts. In other words,

$$(P \Rightarrow Q) \Leftrightarrow (\neg Q \Rightarrow \neg P)$$

Make sure you verify this by truth tables. The fact that a statement and its contrapositive say the same thing causes many students an enormous amount of trouble in reading mathematical proofs.

30.4.3 Example Let's look again at this (true) statement (see Section 10, page 14):

> If the decimal expansion of a real number r has all 0's after a certain point, it is rational.

The contrapositive of this statement is that if r is not rational, then its decimal expansion does not have all 0's after *any* point. In other words, no matter how far out you go in the decimal expansion of a real number that is not rational, you can find a nonzero entry further out. *This statement is true because it is the contrapositive of a true statement.*

30.4.4 Remark Stating the contrapositive of a statement $P \Rightarrow Q$ requires forming the statement $\neg Q \Rightarrow \neg P$, which requires negating each of the statements P and Q. The preceding example shows that this involves subtleties, some of which we consider in Section 77.

30.4.5 Exercise Write the contrapositive and converse of "If $3 \mid n$ then n is prime". Which is true? (Answer on page 243.)

30.4.6 Exercise Write the converse and the contrapositive of each statement in Exercise 26.1.5 without using "$\neg$".

31. Subsets and inclusion

Every integer is a rational number (see Chapter 7). This means that the sets Z and Q have a special relationship to each other: every element of Z is an element of Q. This is the relationship captured by the following definition:

31.1 Definition: inclusion
For all sets A and B, $A \subseteq B$ if and only if $x \in A \Rightarrow x \in B$ is true for all x.

31.1.1 Usage The statement $A \subseteq B$ is read "A is included in B" or "A is a subset of B".

31.1.2 Example $\mathrm{Z} \subseteq \mathrm{Q}$, $\mathrm{Q} \subseteq \mathrm{R}$ and $\mathrm{I} \subseteq \mathrm{R}$.

Definition 31.1 gives an immediate rule of inference and a method:

31.1.3 Method
To show that $A \subseteq B$, prove that every element of A is an element of B.

31.1.4 Remark If $P(x)$ is a predicate whose only variable is x and x is of type S for some set S, then the extension of $P(x)$, namely $\{x \mid P(x)\}$, is a subset of S.

Some useful consequences of Definition 31.1 are included in the following theorem.

31.2 Theorem

a) For any set A, $A \subseteq A$.
b) For any set A, $\emptyset \subseteq A$.
c) For any sets A and B, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

Proof Using Definition 31.1, the statement $A \subseteq A$ translates to the statement $x \in A \Rightarrow x \in A$, which is trivially true. The statement $\emptyset \subseteq A$ is equivalent to the statement $x \in \emptyset \Rightarrow x \in A$, which is vacuously true for any x whatever (the hypothesis is always false). We leave the third statement to you.

31.2.1 Exercise Prove part (c) of Theorem 31.2.

31.3 Definition: strict inclusion

If $A \subseteq B$ but $A \neq B$ then every element of A is in B but there is at least one element of B not in A. This is symbolized by $A \subsetneq B$, and is read "A is **properly included** in B".

31.3.1 Warning Don't confuse the statement "$A \subsetneq B$" with "$\neg(A \subseteq B)$": the latter means that there is an element of A not in B.

31.3.2 Exercise Prove that for all sets A and B, $(A \subsetneq B) \Rightarrow \neg(B \subseteq A)$.

31.4 Inclusion and elementhood

The statement "$A \subseteq B$" must be carefully distinguished from the statement "$A \in B$".

31.4.1 Example Consider these sets:

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

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

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

A and B have three elements each and C has four. $A \in B$ because A occurs in the list which defines B. However, A is not included in B since $3 \in A$ but $3 \notin B$. On the other hand, $A \in C$ and $A \subseteq C$ both.

31.4.2 Exercise Answer each of (i) through (iii) for the sets X and Y as defined:
(i) $X \in Y$,
(ii) $X \subseteq Y$, and
(iii) $X = Y$.
a) X={1,3}, Y={1,3,5}.
b) X={1,2}, Y={1,{1,2}}.
c) X={1,2}, Y={2,1,1}.
d) X={1,2,{1,3}}, Y={1,3,{1,2}}.
e) X={1,2,{1,3}}, Y={1,{1,2},{1,3}}.
(Answer on page 243.)

31.4.3 Remark The fact that $A \subseteq A$ for any set A means that any set is a subset of itself. This may not be what you expected the word "subset" to mean. This leads to the following definition:

31.5 Definition: proper

A **proper subset** of a set A is a set B with the property that $B \subseteq A$ and $B \neq A$. A **nontrivial subset** of A is a set B with the property that $B \subseteq A$ and $B \neq \emptyset$.

31.5.1 Usage

a) The word "contain" is ambiguous as mathematicians usually use it. If $x \in A$, one often says "A contains x", and if $B \subseteq A$, one often says "A contains B"!

One thing that keeps the terminological situation from being worse than it is is that most of the time in practice either none of the elements of a set are sets or all of them are. In fact, sets such as B and C in Example 31.4.1 which have both sets and numbers as elements almost never occur in mathematical writing except as examples in texts such as this which are intended to bring out the difference between "element of" and "included in"!

Nevertheless, when this book uses the word "contain" in one of these senses, one of the phrases "as an element" or "as a subset" is always added.

b) The original notation for "$A \subseteq B$" was "$A \subset B$". In recent years authors of high school and college texts have begun using the symbol '$\subseteq$' by analogy with '$\leq$'. However, the symbol '$\subset$' is still the one used most by research mathematicians. Some authors have used it to mean '$\subsetneqq$', but that is an entirely terrible idea considering that '$\subset$' originally meant and is still widely used to mean '$\subseteq$'. This text avoids the symbol '$\subset$' altogether.

31.5.2 Exercise Explain why each statement is true for all sets A and B, or give an example showing it is false for some sets A and B:

a) $\emptyset \in A$
b) If $A \subseteq \emptyset$, then $A = \emptyset$.
c) If $A = B$, then $A \subseteq B$.
d) If $\emptyset \in A$ then $A \neq \emptyset$.
e) If $A \in B$ and $B \in C$, then $A \in C$.
f) If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.
g) If $A \subsetneqq B$ and $B \subsetneqq C$, then $A \subsetneqq C$.
h) If $A \neq B$ and $B \neq C$, then $A \neq C$.

31.5.3 Exercise Given two sets S and T, how do you show that S is *not* a subset of T? (Answer on page 244.)

32. The powerset of a set

32.1 Definition: powerset
If A is any set, the set of *all* subsets of A is called the **powerset** of A and is denoted $\mathcal{P}A$.

32.1.1 Remark Using setbuilder notation, $\mathcal{P}A = \{X \mid X \subseteq A\}$.

32.1.2 Example The powerset of $\{1,2\}$ is $\{\emptyset, \{1\}, \{2\}, \{1,2\}\}$, and the powerset of $\{1\}$ is $\{\emptyset, \{1\}\}$.

32.1.3 Fact The definition of powerset gives two rules of inference:

$$B \subseteq A \vdash B \in \mathcal{P}A \tag{32.1}$$

and

$$B \in \mathcal{P}A \vdash B \subseteq A \tag{32.2}$$

32.1.4 Example The empty set is an element of the powerset of every set, since it is a subset of every set.

32.1.5 Warning The empty set is *not* an element of every set; for example, it is not an element of $\{1,2\}$.

32.1.6 Exercise How many elements do each of the following sets have?
a) $\{1,2,3,\{1,2,3\}\}$
b) $\emptyset$
c) $\{\emptyset\}$
d) $\{\emptyset, \{\emptyset\}\}$
(Answer on page 244.)

32.1.7 Exercise Write the powerset of $\{5,6,7\}$. (Answer on page 244.)

32.1.8 Exercise State whether each item in the first column is an element of each set in the second column.

a) 1	a) $\mathbb{Z}$
b) 3	b) $\mathbb{R}$
c) π	c) $\{1,3,7\}$
d) $\{1,3\}$	d) $\{x \in \mathbb{R} \mid x = x^2\}$
e) $\{3,\pi\}$	e) $\mathcal{P}(\mathbb{Z})$
f) $\emptyset$	f) $\emptyset$
g) $\mathbb{Z}$	g) $\{\mathbb{Z}, \mathbb{R}\}$

(Answer on page 244.)

33. Union and intersection

33.1 Definition: union
For any sets A and B, the **union** $A \cup B$ of A and B is defined by

$$A \cup B = \{x \mid x \in A \vee x \in B\} \tag{33.1}$$

33.2 Definition: intersection
For any sets A and B, **intersection** $A \cap B$ is defined by

$$A \cap B = \{x \mid x \in A \wedge x \in B\} \tag{33.2}$$

33.2.1 Example Let $A = \{1,2\}$ and $B = \{2,3,4\}$. Then $A \cup B = \{1,2,3,4\}$ and $A \cap B = \{2\}$. If $C = \{3,4,5\}$, then $A \cap C = \emptyset$.

33.2.2 Exercise What are $\{1,2,3\} \cup \{2,3,4,5\}$ and $\{1,2,3\} \cap \{2,3,4,5\}$? (Answer on page 244.)

33.2.3 Exercise What are $\mathrm{N} \cup \mathrm{Z}$ and $\mathrm{N} \cap \mathrm{Z}$? (Answer on page 244.)

33.2.4 Remark Union and intersection mirror the logical connectives '$\vee$' and '$\wedge$' of section 14. The connection is by means of the extensions of the predicates involved. The extension of $P \vee Q$ is the union of the extensions of P and of Q, and the extension of $P \wedge Q$ is the intersection of the extensions of P and of Q.

33.2.5 Example Let S be a set of poker chips, each of which is a single color, either red, green or blue. Let R, G, B be respectively the sets of red, green and blue chips. Then $R \cup B$ is the set of chips which are either red *or* blue; the '$\cup$' symbol mirrors the "or". And $R \cap B = \emptyset$, since it is false that a chip can be both red and blue.

33.2.6 Warning Although union corresponds with "$\vee$", the set $R \cup B$ of the preceding example could also be described as "the set of red chips *and* blue chips"!

33.2.7 Exercise Prove that for any sets A and B, $A \cap B \subseteq A \cup B$. (Answer on page 244.)

33.2.8 Exercise Prove that for any sets A and B, $A \cap B \subseteq A$ and $A \subseteq A \cup B$.

33.3 Definition: disjoint
If A and B are sets and $A \cap B = \emptyset$ then A and B are said to be **disjoint**.

33.3.1 Exercise Name three different subsets of Z that are disjoint from N. (Answer on page 244.)

33.3.2 Exercise If A and B are disjoint, must $\mathcal{P}(A)$ and $\mathcal{P}(B)$ be disjoint?

34. The universal set and complements

Since we cannot talk about the set of all sets, there is no universal way to mirror TRUE as a set. However, in many situations, all elements are of a particular type. For example, all the elements in Example 33.2.5 are chips. The set of all elements of that type constitutes a single set containing as subsets all the sets under consideration. Such a set is called a **universal set**, and is customarily denoted $\mathcal{U}$.

Given a universal set, we can define an operation corresponding to '$\neg$', as in the following defintion.

34.1 Definition: complement
If A is a set, A^c is the set of all elements in $\mathcal{U}$ but not in A. A^c is called the **complement** of A (note the spelling).

34.1.1 Usage A^c may be denoted $\overline{A}$ or A' in other texts.

34.1.2 Example The complement of N in Z is the set of all negative integers.

34.2 Definition: set difference
Let A and B be any two sets. The **set difference** $A-B$ is the set defined by

$$A-B=\{x \mid x\in A \wedge x\notin B\} \tag{34.1}$$

34.2.1 Example Let $A=\{1,2,3\}$ and $B=\{3,4,5\}$; then $A-B=\{1,2\}$.

34.2.2 Exercise What is $\mathrm{Z}-\mathrm{N}$? What is $\mathrm{N}-\mathrm{Z}$? (Answer on page 244.)

34.2.3 Fact If there is a universal set $\mathcal{U}$, then $A^c=\mathcal{U}-A$.

34.2.4 Usage $A-B$ is written $A\backslash B$ in many texts..

34.2.5 Exercise Let $A=\{1,2,3\}$, $B=\{2,3,4,5\}$ and $C=\{1,7,8\}$. Write out the elements of the following sets:

a)	$A\cup B$	f)	$B-C$
b)	$A\cap B$	g)	$A\cap(B\cup C)$
c)	$B\cup C$	h)	$B\cup(A\cap C)$
d)	$B\cap C$	i)	$B\cup(A-C)$
e)	$A-B$		

(Answer on page 244.)

34.2.6 Exercise State whether each item in the first column is an element of each set in the second column. $A = \{1,3,7\}$, $B = \{1,2,3,4,5\}$, and the universal set is $\mathbb{Z}$.

1)	1	1)	$A \cup B$
2)	4	2)	$A \cap B$
3)	7	3)	$A - B$
4)	-2	4)	$A - \mathbb{Z}$
5)	$\emptyset$	5)	B^c
6)	$\{2,4,5\}$	6)	$\mathcal{P}A$
7)	$\{1,3\}$	7)	$\mathcal{P}(A \cap B)$

(Answer on page 244.)

34.2.7 Exercise Explain why the following statements are true for all sets A and B or give examples showing they are false for some A and B.

a) $\mathcal{P}(A) \cap \mathcal{P}(B) = \mathcal{P}(A \cap B)$
b) $\mathcal{P}(A) \cup \mathcal{P}(B) = \mathcal{P}(A \cup B)$
c) $\mathcal{P}(A) - \mathcal{P}(B) = \mathcal{P}(A - B)$

34.2.8 Exercise Show that for any sets A and B included in a universal set $\mathcal{U}$, if $A \cup B = \mathcal{U}$ and $A \cap B = \emptyset$, then $B = A^c$.

35. Ordered pairs

In analytic geometry, one specifies points in the plane by ordered pairs of real numbers, for example $\langle 3,5 \rangle$. (Most books use round parentheses instead of pointy ones.) This is not the same as the two-element set $\{3,5\}$, because in the ordered pair the order matters: $\langle 3,5 \rangle$ is not the same as $\langle 5,3 \rangle$.

In the ordered pair $\langle 3,5 \rangle$, 3 is the **first coordinate** and 5 is the **second coordinate**. Sometimes, the two coordinates are the same: for example, $\langle 4,4 \rangle$ has first and second coordinates both equal to 4.

An ordered pair in general need not have its first and second coordinates of the same type. For example, one might consider ordered pairs whose first coordinate is an integer and whose second coordinate is a letter of the alphabet, such as $\langle 5, \text{'}a\text{'} \rangle$ and $\langle -3, \text{'}d\text{'} \rangle$.

The following specification gives the operational properties of ordered pairs:

35.1 Specification: ordered pair
An ordered pair $\langle x,y \rangle$ is a mathematical object distinct from x and y which is completely determined by the fact that its first coordinate is x and its second coordinate is y.

35.1.1 Remark Specification 35.1 implies that ordered pairs are the same if and only if their coordinates are the same:

$$(\langle x,y \rangle = \langle x',y' \rangle) \Leftrightarrow (x = x' \wedge y = y')$$

Thus we have a method:

35.1.2 Method
To prove two ordered pairs $\langle x, y\rangle$ and $\langle x', y'\rangle$ are the same, prove that $x = x'$ and $y = y'$.

35.1.3 Exercise Which of these pairs of ordered pairs are equal to each other?
a) $\langle 2,3\rangle$, $\langle 3,2\rangle$.
b) $\langle 3,\sqrt{4}\rangle$, $\langle 3,2\rangle$.
c) $\langle 2,\sqrt{4}\rangle$, $\langle \sqrt{4},2\rangle$.
(Answer on page 244.)

35.1.4 Exercise (discussion) In texts on the foundations of mathematics, an ordered pair $\langle a,b\rangle$ is often *defined* to be the set $\Big\{\{a,b\},\{a\}\Big\}$. Prove that at least when a and b are numbers this definition satisfies Specification 35.1 (with a suitable definition of "coordinate").

36. Tuples

In order to generalize the idea of ordered pair to allow more than two coordinates, we need some notation.

36.1 Definition: n
Let n be an integer, $n \geq 1$. Then $\mathbf{n}$ is defined to be the set

$$\{i \in \mathrm{N} \mid 1 \leq i \leq n\}$$

36.1.1 Example $\mathbf{3} = \{1,2,3\}$.

36.1.2 Exercise Let m and n be positive integers. What is $\mathbf{m} \cap \mathbf{n}$? What is $\mathbf{m} \cup \mathbf{n}$? (Answer on page 244.)

A tuple is a generalization of the concept of ordered pair. A tuple satisfies this specification:

36.2 Specification: tuple
A **tuple** of length n, or n-tuple, is a mathematical object which
T.1 has an ith entry for each $i \in \mathbf{n}$, and
T.2 is distinct from its entries, and
T.3 is completely determined by specifying the ith entry for every $i \in \mathbf{n}$.

36.2.1 Example An ordered pair is the same thing as a 2-tuple.

36.2.2 Usage
a) A 3-tuple is usually called an **ordered triple**.
b) The usual way of denoting a tuple is by listing its entries in order inside angle brackets.

36.2.3 Example $\langle 1,3,3,-2\rangle$ is a tuple of integers. It has length 4. The integer 3 occurs as an entry in this 4-tuple twice, for $i=2$ and $i=3$.

36.2.4 Usage Tuples and their coordinates are often named according to a subscripting convention, by which one refers to the ith entry by subscripting i to the name of the tuple. For example, let $a=\langle 1,3,3,-2\rangle$; then $a_2=3$ and $a_4=-2$. One often makes this convention clear by saying, "Let $a=\langle a_i\rangle_{i\in\mathbf{n}}$ be an n-tuple."

Many authors would use curly brackets here: "$\{a_i\}_{i\in\mathbf{n}}$." Nevertheless, a is not a set.

36.2.5 Usage Many computer scientists refer to a tuple as a "vector". Although this usage is widespread, it is not desirable; in mathematics, a vector is a geometric object which can be *represented* as a tuple, but is not itself a tuple.

It follows from Specification 36.2 that two n-tuples are equal if and only if they have the same entries. Formally:

36.3 Theorem: How to tell if tuples are equal

Let a and b be n-tuples. Then

$$a=b\Leftrightarrow(\forall i{:}\mathbf{n})(a_i=b_i)$$

36.3.1 Exercise Which of these pairs of tuples are equal?

a) $\langle 3,3\rangle$, $\langle 3,3,3\rangle$.
b) $\langle 2,3\rangle$, $\langle 2,3,3\rangle$.
c) $\langle 2,3,2\rangle$, $\langle 3,2,2\rangle$.

(Answer on page 244.)

36.4 Special tuples

For formal completeness, one also has the concept of the **null tuple** (or empty tuple) $\langle\rangle$, which has length 0 and no entries, and a 1-tuple, which has length 1 and one entry.

The index set for the null tuple is the empty set. There is only one null tuple. In the context of formal language theory the unique null tuple is often denoted "Λ" (capital lambda) or sometimes "ϵ" (small epsilon). We will use the notation Λ here.

36.4.1 Exercise For each tuple, give the integer n for which it is an n-tuple and also give its second entry.

a)	$\langle 3,4,0\rangle$	d)	$\langle\langle\langle 2,\langle 1,5\rangle\rangle,7\rangle,9\rangle$
b)	$\langle\langle 3,4\rangle,\langle 1,5\rangle\rangle$	e)	$\langle 3,\{1,2\}\rangle$
c)	$\langle 3,\langle 5,\langle 2,1\rangle\rangle\rangle$	f)	$\langle \mathrm{N},\mathrm{Z},\mathrm{Q},\mathrm{R}\rangle$

(Answer on page 244.)

37. Cartesian Products

37.1 Definition: Cartesian product of two sets
Let A and B be sets. $A \times B$ is the set of all ordered pairs whose first coordinate is an element of A and whose second coordinate is an element of B. $A \times B$ is called the **Cartesian product** of A and B (in that order).

37.1.1 Example if $A = \{1,2\}$ and $B = \{2,3,4\}$, then

$$A \times B = \Big\{\langle 1,2\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \langle 2,2\rangle, \langle 2,3\rangle, \langle 2,4\rangle\Big\}$$

and

$$B \times A = \Big\{\langle 2,1\rangle, \langle 2,2\rangle, \langle 3,1\rangle, \langle 3,2\rangle, \langle 4,1\rangle, \langle 4,2\rangle\Big\}$$

37.1.2 Exercise Write out the elements of $\{1,2\} \times \{a,b\}$. (Answer on page 244.)

37.2 Theorem
If A is any set, then $A \times \emptyset = \emptyset \times A = \emptyset$.

37.2.1 Exercise Prove Theorem 37.2.

37.2.2 Example $\mathrm{R} \times \mathrm{R}$ is often called the "real plane", since it consists of all ordered pairs of real numbers, and each ordered pair represents a point in the plane once a coordinate system is given. Graphs of straight lines and curves are subsets of $\mathrm{R} \times \mathrm{R}$. For example, the x-axis is $\Big\{\langle x,0\rangle \mid x \in \mathrm{R}\Big\}$ and the parabola $y = x^2$ is $\Big\{\langle x,y\rangle \mid x \in \mathrm{R} \wedge y = x^2\Big\}$, which could be written $\Big\{\langle x,x^2\rangle \mid x \in \mathrm{R}\Big\}$ (recall the discussion in Section 19.2).

37.3 Definition: diagonal
For any set A, the subset $\Big\{\langle a,a\rangle \mid a \in A\Big\}$ of $A \times A$ of all pairs whose two coordinates are the same is called the **diagonal** of A, denoted Δ_A.

37.3.1 Worked Exercise Write out the diagonal of $\{1,2\} \times \{1,2\}$.
Answer $\{\langle 1,1\rangle, \langle 2,2\rangle\}$.

37.3.2 Example The diagonal Δ_{R} of $\mathrm{R} \times \mathrm{R}$ is the 45-degree line from lower left to upper right. It is the graph of the equation $y = x$.

37.4 Cartesian products in general

The notion of Cartesian product can be generalized to more than two factors using the idea of tuple.

37.5 Definition: Cartesian product
Let $A_1, A_2, \ldots, A_n$ be sets — in other words, let $\langle A_i \rangle_{i \in \mathbf{n}}$ be an n-tuple of sets. Then $A_1 \times A_2 \times \cdots \times A_n$ is the set

$$\Big\{\langle a_1, a_2, \ldots, a_n\rangle \mid (\forall i{:}\mathbf{n})(a_i \in A_i)\Big\} \tag{37.1}$$

of all n-tuples whose ith coordinate lies in A_i.

37.5.1 Example The set $\mathrm{R} \times \mathrm{Z} \times \mathrm{R}$ has triples as elements; it contains as an element the ordered triple $\langle \pi, -2, 3\rangle$, but not, for example, $\langle -2, \pi, 3\rangle$.

37.5.2 Warning Observe that $\mathrm{R} \times \mathrm{N} \times \mathrm{R}$, $(\mathrm{R} \times \mathrm{N}) \times \mathrm{R}$ and $\mathrm{R} \times (\mathrm{N} \times \mathrm{R})$ are three different sets; in fact, any two of them are disjoint. Of course, in an obvious sense they all represent the same data.

37.5.3 Example Consider the set

$$D = \{\langle m, n\rangle \mid m \text{ divides } n\}$$

where m and n are of type integer. Thus $\langle 3, 6\rangle$, $\langle -3, 6\rangle$ and $\langle 5, 0\rangle$ are elements of D but not $\langle 3, 5\rangle$. D is *not* a Cartesian product, although it is a (proper) subset of the cartesian product $\mathrm{Z} \times \mathrm{Z}$. The point is that a pair in $A \times B$ can have any element of A as its first coordinate and any element of B as its second coordinate, *regardless of what the first coordinate is.* In D what the second coordinate is depends on what the first coordinate is.

A set such as D is a relation, a concept discussed later.

37.6 Exercise set

Exercises 37.6.1 through 37.6.6 give "facts" which may or may not be correct for all sets A, B and C. State whether each "fact" is true for all sets A, B and C, or false for some sets A, B or C, and for those that are not true for all sets, give examples of sets for which they are false.

37.6.1 $A \times A = A$. (Answer on page 244.)

37.6.2 $A \times B = B \times A$.

37.6.3 $A \cup (B \times C) = (A \cup B) \times (A \cup C)$.

37.6.4 $A \cap (B \times C) = (A \cap B) \times (A \cap C)$.

37.6.5 $A \times (B \times C) = (A \times B) \times C$.

37.6.6 $\mathcal{P}(A \times B) = \mathcal{P}(A) \times \mathcal{P}(B)$.

37.7 Exercise set

The statements in problems 37.7.1 through 37.7.3 are true for all sets A, B and C, except that in some cases some of the sets A, B and C have to be nonempty if the statement is to be true for *all* other sets named. Amend the statement in each case so that it is true.

37.7.1 For all sets A, B and C, $A \times C = B \times C \Rightarrow A = B$. (Answer on page 244.)

37.7.2 For all sets A and B, $A \times B = B \times A \Rightarrow A = B$.

37.7.3 For all sets A, B and C, $A \subseteq B \Rightarrow (A \times C) \subseteq (B \times C)$.

37.8 Cartesian product in Mathematica

The `dmfuncs.m` package contains the command `CartesianProduct`, which gives the Cartesian product of a sequence of sets. For example,

```
CartesianProduct[{1,2},{a,b,c},{x,y}]
```

produces

```
{{1, a, x}, {1, a, y}, {1, b, x}, {1, b, y}, {1, c, x}, {1, c, y},
 {2, a, x}, {2, a, y}, {2, b, x}, {2, b, y}, {2, c, x}, {2, c, y}}
```

37.8.1 Exercise (Mathematica) The command `CartesianProduct` mentioned in 37.8 works on any lists, not just sets (see 17.2, page 27). Write a precise description of the result given when `CartesianProduct` is applied to a sequence of lists, some of which contain repeated entries.

37.9 Exponential notation

If all the sets in a Cartesian product are the same, exponential notation is also used. Thus $A^2 = A \times A$, $A^3 = A \times A \times A$, and in general

$$A^n = A \times A \times \cdots \times A$$

(n times). These are called **Cartesian powers** of A; in particular, A^2 is the **Cartesian square** of A. This notation is extended to 0 and 1 by setting $A^0 = \{\langle\rangle\}$ (the singleton set containing the null tuple as an element) and $A^1 = A$.

37.9.1 Exercise Let $A = \{1,2\}$ and $B = \{3,4,5\}$. Write all the elements of each set:

a)	A^0	f)	$B \times A$
b)	A^1	g)	$A \times A \times B$
c)	A^2	h)	$A \times (A \times B)$
d)	A^3	i)	$(A \times B) \cup A$
e)	$A \times B$	j)	$(A \times B) \cap A$

(Answer on page 244.)

37.9.2 Exercise For each pair of numbers $\langle m,n\rangle \in \{1,2,3,4,5,6,7\} \times \{1,2,3,4,5,6,7\}$, state whether item m in the first column is an element of the set in item n of the second column. $A = \{1,3,7\}$, $B = \{1,2,3,4,5\}$.

1.	$\langle 3,5\rangle$	1.	$A\times A$
2.	$\langle 3,3\rangle$	2.	A^3
3.	$\langle 3,3,5\rangle$	3.	$A\times B$
4.	$\big\{\langle 3,5\rangle,\langle 7,5\rangle\big\}$	4.	$B\times A$
5.	$\langle\{3,7\},\{3,5\}\rangle$	5.	$\mathcal{P}(A\times B)$
6.	$\emptyset$	6.	$\mathcal{P}A\times\mathcal{P}B$
7.	$\langle 1,7,7\rangle$	7.	B^2

(Answer on page 244.)

38. Extensions of predicates with more than one variable

In section 18.1, page 27, we discussed the extension of a predicate containing one variable; the extension is a subset of the type set of the variable. For example, the extension of "$x<5$" (x real) is the subset $\{x \mid x<5\}$ of R.

38.1.1 Remark A predicate can contain several occurrences of one variable. If it contains no occurrences of other variables, it is still said to contain one variable. For example, "$(x<5)\wedge(x>1)$" contains two occurrences of one variable, namely x. On the other hand, "$(x-y<5)\wedge(x+y>1)$" contains two variables, x and y.

A predicate with more than one variable, such as "$x<y$" (x,y real), has an extension which is a subset of a Cartesian product of its variable types.

38.1.2 Example The extension of "$x<y$" in $\mathrm{R}\times\mathrm{R}$ is

$$\{\langle x,y\rangle \mid x<y\}$$

which is a subset of $\mathrm{R}\times\mathrm{R}$.

38.1.3 Example The extension of the predicate "$x=x$" in R is the subset R of R, whereas the extension of the predicate "$x=y$" in $\mathrm{R}\times\mathrm{R}$ is Δ_{R}, the diagonal subset of $\mathrm{R}\times\mathrm{R}$.

38.1.4 Worked Exercise Write out the extension of the predicate "has the same prime divisors as" in $\{2,3,4,6\}^2$.
Answer $\{\langle 2,2\rangle,\langle 3,3\rangle,\langle 2,4\rangle,\langle 4,2\rangle,\langle 4,4\rangle,\langle 6,6\rangle\}$.

38.1.5 Remark The Cartesian product in which the extension of a predicate lies is not uniquely determined. For example, the predicate "$x<y$" has an extension in the set $\mathrm{R}\times\mathrm{R}\times\mathrm{R}$, namely the subset $\big\{\langle x,y,z\rangle \mid x<y\big\}$. In this case, there is no condition on the variable z. There is a good reason for allowing this situation. For example, the predicate "$y<z$" also has an extension in $\mathrm{R}\times\mathrm{R}\times\mathrm{R}$, namely $\big\{\langle x,y,z\rangle \mid y<z\big\}$. Looking at it this way allows us to say that the extension of "$x<y\wedge y<z$" in $\mathrm{R}\times\mathrm{R}\times\mathrm{R}$ is the intersection of the extensions of "$x<y$" and "$y<z$".

By the way, we could have regarded $\mathrm{R}\times\mathrm{R}\times\mathrm{R}$ as the set of tuples

$$\Big\{\langle y,z,x\rangle \mid x,y,z\in\mathrm{R}\Big\}$$

Then the extension of "$x<y$" would be $\Big\{\langle y,z,x\rangle \mid x<y\Big\}$. Because of this sort of thing, it pays to be careful to describe exactly what set the extension of a predicate lies in.

38.2 Exercise set

In Problems 38.2.1 through 38.2.4, describe explicitly the extensions of the predicates in the given set; x, y and z are real and n is an integer. Associate x,y,z with coordinates in a tuple in alphabetical order.

38.2.1 $x>n$, in $\mathrm{R}\times\mathrm{N}$. (Answer on page 244.)

38.2.2 $x+y=x+1$, in $\mathrm{R}\times\mathrm{R}$. (Answer on page 244.)

38.2.3 $y=1$, in R. (Answer on page 244.)

38.2.4 $x+y=z$, in $\mathrm{R}\times\mathrm{R}\times\mathrm{R}\times\mathrm{R}$. (Answer on page 244.)

39. Functions

39.1 The concept of function

In analytic geometry or calculus class you may have studied a real-valued function such as $G(x)=x^2+2x+5$. This function takes as input a real number and gives a real number as value; for example, the statement that $G(3)=20$ means that an input of 3 gives an output of 20. It also means that the point $\langle 3,20\rangle$ is on the graph of the equation $y=G(x)$.

The concept of function to be studied here is more general than that example in several ways. In the first place, a function F can have one type of input and another type of output. An example is the function which gives the number of characters in an English word: the input is a string of characters, say 'cat', and the output is its length, 3 in this case. Also, the most general sort of function need not be given by a formula the way G is. For example, a price list is a function with input the name of an item and output the price of the item. The relationship between the name and the price is rarely given by a formula.

Here is the specification:

> **39.2 Specification: function**
>
> A function F is a mathematical object which determines and is completely determined by the following data:
>
> F.1 F has a **domain**, which is a set and is denoted $\operatorname{dom} F$.
>
> F.2 F has a **codomain**, which is also a set and is denoted $\operatorname{cod} F$.
>
> F.3 For each element $a\in\operatorname{dom} F$, F has a **value** at a. This value is completely determined by a and F and must be an element of $\operatorname{cod} F$. It is denoted by $F(a)$.

39.2.1 Warning This specification for function is both complicated and subtle and has conceptual traps. One of the *complications* is that the concept of function given here carries more information with it than what is usually given in calculus books. One of the *traps* is that you may tend to think of a function in terms of a formula for computing it, whereas a major aspect of our specification is that what a function is is independent of how you compute it.

39.2.2 Usage The standard notation $F:A\to B$ communicates the information that F is a function with domain A and codomain B. A and B are sets; A consists of exactly those data which you can use as input to (you can "plug into") the function F, and every value of F must lie in B. (Not every element of B has to be a value.)

In the expression "$F(a)$", a is called the **argument** or **independent variable** or **input** to F and $F(a)$ is the **value** or **dependent variable** or **output**. The operation of finding $F(a)$ given F and a is called **evaluation** or **application**.

If $F(a)=b$, one may say "F takes a to b" or "a goes to b under F".

It follows from the definition that we have the following rule of inference:

$$F:A\to B,\ a\in A\ \vdash\ F(a)\in B \tag{39.1}$$

39.2.3 Warning You should distinguish between F, which is the name of the function, and $F(a)$, which is the value of F at the input value a. Nevertheless, a function is often referred to as $F(x)$ — a notation which has the value of telling you what the notation for the input variable is.

39.2.4 Usage Functions are also called **mappings**, although in some texts the word "mapping" is given a special meaning.

39.3 Examples of functions

We give some simple examples of functions to illustrate the basic idea, and then after some more discussion and terminology we will give more substantial examples.

39.3.1 Example The first example is the function $G:\mathrm{R}\to\mathrm{R}$ defined by $G(x)=x^2+2x+5$, which was discussed previously. Referring to it as $G:\mathrm{R}\to\mathrm{R}$ specifies that the domain and codomain are both R.

39.3.2 Usage As is often the case in analytic geometry and calculus texts, we did not formally specify the domain and codomain of G when we defined it at the beginning of this section. In such texts, the domain is often defined implicitly as the set of all real numbers for which the defining formula makes sense. For example, a text might set $S(x)=1/x$, leaving you to see that the domain is $\mathrm{R}-\{0\}$. Normally the codomain is not mentioned at all; it may usually be assumed to be R. In this text, on the other hand, *every function will always have an explicit domain and codomain.*

39.3.3 Example Here is an example of a function with a finite domain and codomain. Let $A=\{1,2,3\}$ and $B=\{2,4,5,6\}$. Let $F:A\to B$ be defined by requiring that $F(1)=4$ and $F(2)=F(3)=5$.

39.3.4 Example Let S be some set of English words, for example the set of words in a given dictionary. Then the length of a word is a function $L : S \to \mathrm{N}$. For example, $L(\text{'}cat\text{'}) = 3$ and $L(\text{'}abbadabbadoo\text{'}) = 12$. (This function in Mathematica is `StringLength`. For example, `StringLength["cat"]` returns `3`.)

39.3.5 Example Let $F : \mathrm{N} \to \mathrm{N}$ be defined by requiring that $F(0) = 0$ and for $n > 0$, $F(n)$ is the nth prime in order. Thus $F(1) = 2$, $F(2) = 3$, $F(3) = 5$, and $F(100) = 541$. (This function is given by the word `Prime` in Mathematica.)

39.3.6 Remarks The preceding examples illustrate several points:

a) You don't have to give a formula to give a function. In the case of Example 39.3.3, we defined F by giving its value explicitly at every element of the domain. Of course, this is possible only when the domain is a small finite set.
b) There must be a value $F(x)$ for every element x of the domain, but not every element of the codomain has to be a value of the function. Thus 4 is not a value of the function in 39.3.5.
c) Different elements of the domain can have the same value (different inputs can give the same output). This happens in Example 39.3.1 too; thus $G(1) = G(-3) = 8$.

39.3.7 Exercise Let $A = \{1,2,3\}$. Let $F : A \to \mathcal{P}\mathcal{P}A$ be defined by requiring that $F(n) = \{B \in \mathcal{P}A \mid n \in B\}$. What are $F(1)$ and $F(2)$? (Answer on page 244.)

39.3.8 Exercise Let A be as in the preceding exercise, and define $G : A \to \mathcal{P}\mathcal{P}A$ by $G(n) = \{B \in \mathcal{P}A \mid n \notin B\}$. What are $G(1)$ and $G(2)$?

39.3.9 Exercise Let $F : \mathrm{Z} \to \mathcal{P}\mathrm{Z}$ be defined by requiring that $F(n)$ be the set of divisors of n. What are $F(0)$, $F(1)$, $F(6)$ and $F(12)$?

39.3.10 Exercise Give an example of a function $F : \mathrm{R} \to \mathrm{R}$ with the property

$$r \text{ is an integer if and only if } F(r) \text{ is not an integer}$$

39.3.11 Exercise Let S be a set and $G : S \to \mathcal{P}S$ a function. Let the subset A of S be defined by

$$A = \{x \mid x \notin G(x)\}$$

Show that there is no element $x \in S$ for which $G(x) = A$.

39.4 Functions in Mathematica

In Mathematica, the name of the function is followed by the input in square brackets. For example, $\sin x$ is entered as `Sin[x]`.

You can define your own functions in Mathematica. The function $G(x) = x^2 + 2x + 5$ of Example 39.3.1 can be defined by typing

```
g[x_] := x^2 + 2 x + 5
```
(39.2)

Then if you typed `g[3]`, Mathematica would return `20`, and if you typed `g[t]`, Mathematica would return `5 + 2 t + t^2`. Comments:

a) All built-in Mathematica functions, such as `Sin`, start with a capital letter. It is customary for the user to use lowercase names so as to avoid overwriting the Mathematica definition of some function. (You could define `Sin` to be anything you want, but that would be undesirable.) Thus there would be no error message if you typed `G` instead of `g` in (39.2), but it is not the Done Thing.

b) On the *left* side of a definition, the variable must be followed by an underline. This is how Mathematica distinguishes between a function and the value of a function.

c) One normally writes `:=` for the equals sign in making a definition. There are occasions when the ordinary equals sign may be used, but a rule of thumb for definitions is to use `:=`.

A function that is defined by giving individual values instead of a formula can be defined in Mathematica by doing just that. For example, the function F in Example 39.3.3 can be defined by entering

```
F[1] := 4; F[2] := 5; F[3] := 5
```
(39.3)

(When commands are strung together with semicolons in this way, Mathematica answers with the last value, `5` in this case. These commands could have been entered on separate lines.)

Mathematica does not keep track of the domain and codomain of the function. If you try to evaluate the function at an input for which it has not been defined, you get back what you typed. For example, if you had entered only the commands in (39.3) and then typed `F[6]`, Mathematica would return `F[6]`.

39.5 More about the input to a function

Let's look at the function G of Example 39.3.1 again. We can calculate that $G(3) = 20$. Since $1+2=3$, it follows that $G(1+2)=20$. Since $\sqrt{9}=3$, it follows that $G(\sqrt{9})=20$. It is the element x (here 3) of the domain (here R) that is being given as input to G, *not* the *name* of the element. It doesn't matter how you represent 3, the value of G at 3 is still 20.

This is summed up by the following theorem:

39.6 Theorem: The Substitution Property

Let $F: A \to B$ be any function, and suppose that $a \in A$ and $a' \in A$. If $a = a'$, then $F(a) = F(a')$.

The last sentence of Specification 39.2 can be stated more precisely this way:

39.7 Theorem: How to tell if functions are equal

If $F_i : A_i \to B_i$, ($i = 1,2$), are two functions, then

$$(F_1 = F_2) \Leftrightarrow \Big(A_1 = A_2 \wedge B_1 = B_2 \wedge (\forall x{:}A_1)\big(F_1(x) = F_2(x)\big)\Big) \qquad (39.4)$$

39.7.1 Method

To show that two functions are the same you have to show they have the same domain, the same codomain and for each element of the domain they have the same value.

39.7.2 Exercise Suppose $F: A \to B$ and $G: A \to B$. What do you have to do to prove that $F \neq G$?

39.7.3 Warning Since Formula (39.4) is an equivalence, this means that the function $S: \mathrm{R} \to \mathrm{R}$ for which $S(x) = x^2$ is not the same as the function $T: \mathrm{R} \to \mathrm{R}^+$ for which $T(x) = x^2$. They have the same domain and the same value at every element of the domain, but they do not have the same codomain. This distinction is often not made in the literature. In some theoretical contexts it is vital to make it, but in others (for example calculus) it makes no difference and is therefore quite rightly ignored. *In this text we are purposely making all the distinctions made in a sizeable fraction of the research literature.*

39.8 The abstract idea of function

As was noted previously, the specification given for functions says nothing about the *formula* for the function. The function G in Example 39.3.1 was defined by the formula $G(x) = x^2 + 2x + 5$, but the definition of the function called F in Example 39.3.3 never mentioned a formula.

Until late in the nineteenth century, functions were usually thought of as defined by formulas. However, problems arose in the theory of Fourier analysis which made mathematicians require a more general notion of function. The definition of function given here is the modern version of that more general concept.It replaces the *algorithmic* and *dynamic* idea of a function as a way of computing an output value given an input value by the *static, abstract* concept of a function as having a domain, a codomain, and a value lying in the codomain for each element of the domain. Of course, often a definition by formula will give a function in this modern sense. However, there is no *requirement* that a function be given by a formula.

The modern concept of function has been obtained from the formula-based idea by *abstracting* basic properties the old concept had and using them as the basis of the new definition. This process of definition by abstracting properties is a major tool in mathematics, and you will see more examples of it later in the book (see Chapter 51, for example).

The concept of function as a formula never disappeared entirely, but was studied mostly by logicians who generalized it to the study of function-as-algorithm. (This is an oversimplification of history.) Of course, the study of algorithms is one of the central topics of modern computer science, so the notion of function-as-formula (more generally, function-as-algorithm) has achieved a new importance in recent years.

Nevertheless, computer science needs the abstract definition of function given here. Functions such as sin may be (and quite often are) programmed to look up their values in a table instead of calculating them by a formula, an arrangement which gains speed at the expense of using more memory.

40. The graph of a function

40.1 Definition: graph of a function
The **graph** of a function $F : A \to B$ is the set

$$\Big\{\langle a, F(a)\rangle \mid a \in A\Big\}$$

of ordered pairs whose first coordinates are all the elements of A with the second coordinate in each case being the value of F at the first coordinate. The graph of F is denoted by $\Gamma(F)$

40.1.1 Fact $\Gamma(F)$ is necessarily a subset of $A \times B$.

40.1.2 Example For the function G of Example 39.3.1, the graph

$$\Gamma(G) = \Big\{\langle x, G(x)\rangle \mid x \in \mathrm{R}\Big\} = \Big\{\langle x, y\rangle \mid x \in \mathrm{R} \wedge y = x^2 + 2x + 5\Big\}$$

which is a subset of $\mathrm{R} \times \mathrm{R}$. $\Gamma(G)$ is of course what is usually called the graph of G in analytic geometry — in this case it is a parabola.

40.1.3 Example The graph of the function F of Example 39.3.3 is

$$\{\langle 1,4\rangle, \langle 2,5\rangle, \langle 3,5\rangle\}$$

40.2 Properties of the graph of a function

Using the notion of graph, Specification 39.2 can be reworded as requiring the following statements to be true about a function $F : A \to B$:
GS.1 $\operatorname{dom} F$ is *exactly* the set of first coordinates of the graph, and
GS.2 For every $a \in A$, there is exactly one element b of B such that $\langle a, b\rangle \in \Gamma(F)$.

40.2.1 Fact GS–2 implies that, for all $a \in A$ and $b \in B$,

$$\Big(\langle a, b\rangle \in \Gamma(F) \wedge \langle a, b'\rangle \in \Gamma(F)\Big) \Rightarrow b = b' \tag{40.1}$$

40.2.2 Usage The requirement of formula (40.1)is sometimes described by saying that functions have to be **single-valued**.

40.2.3 Warning Do not confuse the property of being single-valued with the Substitution Property of Theorem 39.6, which in terms of the graph can be stated this way: For all $a \in A$ and $b \in B$,

$$\Big(\langle a, b\rangle \in \Gamma(F) \wedge a = a'\Big) \Rightarrow \langle a', b\rangle \in \Gamma \tag{40.2}$$

40.2.4 Remark When a function goes from R to R the way the function G of Example 39.3.1 does, its graph can be drawn, and then the single-valued property implies that a vertical line will cross the graph only once. In general, you can't draw the graph of a function (for example, the length function defined on words, as in Example 39.3.4).

40.2.5 Remark Not every set of ordered pairs can be the graph of a function. A set P of ordered pairs is said to be **functional** or to have the **functional property** if

$$\Big(\langle a,b\rangle \in P \wedge \langle a,b'\rangle \in P\Big) \Rightarrow b=b' \tag{40.3}$$

Of course, Formula (40.1) above says that the graph of a function is functional. Conversely, if a set P of ordered pairs is functional, then there are sets A and B and a function $F:A\to B$ for which $\Gamma(F)=P$. F is constructed this way:

FC.1 A must be the set of first coordinates of pairs in P.

FC.2 B can be any set containing as elements all the second coordinates of pairs in P.

FC.3 For each $a\in A$, define $F(a)=b$, where $\langle a,b\rangle$ is the ordered pair in P with a as first coordinate: there is only one such by the functional property.

40.2.6 Exercise For $A=\{1,2,3,4\}$, $B=\{3,4,5,6\}$, which of these sets of ordered pairs is the graph of a function from A to B?

a) $\Big\{\langle 1,3\rangle,\langle 2,3\rangle,\langle 3,4\rangle,\langle 4,6\rangle\Big\}$.
b) $\Big\{\langle 1,3\rangle,\langle 2,3\rangle,\langle 4,5\rangle,\langle 4,6\rangle\Big\}$.
c) $\Big\{\langle 1,3\rangle,\langle 2,3\rangle,\langle 4,6\rangle\Big\}$.
d) $\Big\{\langle 1,3\rangle,\langle 2,4\rangle,\langle 3,5\rangle,\langle 4,6\rangle\Big\}$.

(Answer on page 245.)

40.2.7 Exercise If $P\subseteq A\times B$, then the **opposite** of P is the set $P^{op}=\Big\{\langle b,a\rangle \mid \langle a,b\rangle\in P\Big\}$. Give examples of:

a) a function $F:A\to B$ for which $(\Gamma(F))^{op}$ is the graph of a function.
b) a function $G:A\to B$ for which $(\Gamma(G))^{op}$ is not the graph of a function.

40.2.8 Exercise Create a Mathematica command `InGraphQ` with the property that the expression `InGraphQ[F,{x,y}]` returns `True` if $\langle x,y\rangle\in\Gamma(F)$ and `False` otherwise.

40.2.9 Usage

a) In mathematical texts in complex function theory, and in older texts in general, functions are not always assumed single-valued.
b) As you can see, part FD. 2 requires $\Gamma(F)$ to have the functional property. In texts which do not require that a function's codomain be specified, a function is often defined simply as a set of ordered pairs with the functional property.

40.3 Explicit definitions of function

In many texts, the concept of function is defined explicitly (as opposed to being given a specification) by some such definition as this: A function F is an ordered triple $\langle A, B, \Gamma(F)\rangle$ for which

FD.1 A and B are sets and $\Gamma(F) \subseteq A \times B$, and

FD.2 If $a \in A$, then there is exactly one ordered pair in $\Gamma(F)$ whose first coordinate is a.

41. Some important types of functions

41.1.1 Identity function For any set A, the **identity function** $\mathrm{id}_A : A \to A$ is the function that takes an element to itself; in other words, for every element $a \in A$, $\mathrm{id}_A(a) = a$. Thus its graph is the diagonal of $A \times A$ (see 37.3).

41.1.2 Warning Do not confuse the identity function with the concept of identity for a predicate of Section 13.1.2, or with the concept of identity for a binary operation of Section 45.

41.1.3 Inclusion function If $A \subseteq B$, then there is an **inclusion function** $\mathrm{inc} : A \to B$ which takes every element in A to itself regarded as an element of B. In other words, $\mathrm{inc}(a) = a$ for every element $a \in A$. Observe that the graph of inc is the same as the graph of id_A and they have the same domain, so that the only difference between them is what is considered the codomain (A for id_A, B for the inclusion of A in B).

41.1.4 Constant function If A and B are nonempty sets and b is a specific element of B, then the **constant function** $C_b : A \to B$ is the function that takes every element of A to b; that is, $C_b(a) = b$ for all $a \in A$. A constant function from R to R has a horizontal line as its graph.

41.1.5 Empty function If A is any set, there is exactly one function $E : \emptyset \to A$. Such a function is an **empty function**. Its graph is empty, and it has no values. "An identity function does nothing. An empty function has nothing to do."

41.1.6 Coordinate function If A and B are sets, there are two **coordinate functions** (or **projection functions**) $p_1 : A \times B \to A$ and $p_2 : A \times B \to B$. The function p_i takes an element to its ith coordinate ($i = 1, 2$). Thus for $a \in A$ and $b \in B$, $p_1\langle a, b\rangle = a$ and $p_2\langle a, b\rangle = b$. More generally, for any Cartesian product $\prod_{i=1}^{n} A_i$ there are n coordinate functions; the ith one takes a tuple $\langle a_1, \ldots, a_n\rangle$ to a_i.

41.1.7 Binary operations The operation of adding two real numbers gives a function

$$+ : \mathrm{R} \times \mathrm{R} \to \mathrm{R}$$

which is an example of a binary operation, treated in detail in Chapter 45.

41.1.8 Exercise For each function $F : A \to B$, give $F(2)$ and $F(4)$.
a) $A = B = \mathrm{R}$, F is the identity function.
b) $A = B = \mathrm{R}$, $F = C_{42}$ (the constant function).
c) $A = \mathrm{R}^+, B = \mathrm{R}$, F is the inclusion function.
(Answer on page 245.)

41.1.9 Exercise Give the graphs of these functions. $A = \{1,2,3\}$, $B = \{2,3\}$.
a) id_A.
b) The inclusion of B into A.
c) The inclusion of B into Z.
d) $C_3 : A \to B$.
e) $p_1 : A \times B \to A$.
(Answer on page 245.)

42. Anonymous notation for functions

The curly-brackets notation for sets has the advantage that it allows you to refer to a set without giving it a name. For example, you can say, "$\{1,2,3\}$ has three elements," instead of, "The set A whose elements are 1, 2 and 3 has three elements." This is useful when you only want to refer to it once or twice. A notation which describes without naming is called **anonymous notation**.

The notation we have introduced for functions does not have that advantage. When the two versions of the squaring function were discussed, it was necessary to call them S and T in order to say anything about them.

42.1 Lambda notation

Two types of anonymous notation for functions are used in mathematics. The older one is called **lambda notation** and is used mostly in logic and computer science. To illustrate, the squaring function would be described as "the function $\lambda x.x^2$". The format is: λ, then a letter which is the independent variable, then a period, then a formula in terms of the independent variable which gives the value of the function. In the λ-notation, the variable is bound and so can be changed without changing the function: $\lambda x.x^2$ and $\lambda t.t^2$ denote the same function.

42.1.1 Example The function defined in Example 39.3.1 is $\lambda x.(x^2 + 2x + 5)$.

42.1.2 Example On a set A, the identity function id_A is $\lambda x.x$.

42.2 Barred arrow notation

The other type of anonymous notation is the **barred arrow notation**, which has in recent years gained wide acceptance in pure mathematics and appears in some texts on computer science, too. In this notation, the squaring function would be called the function $x \mapsto x^2 : \mathrm{R} \to \mathrm{R}$, and the function in Example 39.3.1 could be written $x \mapsto x^2 + 2x + 5$.

The barred arrow tells you what an element of the domain is changed to by the function. The straight arrow goes from *domain* to *codomain*, the barred arrow from *element of the domain* to *element of the codomain*.

42.2.1 Example On a set A, the identity function id_A is $x \mapsto x : A \to A$.

42.2.2 Other notations One would write `Function[x,x^2]` or `#^2&` in Mathematica for $x \mapsto x^2$ or $\lambda x.x^2$. The `#` sign stands for the variable and the `&` sign at the end indicates that this is a function rather than an expression to evaluate. More complicated examples require parentheses; for example, $x \mapsto x^2 + 2x + 5$ becomes `(#^2+2 #+5)&`.

42.2.3 Exercise Write the following functions using λ notation and using barred arrow notation. A and B are any sets.

a) $F : \mathrm{R} \to \mathrm{R}$ given by $F(x) = x^3$.
b) $p_1 : A \times B \to A$.
c) Addition on R.

(Answer on page 245.)

43. Predicates determine functions

43.1 Definition: characteristic function

Let A be a set. Any subset B of A determines a **characteristic function** $\chi_B^A : A \to \{\mathrm{TRUE}, \mathrm{FALSE}\}$ defined by requiring that $\chi_B^A(x) = \mathrm{TRUE}$ if $x \in B$ and $\chi_B^A(x) = \mathrm{FALSE}$ if $x \notin B$.

43.1.1 Example If $A = \{1,2,3,4\}$ and $B = \{1,4\}$ then $\chi_B^A(1) = \mathrm{TRUE}$ and $\chi_B^A(2) = \mathrm{FALSE}$.

43.1.2 Fact $\chi_\emptyset^A$ is the constant function which is always FALSE, and χ_A^A is the constant TRUE.

43.1.3 Predicates as characteristic functions Since the extension of a predicate is a subset of its data type, *the truth value of a predicate is the characteristic function of its extension.* For example, the statement "n is even" (about integers) is TRUE if n is even and FALSE otherwise, so that the value of the characteristic function of the subset E of Z consisting of the even integers is the truth value of the predicate "n is even".

Predicates with more than one variable similarly correspond to characteristic functions of subsets of Cartesian products. Thus the truth value of the statement "$m<n$" (about integers) is the characteristic function of the subset

$$\Big\{\langle m,n\rangle \mid m<n\Big\}$$

of $\mathrm{Z}\times\mathrm{Z}$.

43.1.4 Exercise Give the graphs of these functions. $A=\{1,2,3\}$, $B=\{2,3\}$.

a) $\chi_B^A : A \to \{\mathrm{TRUE},\mathrm{FALSE}\}$.
b) The predicate "n is odd" where n is an element of A, regarded as a function to $\{\mathrm{TRUE},\mathrm{FALSE}\}$.
c) $+ : B\times B \to \mathrm{Z}$.

(Answer on page 245.)

43.1.5 Exercise Suppose that a predicate P regarded as the characteristic function of its extension is a constant function. What can you say about P?

44. Sets of functions

As mathematical entities, functions can be elements of sets; in fact the discovery of function spaces, in which functions are regarded as points in a space, was one of the great advances of mathematics.

44.1 Definition: B^A
If A and B are sets, the set of *all* functions $F : A \to B$ is denoted B^A.

44.1.1 Warning The notation B^A refers to the functions from A to B, from the *exponent* to the *base*. It is easy to read this notation backward.

44.1.2 Remark Remark 97.1.5, page 139, and Theorem 122.3, page 188, explain why the notation B^A is used.

44.1.3 Example The function G of Example 39.3.1 is an element of the set R^{R}, and the function of Example 39.3.3 is an element of the set

$$\{2,4,5,6\}^{\{1,2,3\}}$$

44.1.4 Example The function $+ : \mathrm{R}\times\mathrm{R}\to\mathrm{R}$ is an element of $\mathrm{R}^{\mathrm{R}\times\mathrm{R}}$.

44.1.5 Exercise Let $A = \{1,2,3,4,5\}$. For each item in the first column, state which of the items in the second column it is an element of.

a) id_{R}	1) R^{R}
b) the inclusion of A in Z	2) Z^A
c) $\langle 1,2,1\rangle$	3) $\mathrm{R}\times\mathrm{Z}\times\mathrm{R}$
d) $x \mapsto x^2 : \mathrm{R}\to\mathrm{R}$	4) $(\mathrm{R}^+)^{\mathrm{R}}$

(Answer on page 245.)

45. Binary operations

45.1 Definition: binary operation
For any set S, a function $S\times S\to S$ is called a **binary operation** on S.

45.1.1 Remark The domain of a binary operation is the Cartesian square of its codomain. Thus a binary operation on a set S is an element of the function set $S^{S\times S}$. In particular, a function $G : A\times B\to C$ is a binary operation only if $A = B = C$.

45.1.2 Example The function that takes $\langle 1,2\rangle$ to 1, and $\langle 1,1\rangle, \langle 2,1\rangle$ and $\langle 2,2\rangle$ all to 2 is a binary operation on the set $\{1,2\}$.

45.1.3 Example The usual operations of addition, subtraction and multiplication are binary operations on the set R of real numbers. Thus addition is the function

$$\langle x,y\rangle \mapsto x+y : \mathrm{R}\times\mathrm{R}\to\mathrm{R}$$

45.1.4 Example Division is a function from $\mathrm{R}\times(\mathrm{R}-\{0\})$ to R, and so does not fit our definition of binary operation. Restricted to the nonzero reals, however, it is a function from $\mathrm{R}-\{0\}\times\mathrm{R}-\{0\}$ to $\mathrm{R}-\{0\}$ (this says if you divide a nonzero number by another one, you get a nonzero number), and so is a binary operation on $\mathrm{R}-\{0\}$.

45.1.5 Example For any set A, union and intersection are binary operations on $\mathcal{P}A$. This means that each of union and intersection is a function from $\mathcal{P}A\times\mathcal{P}A$ to $\mathcal{P}A$ (not from $A\times A$ to A).

45.1.6 Example For any set A, define the binary operation P on A by requiring that $aPb = b$ for all a and b in A. P is called the **right band** on A.

45.1.7 Unary operations In the context of abstract algebra, a function from a set A to A is called a **unary operation** on A by analogy with the concept of binary operation.

45.1.8 Example Taking the complement of a set is a unary operation on a powerset.

45.1.9 Example The function $-: \mathbf{Z} \to \mathbf{Z}$ (similarly for $\mathbf{R}$) that takes a number r to its negative $-r$ is a unary operation on $\mathbf{R}$. This is distinct from the *binary* operation of subtraction $\langle m, n\rangle \mapsto m - n$.

46. Fixes

46.1 Prefix notation

I have normally written the name of the function to the left of the argument (input value), thus: $F(x)$. This is called **prefix notation** for functions and is familiar from analytic geometry and calculus texts.

46.1.1 Parentheses around the argument Trigonometric functions like $\sin x$ are also written in prefix notation, but it is customary to omit parentheses around the argument. (Pascal and many other computer languages require the parentheses, however, and Mathematica requires square brackets). Many mathematical writers omit the parentheses in other situations too, writing "Fx" instead of "$F(x)$". It is important not to confuse evaluation written like this with multiplication.

46.2 Infix notation

Many common binary operations are normally written *between* their two arguments, "$a + b$" instead of "$+(a, b)$". This is called **infix notation** and naturally applies only to functions with two arguments.

46.2.1 Example The expression $3 - (5 + 2)$ is in infix notation. In prefix notation, the same expression is $-(3, +(5, 2))$.

46.3 Postfix notation

Some authors write functions on the *right*, for example "xF" or "$(x)F$" instead of "$F(x)$". This is called **postfix notation**. This has real advantages which will become apparent when we look at composition in Chapter 98.

46.4 Polish notation

When binary operations are written in either prefix or postfix notation, parentheses are not necessary to resolve ambiguities. In infix notation, for example, parentheses are necessary to distinguish between "$a + b * c$" (which is the same as "$a + (b * c)$") and "$(a + b) * c$". In prefix notation, "$a + b * c$" can be written "$+\ a\ *\ b\ c$" and "$(a + b) * c$" can be written "$*\ +\ a\ b\ c$". Note the use of spaces to separate the items. This is particularly important when multidigit constants are used: for example $35\ 22\ +$ in postfix notation returns 57.

Writing functions of two or more arguments using prefix notation and no parentheses is called **Polish notation** after the eminent Polish logician Jan Łukasiewicz, who invented the notation in the 1920's. Writing functions on the right which are normally infixed, without parentheses, is naturally called **reverse Polish notation**.

Most computer languages use prefix and infix notation similar to that of ordinary algebra. The language Lisp uses prefix notation (with parentheses) and the various dialects of Forth characteristically use reverse Polish notation (no parentheses). Either prefix or postfix notation in a computer language makes it easier to write an interpreter or compiler for the language.

46.4.1 Example The expression of Example 46.2.1 in prefix notation without using parentheses is $-\ 3\ +\ 5\ 2$. In postfix notation it is $3\ 5\ 2\ +\ -$.

46.4.2 Example $a+b+c$ in reverse Polish notation can be written either as $a\ b\ +\ c\ +$ or as $a\ b\ c\ +\ +$.

46.4.3 Exercise Write $(35+22)(6+5)$ in reverse Polish notation. Use $*$ for multiplication. (Answer on page 245.)

46.4.4 Exercise Write $b^2 - 4ad$ in reverse Polish notation. Use $*$ for multiplication and don't use exponents.

Fix notation in Mathematica Mathematica gives the user control over whether a function is written in infix notation or not. For example, we remarked in Section 14.4 that in Mathematica one writes `Xor[p,q]` for the expression $p\,\text{XOR}\,q$. However, by putting tildes before and after the name of a function in Mathematica, you can use it as an infix; thus you can write `p ~Xor~ q` instead of for `Xor[p,q]`.

A function `F` can be used in postfix form by prefixing it with `//`. For example, one can write `Sqrt[2]` or `2 // Sqrt`.

47. More about binary operations

47.1 Notation

In discussing binary operations in general, we will refer to an operation Δ on a set A; thus $\Delta : A \times A \to A$. This operation will be used in infix notation, the way addition and multiplication are normally written, so that we write $a \Delta b$ for $\Delta(a,b)$. Using an unfamiliar symbol like 'Δ' avoids the sneaky way familiar symbols like "+" cause you to fall into habits acquired by long practice in algebra (for example, assuming commutativity) that may not be appropriate for a given situation.

47.1.1 Warning Don't confuse Δ, representing a binary operation, with the diagonal $\Delta_A \subseteq A \times A$ defined in Definition 37.3, page 52.

47.1.2 Multiplication tables We will sometimes give a binary operation Δ on a small finite set by means of a **multiplication table**: For example, here is a binary operation on the set $\{a,b,c\}$.

Δ	a	b	c
a	b	c	a
b	c	c	a
c	a	a	b

The value of $x \Delta y$ is in the row marked $\times$ and the column marked y. This means for example that $a \Delta b = c$ and $c \Delta a = a$.

47.1.3 Example The binary operation of Example 45.1.2 is

Δ	1	2
1	2	1
2	2	2

47.1.4 Exercise Give the multiplication table for the right band on the set $\{1,2,3\}$.

47.1.5 Exercise Give the multiplication table for the operation of union on the powerset of $\{1,2,3\}$.

48. Associativity

48.1 Definition: associative
A binary operation Δ is **associative** if for any elements x, y, z of A,

$$x \Delta (y \Delta z) = (x \Delta y) \Delta z \tag{48.1}$$

48.1.1 Remark In ordinary functional notation (prefix notation), the definition of associative says $\Delta(x, \Delta(y,z) = \Delta(\Delta(x,y,),z))$. In postfix notation: $x\ y\ \Delta\ z\ \Delta = x\ y\ z\ \Delta\ \Delta$.

48.1.2 Example The usual operations of addition and multiplication are associative, but not subtraction; for example, $3-(5-7) \neq (3-5)-7$. The operation given in 47.1.2 is not associative; for example, $(a \Delta a) \Delta c = b \Delta c = a$, but $a \Delta (a \Delta c) = a \Delta a = b$.

48.1.3 Example For any nonempty set X, union and intersection are associative binary operations in $\mathcal{P}X$.

48.1.4 Example For real numbers r and s, let $\max : \mathrm{R} \times \mathrm{R} \to \mathrm{R}$ and $\min : \mathrm{R} \times \mathrm{R} \to \mathrm{R}$ be the functions defined by: $\max(r,s)$ is the larger of r and s and $\min(r,s)$ the smaller. If $r = s$ then $\max(r,s) = \min(r,s) = r = s$. These are both associative binary operations on the set R of real numbers.

48.1.5 Exercise Prove that for any set S, union is an associative binary operation on $\mathcal{P}S$. (Answer on page 245.)

48.1.6 Exercise Prove that for any set S, intersection is an associative binary operation on $\mathcal{P}S$.

48.1.7 Exercise Show that the right band operation on any set A is associative.

48.1.8 Exercise Find a binary operation Δ on some set A with the property that, for some element $a \in A$,

$$(a \,\Delta\, a) \,\Delta\, a) \neq a \,\Delta\, (a \,\Delta\, a)$$

48.1.9 Exercise Is the binary operation Δ given by this table associative? Give reasons for your answer.

Δ	a	b
a	a	a
b	b	a

48.1.10 Exercise Prove that $\max : \mathrm{R} \times \mathrm{R} \to \mathrm{R}$ is associative (see Example 48.1.4).

48.2 The general associative law

If Δ is an associative operation on A, then it is associative in a more general sense, in that it satisfies the **General Associative Law**: Any two meaningful products involving Δ and $a_1, a_2, \ldots, a_n$ (names of elements of A) in that order denote the same element of A.

48.2.1 Example If $a \,\Delta\, (b \,\Delta\, c) = (a \,\Delta\, b) \,\Delta\, c$, then all five meaningful ways of writing the product of four elements are the same:

$$\begin{aligned} a \,\Delta\, (b \,\Delta\, (c \,\Delta\, d)) &= a \,\Delta\, ((b \,\Delta\, c) \,\Delta\, d) = (a \,\Delta\, b) \,\Delta\, (c \,\Delta\, d) \\ &= ((a \,\Delta\, b) \,\Delta\, c) \,\Delta\, d = (a \,\Delta\, (b \,\Delta\, c)) \,\Delta\, d \end{aligned}$$

49. Commutativity

49.1 Definition: commutative
A binary operation Δ on a set A is **commutative** if for all $x, y \in A$, $x \,\Delta\, y = y \,\Delta\, x$.

49.1.1 Example The operations of addition and multiplication, but not subtraction, are commutative operations on R.

49.1.2 Example The binary operations mentioned in Examples 48.1.2, 48.1.3 and 48.1.4 are commutative.

49.1.3 Exercise Let C be a set. Define the binary operation Δ for all subsets A and B of C by

$$A \Delta B = (A \cup B) - (A \cap B)$$

a) Show that Δ is commutative.
b) Show that $A \Delta B = (A - B) \cup (B - A)$.

49.1.4 The General Commutative Law There is a general commutative law analogous to the general associative law: It says that if Δ is commutative *and associative,* then the names $a_1, ..., a_n$ in an expression $a_1 \,\Delta\, a_2 \,\Delta \ldots \Delta\, a_n$ can be rearranged in any way without changing the value of the expression. We will not prove that law here.

50. Identities

50.1 Definition: identity
If Δ is a binary operation on a set A, an element e is a **unity** or **identity** for Δ if for all $x \in A$,

$$x \,\Delta\, e = e \,\Delta\, x = x \tag{50.1}$$

50.1.1 Warning Don't confuse the concept of identity for a binary operation with the concept of an identity function in 41.1.1, page 63. These are two different ideas, but there is a relationship between them (see 98.2.3, page 141).

50.1.2 Example The binary operation of Example 45.1.2 has no identity.

50.1.3 Example The number 1 is an identity for the binary operation of multiplication on R, and 0 is an identity for $+$.

50.1.4 Exercise Which of these binary operations (i) is associative, (ii) is commutative, (iii) has an identity?

Δ	a	b	c
a	a	a	a
b	b	b	b
c	c	c	c

(1)

Δ	a	b	c
a	b	a	a
b	a	c	a
c	a	a	b

(2)

(Answer on page 245.)

50.1.5 Exercise Show that the right band operation on a set with more than one element does not have an identity.

50.1.6 Example The binary operation of multiplication on the set of even integers is associative and commutative, but it has no identity.

50.1.7 Exercise Let S be any set. What is the identity element for the binary operation of union on $\mathcal{P}S$? (Answer on page 245.)

50.1.8 Exercise Let S be any set. What is the identity element for the binary operation of intersection on $\mathcal{P}S$?

50.1.9 Exercise Does $\max : \mathrm{R} \times \mathrm{R} \to \mathrm{R}$ have an identity? What about $\max : \mathrm{R}^+ \times \mathrm{R}^+ \to \mathrm{R}^+$ defined the same way?

The basic fact about identities is:

50.2 Theorem: Uniqueness theorem for identities
If Δ is a binary operation on a set A with identity e, then e is the only identity for Δ.

Proof This follows immediately from Definition 50.1: if e and f are both identities, then $e = e \Delta f$ because f is an identity, and $e \Delta f = f$ because e is an identity.

50.2.1 Exercise Give a rule of inference that allows one to conclude that a certain object is an identity for a binary operation Δ.

50.2.2 Exercise (hard) Find all the binary operations on the set $\{a, b\}$, and state whether each one is associative, is commutative, and has an identity element.

51. Relations

The mathematical concept of relation is an abstraction of the properties of relations such as "$=$" and "$<$" in much the same way as the modern concept of function was abstracted from the concrete functions considered in freshman calculus, as described in Section 39.8.

51.1 Definition: binary relation
A **binary relation** α from a set A to a set B is a subset of $A \times B$. If $\langle a, b\rangle \in \alpha$, then one writes $a\,\alpha\,b$.

51.1.1 Remark *Any* subset of $A \times B$ for *any* sets A and B is a binary relation from A to B.

51.1.2 Fact Definition 51.1 gives the following equivalence, which describes two different ways of writing the same thing:

$$a\,\alpha\,b \Leftrightarrow \langle a, b\rangle \in \alpha \tag{51.1}$$

51.1.3 Usage A relation corresponds to a predicate with two variables, one of type A and the other of type B: the predicate is true if $a\,\alpha\,b$ (that is, if $\langle a, b\rangle \in \alpha$) and false otherwise. Logic texts often define a relation to be a predicate of this type, but the point of view taken here (that a relation is a set of ordered pairs) is most common in mathematics and computer science.

51.1.4 Example Let $A = \{1, 2, 3, 6\}$ and $B = \{1, 2, 3, 4, 5\}$. Then

$$\alpha = \Big\{\langle 2,2\rangle, \langle 1,5\rangle, \langle 1,3\rangle, \langle 2,5\rangle, \langle 2,1\rangle\Big\}$$

is a binary relation from A to B. For this definition, we know $1\,\alpha\,5$ and $2\,\alpha\,1$ but it is not true that $1\,\alpha\,2$.

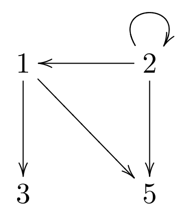
51.1.5 Exercise Write all ordered pairs in the relation from A to B:

a) $A = \{1,2,3\}$, $B = \{1,3,5\}$. α is "$\neq$".
b) $A = \{2,3,5,7\}$, $B = \{1,2,3,4,5,6,7,8,9,10\}$, α is "divides".
c) $A = \{0,1,2,3\}$, $B = \{1,2,3\}$, α is "divides".

(Answer on page 245.)

51.2 Picturing relations

A relation on a small finite set can be exhibited by drawing dots representing the elements of A and B and an arrow from x to y if and only if $x \,\alpha\, y$. Here is the relation in Example 51.1.4 exhibited in this way:



Such a picture is called the **digraph** representing the relation. Digraphs are studied in depth in Chapters 144 and 151.

51.2.1 Example Two not very interesting binary relations from A to B are the **empty relation** $\emptyset \subseteq A \times B$ and the **total relation** $A \times B$. If E denotes the empty relation, then $a \,\mathrm{E}\, b$ is false for any $a \in A$ and $b \in B$, and if T denotes the total relation, $a \,\mathrm{T}\, b$ is *true* for any $a \in A$ and $b \in B$.

51.2.2 Example In a university, the pairs of the form $\langle$student, class$\rangle$ where the student is registered for the class form a relation from the set of students to the set of classes.

> **51.3 Definition:** $\mathrm{Rel}(A,B)$
> The set of all relations from A to B is denoted by $\mathrm{Rel}(A,B)$.

51.3.1 Remark By Definition 51.1, $\mathrm{Rel}(A,B)$ is the same thing as the powerset $\mathcal{P}(A \times B)$; the only difference is in point of view.

51.4 The projections from a relation

A relation α from A to B is a subset of $A \times B$ by definition, so there are functions $p_1^\alpha : \alpha \to A$, $p_2^\alpha : \alpha \to B$, which are the restrictions of the coordinate functioncoordinate (projection) functions (see 41.1.6, page 63) from $A \times B$ to A and to B.

51.4.1 Example If α is defined as in Example 51.1.4, then $p_1^\alpha : \alpha \to \{1,2,3,6\}$ and $p_2^\alpha : \alpha \to \{1,2,3,4,5\}$. In particular, $p_1^\alpha(\langle 1,5 \rangle) = 1$.

52. Relations on a single set

52.1 Definition: relation on a set
If α is a relation from A to A for some set A, then α is a subset of $A \times A$. In that case, α is called a **relation on** A.

52.1.1 Example "$>$" is a relation on R; one element of it is $\langle 5,3\rangle$.

52.1.2 Example A particular relation that any set A has on it is the diagonal Δ_A; $\Delta_A = \{\langle a,a\rangle \mid a \in A\}$. Δ_A is simply the equals relation on A. Don't confuse this with the use of Δ to denote an arbitrary binary operation as in Chapter 45.

52.1.3 Exercise Let $A = \{1,2,3,4\}$. Write out all the ordered pairs in the relation R on A if
a) $a \mathrm{R} b \Leftrightarrow a < b$
b) $a \mathrm{R} b \Leftrightarrow a = b$
c) $a \mathrm{R} b \Leftrightarrow b = 3$.
d) $a \mathrm{R} b \Leftrightarrow a$ and b are both odd.
(Answer on page 245.)

53. Relations and functions

53.1 Functional relations

The graph $\Gamma(F)$ of a function $F : A \to B$ is a binary relation from A to B. It relates $a \in A$ to $b \in B$ precisely when $b = F(a)$. Of course, not any relation can be the graph of a function: to be the graph of a function, a binary relation α from A to B must have the functional property described in 40.2:

$$(a\,\alpha\, b \text{ and } a\,\alpha\, b') \Rightarrow b = b' \tag{53.1}$$

A relation satisfying Equation (53.1) is called a **functional relation**.

This requirement can fail because there are ordered pairs $\langle a,b\rangle$ and $\langle a,b'\rangle$ in α with $b \neq b'$. Even if it is satisfied, α may not be the graph of a function from A to B, since there may be elements $a \in A$ for which there is no ordered pair $\langle a,b\rangle \in \alpha$. However, a functional relation in $\mathrm{Rel}(A,B)$ is always the graph of a function whose domain is some *subset* of A.

53.1.1 Usage For some authors a function is simply a functional relation. For them, the domain and codomain are not part of the definition.

53.1.2 Exercise Which of these are functional relations?
a) $\{\langle 1,3\rangle, \langle 2,3\rangle, \langle 3,4\rangle\}$.
b) $\{\langle 1,1\rangle, \langle 1,2\rangle, \langle 2,3\rangle\}$.
c) $\{\langle x, \sqrt{x}\rangle \mid x \in \mathrm{R}\}$.
d) $\{\langle \sqrt{x}, x\rangle \mid x \in \mathrm{R}\}$.
e) $\{\langle \sqrt{x}, x\rangle \mid x \in \mathrm{R}^+\}$.
(Answer on page 245.)

As we have seen, the concept of relation from A to B is a generalization of the concept of function from A to B. In general, for a given $a \in A$ there may be no ordered pairs $\langle a,b\rangle \in \alpha$ or there may be more than one. Another way of saying this is that for a given element $a \in A$, there is a *set* $\{b \in B \mid \langle a,b\rangle \in \alpha\}$. For α to be a function from A to B, each such set must be a singleton. In general, a relation associates a (possibly empty) *subset* of B to each element of A.

53.2 Definition: relation as function to powerset

If α is a relation from A to B, let $\alpha^* : A \to \mathcal{P}B$ denote the function defined by $\alpha^*(a) = \{b \in B \mid \langle a,b\rangle \in \alpha\}$.

53.2.1 Remark Definition 53.2 gives us a process that constructs a function from A to the powerset of B for each relation from A to B. For any $a \in A$ and $b \in B$,

$$b \in \alpha^*(a) \quad \Leftrightarrow \quad a\alpha b$$

53.2.2 Example For the relation α of Example 51.1.4, we have $\alpha^*(1) = \{3,5\}$, $\alpha^*(2) = \{1,2,5\}$ and $\alpha^*(3) = \emptyset$.

53.2.3 Exercise Write the function $\alpha^* : A \to \mathcal{P}B$ corresponding to the relation in Problem 51.1.5(a). (Answer on page 245.)

Conversely, if we have a function $F : A \to \mathcal{P}B$, we can construct a relation:

53.3 Definition: relation induced by a function to a powerset

Given $F : A \to \mathcal{P}B$, the relation α_F from A to B is defined by $a\,\alpha_F\,b$ if and only if $b \in F(a)$.

53.3.1 Remark In the preceding definition, it makes sense to talk about $b \in F(a)$, because $F(a)$ is a *subset* of B.

53.3.2 Example Let $F : \{1,2,3\} \to \mathcal{P}(\{1,2,3\})$ be defined by $F(1) = \{1,2\}$, $F(2) = \{2\}$ and $F(3) = \emptyset$. Then $\alpha_F = \{\langle 1,1\rangle, \langle 1,2\rangle, \langle 2,2\rangle\}$.

53.3.3 Exercise A function $F : \mathrm{Z} \to \mathcal{P}\mathrm{Z}$ has $F(1) = \{3,4\}$, $F(2) = \{1,3,4\}$, $F(-666) = \{0\}$, and $F(n) = \emptyset$ for all other integers n. List the ordered pairs in the corresponding relation α_F on Z. (Answer on page 245.)

53.3.4 Exercise Let F be the function of Problem 39.3.9, page 58. List the ordered pairs in α_F that have 6 as first element.

54. Operations on relations

54.1 Union and intersection

Since relations from A to B are subsets of $A \times B$, all the usual set operations such as union and intersection can be performed on them.

54.1.1 Example On R, the union of Δ_{R} and "$<$" is (of course!) "$\leq$ ", and the intersection of "$\leq$" and "$\geq$" is Δ_{R}. These statements translate into the obviously true statements

$$r \leq s \Leftrightarrow (r < s \vee r = s)$$

and

$$(r \leq s \wedge r \geq s) \Leftrightarrow r = s$$

> **54.2 Definition: opposite**
> The **opposite** of a relation $\alpha \in \mathrm{Rel}(A,B)$ is the relation $\alpha^{\mathrm{op}} \in \mathrm{Rel}(B,A)$ defined by $\alpha^{\mathrm{op}} = \Big\{\langle b,a\rangle \mid \langle a,b\rangle \in \alpha\Big\}$.

54.2.1 Fact This definition gives an equivalence

$$b\alpha^{\mathrm{op}}a \;\Leftrightarrow\; a\,\alpha\,b$$

It follows that $\alpha \mapsto \alpha^{\mathrm{op}} : \mathrm{Rel}(A,B) \to \mathrm{Rel}(B,A)$ is a function.

54.2.2 Example On R the opposite of "$>$" is "$<$" and the opposite of "$\leq$" is "$\geq$". Of course, for any set A, the opposite of Δ_A is Δ_A.

55. Reflexive relations

> **55.1 Definition: reflexive**
> Let α be a binary relation on A. α is **reflexive** if $a\,\alpha\,a$ for every element $a \in A$.

55.1.1 Example Δ_A is reflexive on any set A, and the relation "$\leq$ " is reflexive on R.

55.1.2 Example On the powerset of any set the relation "$\subseteq$" is reflexive.

55.1.3 Example The relation "$<$" is not reflexive on R, and neither is the relation $\mathcal{S} \Leftrightarrow$ "is the sister of" on the set W of all women, since no one is the sister of herself.

55.1.4 Example Another important type of reflexive relation are the relations like $x\mathcal{N}y \Leftrightarrow |x-y| < 0.1$, defined on R. "$\mathcal{N}$" stands for "**near**". The choice of 0.1 as a criterion for nearness is not important; what is important is that it is a fixed number.

The relations $\mathcal{S}$ and $\mathcal{N}$ will be used several times below in examples.

55.1.5 Fact Let α be a relation on a set A. Then α is reflexive if and only if $\Delta_A \subseteq \alpha$.

55.1.6 Remark The statement that a relation α defined on a set A is reflexive depends on both α and A. For example, the relation

$$\Big\{\langle 1,1\rangle, \langle 1,2\rangle, \langle 2,2\rangle\Big\}$$

is reflexive on $\{1,2\}$ but not on $\{1,2,3\}$.

55.1.7 Warning It is wrong to say that the relation α of 55.1.6 is "reflexive at 1 but not at 3". Reflexivity and irreflexivity are properties of the relation and the set it is defined on, not of particular elements of the set on which the relation is defined. This comment also applies to the other properties of relations discussed in this section.

55.1.8 Fact The digraph of a reflexive relation must have a loop on every node.

55.1.9 Exercise Which of these relations is reflexive?
a) $\alpha = \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle\}$ on $\{1,2,3\}$.
b) $\alpha = \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle\}$ on N.
c) "divides" on Z.
d) α on R defined by $x\alpha y \Leftrightarrow x^2 = y^2$.
(Answer on page 245.)

56. Symmetric relations

56.1 Definition: symmetric
A relation α on a set A is **symmetric** if for all $a,b \in A$,

$$a\,\alpha\,b \;\Rightarrow\; b\,\alpha\,a$$

56.1.1 Example The equals relation on any set is symmetric, and so is the nearness relation $\mathcal{N}$ (see Example 55.1.4). The sister relation $\mathcal{S}$ (Example 55.1.2) is not symmetric on the set of all people, but its restriction to the set of all women is symmetric.

56.1.2 Warning It is important to understand the precise meaning of the definition of symmetric. It is given in the form of an implication: $a\,\alpha\,b \Rightarrow b\,\alpha\,a$. Thus (a) it could be vacuously true (the empty relation is symmetric!) and (b) it does *not* assert that $a\,\alpha\,b$ for any particular elements a and b: *that α is symmetric does not mean* $(a\,\alpha\,b) \wedge (b\,\alpha\,a)$.

56.1.3 Remark The digraph of a symmetric relation has the property that between two distinct nodes there must either be two arrows, one going each way, or no arrow at all.

56.1.4 Exercise Which of these relations is symmetric?

a) $\alpha = \{\langle 1,2\rangle, \langle 2,3\rangle, \langle 1,3\rangle, \langle 2,1\rangle, \langle 4,1\rangle\}$ on $\{1,2,3,4\}$.
b) $\alpha = \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle\}$ on N.
c) The empty relation on N.
d) "is the brother of" on the set of all people.

(Answer on page 245.)

56.1.5 Exercise Show that if a relation α on a set A is not symmetric, then A has at least two distinct elements.

57. Antisymmetric relations

57.1 Definition: antisymmetric

A relation α on a set A is **antisymmetric** if for all $a, b \in A$,

$$(a\,\alpha\,b \wedge b\,\alpha\,a) \Rightarrow a = b$$

57.1.1 Warning Antisymmetry is not the negation of symmetry; there are relations such as Δ which are both symmetric and antisymmetric and others such as "divides" on Z which are neither symmetric nor antisymmetric.

57.1.2 Exercise Prove that on any set A, Δ_A is antisymmetric.

57.1.3 Exercise Prove that on Z, "divides" is neither symmetric nor antisymmetric.

57.1.4 Remark The digraph of an antisymmetric relation may not have arrows going both ways between two distinct nodes.

57.1.5 Example Antisymmetry is typical of many order relations: for example, the relations "$<$" and "$\leq$" on R are antisymmetric. Orderings are covered in Chapter 134.

57.1.6 Example The inclusion relation on the powerset of a set is antisymmetric. This says that for any sets A and B, $A \subseteq B$ and $B \subseteq A$ together imply $A = B$.

57.1.7 Example The relation "$<$" is vacuously antisymmetric, and on any set S, Δ_S is both symmetric and antisymmetric.

57.1.8 Example The nearness relation $\mathcal{N}$ is not antisymmetric; for example, $0.25\,\mathcal{N}\,0.3$ and $0.3\,\mathcal{N}\,0.25$, but $0.25 \neq 0.3$.

57.1.9 Exercise Which of these relations is antisymmetric?

a) $\alpha = \{\langle 1,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle, \langle 2,2\rangle\}$ on $\mathbb{N}$.
b) "divides" on $\mathbb{N}$.
c) $>$ on $\mathbb{R}$.
d) "is the brother of" on the set of all people.

(Answer on page 245.)

57.1.10 Exercise Show that if a relation α on a set A is not antisymmetric, then A has at least two distinct elements.

57.1.11 Exercise Let α be a relation on a set A. Prove that α is antisymmetric if and only if $\alpha \cap \alpha^{\mathrm{op}} \subseteq \Delta_A$. (Another problem like this is Problem 84.2.5, page 124.)

58. Transitive relations

58.1 Definition: transitive

A relation α on A is **transitive** if for all elements a, b and c of A,

$$(a\,\alpha\, b \wedge b\,\alpha\, c) \Rightarrow a\,\alpha\, c$$

58.1.1 Example All the relations Δ_A, "$<$", "$\leq$" and "$\subseteq$" are obviously transitive. That equals is transitive is equivalent to the statement from high-school geometry that two things equal to the same thing are equal to each other.

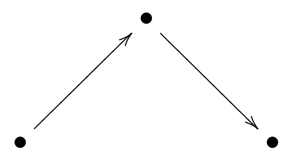
58.1.2 Example The sister relation $\mathcal{S}$ is not transitive, not even on the set of all women. Thus Agatha may be Bertha's sister, whence Bertha is Agatha's sister, but Agatha is not her own sister. This illustrates the general principle that when a definition uses different letters to denote things, they don't *have* to denote different things. In the definition of transitivity, a, b and c *may* be but don't *have* to be different.

58.1.3 Example Nearness relations are not transitive.

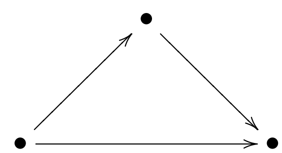
58.1.4 Example Let A be the set $\{\{1,2\},\{3\},2,6,\{\{1,3\},\{1,2\}\}\}$ The relation "$\in$" on A is not transitive, since $2 \in \{1,2\}$ and $\{1,2\} \in \{\{1,3\},\{1,2\}\}$, but $2 \notin \{\{1,3\},\{1,2\}\}$.

58.1.5 Warning Transitivity is defined by an implication and can be vacuously true. In fact, all the properties so far have been defined by implications except reflexivity. And indeed the empty relation is symmetric, antisymmetric and transitive!

58.1.6 Remark The digraph of a transitive relation must have the property that every "path of length two", such as



must be completed to a triangle, like this:



Paths are covered formally in Section 149.

58.1.7 Exercise Give an example of a nonempty, symmetric, transitive relation on the set $\{1,2\}$ that is not reflexive.

58.1.8 Exercise State and prove a theorem similar to Problem 56.1.5 for non-transitive relations.

58.1.9 Exercise Let the relation R be defined on the set $\{x \in \mathrm{R} \mid 0 \le x \le 1\}$ by

$$xRy \Leftrightarrow \exists t\,(x + t = y \text{ and } 0 \le t \le 1)$$

Is R transitive?

58.1.10 Exercise (hard) If possible, give examples of relations on the set $\{1,2,3\}$ which have every possible combination of the properties reflexive, symmetric, antisymmetric and transitive and their negations. (HINT: There are 14 possible combinations and two impossible ones.)

59. Irreflexive relations

59.1 Definition: irreflexive
A relation α is **irreflexive** if $a\,\alpha\,a$ is *false* for every $a \in A$.

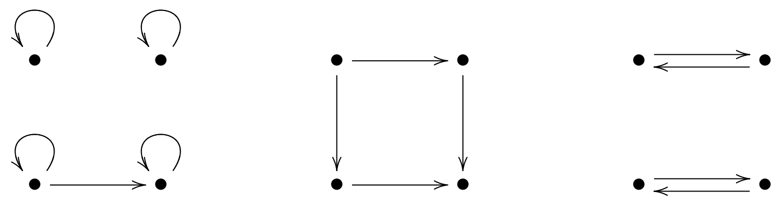
59.1.1 Example The "$<$" relation on R is irreflexive.

59.1.2 Warning Irreflexive is not the negation of reflexive: a relation might be neither reflexive nor irreflexive, such as for example the relation

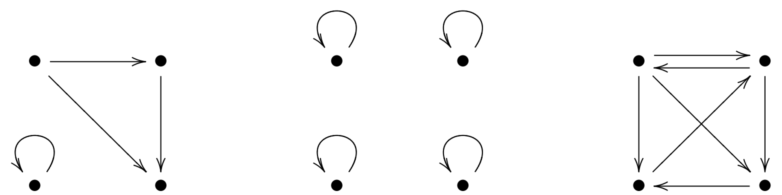
$$\alpha = \Big\{\langle 1,1\rangle, \langle 1,2\rangle, \langle 2,2\rangle\Big\}$$

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

59.1.3 Exercise List the properties (reflexive, symmetric, antisymmetric, transitive, and irreflexive) of the relations given by each picture.



(a) (b) (c)



(d) (e) (f)

(Answer on page 245.)

59.1.4 Exercise List the properties (reflexive, symmetric, antisymmetric, transitive, and irreflexive) of each relation.

a) “not equals” on R.
b) $x\,\alpha\,y \Leftrightarrow x^2 = y^2$ on R
c) $x\,\alpha\,y \Leftrightarrow x = -y$ on R
d) $x\,\alpha\,y \Leftrightarrow x \leq y^2$ on R
e) “divides” on N
f) “leaves the same remainder when divided by 3” on N
g) $\big\{\langle 1,1\rangle, \langle 2,3\rangle, \langle 3,2\rangle, \langle 3,4\rangle\big\}$ on $\{1,2,3,4\}$

(Answer on page 245.)

59.1.5 Exercise Let β be an irreflexive, antisymmetric relation on a set S. Show that at most one of the statements “$a\beta b$” and “$b\beta a$” holds for any pair of elements a,b of S.

60. Quotient and remainder

Let m and n be positive integers with $n \neq 0$. If you divide n into m you get a quotient and a remainder; for example, if you divide 4 into 14 you get a quotient 3 and a remainder 2. We will write the quotient when m is divided by n as $m \operatorname{div} n$ and the remainder as $m \bmod n$, so that $14 \operatorname{div} 4 = 3$ and $14 \bmod 4 = 2$. The basis for the formal definition given below is the property that $14 = 3 \times 4 + 2$.

The following formal definition allows m and n to be negative as well as positive. This has surprising consequences discussed in Section 61.3.

60.1 Definition: quotient and remainder

Let m and n be integers. Then $q = m \operatorname{div} n$ and $r = m \bmod n$ if and only if q and r are integers that satisfy both the following equations:
Q.1 $m = qn + r$, and
Q.2 $0 \le r < |n|$.
If $q = m \operatorname{div} n$, then q is the **quotient (of integers)** when m is divided by n. If $r = m \bmod n$, then r is the **remainder** when m is divided by n.

60.1.1 Remarks

a) It follows from the definition that the equation

$$m = (m \operatorname{div} n)n + (m \bmod n) \tag{60.1}$$

is always true for $n \ne 0$.

b) On the other hand, if $n = 0$, Q.2 cannot be true no matter what r is. In other words, "$m \operatorname{div} 0$" and "$m \bmod 0$" are not defined for any integer m.

60.1.2 Exercise Find the quotient (of integers) and remainder when m is divided by n:

a) $m = 2,\ n = 4$.
b) $m = 0,\ n = 4$.
c) $m = 24,\ n = 12$.
d) $m = 37,\ n = 12$.

(Answer on page 245.)

60.1.3 Warning For q to be $m \operatorname{div} n$ and r to be $m \bmod n$, *both* Q.1 and Q.2 must be true. For example, $14 = 2 \times 4 + 6$ (so Q.1 is satisfied with $q = 2$ and $r = 6$), but $14 \bmod 4 \ne 6$ because Q.2 is not satisfied.

60.1.4 Exercise Suppose that a and b leave the same remainder when divided by m. Show that $a - b$ is divisible by m. (Answer on page 245.)

60.1.5 Exercise Suppose that $a - b$ is divisible by m. Show that a and b leave the same remainder when divided by m.

60.1.6 Exercise Suppose that $a \operatorname{div} m = b \operatorname{div} m$. Show that $|a - b| < |m|$.

60.1.7 Exercise Is the converse of Exercise 60.1.6 true? That is, if $|a - b| < |m|$, must it be true that $a \bmod m = b \bmod m$?

The following theorem is what mathematicians call an "existence and uniqueness" theorem for quotient and remainder.

60.2 Theorem: Existence and Uniqueness Theorem for quotient and remainder

For given integers m and n with $n \neq 0$, there is exactly one pair of integers q and r satisfying the requirements of Definition 60.1.

60.2.1 Remark This theorem says that when $n \neq 0$ there *is* a quotient and a remainder, i.e., there is a pair of numbers q and r satisfying Q.1 and Q.2, and that there is is only *one* such pair.

60.2.2 Worked Exercise Suppose that $m = 3n+5$ and $n > 7$. What is $m \operatorname{div} n$?
Answer $m \operatorname{div} n = 3$. The fact that $m = 3n+5$ and $n > 7$ (hence $n > 5$) means that $q = 3$ and $r = 5$ satisfy the requirements of Definition 60.1.

60.2.3 Exercise Suppose a, b, m and n are integers with m and n nonnegative such that $m = (a+1)n + b + 2$ and $m \operatorname{div} n = a$. Show that b is negative. (Answer on page 245.)

60.2.4 Exercise Suppose $n > 0$, $0 \leq s < n$ and $n \mid s$. Show that $s = 0$. (Answer on page 246.)

There is a connection between these ideas and the idea of "divides" from Definition 4.1 (page 4):

60.3 Theorem
If $n \neq 0$ and $m \bmod n = 0$, then $n \mid m$.

Proof If $m \bmod n = 0$, then by Q.1, $m = (m \operatorname{div} n)n$, so by Definition 4.1 (using $m \operatorname{div} n$ for q), n divides m.

60.4 Mod and div in Mathematica

To compute $m \operatorname{div} n$ in Mathematica, you type `Quotient[m,n]`, and to compute $m \bmod n$, you type `Mod[m,n]`. You can if you wish place either of these function names between the inputs surrounded with tildes: `m ~Quotient~ n` is the same as `Quotient[m,n]`, and `m ~Mod~ n` is the same as `Mod[m,n]`.

60.5 Proof of uniqueness

We will prove that the quotient and remainder *exist* in Section 104.3.2, page 156. It is worthwhile to see the proof that the quotient and remainder are unique, since it shows how it is forced by Definition 60.1.

Suppose $m = qn + r = q'n + r'$ and both pairs $\langle q, r\rangle$ and $\langle q', r'\rangle$ satisfy Q.2. We must show that the two ordered pairs are the same, that is, that $q = q'$ and $r = r'$.

By Q.2 we have $0 \leq r < |n|$ and $0 \leq r' < |n|$. Since r and r' are between 0 and $|n|$ on the number line, the distance between them, which is $|r - r'|$, must also be less than n. A little algebra shows that

$$\left|r - r'\right| = \left|q' - q\right| |n|$$

It then follows from Definition 4.1, page 4, that $|r - r'|$ is divisible by $|n|$. But a nonnegative integer less than $|n|$ which is divisible by $|n|$ must be 0 (Exercise 60.2.4).

So $r = r'$. Since $qn + r = q'n + r'$, it must be that $q = q'$, too. So there can be only one pair of numbers q and r satisfying Q.1 and Q.2.

This proof uses the following method.

60.5.1 Method

To prove that an object that satisfies a certain condition is unique, assume there are two objects A and A' that satisfy the condition and show that $A = A'$.

60.5.2 Exercise Use Definition 60.1 and Theorem 60.2 to prove that when 37 is divided by 5, the quotient is 7 and the remainder is 2. (Answer on page 246.)

60.5.3 Exercise Use Definition 60.1 and Theorem 60.2 to prove that $115 \operatorname{div} 37 = 3$.

60.5.4 Exercise Suppose that $m = 36q + 40$. What is $m \bmod 36$? (Answer on page 246.)

60.5.5 Exercise Prove that if q, m and n are integers and $0 \leq m - qn < |n|$, then $q = m \operatorname{div} n$.

60.5.6 Exercise Show that if a and b are positive integers and $a \bmod 4 = b \bmod 4 = 3$, then $ab \bmod 4 = 1$.

60.5.7 Exercise Prove that for any integer c, $c^2 \bmod 3$ is either 0 or 1.

60.6 More about definitions

Observe that Definition 60.1 defines "$m \operatorname{div} n$" and "$m \bmod n$" without telling you how to compute them. Normally, you would calculate them using long division, but the uniqueness Theorem 60.2 tells you that if you can find them some other way you know you have the right ones. A mathematician would say that Theorem 60.2 ensures that the quotient (of integers) $m \operatorname{div} n$ and the remainder $m \bmod n$ are **well-defined**, or that Definition 60.1 and Theorem 60.2 work together to **characterize** the quotient and remainder.

It is typical of definitions in abstract mathematics that they characterize a concept without telling you how to compute it. The technique of separating the two ideas, "what is it?" and "how do you compute it?", is fundamental in mathematics.

61. Trunc and Floor

Many computer languages have one or both of two operators trunc and floor which are related to div and are confusingly similar. Both are applied to real numbers.

61.1 Definition: floor
Floor(r), or the **greatest integer** in r, is the largest integer n with the property $n \le r$.

61.1.1 Example $\text{floor}(3.1415) = 3$, $\text{floor}(7/8) = 0$, and $\text{floor}(-4.3) = -5$.

61.1.2 Usage Floor(r) is denoted by $\lfloor r \rfloor$ in modern texts, or by $[r]$ in older ones.

61.1.3 Exercise State a rule of inference for $\text{floor}(r)$. (Answer on page 246.)

61.2 Definition: trunc
Trunc(r) is obtained from r by expressing r in decimal notation and dropping all digits after the decimal point.

61.2.1 Fact The function trunc satisfies the equation

$$\text{trunc}(r) = \begin{cases} \text{floor}(r) & r \ge 0 \text{ or } r \text{ an integer} \\ \text{floor}(r) + 1 & r < 0 \text{ and not an integer} \end{cases}$$

61.2.2 Example $\text{trunc}(-4.3) = -4$, but $\text{floor}(-4.3) = -5$. On the other hand, $\text{trunc}(-4) = \text{floor}(-4) = -4$, and if r is any *positive* real number, $\text{trunc}(r) = \text{floor}(r)$.

61.2.3 Exercise Find $\text{trunc}(x)$ and $\text{floor}(x)$ for
a) $x = 7/5$.
b) $x = -7/5$.
c) $x = -7$.
d) $x = -6.7$.
(Answer on page 246.)

61.3 Quotients and remainders for negative integers

61.3.1 Example According to Definition 60.1, $-17 \operatorname{div} 5 = -4$ and $-17 \bmod 5 = 3$, because $-17 = (-4) \cdot 5 + 3$ and $0 \le 3 < 5$. In other words, the quotient is $\text{floor}(-17/5)$, but not $\text{trunc}(-17/5)$.

61.3.2 Usage A computer language which has an integer division (typically called div or "/") which gives this answer for the quotient is said to have **floored division**. Mathematica has floored division.

Other possibilities include allowing the remainder in Definition 60.1 to be negative when m is negative. This results in the quotient being trunc instead of floor, and, when implemented in a computer language, is called **centered division**. That is how many implementations of Pascal behave. When n is negative the situation also allows several possibilities (depending on whether m is negative or not).

In this book, integer division means floored division, so that it conforms to Definition 60.1.

62. Unique factorization for integers

62.1 The Fundamental Theorem of Arithmetic

It is a fact, called **The Fundamental Theorem of Arithmetic**, that a given positive integer $m > 1$ has a unique factorization into a product of positive primes. Thus $12 = 2 \times 2 \times 3$, $111 = 3 \times 37$, and so on. The factorization of a prime is that prime itself: thus the prime factorization of 5 is 5. The Fundamental Theorem of Arithmetic is proved in a series of problems in Chapter 103 as an illustration of the proof techniques discussed there.

The factorization into primes is unique in the sense that different prime factorizations differ only in the order they are written.

Here is the formal statement:

62.2 Theorem

Let m be an integer greater than 1. Then for some integer $n \geq 1$ there is a unique list of primes $p_1, p_2, \ldots, p_n$ and a unique list of integers $k_1, k_2, \ldots, k_n$ such that

FT.1 $p_i < p_{i+1}$ *for* $1 \leq i < n$.

FT.2 $m = p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}$.

62.2.1 Example

$$12 = 2 \times 2 \times 3 = 2 \times 3 \times 2 = 3 \times 2 \times 2$$

Theorem 62.2 specifically gives $12 = 2^2 \times 3^1$. Here $n = 2$, $p_1 = 2$, $p_2 = 3$, $k_1 = 2$ and $k_2 = 1$.

62.2.2 Exercise Give the prime factorizations of 30, 35, 36, 37 and 38. (Answer on page 246.)

62.3 Definition: exponent of a prime in an integer

The largest power of a prime p which divides a positive integer n is the **exponent** of p in n and is denoted $\mathrm{e}_p(n)$.

62.3.1 Example The exponent of 2 in 24 is 3; in other words, $\mathrm{e}_2(24) = 3$. You can check that $\mathrm{e}_{37}(111) = 1$ and $\mathrm{e}_{37}(110) = 0$.

62.3.2 Exercise Find the exponent of each of the primes 3, 7 and 37 in the integers 98, 99, 100, 111, 1332, and 1369. (Answer on page 246.)

The fact that the prime factorization is unique implies the following theorem:

> **62.4 Theorem**
> *Let m and n be positive integers. If $m \mid n$ and p is a prime, then $\mathrm{e}_p(m) \leq \mathrm{e}_p(n)$. Conversely, if for every prime p, $\mathrm{e}_p(m) \leq \mathrm{e}_p(n)$, then $m \mid n$.*

62.5 Prime factorization in Mathematica

`FactorInteger` is the Mathematica command for finding the factors of an integer. The answer is given as a list of pairs; the first coordinate in each pair is a prime and the second coordinate is the exponent of the prime in the number being factored. Thus if you type `FactorInteger[360]`, the answer will be `{{2,3},{3,2},{5,1}}`, meaning that $360 = 2^3 \cdot 3^2 \cdot 5$.

62.5.1 Exercise Factor all the two-digit positive integers that begin with 9. (Answer on page 246.)

62.5.2 Exercise Show that for every positive integer k, there is an integer n that has exactly k positive divisors.

62.5.3 Exercise (hard) Prove Theorem 62.4.

62.5.4 Exercise (discussion) Type `FactorInteger[6/7]` in Mathematica. Explain the answer you get. Should the name "`FactorInteger`" be changed to some other phrase?

63. The GCD

> **63.1 Definition: greatest common divisor**
> The **greatest common divisor** or **GCD** of two nonnegative integers m and n is 0 if $m = n = 0$; otherwise the GCD is the largest number which divides both of them.

> **63.2 Definition: least common multiple**
> The **least common multiple** (LCM) of two nonnegative integers m and n is 0 if either m or n is 0; otherwise it is the smallest *positive* integer which both m and n divide.

63.2.1 Example It follows from the definition that $\mathrm{GCD}(0,0) = 0$, $\mathrm{GCD}(0,4) = \mathrm{GCD}(4,0) = 4$, $\mathrm{GCD}(16,24) = 8$, and $\mathrm{GCD}(5,6) = 1$. Similarly, $\mathrm{LCM}(0,0) = 0$, $\mathrm{LCM}(1,1) = 1$, $\mathrm{LCM}(8,12) = 24$ and $\mathrm{LCM}(5,6) = 30$.

63.2.2 Exercise Find GCD(12,12), GCD(12,13), GCD(12,14), GCD(12,24), and also find the LCM's of the same pairs of numbers. (Answer on page 246.)

63.2.3 Exercise Compute GCD(48,72) and LCM(48,72).

63.2.4 Exercise If m and n are positive integers and $d = \mathrm{GCD}(m,n)$, must $\mathrm{GCD}(m/d,n) = 1$? Explain your answer. (Answer on page 246.)

63.2.5 Exercise Let $A = \{1,2,3,4\}$. Write out all the ordered pairs in the relation α on A where α is defined by: $a\alpha b \Leftrightarrow \mathrm{GCD}(a,b) = 1$. (Answer on page 246.)

63.2.6 Exercise Let α be the relation on Z defined by $a\alpha b \Leftrightarrow \mathrm{GCD}(a,b) = 1$. Determine which of these properties α satisfies: Reflexive, symmetric, transitive, antisymmetric.

63.2.7 Usage Some texts call the GCD the Greatest Common Factor (GCF).

63.2.8 Remark In general, $\mathrm{GCD}(0,m) = \mathrm{GCD}(m,0) = m$ for any nonnegative integer m. Note that Definition 63.1 defined GCD(0,0) as a special case. This is necessary because every integer divides 0, so there is no largest integer that divides 0. This awkward detail occurs because our definition is in a certain sense not the best definition. (See Corollary 64.2 below.)

63.3 Definition: relatively prime
If $\mathrm{GCD}(m,n) = 1$, then m and n are **relatively prime**.

63.3.1 Example 5 and 6 are relatively prime, but 74 and 111 are not relatively prime since their GCD is 37.

63.3.2 Exercise Show that for any integer n, n and $n+1$ are relatively prime. (Answer on page 246.)

63.3.3 Exercise
a) Show that if $n+1$ distinct integers are chosen from the set $\{1,2,\dots,2n\}$, then two of them are relatively prime.
b) Show that there is a way to choose n integers from $\{1,2,\dots,2n\}$ so that no two different ones are relatively prime.

63.3.4 Warning The property "relatively prime" concerns *two* integers. It makes no sense to speak of a single integer as being "relatively prime".

63.4 Definition: lowest terms
A rational number m/n is in lowest terms (see Definition 7.3, page 11) if m and n are relatively prime.

63.4.1 Exercise Prove that if m/n and r/s are rational numbers represented in lowest terms and $m/n = r/s$, then $|m| = |r|$ and $|n| = |s|$.

64. Properties of the GCD

If $m > 1$ and $n > 1$, and you know the prime factorizations of both of them, the GCD and LCM can be calculated using the following theorem, in which $\mathrm{e}_p(m)$ denotes the exponent of p in m (Definition 62.3), $\min(r,s)$ denotes the smaller of r and s and $\max(r,s)$ the larger.

64.1 Theorem
Let p be a prime and m and n positive integers. Then

$$\mathrm{e}_p(\mathrm{GCD}(m,n)) = \min(\mathrm{e}_p(m), \mathrm{e}_p(n))$$

and

$$\mathrm{e}_p(\mathrm{LCM}(m,n)) = \max(\mathrm{e}_p(m), \mathrm{e}_p(n))$$

64.1.1 Example $60 = 2^2 \times 3 \times 5$ and $72 = 2^3 \times 3^2$. Their GCD is $12 = 2^2 \times 3$, in which 2 occurs $\min(2,3)$ times, 3 occurs $\min(1,2)$ times, and 5 occurs $\min(1,0)$ times. Their LCM is $360 = 2^3 \times 3^2 \times 5$.

64.2 Corollary
Let m and n be nonnegative integers. $\mathrm{GCD}(m,n)$ *is the unique nonnegative integer with these properties:*
- *a)* $\mathrm{GCD}(m,n)$ *divides both m and n.*
- *b) Any integer e which divides both m and n must divide* $\mathrm{GCD}(m,n)$.

64.2.1 Remark The property of GCD given in this corollary is often taken as the definition of GCD. Note that no special consideration has to be given to the case $m = n = 0$.

64.2.2 Exercise Prove Corollary 64.2. (This corollary can be proved without using the Fundamental Theorem of Arithmetic. See Exercise 88.3.8, page 130.) (Answer on page 246.)

64.2.3 Exercise Use Theorems 62.4 and 64.1 to prove these facts about the GCD and the LCM:
- a) $\mathrm{GCD}(m,n)\,\mathrm{LCM}(m,n) = mn$ for any positive integers m and n.
- b) If m and n are relatively prime, then $\mathrm{LCM}(m,n) = mn$.

64.2.4 Exercise Prove that if $d = \mathrm{GCD}(m,n)$, then m/d and n/d are relatively prime. (Answer on page 246.)

64.2.5 Exercise Prove that every rational number has a representation in lowest terms.

64.2.6 Exercise Prove that GCD is commutative: for all integers m and n, $\mathrm{GCD}(m,n) = \mathrm{GCD}(n,m)$.

64.2.7 Exercise Prove that GCD is associative:

$$\mathrm{GCD}((\mathrm{GCD}(k,m),n) = \mathrm{GCD}(k,\mathrm{GCD}(m,n))$$

Hint: Use Theorem 64.1 and the fact that the smallest of the numbers x, y and z is

$$\min(x,\min(y,z)) = \min(\min(x,y),z) = \min(x,y,z)$$

64.2.8 Exercise (Mathematica)

a) Use Mathematica to determine which ordered pairs $\langle a,b\rangle$ of integers, with $a \in \{1,\ldots,10\}$, $b \in \{1,\ldots,10\}$, have the property that the sequence $a+b, 2a+b, \ldots, 10a+b$ contains a prime.

b) Let (C) be the statement:

> There is an integer $k > 0$ for which $ak+b$ is prime.

(The integer k does not have to be less than or equal to 10.) Based on the results, formulate a predicate $P(a,b)$ such that the condition (C) implies $P(a,b)$. The predicate P should not mention k.

c) Prove that (C) implies $P(a,b)$.

Note: Define a function by typing `t[a_,b_] := Table[a k + b,{k,1,10}]` (notice the spacing and the underlines). Then if you type, for example, `t[3,5]`, you will get `{8,11,14,17,20,23,26,29,32,35}`. If `L` is a list, `Select[L,PrimeQ]` produces a list of primes occurring in `L`.

64.3 Extensions of the definition of GCD

GCD is often defined for all integers, so that $\mathrm{GCD}(m,n)$ is $\mathrm{GCD}(|m|,|n|)$. For example, $\mathrm{GCD}(-6,4) = \mathrm{GCD}(6,-4) = \mathrm{GCD}(-6,-4) = 2$. With this extended definition, GCD is an associative and commutative binary operation on Z (Section 143.2.1). Associativity means it is unambiguous to talk about the GCD of more than two integers. In fact, we can define that directly:

64.4 Definition: generalized GCD

Let n_1, $n_2, \ldots, n_k$ be integers. Then $\mathrm{GCD}(n_1,\ldots,n_k)$ is the largest integer that divides all the numbers $|n_1|,|n_2|,\ldots,|n_k|$.

64.4.1 Example $\mathrm{GCD}(4,6,-8,12) = 2$.

64.4.2 Remarks

a) Similar remarks can be made about the LCM.

b) These functions are implemented in Mathematica using the same names. For example, `GCD[4,6,-8,12]` returns `2`.

65. Euclid's Algorithm

Theorem 64.1 is fine for finding the GCD or LCM of two numbers when you know their prime factorization. Unfortunately, the known algorithms for finding the prime factorization are slow for large numbers. There is another, more efficient method for finding the GCD of two numbers which does not require knowledge of the prime factorization. It is based on this theorem:

> **65.1 Theorem: Euclid's Algorithm**
> *For all nonnegative integers m and n:*
> EA.1 $\mathrm{GCD}(m,0) = m$ *and* $\mathrm{GCD}(0,n) = n$.
> EA.2 *Let r be the remainder when m is divided by n. Then*
> $$\mathrm{GCD}(m,n) = \mathrm{GCD}(n,r)$$

Proof Both parts of Theorem 65.1 follow from Definition 6.1, page 10. EA.1 follows because *every* integer divides 0 (Theorem 5.1(2)), so that if $m \neq 0$, then largest integer dividing m and 0 is the same as the largest integer dividing m, which of course is m.

To prove EA.2, suppose d is an integer that divides both m and n. Since $r = m - qn$, where $q = m \operatorname{div} n$, it follows from Theorem 5.4, page 8, that d divides r. Thus d divides both n and r.

Now suppose e divides both n and r. Since $m = qn + r$, it follows that e divides m. Thus e divides both m and n.

In the preceding two paragraphs, I have shown that m and n have the *same common divisors* as n and r. It follows that m and n have the same *greatest* common divisor as n and r, in other words $\mathrm{GCD}(m,n) = \mathrm{GCD}(n,r)$.

65.1.1 How to compute the GCD Theorem 65.1 provides a computational process for determining the GCD. This process is the **Euclidean algorithm**. The process always terminates because every time EA.2 is used, the integers involved are replaced by smaller ones (because of Definition 60.1(Q.2), page 83) until one of them becomes 0 and EA.1 applies.

65.1.2 Example

$$\mathrm{GCD}(164,48) = \mathrm{GCD}(48,20) = \mathrm{GCD}(20,8) = \mathrm{GCD}(8,4) = \mathrm{GCD}(4,0) = 4$$

65.2 Pascal program for Euclid's algorithm

A fragment of a Pascal program implementing the Euclidean algorithm is given formally in Program 65.1.

```
{M>0, N>0, K=M, L=N}
while N <> 0 do
  begin
    rem := M mod N;
    M := N;
    N := rem;
  end;
{M=GCD(K,L)}
```

Program 65.1: Pascal Program for GCD

66. Bases for representing integers

66.1 Characters and strings

The number of states in the United States of America is an integer. In the usual notation, that integer is written '50'.

In this section, we discuss other, related ways of expressing integers which are useful in applications to computer science. In doing this it is important to distinguish between numerals like '5' and '0' and the integers they represent. In particular, the sequence of numerals '50' represents the integer which is the number of states in the USA, but it is *not the same thing as that integer*.

Numerals, as well as letters of the alphabet and punctuation marks, are **characters**. Characters are a type of data, distinct from integers or other numerical types. In order to distinguish between a character like '5' and the number 5 we put characters which we are discussing in single quotes. Pascal has a data type CHAR of which numerals and letters of the alphabet are subtypes. Single quotes are used in Pascal as we use them.

66.2 Specification: string

A sequence of characters, such as '50' or 'cat', is also a type of data called a **string**.

66.2.1 Remarks

a) Strings will be discussed from a theoretical point of view in Chapter 109.
b) In this book we put strings in single quotes when we discuss them. Thus 'cat' is a string of characters whereas "cat" is an English word (and a cat is an animal!).

66.3 Bases

The decimal notation we usually use expresses an integer as a string formed of the numerals '0', '1', ... ,'9'. These numerals are the **decimal digits**. The word "digit" is often used for the integers they represent, as well. The notation is based on the fact that any positive integer can be expressed as a sum of numbers, each of which is the value of a digit times a power of ten. Thus

$$258 = 2 \times 10^2 + 5 \times 10^1 + 8 \times 10^0 .$$

The expression '258' gives you the digits multiplying each power of 10 in decreasing order, the rightmost numeral giving the digit which multiplies $1 = 10^0$.

Any integer greater than 1 can be used instead of 10 in an analogous way to express integers. The integer which is used is the **base** or **radix** of the notation. In **octal notation**, for example, the base is 8, and the octal digits are '0', '1', ... , '7'. For example,

$$258 = 4 \times 8^2 + 0 \times 8^1 + 2 \times 8^0$$

so the number represented by '258' in decimal notation is represented in octal notation by '402'.

Here is the general definition for the representation of an integer in base b.

66.4 Definition: base

If n and b are nonnegative integers and $b > 1$, then the expression

$$\text{`}d_m d_{m-1} d_{m-2} \cdots d_1 d_0\text{'} \tag{66.1}$$

represents n in base b notation if for each i, d_i is a symbol (base-b digit) representing the integer n_i,

$$n = n_m b^m + n_{m-1} b^{m-1} + \cdots + n_0 b^0 \tag{66.2}$$

and for all i,

$$0 \le n_i \le b - 1 \tag{66.3}$$

66.4.1 Remarks

a) We will say more about the symbols d_i below. For bases $b \le 10$ these symbols are normally the usual decimal digits,

$$d_0 = \text{`0'},\ d_1 = \text{`1'}, \ldots, d_9 = \text{`9'}$$

as illustrated in the preceding discussion.

b) Efficient ways of determining the base-b representation of some integer are discussed in Chapter 68. Note that you can do the exercises in this section without knowing how to find the base-b representation of an integer — all you need to know is its definition.

66.4.2 Notation When necessary, we will use the base as a subscript to make it clear which base is being used. Thus $258_{10} = 402_8$, meaning that the number represented by '258' in base 10 is represented by '402' in base 8.

66.5 Definition: significance

The digit d_i is **more significant** than d_j if $i > j$. Thus, if a number n is represented by '$d_m d_{m-1} \ldots d_1 d_0$', then d_0 is the **least significant digit** and, if d_m does not denote 0, it is the **most significant digit**.

66.5.1 Example The least significant digit in 258_{10} is 8 and the most significant is 2.

66.5.2 Remark For a given b and n, the following theorem says that the representation given by definition 66.4 is unique, except for the choice of the symbols representing the n_i. We will take this theorem as known.

66.6 Theorem

If n and b are positive integers with $b > 1$, then there is only one sequence $n_0, n_1, \ldots, n_m$ of integers for which $n_m \neq 0$ and formulas (66.2) and (66.3) are true.

66.6.1 Worked Exercise Prove that the base 4 representation of 365 is 11231.
Answer $365 = 1\cdot 4^4 + 1\cdot 4^3 + 2\cdot 4^2 + 3\cdot 4^1 + 1\cdot 4^0$, and $1, 2, 3$ are all less than 4, so the result follows from Theorem 66.6.

Note that in this answer we merely showed that 11231 fit the definition. *That is all that is necessary.* Of course, if you are not given the digits as you were in this problem, you need some way of calculating them. We will describe ways of doing that in Chapter 68.

66.6.2 Exercise Prove that the base 8 representation of 365 is 555.

66.6.3 Exercise Prove that if an integer n is represented by '$d_m d_{m-1} \cdots d_1$' in base b, then '$d_m d_{m-1} \cdots d_1 0$' represent bn in base b notation. (Answer on page 246.)

66.6.4 Exercise Suppose b is an integer greater than 1 and suppose n is an integer such that the base b representation of n is 352. Prove using *only the definition of representation to base* b that the base b representation of $b^2 n + 1$ is 35201.

66.7 Specific bases

66.7.1 Base 2 The digits for base 2 are '0' and '1' and are called **bits**. Base 2 notation is called **binary notation**.

66.7.2 Bases larger than 10 For bases $b \leq 10$, the usual numerals are used, as mentioned before. A problem arises for bases bigger than 10: you need single symbols for the integers 10, 11, ... Standard practice is to use the letters of the alphabet (lowercase here, uppercase in many texts): 'a' denotes 10, 'b' denotes 11, and so on. This allows bases up through 36.

66.7.3 Base 16 Base 16 (giving **hexadecimal notation**) is very commonly used in computing. For example, 95_{10} is $5\mathrm{f}_{16}$, and 266_{10} is hexadecimal $10\mathrm{a}_{16}$ (read this "one zero a", not "ten a"!) In texts in which decimal and nondecimal bases are mixed, the numbers expressed nondecimally are often preceded or followed by some symbol; for example, many authors write $10a or H10a to indicate 266_{10} expressed hexadecimally.

66.8 About representations

(This continues the discussion of representations in Section 10.2 and Remark 17.1.3.) It is important to distinguish between the (abstract) integer and any representation of it. The number of states in the U.S.A is represented as '50' in decimal notation, as '110010' in binary, and as a pattern of electrical charges in in a computer. These are all representations or **realizations** of the abstract integer. (The word "realization" here has a technical meaning, roughly made real or made concrete.) All the representations are matters of convention, in other words, are based on agreement rather than intrinsic properties. Moreover, no one representation is more fundamental or correct than another, although one may be more familiar or more convenient than another.

There is also a distinction to be made between properties of an integer and properties of the representation of an integer. For example, being prime is a property of the integer; whether it is written in decimal or binary is irrelevant. Whether its least significant digit is 0, on the other hand, is a property of the representation: the number of states in the USA written in base 10 ends in '0', but in base 3 it ends in '2'.

66.8.1 Exercise Suppose b is an integer greater than 1, a is an integer dividing b, and n is an integer. When n is written in base b, how do you tell from the digits of n whether n is divisible by a? Prove that your answer is correct.

66.8.2 Exercise Would Theorem 66.6 still be true if the requirement that $0 \leq n_i \leq b-1$ for all i were replaced by the requirement that the n_i be nonnegative?

66.8.3 Exercise (Mathematica) A positive integer is a **repunit** if all its decimal digits are 1.

a) Use Mathematica to determine which of the repunits up to a billion are divisible by 3.
b) Based on the results of part (a), formulate a conjecture as to which repunits are divisible by 3. The conjecture should apply to all repunits, not just those less than a billion.
c) Prove the conjecture.

66.8.4 Exercise (discussion) Some computer languages (FORTH is an example) have a built-in integer variable `BASE`. Whatever integer you set `BASE` to will be used as the base for all numbers output. How would you discover the current value of `BASE` in such a language? (Assume you print the value of a variable `X` by writing `PRINT(X)`).

67. Algorithms and bases

Among the first algorithms of any complexity that most people learn as children are the algorithms for adding, subtracting, multiplying and dividing integers written in decimal notation. In medieval times, the word "algorithm" referred specifically to these processes.

67.1 Addition

The usual algorithm for addition you learned in grade school works for numbers in other bases than 10 as well. The only difference is that you have to use a different addition table for the digits.

67.1.1 Example To add 95a and b87 in hexadecimal you write them one above the other:

$$\begin{array}{r} \texttt{95a} \\ \underline{+\texttt{b87}} \\ \texttt{14e1} \end{array}$$

Here is a detailed description of how this is done, all in base 16.

- Calculate $a + 7 = 11_{16}$, with a carry of 1 since $11_{16} \geq 10_{16}$. (Pronounce 10_{16} as "one-zero", not "ten", since it denotes sixteen, and similarly for 11_{16} which denotes seventeen. By the way, the easiest way to figure out what $a + 7$ is is to count on your fingers!)
- Then add 5 and 8 and get d (not 13!) and the carry makes e. $e < 10_{16}$ so there is no carry.
- Finally, $9 + b = 14_{16}$.

So the answer is $14e1_{16}$. *The whole process is carried out in hexadecimal* without any conversion to decimal notation.

67.1.2 Addition in binary The addition table for binary notation is especially simple: $0 + 0 = 0$ without carry, $1 + 0 = 0 + 1 = 1$ without carry, and $1 + 1 = 0$ with carry.

67.2 Multiplication

The multiplication algorithm similarly carries over to other bases. Normally in a multiplication like

$$\begin{array}{rl} \texttt{346} & \text{(multiplicand)} \\ \underline{\times\texttt{527}} & \text{(multiplier)} \\ \texttt{2422} & \\ \texttt{6920} & \text{(partial products)} \\ \underline{\texttt{173000}} & \\ \texttt{182342} & \text{(product)} \end{array}$$

you produce successive partial products, and then you add them. The partial product resulting from multiplying by the ith digit of the multiplier is

$$\text{digit} \times \text{multiplicand} \times 10^i$$

base 94
digit 93
hexadecimal notation 95

(Most people are taught in grade school to suppress the zeroes to the right of the multiplying digit.)

67.2.1 Binary multiplication Multiplication in binary has a drastic simplification. In binary notation, the only digits are 0, which causes a missing line, and 1, which involves only shifting the top number. So multiplying one number by another in binary consists merely of shifting the first number once for each 1 in the second number and adding.

67.2.2 Example With trailing zeroes suppressed:

```
    1101
   ×1101
   -----
    1101
  1101
 1101
 -------
10101001
```

67.2.3 Exercise Perform these additions and multiplications in binary.

```
a)  110001   b)  1011101   c)  10011   d)  11100
   +101111      +1110101      ×10101      ×11001
```

(Answer on page 246.)

67.2.4 Exercise Perform these additions in hexadecimal:

```
a)  9ae   b)  389   c)  feed
   +b77      +777      +dad
```

(Answer on page 246.)

67.2.5 Exercise Show that in adding two numbers in base b, the carry is never more than 1, and in multiplying in base b, the carry is never more than $b-2$.

67.2.6 Exercise (discussion) Because subtracting two numbers using pencil and paper is essentially a solitary endeavor, most people are not aware that there are two different algorithms taught in different public school systems. Most American states' school systems teach one algorithm (Georgia used to be an exception),and many European countries teach another one. Ask friends from different parts of the world to subtract 365 from 723 while you watch, explaining each step, and see if you detect anyone doing it differently from the way you do it.

68. Computing integers to different bases

68.1 Representing an integer

68.1.1 Remark Given a nonnegative integer n and a base b, the most significant nonzero digit of n when it is represented in base b is the quotient when n is divided by the largest power of b less than n. For example, in base 10, the most significant digit of 568 is 5, and indeed $5 = 568 \operatorname{div} 100$ (100 is the largest power of 10 less than 568). Furthermore, 68 is the remainder when 568 is divided by 500.

This observation provides a way of computing the base-b representation of an integer.

68.1.2 Method
Suppose the representation for n to base b is '$d_m d_{m-1} \cdots d_0$', where d_i represents the integer n_i in base b. Then

$$d_m = n \operatorname{div} b^m$$

and

$$d_{m-1} = (n - d_m b^m) \operatorname{div} b^{m-1}$$

In general, for all $i = 0, 1, \ldots, m-1$,

$$d_i = (n_{i+1} - d_{i+1} b^{i+1}) \operatorname{div} b^i \tag{68.1}$$

where

$$n_m = n \tag{68.2}$$

and for $i = 0, 1, \ldots, m-1$,

$$n_i = n_{i+1} - d_{i+1} b^{i+1} \tag{68.3}$$

68.1.3 Example The '6' in 568 is

$$(568 - 5 \cdot 100) \operatorname{div} 10$$

(here $m = 2$: note that the 5 in 568 is d_2 since we start counting on the right at 0).

68.1.4 Remark Observe that (68.1) can be written

$$d_i = (n \bmod b^{i+1}) \operatorname{div} b^i \tag{68.4}$$

which is correct for all $i = 0, 1, \ldots, m$. The way (68.1) is written shows that the computation of $n \bmod b^{i+1}$ uses the previously-calculated digit d_{i+1}.

68.1.5 Example We illustrate this process by determining the representation of 775 to base 8. Note that $512 = 8^3$:

a) $775 \operatorname{div} 512 = 1$.
b) $775 - 1 \times 512 = 263$.
c) $263 \operatorname{div} 64 = 4$.
d) $263 - 4 \times 64 = 7$.

e) $7 \operatorname{div} 8 = 0$.
f) $7 - 0 \times 8 = 7$.
g) $7 \operatorname{div} 1 = 7$.

And 775 in octal is indeed 1407.

68.2 The algorithm in Pascal

The algorithm just described is expressed in Pascal in Program 68.1. This algorithm is perhaps the most efficient for pencil-and-paper computation. As given, it only works as written for bases up to and including 10; to have it print out 'a' for 11, 'b' for 12 and so on would require modifying the "`write(place)`" statement.

```
var N, base, count, power, limit, place: integer;
(* Requires B > 0 and base > 1 *)
   begin
      power := 1; limit := N div base;
      (*calculate the highest power of the base less than N*)
        while power <= limit do
          begin
            power := power*base
          end;
        while power > 1 do
          begin
            place := N div power; write(place);
            n := n-place*power; power := power div base
          end
   end
```

Program 68.1: Program for Base Conversion

68.3 Another base conversion algorithm

Another algorithm, which computes the digits backwards, stores them in an array, and then prints them out in the correct order, is given in Program 68.2. It is more efficient because it is unnecessary to calculate the highest power of the base less than N first. This program starts with the observation that the *least* significant digit in a number n expressed in base b notation is $n \bmod b$. The other digits in the representation of n represent $(n - (n \bmod b))/b$. For example, $568 \bmod 10 = 8$, and the number represented by the other digits, 56, is $(568 - 8)/10$.

In the program in Program 68.2, `count` and `u` are auxiliary variables of type integer. The size `longest` of the array `D` has to be known in advance, so there is a bound on the size of integer this program can compute, in contrast to the previous algorithm. It is instructive to carry out the operations of the program in Program 68.2 by hand to see how it works.

68.4 Comments on the notation for integers

Suppose n is written '$d_m d_{m-1} \ldots d_0$' in base b. Then the exact significance of d_m, namely the power b^m that its value n_m is multiplied by in Equation (66.2) of Definition 66.4 (page 94), depends on the length of the string of digits representing

```
var count, u, N, base: integer;
var D:array [0..longest] of integer;
  begin
    count := 0; u := N;
    while u<>0 do
      begin
        D[count] := u mod base;
        u := (u-D[count]) div base;
        count := count+1
      end;
    while count<>0 do
      begin
        count := count-1;
        write D[count]
      end
  end;
```

Program 68.2: Faster Program for Base Conversion

n (the length is $m+1$ because the count starts at 0). If you read the digits from left to right, as is usual in English, you have to read to the end before you know what m is. On the other hand, the significance of the *right* digit d_0 is known without knowing the length m. In particular, the program in Program 68.1 has to read to the end of the representation to know the power b^m to start with.

The fact that the significance of a digit is determined by its distance from the right is the reason a column of integers you want to add is always lined up with the right side straight. In contrast to this, the sentences on a typewritten page are lined up with the *left* margin straight.

There is a good reason for this state of affairs: this notation was invented by Arab mathematicians, and Arabic is written from right to left.

68.4.1 Exercise Represent the numbers 100, 111, 127 and 128 in binary, octal, hexadecimal and base 36. (Answer on page 246.)

68.4.2 Exercise Represent the numbers 3501, 29398 and 602346 in hexadecimal and base 36.

68.5 Exercise set

Exercises 68.5.1 through 68.5.4 are designed to give a proof of Formula (68.4), page 99, so they should be carried out without using facts about how numbers are represented in base b. In these exercises, all the variables are of type integer.

68.5.1 Exercise Let $b > 1$. Prove that if for all $i \geq 0$, $0 \leq d_i < b$, then

$$d_m b^m + d_{m-1} b^{m-1} + \cdots + d_1 b + d_0 < b^{m+1}$$

68.5.2 Exercise Let $b > 1$ and $n > 0$. Let $n = d_m b^m + \cdots + d_1 b + d_0$ with $0 \leq d_i < b$ for $i = 0, 1, \ldots, m$. Prove that for any $i \geq 0$,

$$n = b^i [d_m b^{m-i} + d_{m-1} b^{m-i-1} + \cdots + d_i] + d_{i-1} b^{i-1} + \cdots + d_1 b + d_0$$

and $0 \le d_{i-1}b^{i-1} + \cdots + d_1 b + d_0 < b^i$. (Hint: Use Exercise 68.5.1.)

68.5.3 Exercise Let $b > 1$ and $n > 0$ and let $n = d_m b^m + \cdots + d_1 b + d_0$ with $0 \le d_i < b$ for $i = 0, 1, \ldots, m$. Prove that for any $i \ge 0$,

$$d_m b^{m-i} + d_{m-1} b^{m-i-1} + \cdots + d_i = n \operatorname{div} b^i$$

and

$$d_i b^i + \cdots + d_1 b + d_0 = n \bmod b^{i+1}$$

68.5.4 Exercise Prove Equation (68.4), page 99.

69. The DeMorgan Laws

Consider what happens when you negate a conjunction. The statement $\neg(P \wedge Q)$ means that it is false that P and Q are both true; thus one of them must be false. In other words, either $\neg P$ is true or $\neg Q$ is true. This is one of the two DeMorgan Laws:

69.1 Definition: DeMorgan Laws
The **DeMorgan Laws** are:
DM.1 $\neg(P \wedge Q) \Leftrightarrow \neg P \vee \neg Q$
DM.2 $\neg(P \vee Q) \Leftrightarrow \neg P \wedge \neg Q$.
These laws are true *no matter what propositions P and Q are.*

69.1.1 Remark The DeMorgan Laws give rules of inference

$$\neg(P \wedge Q) \vdash \neg P \vee \neg Q \text{ and } \neg P \vee \neg Q \vdash \neg(P \wedge Q) \tag{69.1}$$

and

$$\neg(P \vee Q) \vdash \neg P \wedge \neg Q \text{ and } \neg P \wedge \neg Q \vdash \neg(P \vee Q) \tag{69.2}$$

69.1.2 Example The negation of $(x + y = 10) \wedge (x < 7)$ is $(x + y \ne 10) \vee \neg(x < 7)$. Of course, $\neg(x < 7)$ is the same as $x \ge 7$.

69.2 Using the DeMorgan Laws in proofs

The unit interval $\mathrm{I} = \{x \mid 0 \le x \le 1\}$, which means that $x \in \mathrm{I}$ if and only if $0 \le x$ *and* $x \le 1$. Therefore to prove that some number a is *not* in I, you must prove the *negation* of the defining condition, namely that it is not true that $0 \le x$ and $x \le 1$. By the DeMorgan Laws, this means you must prove

$$\neg(0 \le x) \vee \neg(x \le 1)$$

which is the same as proving that $(0 > x) \vee (x > 1)$.

69.2.1 Warning When proving that a conjunction is false, it is easy to forget the DeMorgan Laws and try to prove that both negatives are true. In the preceding example, this would require showing that both $0 > x$ and $x > 1$, which is obviously impossible.

In contrast, if you must prove that a disjunction $P \vee Q$ is false, you must show that *both* P and Q are false. An error here is even more insidious, because if you are tempted to prove that only one of P and Q is false, you often can do that without noticing that you have not done everything required.

69.2.2 Example Consider the statement, "A positive integer is either even or it is prime". This statement is false. To show it is false, you must find a positive integer such as 9 which is *both* odd and nonprime.

69.2.3 Method

To prove that $P \vee Q$ is false, prove that $\neg P \wedge \neg Q$ is true. To prove that $P \wedge Q$ is false, prove that $\neg P \vee \neg Q$ is true.

69.2.4 Example Given two sets A and B, how does one show that $A \neq B$? By Method 21.2.1 on page 32, $A = B$ means that every element of A is an element of B *and* every element of B is an element of A. By DeMorgan, to prove $A \neq B$ you must show that one of those two statements is false: you must show *either* that there is an element of A that is not an element of B or that there is an element of B that is not an element of A. You needn't show both, and indeed you often *can't* show both. For example, $\{1,2\} \neq \{1,2,3\}$, yet every element of the first one is an element of the second one.

69.2.5 Worked Exercise Let A and B be sets. How do you prove $x \notin A \cup B$? How do you prove $x \notin A \cap B$?

Answer To prove that $x \notin A \cup B$, you must prove both that $x \notin A$ and that $x \notin B$. This follows from the DeMorgan Law and the definition of union. To prove $x \notin A \cap B$, you need only show $x \notin A$ *or* $x \notin B$.

69.3 Exercise set

Reword the predicates in Exercises 69.3.1 through 69.3.3 so that they do not begin with "$\neg$". x is real.

69.3.1 $\neg(x < 10) \wedge (x > 12)$. (Answer on page 246.)

69.3.2 $\neg(x < 10) \wedge (x < 12)$. (Answer on page 247.)

69.3.3 $\neg(\neg(x > 5) \wedge \neg(x < 6))$.

70. Propositional forms

The letters P and Q in the DeMorgan Laws are called **propositional variables**. They are like variables in algebra except that you substitute propositions or predicates for them instead of numbers. Don't confuse propositional variables with the variables which occur in predicates such as "$x < y$". The variables in predicates are of the type of whatever you are talking about, presumably numbers in the case of "$x < y$". Propositional variables are of type "proposition": they vary over propositions in the same way that x and y in the statement "$x < y$" vary over numbers.

70.1.1 Worked Exercise Write the result of substituting $x = 7$ for P and $x \neq 5$ for Q in the expression $\neg P \vee (P \wedge Q)$.
Answer $x \neq 7 \vee (x = 7 \wedge x \neq 5)$.

70.2 Variables in Pascal

Pascal does not have variables or expressions of type proposition. It does have Boolean variables, which have `TRUE` and `FALSE` as their only possible values.

An expression such as '$X < Y$' has numerical variables, and a Boolean value — `TRUE` or `FALSE`, so it might correctly be described as a proposition (assuming the program has already given values to X and Y). However, if B is a Boolean variable, an assignment statement of the form $B := X < Y$ sets B equal to the *truth value* of the statement '$X < Y$' at that point on the program; B is not set equal to the *proposition* '$X < Y$'. If X and Y are later changed, changing the truth value of '$X < Y$', the value of B will not automatically be changed.

70.2.1 Example The following program prints TRUE. Here `B` is type `BOOLEAN` and `X` is of type `INTEGER`:

```
X := 3;
B := X < 5;
X:= 7;
PRINT(B);
```

70.3 Propositional forms

Meaningful expressions made up of propositional variables and logical connectives are called **propositional forms** or **propositional expressions**. The expressions in DM.1 and DM.2 are examples of propositional forms. Two simpler ones are

$$P \vee \neg P \tag{70.1}$$

and

$$\neg P \vee Q \tag{70.2}$$

70.3.1 Substituting in propositional forms If you substitute propositions for each of the variables in a propositional form you get a proposition.

You may also substitute *predicates* for the propositional variables in a propositional form and the result will be a predicate.

70.3.2 Example If you substitute the proposition "$3 < 5$" in formula (70.1) you get (after a little rewording) "$3 < 5$ or $3 \geq 5$" which is a proposition (a true one, in fact).

If you substitute $x < 5$ for P in formula (70.1) you get "$x < 5$ or $x \geq 5$", which is true for any real number x. This is not surprising because formula (70.1) is a tautology (discussed later).

If you substitute $x < 5$ for P and $x \neq 6$ for Q in $\neg P \vee Q$ you get "$x \geq 5$ or $x \neq 6$", which is true for some x and false for others.

70.3.3 Remarks

a) This would be a good time to reread Section 12.1.4. Propositional forms are a third type of expression beside algebraic expressions and predicates. In an algebraic expression the variables are some type of number and the output when you substitute the correct type of data for the variables is a number. In a predicate the output is a proposition: a statement that is either true or false. And now in propositional forms the variables are propositions and when you substitute a proposition for each propositional variable the output is a proposition.

b) We have not given a formal definition of "meaningful expression". This is done in texts on formal logic using definitions which essentially constitute a context-free grammar.

71. Tautologies

71.1 Discussion

Each DeMorgan Law is the assertion that a certain propositional form is true *no matter what propositions are plugged in for the variables.* For example, the first DeMorgan Law is

$$\neg(P \wedge Q) \Leftrightarrow \neg P \vee \neg Q$$

No matter which predicates we let P and Q be in this statement, the result is a true statement.

71.1.1 Example let P be the statement $x < 5$ and Q be $x = 42$. Then the first DeMorgan Law implies that

$$\neg\big((x < 5) \wedge (x = 42)\big) \Leftrightarrow \big((x \geq 5) \vee (x \neq 42)\big)$$

is a true statement.

71.2 Definition: tautology

A propositional form which is true for all possible substitutions of propositional variables is called a **tautology**.

71.2.1 Fact The truth table for a tautology S has all T's in the column under S.

71.2.2 Example Both DeMorgan laws are tautologies, and so is the formula (70.1), which is called **The law of the excluded middle**. Both lines of its truth table have T.

P	$\neg P$	$P \vee \neg P$
T	F	T
F	T	T

71.2.3 Warning Don't confuse tautologies with predicates all of whose instances are true. A tautology is an expression containing propositional variables which is true no matter which propositions are substituted for the variables. Expression (70.2) is not a tautology, but some instances of it, for example "not $x > 5$ or $x > 3$" are predicates which are true for all values (of the correct type) of the variables.

71.2.4 Example Formula (70.2) (page 104) is not a tautology. For example, let P be "$4 > 3$" and Q be "$4 > 5$", where x ranges over real numbers; then Formula (70.2) becomes the proposition"(not $4 > 3$) or $4 > 5$", i.e., "$4 \leq 3$ or $4 > 5$", which is false.

71.2.5 Exercise Show that $P \vee Q \Leftrightarrow \neg(\neg P \wedge \neg Q)$ is a tautology. (Answer on page 247.)

71.2.6 Exercise Show that the following are tautologies.

a) $P \wedge Q \Leftrightarrow \neg(\neg P \vee \neg Q)$
b) $(P \wedge \neg P) \Rightarrow Q$
c) $P \Rightarrow (Q \vee \neg Q)$
d) $P \vee (P \Rightarrow Q)$
e) $\Big((P \wedge Q) \Rightarrow R\Big) \Leftrightarrow \Big(P \Rightarrow (Q \Rightarrow R)\Big)$
f) $P \wedge (Q \vee R) \Rightarrow P \vee (Q \wedge R)$

71.2.7 Remark Many laws of logic are equivalences like the DeMorgan laws. By Theorem 29.2, an equivalence between two expressions is a tautology if the truth tables for the two expressions are identical. Thus the truth tables for $\neg(P \wedge Q)$ and $\neg P \vee \neg Q$ are identical:

P	Q	$P \wedge Q$	$\neg(P \wedge Q)$	$\neg P$	$\neg Q$	$\neg P \vee \neg Q$
T	T	T	F	F	F	F
T	F	F	T	F	T	T
F	T	F	T	T	F	T
F	F	F	T	T	T	T

71.2.8 Example You can check using this method that $\neg P \vee Q$ (i.e., Formula (70.2)) is equivalent to $P \Rightarrow Q$.

71.2.9 Exercise Prove by using Theorem 29.2 that the propositional forms $P \Rightarrow Q$, $\neg P \vee Q$ and $\neg(P \wedge \neg Q)$ are all equivalent. (Answer on page 247.)

71.2.10 Exercise Prove that $(P \Rightarrow Q) \Rightarrow Q$ is equivalent to $P \vee Q$.

72. Contradictions

> **72.1 Definition: contradiction**
> A propositional form is a **contradiction** if it is false for all possible substitutions of propositional variables.

72.1.1 Fact The truth table for a contradiction has all F's.

72.1.2 Example The most elementary example of a contradiction is "$P \wedge \neg P$".

72.1.3 Exercise Show that the following are contradictions.
a) $\neg(P \vee \neg P)$.
b) $\neg(P \vee (P \Rightarrow Q))$.
c) $Q \wedge \neg(P \Rightarrow Q)$.

72.1.4 Exercise If possible, give an example of a propositional form involving "$\Rightarrow$" that is neither a tautology nor a contradiction.

73. Lists of tautologies

Tables 72.1 and 72.2 give lists of tautologies. Table 72.1 is a list of tautologies involving "and", "or" and "not". Because union, intersection and complementation for sets are defined in terms of "and", "or" and "not", the tautologies correspond to universally true statements about sets, which are given alongside the tautologies.

Table 72.2 is a list of tautologies involving implication. Because of the modus ponens rule, the major role implication plays in logic is to provide successive steps in proofs. These laws can be proved using truth tables or be deriving them from the laws in Table 72.1 and the first law in Table 72.2, which allows you to define '$\Rightarrow$' in terms of '$\neg$' and '$\vee$'. It is an excellent exercise to try to understand why the tautologies in both lists are true, either directly or by using truth tables.

73.1 The propositional calculus

The laws in Tables 72.1 and 72.2 allow a sort of computation with propositions in the way that the rules of ordinary algebra allow computation with numbers, such as the distributive law for multiplication over addition which says that $3(x+5) = 3x+15$. This system of computation is called the **propositional calculus**, a phrase which uses the word "calculus" in its older meaning "computational system". (What is called "calculus" in school used to be taught in two parts called the "differential calculus" and the "integral calculus".)

Recall that every predicate becomes a proposition (called an "instance" of the predicate) when constants are substituted for all its variables. Thus *when predicates are substituted for the propositional variables in these laws, they become predicates which are true in every instance.*

equivalent 40

(consistency)	$\neg T \Leftrightarrow F$	$\mathcal{U}^c = \emptyset$
	$\neg F \Leftrightarrow T$	$\emptyset^c = \mathcal{U}$
(unity)	$P \wedge T \Leftrightarrow P$	$A \cap \mathcal{U} = A$
	$P \vee F \Leftrightarrow P$	$A \cup \emptyset = A$
(nullity)	$P \wedge F \Leftrightarrow F$	$A \cap \emptyset = \emptyset$
	$P \vee T \Leftrightarrow T$	$A \cup \mathcal{U} = \mathcal{U}$
(idempotence)	$P \wedge P \Leftrightarrow P$	$A \cap A = A$
	$P \vee P \Leftrightarrow P$	$A \cup A = A$
(commutativity)	$P \wedge Q \Leftrightarrow Q \wedge P$	$A \cap B = B \cap A$
	$P \vee Q \Leftrightarrow Q \vee P$	$A \cup B = B \cup A$
(associativity)	$P \wedge (Q \wedge R)$	$A \cap (B \cap C)$
	$\Leftrightarrow (P \wedge Q) \wedge R$	$= (A \cap B) \cap C$
	$P \vee (Q \vee R)$	$A \cup (B \cup C)$
	$\Leftrightarrow (P \vee Q) \vee R$	$= (A \cup B) \cup C$
(distributivity)	$P \wedge (Q \vee R)$	$A \cap (B \cup C)$
	$\Leftrightarrow (P \wedge Q) \vee (P \wedge R)$	$= (A \cap B) \cup (A \cap C)$
	$P \vee (Q \wedge R)$	$A \cup (B \cap C)$
	$\Leftrightarrow (P \vee Q) \wedge (P \vee R)$	$= (A \cup B) \cap (A \cup C)$
(complement)	$P \vee \neg P \Leftrightarrow T$	$A \cup A^c = \mathcal{U}$
	$P \wedge \neg P \Leftrightarrow F$	$A \cap A^c = \emptyset$
(double negation)	$\neg\neg P \Leftrightarrow P$	$(A^c)^c = A$
(absorption)	$P \wedge (P \vee Q) \Leftrightarrow P$	$A \cap (A \cup B) = A$
	$P \vee (P \wedge Q) \Leftrightarrow P$	$A \cup (A \cap B) = A$
(DeMorgan)	$\neg(P \vee Q) \Leftrightarrow \neg P \wedge \neg Q$	$(A \cup B)^c = (A^c) \cap (B^c)$
	$\neg(P \wedge Q) \Leftrightarrow \neg P \vee \neg Q$	$(A \cap B)^c = (A^c \cup B^c)$

Table 72.1: Boolean Laws

('$\Rightarrow$'-elimination)	$(P \Rightarrow Q) \Leftrightarrow (\neg P \vee Q)$
(transitivity)	$((P \Rightarrow Q) \wedge (Q \Rightarrow R)) \Rightarrow (P \Rightarrow R)$
(modus ponens)	$(P \wedge (P \Rightarrow Q)) \Rightarrow Q$
(modus tollens)	$(\neg Q \wedge (P \Rightarrow Q)) \Rightarrow \neg P$
(inclusion)	$P \Rightarrow (P \vee Q)$
(simplification)	$(P \wedge Q) \Rightarrow P$
(cases)	$(\neg P \wedge (P \vee Q)) \Rightarrow Q$
(everything implies true)	$Q \Rightarrow (P \Rightarrow Q)$
(false implies everything)	$\neg P \Rightarrow (P \Rightarrow Q)$

Table 72.2: Laws of Implication

73.1.1 Example When you substitute $x > 7$ for P and $x = 5$ for Q in the second absorption law $P \vee (P \wedge Q) \Leftrightarrow P$ you get, in words, "Either $x > 7$ or both $x > 7$ and $x = 5$" is the same thing as saying "$x > 7$". This statement is certainly true: it is true *by its form*, not because of anything to do with the individual statements "$x > 7$" and "$x = 5$".

73.1.2 Exercise Define the logical connective NAND by requiring that P NAND Q be true provided at least one of P and Q is false.

a) Give the truth table for NAND.
b) Write a statement equivalent to "P NAND Q" using only '$\wedge$', '$\vee$', '$\neg$', 'P', 'Q' and parentheses.
c) Give statements equivalent to "$\neg P$", "$P \wedge Q$" and "$P \vee Q$" using only 'P', 'Q', 'NAND', parentheses and spaces.

73.1.3 Exercise Do the same as Problem 73.1.2 for the connective NOR, where P NOR Q is true only if both P and Q are false.

73.1.4 Exercise Show how to define implication in terms of each of the connectives NAND and NOR of exercises 73.1.2 and 73.1.3.

73.1.5 Exercise Let '$*$' denote the operation XOR discussed in Chapter 11. Prove the following laws:

a) $P * Q \Leftrightarrow Q * P$.
b) $P * (Q * R) \Leftrightarrow (P * Q) * R$.
c) $P \wedge (Q * R) \Leftrightarrow (P \wedge Q) * (P \wedge R)$.

73.1.6 Exercise (Mathematica)

a) Show that there are 16 possible truth tables for a Boolean expression with two variables.

b) Produce Boolean expressions with "$\neg$" and "$\Rightarrow$" as the only logical connectives that give each of the possible truth tables. Both variables must appear in each expression. Include a printout of Mathematica commands that verify that each expression gives the table claimed.

(Enter $p \Rightarrow q$ as `p ~Implies~ q`.)

73.1.7 Exercise (hard) A **distributive law** involving binary operations 'Δ' and '∇' is a tautology of the form

$$P\nabla(Q\Delta R) \Leftrightarrow (P\nabla Q)\Delta(P\nabla R)$$

Let '$*$' be defined as in Problem 73.1.5. Give examples showing that of the four possible distributive laws combining '$*$' with '$\wedge$' or '$\vee$', the only correct one is that in Problem 73.1.5(c).

74. The tautology theorem

In Section 28, we discussed the rule of inference called "modus ponens":

$$P,\, P \Rightarrow Q \vdash Q$$

This rule is closely related to the tautology also called modus ponens in section 71:

$$\Big(P \wedge (P \Rightarrow Q)\Big) \Rightarrow Q$$

This tautology is a propositional form which is true for any proposition P and Q. This is a special case of the general fact that, roughly speaking, any implication involving propositional forms which is a tautology is equivalent to a rule of inference:

74.1 Theorem: The Tautology Theorem
Suppose that $F_1, \ldots, F_n$ and G are propositional forms. Then

$$F_1, \ldots, F_n \vdash G \tag{74.1}$$

is a valid rule of inference if and only if

$$(F_1 \wedge \ldots \wedge F_n) \Rightarrow G \tag{74.2}$$

is a tautology.

Proof If the rule of inference (74.1) is correct, then whenever all the propositions $F_1, \ldots, F_n$ are true, G must be true, too. Then if $F_1 \wedge \cdots \wedge F_n$ is true, then every one of $F_1, \ldots, F_n$ is true, so G must be true. This means that (74.2) must be a tautology, for the only way it could be false is if $F_1 \wedge \cdots \wedge F_n$ is true and G is false. (This is because any implication $P \Rightarrow Q$ is equivalent to $\neg(P \wedge \neg Q)$.)

On the other hand, if (74.2) is a tautology, then whenever $F_1, \ldots, F_n$ are all true, then $F_1 \wedge \cdots \wedge F_n$ is true, so that G has to be true, too. That means that (74.1) is a valid rule of inference.

74.1.1 Example The preceding theorem applies to modus ponens: Take F_1 to be the formula P, F_2 to be "$P \Rightarrow Q$", and G to be Q. Since $\Big(P \wedge (P \Rightarrow Q)\Big) \Rightarrow Q$ is a tautology, the validity of the rule of inference called modus ponens follows by the Tautology Theorem from the tautology called modus ponens.

74.1.2 Remark Not all rules of inference come from tautologies – only those involving propositional forms. We have already seen examples of rules of inference not involving propositional forms in 18.1.11, page 29.

74.1.3 Warning The Tautology Theorem does not say that "$\vdash$" is the same thing as "$\Rightarrow$". "$\vdash$" is not a logical connective and cannot be used in formulas the way "$\Rightarrow$" can be. For example you may write $P \wedge (P \Rightarrow Q)$ but not $P \wedge (P \vdash Q)$. "$\vdash$" may be used *only in rules of inference.*

74.2 Exercise set

For problems 74.2.1 to 74.2.6, state whether the given rule is a valid rule of inference.

74.2.1 $\neg P,\ P \vee Q \vdash Q$ (Answer on page 247.)

74.2.2 $\neg Q,\ P \Rightarrow (Q \wedge R) \vdash \neg P$ (Answer on page 247.)

74.2.3 $\neg P,\ (P \wedge Q) \Rightarrow R \vdash \neg R$ (Answer on page 247.)

74.2.4 $\neg P \wedge Q,\ Q \vdash \neg P$

74.2.5 $(P \vee Q) \Rightarrow R,\ P \vdash R$

74.2.6 $(P \wedge Q) \Rightarrow R,\ \neg R \vdash \neg P \wedge \neg Q$

74.2.7 Exercise Show that the statement $(P \Rightarrow Q) \Rightarrow Q$ is not a tautology by giving an example of statements P and Q for which it is false. (Answer on page 247.)

74.2.8 Exercise Show that the following statements are not tautologies by giving examples of statements P and Q for which they are false.

a) $(P \Leftrightarrow Q) \Rightarrow P$
b) $\Big((P \Rightarrow Q) \Rightarrow R\Big) \Leftrightarrow \Big(P \Rightarrow (Q \Rightarrow R)\Big)$

74.2.9 Exercise Use the Tautology Theorem to prove that the following rules of inference are valid:

a) $Q \vdash P \Rightarrow Q$
b) $P, Q \vdash P \wedge Q$
c) $P \wedge Q \vdash P$
d) $\neg P \vdash P \Rightarrow Q$
e) $\neg Q,\ P \Rightarrow Q \vdash \neg P$

75. Quantifiers

75.1 Definition: universal quantifier
Let $Q(x)$ be a predicate. The statement $(\forall x)Q(x)$ is true if and only if $Q(x)$ is true for every value of the variable x. The symbol $\forall$ is called the **universal quantifier**.

75.1.1 Example Let $P(x)$ be the statement $(x > 5) \Rightarrow (x > 3)$. $P(x)$ is universally true, that is, it is true for every real number x. Therefore, the expression $(\forall x)P(x)$ is true.

We defined $\forall$ in 13.2; now we will go into more detail.

75.1.2 Showing the types of the variables A short way of saying that x is of type real and that $(\forall x)Q(x)$ is to write $(\forall x{:}\mathrm{R})Q(x)$, read "for all x of type R, $Q(x)$" or "for all real numbers x, $Q(x)$".

75.1.3 Example The statement $(\forall n{:}\mathrm{Z})((n > 5)$ is false because "$n > 5$" is false for $n = 3$ (and for an infinite number of other values of n).

75.1.4 Example The statement $(\forall n{:}\mathrm{Z})((n > 5) \vee (n < 5))$ is false because the statement "$(n > 5) \vee (n < 5)$" is false when $n = 5$. Note that in contrast to Example 75.1.3, $n = 5$ is the only value for which the statement "$(n > 5) \vee (n < 5)$" is false.

A statement like $(\forall x)Q(x)$ is true if $Q(x)$ is true no matter what is substituted for x (so long as it is of the correct type). *If there is even one x for which $Q(x)$ is false, then $(\forall x)Q(x)$ is false.* A value of x with this property is important enought to have a name:

75.2 Definition: counterexample
Let $Q(x)$ denote a predicate. An instance of x for which $Q(x)$ is false is called a **counterexample** to the statement $(\forall x)Q(x)$. If there is a counterexample to the statement $(\forall x)Q(x)$, then that statement is false.

75.2.1 Example $(\forall x{:}\mathrm{N})((x \leq 5) \vee (x \geq 6))$ is true, but $(\forall x{:}\mathrm{R})((x \leq 5) \vee (x \geq 6))$ is false (counterexample: $\frac{11}{2}$).

75.2.2 Example A counterexample to the statement $(\forall n{:}\mathrm{Z})((n > 5)$ is 3; in fact there are an infinite number of counterexamples to this statement. In contrast, the statement $(\forall n{:}\mathrm{Z})((n > 5) \vee (n < 5))$ has exactly one counterexample.

75.2.3 Exercise Find a universal statement about integers that has exactly 42 counterexamples.

75.2.4 Exercise Find a universal statement about real numbers that has exactly 42 counterexamples.

75.3 Definition: existential quantifier
Let $Q(x)$ be a predicate. The statement $(\exists x)Q(x)$ means there is some value of x for which the predicate $Q(x)$ is true. The symbol $\exists$ is called an **existential quantifier**, and a statement of the form $(\exists x)Q(x)$ is called an **existential statement**. A value c for which $Q(c)$ is true is called a **witness** to the statement $(\exists x)Q(x)$.

75.3.1 Remark One may indicate the type of the variable in an existential statement in the same way as in a universal statement.

75.3.2 Example Let x be a real variable and let $Q(x)$ be the predicate $x > 50$. This is certainly *not* true for all integers x. $Q(40)$ is false, for example. However, $Q(62)$ is true. Thus there are *some* integers x for which $Q(x)$ is true. Therefore $(\exists x{:}\mathrm{R})Q(x)$ is true, and 62 is a witness.

75.3.3 Exercise Find an existential statement about real numbers with exactly 42 witnesses.

75.3.4 Exercise In the following sentences, the variables are always natural numbers. $P(n)$ means n is a prime, $E(n)$ means n is even. State which are true and which are false. Give reasons for your answers.

a) $(\exists n)\big(E(n) \wedge P(n)\big)$
b) $(\forall n)\big(E(n) \vee P(n)\big)$
c) $(\exists n)(E(n) \Rightarrow P(n))$
d) $(\forall n)(E(n) \Rightarrow P(n))$

(Answer on page 247.)

75.3.5 Exercise Which of these statements are true for all possible one-variable predicates $P(x)$ and $Q(x)$? Give counterexamples for those which are not always true.

a) $(\forall x)(P(x) \wedge Q(x)) \Rightarrow (\forall x)P(x) \wedge (\forall x)Q(x)$
b) $(\forall x)P(x) \wedge (\forall x)Q(x) \Rightarrow (\forall x)(P(x) \wedge Q(x))$
c) $(\exists x)(P(x) \wedge Q(x)) \Rightarrow (\exists x)P(x) \wedge (\exists x)Q(x)$
d) $(\exists x)P(x) \wedge (\exists x)Q(x) \Rightarrow (\exists x)(P(x) \wedge Q(x))$

(Answer on page 247.)

75.3.6 Exercise Do the same as for Problem 75.3.5 with '$\vee$' in the statements in place of '$\wedge$'.

75.3.7 Exercise Do the same as for Problem 75.3.5 with '$\Rightarrow$' in the statements in place of '$\wedge$'.

75.3.8 Usage The symbols $\forall$ and $\exists$ are called **quantifiers**. The use of quantifiers makes an extension of the propositional calculus called the **predicate calculus** which allows one to say things about an infinite number of instances in a way that the propositional calculus does not.

76. Variables and quantifiers

If a predicate $P(x)$ has only one variable x in it, then using any quantifier in front of $P(x)$ with respect to that variable turns the statement into one which is either true or false — in other words, into a *proposition*.

76.1.1 Example If we let $P(n)$ be the statement $(n > 4) \wedge (n < 6)$, for n ranging over the integers, then $(\exists n)P(n)$, since $P(5)$ is true (5 is a witness). However, $(\forall n)P(n)$ is false, because for example $P(6)$ is false (6 is a counterexample). Both statements $(\exists n)P(n)$ and $(\forall n)P(n)$ are propositions; propositions, unlike predicates, are statements which are definitely true or false.

76.1.2 Predicates with more than one variable When a predicate has more than one variable, complications ensue. Let $P(x, y)$ be the predicate $(x > 5) \vee (5 > y)$. Let $Q(y)$ be the predicate $(\forall x{:}\mathrm{N})P(x, y)$. Then $Q(y)$ is the statement: "For every integer x, $x > 5$ or $5 > y$." This is still not a proposition. It contains one variable y, for which you can substitute an integer. It makes no sense to substitute an integer for x in $Q(y)$ (what would "For all 14, $14 > 5$ or $5 > y$" mean?) which is why x is not shown in the expression "$Q(y)$".

76.1.3 Bound and free A variable which is controlled by a quantifier in an expression is bound in the sense of 20.2. A logical expression in which all variables are bound is a proposition which is either true or false. If there are one or more free variables, it is not a proposition, but it is still a predicate.

76.1.4 Exercise Let $P(x, y)$ be the predicate

$$(x = y) \vee (x > 5)$$

If possible, find a counterexample to $(\forall y)P(14, y)$ and find a witness to $(\exists x)P(x, 3)$. (Answer on page 247.)

76.1.5 Exercise Let $Q(m, n)$ be each of the following statements. Determine in each case if $(\forall m{:}\mathrm{N})Q(m, 12)$ and $(\exists n{:}\mathrm{Z})Q(3, n)$ are true and give a counterexample or witness when appropriate.

a) $m \mid n$.
b) $\mathrm{GCD}(m, n) = 1$.
c) $(m \mid n) \Rightarrow (m \mid 2n)$.
d) $(m \mid n) \Rightarrow (mn = 12)$.

77. Order of quantifiers

Many important mathematical principles are statements with several quantified variables. The ordering of the quantifiers matters. The subtleties involved can be confusing.

77.1.1 Example The following statement is the **Archimedean property of the real numbers**.

$$(\forall x{:}\mathrm{R})(\exists n{:}\mathrm{N})(x < n) \tag{77.1}$$

In other words, "For any real number x there is an integer n bigger than x."

Proof If you are given a real number x, then $\operatorname{trunc}(x)+1$ is an integer bigger than x.

77.1.2 Example On the other hand, the statement

$$(\exists n{:}\mathrm{N})(\forall x{:}\mathrm{R})(x < n) \tag{77.2}$$

is *false*. It says there is an integer which is bigger than any real number. That is certainly not true: if you think 456,789 is bigger than any real number, then I reply, "It is not bigger than 456,790". In general, for any integer n, $n+1$ is bigger — and of course it is a real number, like any integer.

As these examples illustrate, in general, $(\forall x)(\exists y)P(x,y)$ does not mean the same as $(\exists y)(\forall x)P(x,y)$, although of course for particular statements both might be true.

On the other hand, two occurrences of the *same* quantifier in a row *can* be interchanged:

77.2 Theorem
For any statement $P(x,y)$,

$$(\forall x)(\forall y)P(x,y) \vdash (\forall y)(\forall x)P(x,y) \tag{77.3}$$

and

$$(\forall y)(\forall x)P(x,y) \vdash (\forall x)(\forall y)P(x,y) \tag{77.4}$$

and similarly

$$(\exists x)(\exists y)P(x,y) \vdash (\exists y)(\exists x)P(x,y) \tag{77.5}$$

and

$$(\exists y)(\exists x)P(x,y) \vdash (\exists x)(\exists y)P(x,y) \tag{77.6}$$

77.2.1 Exercise Are these statements true or false? Explain your answers. All variables are real.

a) $(\forall x)(\exists y)(x > y)$.
b) $(\exists x)(\forall y)(x > y)$
c) $(\exists x)(\exists y)((x > y) \Rightarrow (x = y))$.

(Answer on page 247.)

77.2.2 Exercise Are these statements true or false? Explain your answers. All variables are of type integer.

a) $(\forall m)(\exists n)(m \mid n)$.
b) $(\exists m)(\forall n)(m \mid n)$.
c) $(\forall m)(\exists n)((m \mid n) \Rightarrow (m \mid mn))$.
d) $(\exists m)(\forall n)((m \mid n) \Rightarrow (m \mid mn))$.

77.2.3 Exercise Are these statements true or false? Give counterexamples if they are false. In these statements, p and q are primes and m and n are positive integers.

a) $(\forall p)(\forall m)(\forall n)\big((p \mid m \Rightarrow p \mid n) \Rightarrow m \mid n\big)$
b) $(\forall m)(\forall n)\big(m \mid n \Rightarrow (\exists p)(p \mid m \wedge p \mid n)\big)$

77.2.4 Exercise (hard) Are these equivalences true for all predicates P and Q? Assume that the only variable in P is x and the only variables in Q are x and y. Give reasons for your answer.

a) $(\forall x)(\exists y)\big(P(x) \Rightarrow Q(x,y)\big) \Leftrightarrow (\forall x)\big(P(x) \Rightarrow (\exists y)Q(x,y)\big)$
b) $(\exists x)(\forall y)\big(P(x) \Rightarrow Q(x,y)\big) \Leftrightarrow (\exists x)\big(P(x) \Rightarrow (\forall y)Q(x,y)\big)$

78. Negating quantifiers

Negating quantifiers must be handled with care, too:

78.1 Theorem: Moving "not" past a quantifier
For any predicate P,
Q.1 $\neg((\exists x)P(x)) \Leftrightarrow (\forall x)(\neg P(x))$
Q.2 $\neg((\forall x)P(x)) \Leftrightarrow (\exists x)(\neg P(x))$.

Proof We give the argument for Q.1; the argument for Q.2 is similar.

For $(\exists x{:}A)P(x)$ to be false requires that $P(x)$ be false for every x of type A; in other words, that $\neg P(x)$ be *true* for every x of type A. For example, if $P(x)$ is the predicate $(x > 5) \wedge (x < 3)$, then $(\exists x{:}\mathrm{R})P(x)$ is false. In other words, the rule Q.1 is valid.

78.1.1 Remark Finding the negation of a proposition with several quantifiers can be done mechanically by applying the rules (Q.1) and (Q.2) over and over.

78.1.2 Example The negation of the Archimedean property can take any of the following equivalent forms:

a) $\neg\big((\forall x{:}\mathrm{R})(\exists n{:}\mathrm{N})(x < n)\big)$
b) $(\exists x{:}\mathrm{R})\neg\big((\exists n{:}\mathrm{N})(x < n)\big)$
c) $(\exists x{:}\mathrm{R})(\forall n{:}\mathrm{N})(x \geq n)$

The last version is easiest to read, and clearly false — there is no real number bigger than any integer. It is usually true that the easiest form to understand is the one with the '$\neg$ ' as "far in as possible".

78.1.3 Worked Exercise Express the negation of $(\forall x)(x < 7)$ without using a word or symbol meaning "not".
Answer $(\exists x)(x \geq 7)$.

78.1.4 Exercise Express the negation of $(\exists x)(x \leq 7)$ without using a word or symbol meaning "not".

78.1.5 Exercise Write a statement in symbolic form equivalent to the negation of

$$(\forall x)(P(x) \Rightarrow Q(x))$$

without using the '$\forall$' symbol.

78.1.6 Exercise Write a statement in symbolic form equivalent to the negation of the expression "$(\exists x)(P(x) \Rightarrow \neg Q(x))$" without using '$\exists$', ' $\Rightarrow$ ' or '$\neg$'.

79. Reading and writing quantified statements

An annoying fact about the predicate calculus is that even when you get pretty good at disentangling complicated logical statements, you may still have trouble reading mathematical proofs. One reason for this may be unfamiliarity with certain techniques of proof, some of which are discussed in the next chapter. Another is the variety of ways a statement in logic can be written in English prose. You have already seen the many ways an implication can be written (Section 27).

Much more about reading mathematical writing may be found in the author's works [Wells, 1995], [Bagchi and Wells, 1998b], [Bagchi and Wells, 1998a], and [Wells, 1998].

79.1.1 Example The true statement, for real numbers,

$$(\forall x)\big(x \geq 0 \Rightarrow (\exists y)(y^2 = x)\big) \tag{79.1}$$

could be written in a math text in any of the following ways:

a) If $x \geq 0$, then there is a y for which $y^2 = x$.
b) For any $x \geq 0$, there is some y such that $y^2 = x$.
c) If x is nonnegative, then it is the square of some real number.
d) Any nonnegative real number is the square of another one.
e) A nonnegative real number has a square root.

Or it could be set off this way

$$x \geq 0 \Rightarrow (\exists y)(y^2 = x) \qquad\qquad (x)$$

with the (x) on the far right side denoting "$\forall x$". Sometimes (x) is used instead of $\forall x$ next to the predicate, too:

$$(x)(x \geq 0 \Rightarrow (\exists y)(y^2 = x))$$

79.1.2 Warning The words “any”, “all” and “every” have rather delicate rules of usage, as well. Sometimes they are interchangeable and sometimes not. The Archimedean axiom could be stated, “For every real x there is an integer $n > x$,” or “For any real x there is an integer $n > x$.” But it would be misleading, although perhaps not strictly wrong, to say, “For all real numbers x there is an integer $n > x$,” which could be misread as claiming that there is one integer n that works for all x.

79.1.3 Warning Observe that the statements in (a), (c) and (e) have no obvious English word corresponding to the quantifier. This usage there is somewhat similar to the use of the word “dog” in a sentence such as, “A wolf mates for life”, meaning every wolf mates for life.

Students sometimes respond to a question such as, “Prove that an integer divisible by 4 is even” with an answer such as, “The integer 12 is divisible by 4 and it is even”. However, the question means, “Prove that *every* integer divisible by 4 is even.” This blunder is the result of not understanding the way a universal quantifier can be signaled by the indefinite article.

79.1.4 Example Consider the well-known remark, “All that glitters is not gold.” This statement means

$$\neg(\forall x)(\mathrm{GLITTER}(x) \Rightarrow \mathrm{GOLD}(x))$$

rather than

$$(\forall x)(\mathrm{GLITTER}(x) \Rightarrow \neg\mathrm{GOLD}(x))$$

In other words, it means, “Not all that glitters is gold.” (We do *not* say the statement is incorrect English or correct English with a different meaning; we only give it as an illustration of the subtleties involved in translating from English to logic.)

79.1.5 Worked Exercise Write these statements in logical notation. Make up suitable names for the predicates.

a) All people are mortal.
b) Some people are not mortal.
c) All people are not mortal.

Answer (a) $(\forall x)\big(\mathrm{Person}(x) \Rightarrow \mathrm{Mortal}(x)\big)$
(b) $(\exists x)\big(\mathrm{Person}(x) \wedge \neg\mathrm{Mortal}(x)\big)$
(c) $(\forall x)\big(\mathrm{Person}(x) \Rightarrow \neg\mathrm{Mortal}(x)\big)$

79.1.6 Exercise Write these statements in logical notation.

a) Everybody likes somebody.
b) Everybody doesn’t like something.
c) Nobody likes everything.
d) You can fool all of the people some of the time and some of the people all of the time, but you can’t fool all of the people all of the time.

79.1.7 Exercise Write the statement in GS.2, page 61, using quantifiers.

80. Proving implications: the Direct Method

Because so many mathematical theorems are implications, it is worthwhile considering the ways in which an implication can be proved. We consider two common approaches in this chapter.

80.1 The direct method

If you can deduce Q from P, then $P \Rightarrow Q$ must be true. That is because the only line of the truth table for ‘ $\Rightarrow$ ’ (Table 25.1) which has an ‘F’ is the line for which P is true and Q is false, which cannot happen if you can deduce Q from P. This gives:

> **80.1.1 Method: Direct Method**
> To prove $P \Rightarrow Q$, assume P is true and deduce Q.

80.1.2 Remark Normally, in proving Q, you would use other facts at your disposal as well as the assumption that P is true. As an illustration of the direct method, we prove the following theorem.

80.2 Theorem
If a positive integer is divisible by 2 *then* 2 *occurs in its prime factorization.*

Proof Let n be divisible by 2. (Thus we assume the hypothesis is true.) Then 2 divides n, so that by definition of division $n = 2m$ for some integer m. Let

$$m = p_1^{e_1} \times \ldots \times p_n^{e_n}$$

be the prime factorization of m. Then

$$n = 2 \times p_1^{e_1} \times \ldots \times p_n^{e_n}$$

is a factorization of n into primes (since 2 is a prime), so is *the* prime factorization of n because the prime factorization is unique by the Fundamental Theorem of Arithmetic.

80.2.1 Coming up with proofs In a more complicated situation, you might have to prove $P \Rightarrow P_1$, $P_1 \Rightarrow P_2$, $\ldots$, $P_k \Rightarrow Q$ in a series of deductions.

Normally, although your final proof would be written up in that order, *you would not think up the proof by thinking up* $P_1, P_2, \ldots$ *in order.* What happens usually is that you think of statements which imply Q, statements which imply *them* (backing up), and at the same time you think of statements which P implies, statements which *they* imply (going forward), and so on, until your chain meets in the middle (if you are lucky). Thinking up a proof is thus a creative act rather than the cut-and-dried one of grinding out conclusions from hypotheses.

80.2.2 Exercise Prove by the direct method that for any integer n, if n is even so is n^2.

81. Proving implications: the Contrapositive Method

It is very common to use the contrapositive to prove an implication. Since "$P \Rightarrow Q$" is equivalent to "$\neg Q \Rightarrow \neg P$", you can prove "$P \Rightarrow Q$" by using the direct method to prove "$\neg Q \Rightarrow \neg P$". In detail:

81.0.3 Method: Contrapositive Method
(The contrapositive method) To prove $P \Rightarrow Q$, assume Q is false and deduce that P is false.

81.0.4 Warning This method is typically used in math texts without mentioning that the contrapositive is being used. You have to realize that yourself.

81.0.5 Example The proof of the following theorem is an illustration of the use of the contrapositive, written the way it might be written in a math text. Recall that an integer k is even if $2 \mid k$.

81.1 Theorem
For all positive integers n, if n^2 is even, so is n.

Proof Let n be odd. Then 2 does not occur in the prime factorization of n. But the prime factorization of n^2 merely repeats each prime occurring in the factorization of n, so no new primes occur. So 2 does not occur in the factorization of n^2 either, so by Theorem 80.2, n^2 is odd. This proves the theorem.

81.1.1 Remarks

a) If you didn't think of proving the contrapositive, you might be dumbfounded when you saw that a proof of a theorem which says "if n^2 is even then n is even" begins with, "Let n be odd..." The contrapositive of the statement to be proved is, "If n is odd, then n^2 is odd." The proof of the contrapositive proceeds like any direct-method proof, by assuming the hypothesis (n is odd).
b) The contrapositive of Theorem 80.2 is used in the proof of Theorem 81.1. That theorem says that if n is even, then its prime factorization contains 2. Here we are using it in its contrapositive form: if 2 does not occur in the prime factorization of n, then n is not even, i.e., n is odd. Again, the proof does not mention the fact that it is using Theorem 80.2 in the contrapositive form.
c) Theorem 81.1, like most theorems in mathematics, is a universally quantified implication, so using universal generalization we showed that if n is an *arbitrary* positive integer satisfying the hypothesis, then it must satisfy the conclusion. In such a proof, we are not allowed to make any special assumptions about n except that it satisfies the hypothesis. On the other hand, if we suspected that the theorem were false, we could prove that it is false merely by finding a *single* positive integer n satisfying the hypothesis but not the conclusion. (Consider the statement, "If n is prime, then it is odd.") This phenomenon has been known to give students the impression that proving statements is much harder than disproving them, which somehow doesn't seem fair.

81.1.2 Exercise Prove by the contrapositive method that if n^2 is odd then so is n.

81.2 Exercise set

Exercises 81.2.1 through 81.2.3 provide other methods of proof.

81.2.1 Exercise Prove that

$$(P \wedge \neg Q) \Leftrightarrow \neg(P \Rightarrow Q) \tag{81.1}$$

is a tautology. Thus to prove that an implication is false, you must show that its hypothesis is true and its conclusion is false. In particular, *the negation of an implication is not an implication.*

81.2.2 Exercise Prove that the rule

$$\neg P \Rightarrow Q \vdash P \vee Q \tag{81.2}$$

is a valid inference rule. (A proof using this rule would typically begin the proof of $P \vee Q$ by saying, "Assume $\neg P$..." and then proceed to deduce Q.)

81.2.3 Exercise Prove that the rule

$$P \Rightarrow Q, Q \Rightarrow R \vdash P \Rightarrow R$$

is a valid inference rule. (This allows proofs to be strung together.)

81.2.4 Exercise (hard) Use the methods of this chapter to prove that n is prime if and only if $n > 1$ and there is no divisor k of n satisfying $1 < k \leq \sqrt{n}$.

82. Fallacies connected with implication

82.1 Definition: fallacy
An argument which does not use correct rules of inference is called a **fallacy**.

82.1.1 Example Two very common fallacies concerning implications are

F.1 assuming that from $P \Rightarrow Q$ and Q you can derive P ("A cow eats grass. This animal eats grass, so it must be a cow.") and

F.2 assuming that from $P \Rightarrow Q$ and $\neg P$ that you can derive $\neg Q$ ("A cow eats grass. This animal is not a cow, so it won't eat grass.")

82.1.2 Remark You will sometimes hear these fallacies used in political arguments. F.1 is called **affirming the hypothesis** and F.2 is called **denying the consequent**.

82.1.3 Remark Fallacious arguments involve an incorrect use of logic, although both the hypothesis and the conclusion might accidentally be correct. Fallacious arguments should be distinguished from correct arguments based on faulty assumptions.

82.1.4 Example The statement, "A prime number bigger than 2 is odd. 5 is odd, so 5 is prime" is fallacious, even though the conclusion is true. (The hypothesis is true, too!). It is an example of affirming the hypothesis.

82.1.5 Example The statement "An odd number is prime, 15 is odd, so 15 is prime" is *not* fallacious— it is a logically correct argument based on an incorrect hypothesis ("garbage in, garbage out").

82.1.6 Example The argument, "Any prime is odd, 16 is even, so 16 is not a prime" is a logically correct argument with a correct conclusion, but the hypothesis, "Any prime is odd", is false. The latter is a case of "getting the right answer for the wrong reason," which is a frequent source of friction between students and math teachers.

82.2 Exercise set

In Problems 82.2.1 through 82.2.5, some arguments are valid and some are fallacious. Some of the valid ones have false hypotheses and some do not. (The hypothesis is in square brackets.) State the method of proof used in those that are valid and explain the fallacy in the others. The variable n is of positive integer type.

82.2.1 [$n > 5$ only if $n > 3$]. Since $17 > 5$, it must be that $17 > 3$. (Answer on page 247.)

82.2.2 [$n > 5$ only if $n > 3$]. Since $4 > 3$, it must be that $4 > 5$. (Answer on page 247.)

82.2.3 [If n is odd, then $n \neq 2$]. 6 is not odd, so $6 = 2$. (Answer on page 247.)

82.2.4 [n is odd only if it is prime]. 17 is odd, so 17 is a prime. (Answer on page 247.)

82.2.5 [If n is even and $n > 2$, then n is not prime]. 15 is odd, so 15 is prime. (Answer on page 247.)

83. Proving equivalences

> **83.1.1 Method**
> An equivalence "$P \Leftrightarrow Q$" is proved by proving both $P \Rightarrow Q$ and $Q \Rightarrow P$.

83.1.2 Remark Remember the slogan: *To prove an equivalence you must prove two implications.*

83.1.3 Remark Quite commonly the actual proof proves (for example) $P \Rightarrow Q$ and $\neg P \Rightarrow \neg Q$ (the contrapositive of $Q \Rightarrow P$), so the proof has two parts: the first part begins, "Assume P", and the second part begins, "Assume $\neg P$..."

83.1.4 Example Here is an example of a theorem with such a proof. The proof avoids the use of the Fundamental Theorem of Arithmetic, which would make it easier, so as to provide a reasonable example of the discussion in the preceding paragraph.

83.2 Theorem
For any integer n, $2\,|\,n$ if and only if $4\,|\,n^2$.

Proof If $2\,|\,n$ then by definition there is an integer k for which $n = 2k$. Then $n^2 = 4k^2$, so n^2 is divisible by 4.

Now suppose 2 does not divide n, so that n is odd. That means that $n = 2k+1$ for some integer k. Then $n^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$ which is odd, so is not divisible by 2, much less by 4.

83.2.1 Remark The preceding proof is written the way such proofs commonly appear in number theory texts: no overt statement is made concerning the structure of the proof. You have to deduce the structure by the way it proceeds. In this proof, P is the statement "$2\,|\,n$" and Q is the statement "$4\,|\,n^2$". To prove $P \Leftrightarrow Q$, the proof proceeds to prove first (before the phrase "Now suppose") that $P \Rightarrow Q$ by the direct method, and then to prove that $Q \Rightarrow P$ by the contrapositive method, that is, by proving $\neg P \Rightarrow \neg Q$ by the direct method.

83.2.2 Exercise Prove that for all integers m and n, $m+n$ is even if and only if $m-n$ is even.

83.2.3 Exercise Let α be a relation on a set A. Prove that α is reflexive if and only if $\Delta_A \subseteq \alpha$.

83.2.4 Exercise Let α be a relation on a set A. Prove that α is antisymmetric if and only if

$$\alpha \cap \alpha^{\mathrm{op}} \subseteq \Delta_A$$

84. Multiple equivalences

Some theorems are in the form of assertions that three or more statements are equivalent.

This theorem provides an example:

84.2 Theorem
The following are equivalent for a positive integer n:
D.1 *n is divisible by 4.*
D.2 *$n/2$ is an even integer.*
D.3 *$n/4$ is an integer.*

84.2.1 Remark In proving such a theorem, it is only necessary to prove three implications, not six, provided the three are chosen correctly. For example, it would be sufficient to prove $P \Rightarrow Q$, $Q \Rightarrow R$ and $R \Rightarrow P$. Then for example $Q \Rightarrow P$ follows from $Q \Rightarrow R$ and $R \Rightarrow P$. (See Problem 84.2.3).

84.2.2 Warning Theorem 84.2 does *not* say that n is divisible by 4. It says that if one of the statements is true, the other two must be true also (so if one is false the other two must be false). It therefore says

$$(P \Leftrightarrow Q) \wedge (Q \Leftrightarrow R) \wedge (P \Leftrightarrow R)$$

for certain statements P, Q and R. That is the same as asserting *six* implications, $P \Rightarrow Q$, $Q \Rightarrow P$, $P \Rightarrow R$, $R \Rightarrow P$, $Q \Rightarrow R$, and $R \Rightarrow Q$.

84.2.3 Exercise Write out careful proofs of Theorem 84.2 in two ways:

a) (D.1) $\Rightarrow$ (D.2), (D.2) $\Rightarrow$ (D.3), and (D.3) $\Rightarrow$ (D.1), and
b) (D.1) $\Rightarrow$ (D.3), (D.3) $\Rightarrow$ (D.2), and (D.2) $\Rightarrow$ (D.1).

84.2.4 Exercise Prove that the following three statements are equivalent for any sets A and B:

a) $A \subseteq B$
b) $A \cup B = B$
c) $A \cap B = A$

84.2.5 Exercise Let α be a relation on a set A. Prove that the following three statements are equivalent.

a) α is symmetric.
b) $\alpha \subseteq \alpha^{\mathrm{op}}$.
c) $\alpha = \alpha^{\mathrm{op}}$.

85. Uniqueness theorems

In a particular system such as the positive integers, *any uniqueness theorem gives a rule of inference.* Such a rule only applies to the data type for which the uniqueness theorem is stated.

85.1.1 Example Theorem 60.2 says that the quotient and remainder are uniquely determined by Definition 60.1. This provides a rule of inference for nonnegative integers:

$$m = qn + r,\ 0 \le r < n \ \vdash\ (q = m \operatorname{div} n) \wedge (r = m \bmod n) \tag{85.1}$$

85.1.2 Remark The conclusion of this rule of inference can be worded this way: q is the quotient and r is the remainder when m is divided by n. For example, because $m = 50$, $n = 12$, $q = 4$ and $r = 2$ satisfy Rule (85.1), $4 = 50 \operatorname{div} 12$ and $2 = 50 \bmod 12$. You do not have to do a long division to verify that; it follows from Rule (85.1).

85.1.3 Exercise Prove that if $0 \le m - qn < n$, then $q = m \operatorname{div} n$. (Answer on page 247.)

85.1.4 Exercise Use Rule (85.1) to prove that if $r = m \bmod n$ and $r' = m' \bmod n$, then $(m+m') \bmod n$ is either $r+r'$ or $r+r'-n$.

85.1.5 Exercise Use Rule (85.1) to prove that if $r = m \bmod n$, then $n \mid m-r$.

85.1.6 A rule for GCD's For positive integers m and n, the greatest common divisor $\mathrm{GCD}(m,n)$ is the largest integer dividing both m and n; this definition was also given in Chapter 60. This obviously determines the GCD uniquely — there cannot be two largest integers which divide both m and n. This can be translated into a rule of inference:

$$(\forall e)\Big((d \mid m) \wedge (d \mid n) \wedge ((e \mid m \wedge e \mid n) \Rightarrow e \le d)\Big) \vdash d = \mathrm{GCD}(m,n) \qquad (85.2)$$

85.1.7 Example Let's use the rule just given to prove Theorem 64.1. We must prove that the number d which is the product of all the numbers $p^{\min(\mathrm{e}_p(m),\mathrm{e}_p(n))}$ for all primes p which divide m or n or both is $\mathrm{GCD}(m,n)$.

First, $d \mid m$ and $d \mid n$, since the exponent of any prime p in d, which is

$$\min(\mathrm{e}_p(m), \mathrm{e}_p(n))$$

is obviously less than or equal to $\mathrm{e}_p(m)$ and to $\mathrm{e}_p(n)$, so Theorem 62.4 applies. Thus we have verified two of the three hypotheses of Rule (85.2). As for the third, suppose $e \mid m$ and $e \mid n$. Then $\mathrm{e}_p(e) \le \mathrm{e}_p(m)$ and $\mathrm{e}_p(e) \le \mathrm{e}_p(n)$, so $\mathrm{e}_p(e) \le \min(\mathrm{e}_p(m), \mathrm{e}_p(n))$, so $e \mid d$. But if $e \mid d$, then $e \le d$, so the third part of Rule (85.2) is correct. Hence the conclusion that $d = \mathrm{GCD}(m,n)$ must be true.

85.1.8 Exercise State and prove a rule like Rule (85.2) for $\mathrm{LCM}(m,n)$.

86. Proof by Contradiction

Another hard-to-understand method of proof is proof by contradiction, one form of which is expressed by this rule of inference:

86.1 Theorem

$$\neg Q,\ P \Rightarrow Q \vdash \neg P \qquad (86.1)$$

86.1.1 Remarks

a) Theorem 86.1 follows from the tautology

$$\big(\neg Q \wedge (P \Rightarrow Q)\big) \Rightarrow \neg P$$

b) This rule says that to prove $\neg P$ it suffices to prove $\neg Q$ and that $P \Rightarrow Q$.

86.1.2 Usage A proof using the inference rule of Theorem 86.1 is called **proof by contradiction**, or **reductio ad absurdum** ("r.a.a").

86.1.3 Remarks

a) In practice it frequently happens that Q is obviously false so that the work goes into proving $P \Rightarrow Q$. Thus a proof of $\neg P$ by contradiction might begin, "Suppose P is true ... "!
b) Authors typically don't tell the reader they are doing a proof by contradiction. It is generally true that mathematical authors are very careful to tell the reader which previous or known theorems his proof depends on, but says nothing at all about the rule of inference or method of proof being used.

As an illustration of proof by contradiction, we will prove this famous theorem:

86.2 Theorem
$\sqrt{2}$ is not rational.

86.2.1 Remarks

a) The discovery of this theorem by an unknown person in Pythagoras' religious colony in ancient Italy caused quite a scandal, because the "fact" that any real number could be expressed as a fraction of integers was one of the beliefs of their religion (another was that beans were holy).
b) Theorem 86.2 is a remarkable statement: it says that there is *no fraction* m/n for which $(m/n)^2 = 2$. Although $\sqrt{2}$ is approximately equal to $1414/1000$, it is not *exactly* equal to any fraction of integers whatever. The fact that $\sqrt{2}$ has a nonterminating decimal expansion does not of course prove this, since plenty of fractions (e.g., $1/3$) have nonterminating decimal expansions.

   How on earth do you prove an impossibility statement like that? After all, you can't go through the integers checking every fraction m/n. It is that sort of situation that demands a proof by contradiction.

Proof Here is the proof, using the Fundamental Theorem of Arithmetic. Suppose $\sqrt{2}$ is rational, so that for some integers m and n, $2 = (m/n)^2$. Then $2n^2 = m^2$. Every prime factor in the square of an integer must occur an even number of times. Thus $\mathrm{e}_2(m^2)$ is even and $\mathrm{e}_2(n^2)$ is even. But $\mathrm{e}_2(2n^2) = 1 + \mathrm{e}_2(n^2)$, so $\mathrm{e}_2(2n^2)$ is odd, a contradiction.

86.2.2 Remark In fact, π (and many other numbers used in calculus) is not rational either, but the proof is harder.

86.2.3 Worked Exercise Use the Fundamental Theorem of Arithmetic to prove that there are an infinite number of primes.

Answer This will be a proof by contradiction. Suppose there is a finite number of primes: suppose that $p_1, p_2, \ldots, p_k$ are *all* the primes. Let $m = p_1 \cdot p_2 \cdots p_k + 1$. Then the remainder when m is divided by *any* prime is 1. Since no prime divides m, it cannot have a prime factorization, contradicting the Fundamental Theorem of Arithmetic.

86.2.4 Exercise Use proof by contradiction to prove that if p is a prime and $p > 2$, then p is odd. (Answer on page 247.)

86.2.5 Exercise Prove that for all rational numbers x, $(x^2 < 2) \Leftrightarrow (x^2 \leq 2)$.

86.2.6 Exercise Give an example of a pair of distinct irrational numbers r and s with the property that $r+s$ is rational.

86.2.7 Exercise Use proof by contradiction to prove that if r and s are real numbers and r is rational and s is not rational, then $r+s$ is not rational.

86.2.8 Exercise Use proof by contradiction to prove that for any integer $k > 1$ and prime p, the kth root of p is not rational.

86.2.9 Exercise (hard) Use Problem 86.2.8 to prove that the kth root of a positive integer is either an integer or is not rational.

86.2.10 Exercise (hard) Show that there are infinitely many primes p such that $p \bmod 4 = 3$. Hint: Use proof by contradiction. Assume there are only finitely many such primes, and consider the number m which is the product of all of them. Consider two cases, $m \bmod 4 = 1$ and $m \bmod 4 = 3$, and ask what primes can divide $m+2$ or $m+4$. Use problem 60.5.6, page 85 and other similar facts. Note that the similar statement about $p \bmod 4 = 1$ is also true but *much* harder to prove.

87. Bézout's Lemma

The Fundamental Theorem of Arithmetic, that every integer greater than one has a unique factorization as a product of primes, was stated without proof in Chapter 62. It actually follows from certain facts about the GCD by a fairly complicated proof by contradiction. This proof is based on Theorem 87.2 below, a theorem which is worth knowing for its own sake. The proof of the Fundamental Theorem is completed in Problems 104.4.1 through 104.4.4.

87.1 Definition: integral linear combination
If m and n are integers, an **integral linear combination** of m and n is an integer d which is expressible in the form $d = am + bn$, where a and b are integers.

87.1.1 Example 2 is an integral linear combination of 10 and 14, since

$$3 \times 10 - 2 \times 14 = 2$$

However, 1 is not an integral linear combination of 10 and 14, since any integral linear combination of 10 and 14 must clearly be even.

87.1.2 Remark Note that in the definition of integral linear combination, the expression $d = am + bn$ does not determine a and b uniquely for a given m and n.

87.1.3 Example $3 \times 10 - 2 \times 14 = 2$ and $-4 \times 10 + 3 \times 14 = 2$. (See Exercise 88.3.7.)

87.1.4 Exercise Show that if $d \mid m$ and $d \mid n$ then d divides any integral linear combination of m and n.

> **87.2 Theorem: Bézout's Lemma**
> *If m and n are positive integers, then* $\operatorname{GCD}(m,n)$ *is the smallest positive integral linear combination of m and n.*

87.2.1 Remark Bézout's Lemma should not be confused with Bézout's Theorem, which is a much more substantial mathematical result concerning intersections of surfaces defined by polynomial equations.

87.2.2 Example $\operatorname{GCD}(10,14) = 2$, and 2 is an integral linear combination of 10 and 14 $(2 = 3 \cdot 10 + (-2) \cdot 14)$ but 1 is not, so 2 is the smallest positive integral linear combination of 10 and 14.

87.2.3 Proof of Bézout's Lemma We prove this without using the Fundamental Theorem of Arithmetic, since the lemma will be used later to prove the Fundamental Theorem.

Let e be the smallest positive integral linear combination of m and n. Suppose $e = am + bn$. Let $d = \operatorname{GCD}(m,n)$.

First, we show that $d \le e$. We know that $d \mid m$ and $d \mid n$, so there are integers h and k for which $m = dh$ and $n = dk$. Then $e = am + bn = adh + bdk = d(ah + bk)$ is divisible by d. It follows that $d \le e$.

Now we show that $e \mid m$ and $e \mid n$. Let $m = eq + r$ with $0 \le r < e$. Then

$$r = m - eq = m - (am + bn)q = (1 - aq)m - bqn$$

so r is an integral linear combination of m and n. Since e is the smallest positive integral linear combination of m and n and $r < e$, this means $r = 0$, so $e \mid m$. A similar argument shows that $e \mid n$.

It follows that e is a *common divisor* of m and n and d is the *greatest common divisor*; hence $e \le d$. Combined with the previous result that $d \le e$, we see that $d = e$, as required.

88. A constructive proof of Bézout's Lemma

The preceding proof of Bézout's Lemma does not tell us *how to calculate* the integers a and b for which $am + bn = \operatorname{GCD}(m,n)$. For example, see how fast you can find integers a and b for which $13a + 21b = 1$. (See Exercise 107.3.4.)

We now give a modification of the Euclidean algorithm which constructs integers a and b for which $\operatorname{GCD}(m,n) = am + bn$. The Euclidean algorithm is given as program 65.1, page 93, based on Theorem 65.1, which says that for any integers m and n, $\operatorname{GCD}(m,n) = \operatorname{GCD}(n, m \bmod n)$. Program 65.1 starts with M and N and repeatedly replaces N by $M \bmod N$ and M by N. The last value of N before it becomes 0 is the GCD. This lemma shows how being an integral linear combination is preserved by that process:

88.1 Lemma

Let m and n be positive integers.

B.1 *The integers m and n are integral linear combinations of m and n.*

B.2 *If u and v are integral linear combinations of m and n and $v \neq 0$, then $u \bmod v$ is also an integral linear combination of m and n.*

Proof B.1 is trivial: $m = 1 \times m + 0 \times n$ and $n = 0 \times m + 1 \times n$. As for B.2, suppose $u = wm + xn$ and $v = ym + zn$. Let $u = qv + r$ with $0 \leq r < v$, so $r = u \bmod v$. Then

$$r = u - qv = wm + xn - q(ym + zn) = (w - qy)m + (x - qz)n$$

so r is an integral linear combination of m and n, too.

88.2 A method for calculating the Bézout coefficients

We now describe a method for calculating the Bézout coefficients based on Lemma 88.1. Given positive integers m and n with $d = \text{GCD}(m, n)$, we calculate integers a and b for which $am + bn = d$ as follows: Make a table with columns labeled u, v, w and $w = am + bn$.

1. Put $u = m$, $v = n$, $w = m \bmod n$ in the first row, and in the last column put the equation $w = m - (m \operatorname{div} n)n$. Note that this equation expresses $m \bmod n$ in the form $am + bn$ (here $a = 1$ and $b = -m \operatorname{div} n$).

2. Make each succeeding row u', v', w', $w' = a'm + b'n$ by setting $u' = v$ (the entry under v in the preceding row), $v' = w$ and $w' = v \bmod w$, and solving for a' and b' by using the equation $w' = u' - (u' \operatorname{div} v')v'$ and the equations in the preceding rows. Note that the entry in the last column always expresses w in terms of the original m and n, *not* in terms of the u and v in that row.

3. Continue this process until the entry under w is $\text{GCD}(m, n)$ (this always happens because the first three columns in the process constitute the Euclidean algorithm).

88.2.1 Example The following table shows the calculation of integers a and b for which $100a + 36b = 4$.

u	v	w	
100	36	28	$28 = 100 - 2 \cdot 36$ Note that $100 \operatorname{div} 36 = 2$
36	28	8	$8 = 36 - 28 = 36 - (100 - 2 \cdot 36) = 3 \cdot 36 - 100$
28	8	4	$4 = 28 - 3 \cdot 8 = 100 - 2 \cdot 36 - 3(3 \cdot 36 - 100) = 4 \cdot 100 - 11 \cdot 36$

so that $a = 4$, $b = -11$.

88.3 Constructive and nonconstructive

The two proofs we have given for Theorem 87.2 illustrate a common phenomenon in mathematics. The first proof is **nonconstructive**; it shows that the requisite integers a and b exist but does not tell you how to get them. The second proof is **constructive**; it is more complicated but gives an explicit way of constructing a and b.

88.3.1 Exercise Express a as an integral linear combination of b and c, or explain why this cannot be done.

a	b	c
2	12	16
4	12	16
2	26	30
4	26	30
−2	26	30
1	51	100

(Answer on page 247.)

88.3.2 Exercise Express 1 as an integral linear combination of 13 and 21.

88.3.3 Exercise (M. Leitman) Suppose a, b, m and n are integers. Prove that if m and n are relatively prime and $am+bn=e$, then there are integers a' and b' for which $a'm+b'n=e+1$. (Answer on page 247.)

88.3.4 Exercise Prove without using the Fundamental Theorem of Arithmetic that if $\mathrm{GCD}(m,n)=1$ and $m \mid nr$ then $m \mid r$. (Use Bézout's Lemma, page 128.)

88.3.5 Exercise Suppose that a, b and c are positive integers for which $c=12a-8b$. Show that $\mathrm{GCD}(a,b) \leq \frac{c}{4}$.

88.3.6 Exercise Prove that the following rule of inference is valid (use Bézout's Lemma, page 128).

$$e \mid m,\ e \mid n \ \vdash\ e \mid \mathrm{GCD}(m,n)$$

(It follows that the statement "$e \leq d$" in Rule (85.2) can be replaced by "$e \mid d$".)

88.3.7 Exercise (hard) Prove that if d is an integral linear combination of m and n then there are an infinite number of different pairs of integers a and b for which $d=am+bn$.

88.3.8 Exercise Use Bézout's Lemma (page 128) to prove Corollary 64.2 on page 90 without using the Fundamental Theorem of Arithmetic.

89. The image of a function

If $F: A \to B$ is a function, it can easily happen that not every element of B is a value of F. For example, the function $x \mapsto x^2 : \mathrm{R} \to \mathrm{R}$ takes only nonnegative values.

89.1 Definition: image of a function
The **image** of $F: A \to B$ is the set of all values of F, in other words the set $\{b \in B \mid (\exists a : A)(F(a) = b)\}$. The image of F is also denoted $\mathrm{Im}(F)$.

89.1.1 Fact This definition gives the equivalence:

$$(\exists a)(F(a) = b) \Leftrightarrow b \in \mathrm{Im}\, F$$

89.1.2 Fact For any function F, $\mathrm{Im}(F) \subseteq \mathrm{cod}\, F$.

89.1.3 Usage Many authors use the word “range” for the image, but others use “range” for the codomain.

89.1.4 Example The image of the squaring function $x \mapsto x^2 : \mathrm{R} \to \mathrm{R}$ is the set of nonnegative real numbers.

89.1.5 Example Let the function $F: \{1,2,3\} \to \{2,4,5,6\}$ be defined by $F(1) = 4$ and $F(2) = F(3) = 5$. Then F has image $\{4,5\}$.

89.1.6 Remark The image of a function can be difficult to determine if it is given by a formula; for example it requires a certain amount of analytic geometry (or calculus) to determine that the image of the function $G(x) = x^2 + 2x + 5$ is the set of real numbers ≥ 4, and determining the image of more complicated functions can be very difficult indeed.

89.1.7 Exercise Find the image of the function $n \mapsto n + 1 : \mathrm{N} \to \mathrm{N}$. (Answer on page 247.)

89.1.8 Exercise Find the image of the function $n \mapsto n - 1 : \mathrm{Z} \to \mathrm{Z}$.

89.1.9 Exercise Find the image of the function $x \mapsto x^2 - 1 : \mathrm{R} \to \mathrm{R}$.

89.1.10 Exercise Find the image of the function $x \mapsto x^2 + x + 1 : \mathrm{R} \to \mathrm{R}$.

90. The image of a subset of the domain

The word "image" is used in a more general way which actually makes the image a function itself.

90.1 Definition: Image of a subset
Let $F : A \to B$ is a function, and suppose $C \subseteq A$. Then $F(C)$ denotes the set $\{F(x) \mid x \in C\}$, and is called the **image of** C **under** F. The map $C \mapsto F(C)$ defines a function from $\mathcal{P}A$ to $\mathcal{P}B$ called the **image function** of F.

90.1.1 Remark In particular, $F(A)$ is what we called $\operatorname{Im}(F)$ in Chapter 89.

90.1.2 Example If $F : \{1,2,3\} \to \{2,4,5,6\}$ is defined as in 89.1.5 by $F(1) = 4$ and $F(2) = F(3) = 5$, then $F(\{1,2\}) = \{4,5\}$ and $F(\emptyset) = \emptyset$. Thus the image of $\{1,2\}$ under F is $\{4,5\}$.

90.1.3 Warning The image function is not usually distinguished from F in notation. A few texts use $F_* : \mathcal{P}A \to \mathcal{P}B$, and so would write $F(x)$ for $x \in A$ but $F_*(C)$ for a subset $C \subseteq A$. In this text, as in almost all mathematics texts, we simply write $F(C)$. Context usually disambiguates this notation (but there are exceptions!).

90.1.4 Exercise Describe a function where our notation $F(C)$ is ambiguous.

90.1.5 Exercise Let F be defined as in Example 90.1.2. What are $F(\{2,3\})$ and $F(\{3\})$? (Answer on page 247.)

90.1.6 Exercise Let $F : \mathrm{R} \to \mathrm{R}$ be defined by $F(x) = x^2 + 1$. What is $F((3\,..\,4))$? What is $F([-1\,..\,1])$?

90.1.7 Exercise Let F be defined as in Example 90.1.2. How many ordered pairs are in the graph of the image function of F?

91. Inverse images

91.1 Definition: Inverse image
Let $F : A \to B$ be a function. For any subset $C \subseteq B$, the set

$$\{a \in A \mid F(a) \in C\}$$

is called the **inverse image** of C under F, also written $F^{-1}(C)$.

91.1.1 Example Let $F : \{1,2,3\} \to \{2,4,5,6\}$ be defined (as in Example 89.1.5) by $F(1) = 4$ and $F(2) = F(3) = 5$. Then $F^{-1}(\{4,6\}) = \{1\}$, $F^{-1}(\{5\}) = \{2,3\}$, and $F^{-1}(\{2,6\}) = \emptyset$.

91.1.2 Example For the function $F : \mathrm{R} \to \mathrm{R}$ defined by $F(x) = x^2 + 1$,

$$F^{-1}([2\,..\,3]) = [1\,..\,\sqrt{2}] \cup [-\sqrt{2}\,..\,-1]$$

and

$$F^{-1}([0\,..\,1]) = \{0\}$$

91.1.3 Inverse image as function Like the image function, this inverse image function can also be defined as a function $F^{-1} : \mathcal{P}B \to \mathcal{P}A$ (note the reversal), where

$$F^{-1}(D) = \{x \in A \mid F(x) \in D\}$$

for any $D \subseteq B$. F^{-1} is sometimes denoted F^*.

91.1.4 Usage It is quite common to write $F^{-1}(x)$ instead of $F^{-1}(\{x\})$.

91.1.5 Example For the function of Example 91.1.2, $F^{-1}(3) = \{-\sqrt{2}, \sqrt{2})\}$.

91.1.6 Exercise Let $F : \mathrm{R} \to \mathrm{R}$ be defined by $F(x) = x^2 + 1$. What is $F^{-1}(\{1, 2\})$? What is $F^{-1}((1\,..\,2))$?

91.1.7 Exercise For any function $F : A \to B$, what is $F^{-1}(\emptyset)$? What is $F^{-1}(B)$?

92. Surjectivity

92.1 Definition: surjective
Let $F : A \to B$ be a function. F is said to be **surjective** if and only if $\mathrm{Im}(F) = B$.

92.1.1 Fact $F : A \to B$ is surjective if and only if for every element element $b \in B$ there is an element $a \in A$ for which $F(a) = b$.

92.1.2 Usage If F is surjective, it is said to be a **surjection** or to be **onto**.

92.1.3 Warning Whether a function is surjective or not depends on the codomain you specify for it.

92.1.4 Example For the two functions $S : \mathrm{R} \to \mathrm{R}$ and $T : \mathrm{R} \to \mathrm{R}^+$ of 39.7.3, with $S(x) = T(x) = x^2$, S is not surjective but T is. To say that T is surjective is to say that every nonnegative real number has a square root. Authors who do not normally specify codomains have to say, "T is surjective onto R^+."

92.1.5 Example A function $F : \mathrm{R} \to \mathrm{R}$ is surjective if every horizontal line crosses its graph.

92.1.6 Exercise How do you prove that a function $F: A \to B$ is not surjective?

92.1.7 Exercise Let α be a relation on A.

a) Show that if α is reflexive, then the coordinate functions $p_1^\alpha : \alpha \to A$ and $p_2^\alpha : \alpha \to A$ are surjective.
b) Show that the converse of (a) need not be true.

92.1.8 Exercise (hard) Show that there for any set S, no function from S to $\mathcal{P}S$ is surjective. Do not assume S is finite.

Extended hint: If $F: S \to \mathcal{P}S$ is a function, consider the subset

$$\{x \mid x \text{ is not an element of } F(x)\}$$

No argument that says anything like "the powerset of a set has more elements than the set" can possibly work for this problem, and therefore such arguments will not be given even part credit. The reason is that we have developed none of the theory of what it means to talk about the number of elements of an infinite set, and in any case this problem is a basic theorem of that theory.

Let's be more specific: One such invalid argument is that the function that takes x to $\{x\}$ is an injective function from S to $\mathcal{P}S$, and it clearly leaves out the empty set (and many others) so $\mathcal{P}S$ has "more elements" than S. This is an invalid argument. Consider the function from N to N that takes n to $42n$. This is injective and leaves out lots of integers, so does N have more elements than itself?? (In any case you can come up with other functions from N to N that don't leave out elements.)

93. Injectivity

93.1 Definition: injective
$F: A \to B$ is **injective** if and only if different inputs give different outputs, in other words if $a \neq a' \Rightarrow F(a) \neq F(a')$ for all $a, a' \in A$.

93.1.1 Fact To say $F: A \to B$ is injective is equivalent to saying that $F(a) = F(a') \Rightarrow a = a'$ for all $a, a' \in A$ (the contrapositive of the definition).

93.1.2 Usage An injective function is called an **injection** or is said to be **one to one**.

93.1.3 Example The squaring function $S: \mathrm{R} \to \mathrm{R}$ is not injective since $S(x) = S(-x)$ for every $x \in \mathrm{R}$. The cubing function $x \mapsto x^3 : \mathrm{R} \to \mathrm{R}$ of course is injective, and so is any identity function or inclusion function on any set.

93.1.4 Exercise In this problem, $A = \{1,2,3,4\}$ and $B = \{2,3,4\}$. For each of these functions, state whether the function is injective, whether it is surjective, and give its image explicitly.

a) $F: A \to B$, $\Gamma(F) = \Big\{\langle 1,4\rangle, \langle 2,4\rangle, \langle 3,2\rangle, \langle 4,3\rangle\Big\}$.
b) $F: A \to B$, $\Gamma(F) = \Big\{\langle 1,3\rangle, \langle 2,2\rangle, \langle 3,2\rangle, \langle 4,3\rangle\Big\}$.
c) id_A.

d) The inclusion of B into A.
e) The inclusion of B into Z.
f) $C_3 : A \to B$ (the constant function).
g) $\chi_B^A : A \to \{\text{TRUE}, \text{FALSE}\}$.
h) $p_1 : A \times B \to A$.
i) $+ : B \times B \to \mathrm{Z}$.
j) The predicate "n is even" regarded as a characteristic function with domain A.

(Answer on page 247.)

93.1.5 Exercise Same instructions as for Exercise 93.1.4

a) $x \mapsto 3x - 4 : \mathrm{R} \to \mathrm{R}$.
b) $x \mapsto x^3 : \mathrm{R} \to \mathrm{R}$.
c) $F = \lambda x.(x^2 + 1) : \mathrm{R} \to \mathrm{R}$.
d) $x \mapsto 2 - x^2 : \mathrm{R} \to \mathrm{R}$.

(Answer on page 248.)

93.1.6 Exercise Let $F : A \to B$ be a function of the type indicated. Give a precise description of all the sets A and B for which F is injective, and a precise description of all the sets A and B for which F is surjective.

a) An identity function.
b) An inclusion function.
c) A constant function.
d) An empty function.
e) A coordinate function.

93.1.7 Exercise How do you prove that a function $F : A \to B$ is not injective? (Answer on page 248.)

93.1.8 Exercise Prove that the function $\langle m, n \rangle \mapsto 2^m 3^n - 1 : \mathrm{N} \times \mathrm{N} \to \mathrm{N}$ is injective.

93.1.9 Exercise Give an example of a function $F : \mathrm{R} \to \mathrm{R}$ with the property that F is not injective but $F|\mathrm{N}$ is injective.

93.1.10 Exercise (calculus)

a) Show that if a cubic polynomial function $x \mapsto ax^3 + bx^2 + cx + d$ is not injective, then $b^2 - 3ac \geq 0$. (The "3" is not a misprint.)
b) Show that the converse of the statement in (a) is not true.
c) Think of a more sophisticated condition involving a, b, c and d that is true *if and only if* the function is injective.

94. Bijectivity

94.1 Definition: bijective
A function which is both injective and surjective is **bijective**.

94.1.1 Remark A bijection $F : A \to B$ matches up the elements of A and B — each element of A corresponds to exactly one element of B and each element of B corresponds to exactly one element of A.

94.1.2 Usage A bijective function is called a **bijection** and is said to be a **one to one correspondence**.

94.1.3 Example For any set A, $\mathrm{id}_A : A \to A$ is bijective. Another example is the function $F : \{1,2,3\} \to \{2,3,4\}$ defined by $F(1) = 3$, $F(2) = 2$, $F(3) = 4$.

94.1.4 Exercise Show that the function $G : \mathrm{N} \to \mathrm{Z}$ defined by

$$G(n) = \begin{cases} -\frac{n}{2} & n \text{ even} \\ \frac{n+1}{2} & n \text{ odd} \end{cases}$$

is a bijection.

94.1.5 Exercise Show how to construct bijections β as follows for any sets A, B and C.

a) $\beta : A \times B \to B \times A$.
b) $\beta : (A \times B) \times C \to A \times (B \times C)$.
c) $\beta : \{1\} \times A \to A$.

94.1.6 Exercise Let α be a relation from A to B.

a) Prove that α is functional if and only if the first coordinate function p_1^α is injective. (See Section 51.4.)
b) Prove that α is the graph of a function from A to B if and only if the first coordinate function is bijective.

94.1.7 Exercise Give an example of a function $F : \mathrm{R} \to \mathrm{R}^{++}$ for which F is bijective. (R^{++} is the set of positive real numbers.)

94.1.8 Exercise (hard) Give an example of a function $F : \mathrm{R} \to \mathrm{R}^{+}$ for which F is bijective. (R^{+} is the set of nonnegative real numbers.)

94.1.9 Exercise (hard) Let $F : A \to B$ be a function. Prove that the restriction to $\Gamma(F)$ of the first coordinate function from $A \times B$ is a bijection.

94.1.10 Exercise (hard) Prove that a subset C of $A \times B$ is the graph of a function from A to B if and only if the restriction to C of the first coordinate function is a bijection.

94.1.11 Exercise (hard) Let $\beta : \mathrm{Rel}(A,B) \to (\mathcal{P}B)^A$ be the function which takes a relation α to the function $\alpha^* : A \to \mathcal{P}B$ defined by $\alpha^*(a) = \{b \in B \mid a\alpha b\}$ (see Definition 53.2). Show that β is a bijection. (This function is studied further in Problem 100.1.8, page 145, and in Problem 101.5.10, page 150.)

94.1.12 Exercise (hard) Let A, B and C be sets. In this exercise we define a particular function β from the set $B^A \times C^A$ to the set $(B \times C)^A$, so that β as input a *pair* of functions $\langle f, g\rangle$, with $f : A \to B$ and $g : A \to C$, and outputs a function $\beta(f,g)$ from A to $B \times C$. Here is the definition of β: for all $a \in A$,

$$\beta(f,g)(a) = \langle f(a), g(a)\rangle$$

Prove that β is a bijection.

95. Permutations

95.1 Definition: permutation
A **permutation** of a set A is a bijection $\beta : A \to A$.

95.1.1 Example The fact just noted that id_A is a bijection says that id_A is a (not very interesting) permutation of A for any set A.

95.1.2 Example The function $F : \{1,2,3\} \to \{1,2,3\}$ that takes 1 to 2, 2 to 1 and 3 to 3 is a permutation of $\{1,2,3\}$.

95.1.3 Usage Many books define a permutation to be a list exhibiting a rearrangement of the set $\{1,2,\ldots,n\}$ for some n. If the ith entry in the list is a_i that indicates that the permutation takes i to a_i.

95.1.4 Example The permutation of Example 95.1.2 would be given in the list notation as $\langle 2,1,3\rangle$.

95.1.5 Worked Exercise List all the permutations of $\{1,2,3,4\}$ that take 1 to 3 and 2 to 4.
Answer $\langle 3,4,1,2\rangle$ and $\langle 3,4,2,1\rangle$,

95.1.6 Exercise List all six permutations of $\{1,2,3\}$.

96. Restrictions and extensions

96.1 Definition: restriction
Suppose $F : A \to B$ is a function and $A' \subseteq A$. The **restriction** of F to A' is a function denoted $F|A' : A' \to B$, whose value $(F|A')(a)$ for $a \in A'$ is $F(a)$.

96.1.1 Remark Note that the codomain of the restriction is the codomain of the function.

96.1.2 Example Let $F:\{1,2,3\}\to\{2,4,5,6\}$ be defined by $F(1)=4$ and $F(2)=F(3)=5$, as before. Then F restricted to $\{2,3\}$ has graph $\bigl\{\langle 2,5\rangle,\langle 3,5\rangle\bigr\}$ and $F|\{1,3\}$ has graph $\bigl\{\langle 1,4\rangle,\langle 3,5\rangle\bigr\}$. Observe that $F|\{2,3\}$ is a constant function and $F|\{1,3\}$ is injective, whereas F is neither constant nor injective.

96.2 Definition: extension of a function
Let $F:A\to B$ and let C be a set *containing* A as a subset. Any function $G:C\to B$ for which $G|A=F$ is called an **extension** of F to C.

96.2.1 Remark Note that both "restriction" and "extension" have to do with the *domain.*

96.2.2 Example Let $F:\{1,2,3\}\to\{2,4,5,6\}$ be defined by $F(1)=4$ and $F(2)=F(3)=5$, as before. Then F has four extensions F_1, F_2, F_3, and F_4, to $\{1,2,3,7\}$, defined by $F_1(7)=2, F_2(7)=4$ $, F_3(7)=5$ and $F_4(7)=6$. (Of course in all cases $F_i(n)$ is the same as $F(n)$ for $n=1,2,3$).

96.2.3 Example The absolute value function $\mathrm{ABS}:\mathrm{R}\to\mathrm{R}$ is an extension of the inclusion of R^+ into R, and id_R is a *different* extension of the same function.

96.2.4 Usage The meaning just given of "extension" is a different usage of the word from the meaning used in Definition 18.1 of the set of data items for which a predicate is true.

You may wonder how the word "extension" got two such different meanings. The answer is that the concept of extension of a predicate was named by logicians, whereas the concept of extension of a function was named by mathematicians.

96.2.5 Exercise For each of these functions from R to R, state whether the function is injective or surjective, and state whether its restriction to $\mathrm{R}^+=\{r\in\mathrm{R}\mid r\geq 0\}$ is injective or surjective.

a) $x\mapsto x^2$.
b) $\lambda x.x+1$.
c) $\lambda x.1-x$.

(Answer on page 248.)

97. Tuples as functions

Let n be a positive integer, and let

$$\mathbf{n}=\{1,2,\ldots,n\}$$

An n-tuple

$$\mathbf{a}=\langle a_1,\ldots,a_n\rangle$$

in A^n associates to each element i of $\mathbf{n}$ an element a_i of A. This determines a function $i\mapsto a_i$ with domain $\mathbf{n}$ and codomain A. Conversely, any such function determines an n-tuple in A^n by setting its coordinate at i to be its value at i.

When $\mathbf{a} \in A_1 \times A_2 \times \cdots \times A_n$, so that different components are in different sets, this way of looking at n-tuples is more complicated. Every coordinate a_i is an element of the union $C = A_1 \cup A_2 \cup \cdots \cup A_n$, so that $\mathbf{a}$ can be thought of as a function from $\mathbf{n} \to C$. In this case, however, not every such function is a tuple in $A_1 \times A_2 \times \cdots \times A_n$: we must impose the additional requirement that $a_i \in A_i$.

We sum all this up in an alternative definition of tuple:

97.1 Definition: tuple as function

A tuple in $\prod_{i=1}^{n} A_i$ is a function

$$\mathbf{a} : \mathbf{n} \to A_1 \cup A_2 \cup \cdots \cup A_n$$

with the property that for each i, $a(i) \in A_i$.

97.1.1 Example The tuple $\langle 2,1,3\rangle$ is the function $1 \mapsto 2$, $2 \mapsto 1$, $3 \mapsto 3$ (compare Section 95.1.3).

97.1.2 Example The tuple $\langle 5,5,5,5\rangle$ is the constant function $C_5 : \{1,2,3,4\} \to \mathrm{Z}$.

97.1.3 Exercise Write the domain and the graph of these tuples regarded as functions on the index set.

a) $\langle 2,5,-1,3,6\rangle$.
b) $\langle \pi,5,\pi-1,\sqrt{2}\rangle$.
c) $\langle\langle 3,5\rangle,\langle 8,-7\rangle,\langle 5,5\rangle\rangle$.

(Answer on page 248.)

97.1.4 Example A simple database might have records each of which consists of the name of a student, the student's student number, and the number of classes the student takes. Such a record would be a triple $\langle w,x,n\rangle$, where w is an element of the set A^* of strings of English letters and spaces (this notation is introduced formally in Definitions 109.2 and 110.1), x is an element of the set D^* of strings of decimal digits, and $n \in \mathrm{N}$. This triple corresponds to a function $F : \{1,2,3\} \to A^* \times D^* \times \mathrm{N}$ with the property that $F(1) \in A^*$, $F(2) \in D^*$ and $F(3) \in \mathrm{N}$.

Modeling detabases this way is the principle behind relational database theory.

97.1.5 Remark In the case that all the A_i are the same, so that $\mathbf{a} \in A^n$, we now have the situation that $A^{\mathbf{n}}$ (the set of functions from $\mathbf{n}$ to A, where $\mathbf{n} = \{1,2,\ldots,n\}$) and A^n (the set of n-tuples in A) are essentially the same thing. That is the origin of the notation B^A.

97.2 Tuples with other index sets

The discussion above suggests that by regarding a tuple as a function set, we can use *any* set as index set.

97.2.1 Example In computer science it is often convenient to start a list at 0 instead of at 1, giving a tuple $\langle a_0,a_1,\ldots,a_n\rangle$. This is then a tuple indexed by the set $\{0,1,\ldots,n\}$ for some n (so it has $n+1$ entries!).

97.2.2 Example An infinite sequence of integers is indexed by N^+, so it is an element of $\mathrm{Z}^{\mathrm{N}^+}$.

97.2.3 Example This is another look at Example 97.1.4. The point of view that a triple $\langle \text{Jones}, 1235551212, 4\rangle$ is a function with domain $\{1,2,3\}$ has an arbitrary nature: it doesn't matter that the name is first, the student number second and the number of classes third. What matters is that Jones is the name, 1235551212 is the student number and 4 is the number of classes. Thus it would be conceptually better to regard the triple as a function whose domain is the set $\{\text{Name}, \text{StudentNumber}, \text{NumberOfClasses}\}$, with the property that $f(\text{Name}) \in A^*$, $F(\text{StudentNumber}) \in D^*$ and $F(\text{NumberOfClasses}) \in \mathrm{N}$. This eliminates the spurious ordering of data imposed by using the set $\{1,2,3\}$ as domain.

In this context, the elements of a set such as

$$\{\text{Name}, \text{StudentNumber}, \text{NumberOfClasses}\}$$

are called the **field names** of the database.

97.3 Definition: function as tuple
A function $T : S \to A$ is is also called an S**-tuple** or a **family of elements of** A **indexed by** S.

97.3.1 Exercise Write each of these functions as tuples.

a) $F : \{1,2,3,4,5\} \to \mathrm{R}$, $\Gamma(F) = \Big\{\langle 2,5\rangle, \langle 1,5\rangle, \langle 3,3\rangle, \langle 5,-1\rangle, \langle 4,17\rangle\Big\}$.
b) $F : \{1,2,3,4,5\} \to \mathrm{R}$, $F(n) = (n+1)\pi$.
c) $x \mapsto x^2 : \{1,2,3,4,5,6\} \to \mathrm{R}$.

(Answer on page 248.)

98. Functional composition

98.1 Definition: composition of functions
If $F : A \to B$ and $G : B \to C$, then $G \circ F : A \to C$ is the function defined for all $a \in A$ by $(G \circ F)(a) = G(xxF(a))$. $G \circ F$ is the **composite** of F and G, and the operation "$\circ$" is called **functional composition**.

98.1.1 How to think about composition The composite of two functions is obtained by feeding the output of one into the input of the other. Suppose $F : A \to B$ and $G : B \to C$ are functions. If a is any element of A, then $F(a)$ is an element of B, and so $G(F(a))$ is an element of C. Thus applying F, then G, gives a function from A to C, and that is the composite $G \circ F : A \to C$.

98.1.2 Remarks

a) You may be familiar with the idea of functional composition in connection with the chain rule in calculus.

b) Our definition of $G \circ F$ requires that the codomain of F be the domain of G. Actually, the expression $G(F(a))$ makes sense even if $\operatorname{cod} F$ is only included in $\operatorname{dom} G$, and many authors allow the composite $G \circ F$ to be formed in that case, too. We will not follow that practice here.

98.1.3 Example If $A = \{1,2,3,4\}$, $B = \{3,4,5,6\}$, $C = \{1,3,5,7\}$, F is defined by $F(1) = F(3) = 5$, $F(2) = 3$ and $F(4) = 6$, and G is defined by $G(3) = 7$, $G(4) = 5$, $G(5) = 1$ and $G(6) = 3$, then $G \circ F$ takes $1 \mapsto 1$, $2 \mapsto 7$, $3 \mapsto 1$ and $4 \mapsto 3$.

98.1.4 Warning Applying the function $G \circ F$ to an element of A involves applying F, then G — in other words, the notation "$G \circ F$" is read from *right* to *left*.

Functional composition is associative when it is defined:

98.2 Theorem
If $F : A \to B$, $G : B \to C$ and $H : C \to D$ are all functions, then $H \circ (G \circ F)$ and $(H \circ G) \circ F$ are both defined and

$$H \circ (G \circ F) = (H \circ G) \circ F$$

Proof Let $a \in A$. Then by applying Definition 98.1 twice,

$$\Big(H \circ (G \circ F)\Big)(a) = H\Big((G \circ F)(a)\Big) = H(G(F(a)))$$

and similarly

$$\Big((H \circ G) \circ F\Big)(a) = (H \circ G)(F(a)) = H(G(F(a)))$$

so $H \circ (G \circ F) = (H \circ G) \circ F$.

98.2.1 Warning Commutativity is a different story. If $F : A \to B$ and $G : B \to C$, $G \circ F$ is defined, but $F \circ G$ is not defined unless $A = C$. If $A = C$, then $G \circ F : A \to C$ and $F \circ G : C \to A$, so normally $F \circ G \neq G \circ F$. Commutativity may fail even when $A = B = C$: For example, let $S = x \mapsto x^2 : \mathrm{R} \to \mathrm{R}$ and $T = x \mapsto x + 1 : \mathrm{R} \to \mathrm{R}$. Then for any $x \in \mathrm{R}$, $S(T(x)) = (x+1)^2$ and $T(S(x)) = x^2 + 1$, so $S \circ T \neq T \circ S$.

Pondering the following examples of functional composition may be helpful in understanding the idea of composition.

98.2.2 Example Let $\mathrm{SQ} = x \mapsto x^2 : \mathrm{R} \to \mathrm{R}^+$ and $\mathrm{SQRT} = x \mapsto \sqrt{x} : \mathrm{R}^+ \to \mathrm{R}$. ($\sqrt{x}$ denotes the *nonnegative* square root of x.) Let ABS denote the absolute value function from R to R. Then the following are true.

(i) $\mathrm{SQRT} \circ \mathrm{SQ} = \mathrm{ABS} : \mathrm{R} \to \mathrm{R}$.
(ii) $(\mathrm{SQRT} \circ \mathrm{SQ})|\mathrm{R}^+ = \mathrm{id}_{\mathrm{R}^+}$.
(iii) $\mathrm{SQ} \circ \mathrm{SQRT} = \mathrm{id}_{\mathrm{R}^+}$.

98.2.3 Example If $F : A \to B$ is any function, then

(i) $F \circ \mathrm{id}_A = F$ and
(ii) $\mathrm{id}_B \circ F = F$.

This is analogous to the property that an identity element for a binary operation has (see 50.1), but in fact composition of functions is not a binary operation since it is not defined for all pairs of functions.

98.2.4 Example If $A \subseteq B$ and $B \subseteq C$, and $i : A \to B$ and $j : B \to C$ are the corresponding inclusion functions, then $j \circ i$ is the inclusion of A into C.

98.2.5 Example If $F : A \to B$ and $C \subseteq A$ with inclusion map $i : C \to A$, then $F|C = F \circ i$. In other words, *restriction is composition with inclusion.*

98.2.6 Exercise Describe explicitly (give the domain and codomain and either the graph or a formula) the composite $G \circ F$ if

a) $F : \{1,2,3,4\} \to \{3,4,5,6\}$, with $1 \mapsto 3$, $2 \mapsto 4$, $3 \mapsto 6$, and $4 \mapsto 5$, and $G : \{3,4,5,6\} \to \{1,3,5,7,9\}$ with $3 \mapsto 1$, $4 \mapsto 7$, $5 \mapsto 7$ and $6 \mapsto 3$.
b) $F : x \mapsto x^3 : \mathrm{R} \to \mathrm{R}$, $G : x \mapsto 2x : \mathrm{R} \to \mathrm{R}$.
c) $F : x \mapsto 2x : \mathrm{R} \to \mathrm{R}$. $G : x \mapsto x^3 : \mathrm{R} \to \mathrm{R}$,
d) $F = \text{inclusion} : \mathrm{N} \to \mathrm{R}$, $G : x \mapsto (x/2) : \mathrm{R} \to \mathrm{R}$.
e) $F = p_1 : \mathrm{R} \times \mathrm{R} \to \mathrm{R}$, $G : x \mapsto (3, x) : \mathrm{R} \to \mathrm{R} \times \mathrm{R}$.

(Answer on page 248.)

98.2.7 Exercise Let $F : A \to B$, $G : B \to C$. Show the following facts:

a) If F and G are both injective, so is $G \circ F$.
b) If F and G are both surjective, so is $G \circ F$.
c) If F and G are both bijective, so is $G \circ F$.
d) If $G \circ F$ is surjective, so is G.
e) If $G \circ F$ is injective, so is F.

98.2.8 Exercise Give examples of functions F and G for which $G \circ F$ is defined and

a) F is injective but $G \circ F$ is not.
b) G is surjective but $G \circ F$ is not.
c) $G \circ F$ is injective but G is not.
d) $G \circ F$ is surjective but F is not.

98.2.9 Exercise (hard) Let A, B and C be sets.

a) Prove that if $F : A \to B$ is a function and C is nonempty, then $G \mapsto F \circ G : A^C \to A^C$ is a function which is injective if and only if F is injective, and surjective if and only if F is surjective.
b) Prove that if $H : B \to C$ is a function and A has more than one element, then $G \mapsto (G \circ H) : A^C \to A^B$ is a function which is injective if and only if H is surjective, and surjective if and only if H is injective.

99. Idempotent functions

99.1 Definition: Idempotent function
A function $F:A\to A$ is **idempotent** if $F\circ F=F$.

99.1.1 How to think about idempotent functions F is idempotent if doing F twice is the same as doing it once: If you do F, then do it again, the second time nothing happens.

99.1.2 Example The function $\langle x,y\rangle \mapsto \langle x,0\rangle : \mathrm{R}\times\mathrm{R}\to\mathrm{R}\times\mathrm{R}$ is idempotent. Note the close connection between this function and the first coordinate function $p_1:\mathrm{R}\times\mathrm{R}\to\mathrm{R}$.

99.1.3 Example Let S be a set of files that contains a sorted version of every file in the set. Then "sort" is a function that takes each file in the set to a possibly different file. Sorting a file that is already sorted does not change it (that is true of many sorting functions found on computers, but not all). Thus sorting and then sorting again is the same as sorting once, so sorting is idempotent.

99.1.4 Usage Following Example 99.1.2, the word "projection" is used in some branches of mathematics to mean "idempotent function". In other brances, "projection" means "coordinate function".)

99.1.5 Exercise Let $A=\{1,2,3\}$. Give an example of an idempotent function $F:A\to A$ that is not id_A. (Answer on page 248.)

99.1.6 Exercise Show that if $F:A\to A$ is injective and idempotent, then $F=\mathrm{id}_A$.

99.2 Definition: Fixed point
Let $F:A\to A$ be any function. An element $x\in A$ is a **fixed point** of F if $F(x)=x$.

This is the fundamental theorem on idempotent functions:

99.3 Theorem
A function $F:A\to A$ is idempotent if and only if every element of $\mathrm{Im}\,F$ is a fixed point of F.

99.3.1 Exercise Prove Theorem 99.3.

99.3.2 Exercise Use Theorem 99.3 to show that if $F:A\to A$ is surjective and idempotent, then $F=\mathrm{id}_A$.

100. Commutative diagrams

$F : A \to B$ and $G : B \to C$ can be illustrated by this diagram:

$$\begin{tikzcd} A \arrow[r, "F"] \arrow[dr, "H"'] & B \arrow[d, "G"] \\ & C \end{tikzcd} \tag{100.1}$$

If the two ways of evaluating functions along paths from A to C in this diagram give the same result, then, by definition of composition, $H = G \circ F$.

> **100.1 Definition: commutative diagram**
> A diagram with the property that any two paths between the same two nodes compose to give the same function is called a **commutative diagram**.

100.1.1 Example To say that the following diagram commutes is to say that $H \circ F = K \circ G$; in other words, that for all $a \in A$, $H\Big(F(a)\Big) = K\Big(G(a)\Big)$.

$$\begin{tikzcd} A \arrow[r, "F"] \arrow[d, "G"'] & B \arrow[d, "H"] \\ C \arrow[r, "K"'] & D \end{tikzcd} \tag{100.2}$$

100.1.2 Remarks

a) Commutative diagrams exhibit more of the data involved in a statement such as "$H \circ F = K \circ G$" than the statement itself shows (in particular it shows what the domains and codomains are), and moreover they exhibit it in a geometric way which is easily grasped.

b) **Warning:** The concept of commutativity of diagrams and the idea of the commutative law for operations such as addition are distinct and not very closely related ideas, in spite of their similar names.

100.1.3 Example Example 98.2.3 on page 141 says that for any function F, this diagram commutes:

$$\begin{tikzcd} A \arrow[r, "F"] \arrow[d, "\mathrm{id}_A"'] \arrow[dr, "F"] & B \arrow[d, "\mathrm{id}_B"] \\ A \arrow[r, "F"'] & B \end{tikzcd} \tag{100.3}$$

100.1.4 Example Theorem 98.2 says that if both triangles in this diagram commute,

$$\begin{tikzcd} A \arrow[r, "F"] \arrow[d] & B \arrow[d] \arrow[dl, "G"] \\ C \arrow[r, "H"'] & D \end{tikzcd} \tag{100.4}$$

then the whole diagram commutes. Thus the associative law for for functional composition becomes a statement that commutative triangles can be pasted together in a certain way.

100.1.5 Exercise Draw commutative diagrams expressing these facts:

a) The square of the square root of a nonnegative real number is the number itself.
b) The positive square root of the square of a real number is the absolute value of the number.

(Answer on page 248.)

100.1.6 Exercise Draw commutative diagrams which express each of the following facts. No one arrow should be labeled with a composite of two functions — draw a separate arrow for each function.

a) Addition, as a binary operation on Z, is commutative.
b) Addition as in (a) is associative.

100.1.7 Exercise Prove that if $F : A \to A$ is an idempotent function, then there is a set B and functions $G : A \to B$ and $H : B \to A$ such that both the following diagrams commute:

$$\begin{tikzcd} A \arrow[rr, "F"] \arrow[dr, "G"'] & & A \\ & B \arrow[ur, "H"'] & \end{tikzcd} \qquad \begin{tikzcd} B \arrow[rr, "\mathrm{id}_B"] \arrow[dr, "H"'] & & B \\ & A \arrow[ur, "G"'] & \end{tikzcd}$$

100.1.8 Exercise (hard) Let β be defined as in Problem 94.1.11, page 137. Let $F : A' \to A$ and $G : B' \to B$. Let

$$H : \mathrm{Rel}(A, B) \to \mathrm{Rel}(A', B')$$

be the function which takes α to the relation α' defined by $a' \, \alpha' \, b' \Leftrightarrow F(a') \, \alpha \, G(b')$. Let

$$K : (\mathcal{P}B)^A \to (\mathcal{P}B')^{A'}$$

be the function which takes $r : A \to \mathcal{P}B$ to the function $r' : A' \to \mathcal{P}B'$ defined by

$$r'(a') = G^{-1}(r(F(a')))$$

Show that the following diagram commutes.

$$\begin{tikzcd}
\mathrm{Rel}(A,B) \arrow[r, "\beta"] \arrow[d, "H"'] & (\mathcal{P}B)^A \arrow[d, "K"] \\
\mathrm{Rel}(A',B') \arrow[r, "\beta"'] & (\mathcal{P}B')^{A'}
\end{tikzcd} \tag{100.5}$$

101. Inverses of functions

The number $1/2$ is the "multiplicative inverse" of the number 2 because their product is 1. A similar relationship can hold between functions, but because functional composition is not normally commutative, one has to specify which way the composite is taken.

101.1 Definition: inverse of a function
If $F : A \to B$ and $G : B \to A$, then G is a **left inverse** to F, and F is a **right inverse** to G, if

$$G \circ F = \mathrm{id}_A \tag{101.1}$$

If G is both a left and a right inverse to F, in other words if both Equation (101.1) and

$$F \circ G = \mathrm{id}_B \tag{101.2}$$

hold, then G is an **inverse** to F.

101.1.1 Usage A function that has an inverse is said to be **invertible**.

101.1.2 Fact It follows from the definition that if G is an inverse to F, then F is an inverse to G.

101.1.3 Fact The definition of inverse function can be expressed in other ways equivalent to Definition 101.1.

a) G is the inverse of F if and only if for all $a \in A$, $G(F(a)) = a$ and for all $b \in B$, $F(G(b)) = b$. (*Both* equations must hold.)
b) G is the inverse of F if and only if the following diagrams commute:

$$\begin{tikzcd}
A \arrow[r, "F"] \arrow[dr, "\mathrm{id}_A"'] & B \arrow[d, "G"] \\
& A
\end{tikzcd}
\qquad
\begin{tikzcd}
B \arrow[r, "G"] \arrow[dr, "\mathrm{id}_B"'] & A \arrow[d, "F"] \\
& B
\end{tikzcd} \tag{101.3}$$

101.1.4 Example Let $F:\{1,3,5\} \to \{2,3,4\}$ be the function that takes 1 to 3, 3 to 4 and 5 to 2. Then the function $G:\{2,3,4\} \to \{1,3,5\}$ that takes 2 to 5, 3 to 1 and 4 to 3 is the inverse of F. (And F is the inverse of G.)

101.1.5 Example Example 98.2.2(3) above says that the squaring function is a left inverse to the square root function: squaring the positive square root gives you what you started with. It is not the inverse, however: taking the square root of the square won't give you the number you started with if it is negative. On the other hand, the cubing function *is* the inverse of the cube root function.

A function can have more than one left inverse (Problem 101.2.6) but not more than one inverse:

101.2 Theorem: Uniqueness Theorem for Inverses
If $F:A \to B$ has an inverse $G:B \to A$, then G is the only inverse to F.

Proof The proof uses the definition of inverse, Theorem 98.2 and Example 98.2.3: If $H:B \to A$ is another inverse of F, then

$$H = H \circ \mathrm{id}_B = H \circ (F \circ G) = (H \circ F) \circ G = \mathrm{id}_A \circ G = G$$

101.2.1 Usage The fact that if a function has an inverse, it has only one, means that we can give the inverse a name: The inverse of F, if it exists, is denoted F^{-1}.

101.2.2 Remark The uniqueness theorem also means we have a rule of inference: Given $F:A \to B$ and $G:B \to A$,

$$G \circ F = \mathrm{id}_A,\ F \circ G = \mathrm{id}_B \ \models\ G = F^{-1} \tag{101.4}$$

101.2.3 Exercise Which of the following functions have inverses? If it has one, give the inverse (by describing its graph or by a formula).
a) $F:\{1,2,3,4\} \to \{3,4,5,6\}$, with $1 \mapsto 3,\ 2 \mapsto 4,\ 3 \mapsto 6$, and $4 \mapsto 5$.
b) $G:\{1,2,3,4\} \to \{3,4,5,6,7\}$, with $1 \mapsto 3,\ 2 \mapsto 4,\ 3 \mapsto 6$, and $4 \mapsto 5$.
c) $H:\{1,2,3,4\} \to \{3,4,5,6\}$ with $1 \mapsto 3,\ 2 \mapsto 5,\ 3 \mapsto 6$, and $4 \mapsto 5$.
d) $n \mapsto 2n : \mathrm{N} \to \mathrm{N}$.
e) $n \mapsto n+1 : \mathrm{N} \to \mathrm{N}$.
f) $n \mapsto n+1 : \mathrm{Z} \to \mathrm{Z}$.
g) $n \mapsto (1/n) : \mathrm{N} - \{0\} \to \mathrm{R}$.
h) $K:\{1,2,3,4,5\} \to \{1,2,3\}$ with $K(n) = \mathrm{floor}((n+1)/2)$.
(Answer on page 248.)

101.2.4 Exercise Which of the functions in Exercise 101.2.3 have (a) left inverses, (b) right inverses? (Answer on page 248.)

101.2.5 Exercise Show that if a function G has an inverse F, then it has only one left inverse and that is F. (Answer on page 248.)

101.2.6 Exercise Let $I : \mathrm{R}^+ \to \mathrm{R}$ denote the inclusion function. Show that I has infinitely many left inverses.

The inverse of the composite of two functions which have inverses is the composite of the inverses, only in the reverse order:

101.3 Theorem: The Shoe-Sock Theorem
If $F : A \to B$ and $G : B \to C$ both have inverses, then

$$(G \circ F)^{-1} = F^{-1} \circ G^{-1}$$

101.3.1 Remark The name comes from the fact that the inverse of putting on your socks and then putting on your shoes is taking off your *shoes* and then taking off your *socks*.

Proof To prove the Shoe-Sock Theorem, we will prove that

$$(F^{-1} \circ G^{-1}) \circ (G \circ F) = \mathrm{id}_A \tag{101.5}$$

and

$$(G \circ F) \circ (F^{-1} \circ G^{-1}) = \mathrm{id}_C \tag{101.6}$$

and then apply Rule (101.4), page 147. To prove Equation (101.5), we note that the following diagram commutes: the left and right triangles are the diagrams in Figure (101.3), and the middle triangle is the left triangle in Figure (100.3).

$$\begin{tikzcd}
A \arrow[r, "F"] \arrow[d, "\mathrm{id}_A"'] & B \arrow[r, "G"] \arrow[dl, "F^{-1}"] \arrow[d, "\mathrm{id}_B"] & C \arrow[dl, "G^{-1}"] \\
A & B \arrow[l, "F^{-1}"]
\end{tikzcd} \tag{101.7}$$

Equation (101.6) is proved similarly.

Another fact with a similar proof (left as an exercise) is:

101.4 Theorem
If F has an inverse, then $(F^{-1})^{-1} = F$.

101.4.1 Remark Another way of saying this is that a function is the inverse of its own inverse.

101.4.2 Exercise Prove Theorem 101.4.

A final fact about inverses is the very important:

101.5 Theorem: Characterization of invertible functions
A function $F: A \to B$ has an inverse if and only if it is a bijection.

101.5.1 Remark The importance of Theorem 101.5 lies in the fact that having an inverse is defined in terms of functional composition but being a bijection is defined in terms of application of the function to an element of its domain. Any time a mathematical fact connects two such differently-described ideas it is probably useful.

Proof I will go through the proof in some detail since it ties together several of the ideas of this chapter. We have to prove an equivalence, which means two implications.

First we show that if F has an inverse then it is a bijection. Suppose F has an inverse. We must show that it is injective and surjective. To show that it is injective, suppose $F(a) = F(a')$. Then

$$a = F^{-1}(F(a)) = F^{-1}(F(a')) = a'$$

(The first and last equations follow from 101.1.3a and the middle equation from the substitution property, Theorem 39.6.) So F is injective.

To show F is surjective, let $b \in B$. We must find an element $a \in A$ for which $F(a) = b$. The element is $F^{-1}(b)$, since $F(F^{-1}(b)) = b$. Thus F is bijective.

Now we must show that if F is bijective, then it has an inverse. Suppose F is bijective. We must define a function $G: B \to A$ which is the inverse of F. Let $b \in B$. Then, since F is surjective, there is an element $a \in A$ for which $F(a) = b$. Since F is injective there is exactly one such a. Let $G(b) = a$. Since $F(a) = b$, that says that $G(F(a)) = a$, which is half of what we need to show to prove (using Definition 101.1) that $G = F^{-1}$. The other thing needed is that $F(G(b)) = b$. But by definition of G, $G(b)$ is the element a for which $F(a) = b$, so $F(G(b)) = b$. That finishes the proof.

101.5.2 Remarks

a) The second part of the proof says this: If $F(a) = b$, then $F^{-1}(b) = a$, and if $F^{-1}(b) = a$, then $F(a) = b$.
b) You might experiment with proving the contrapositives of the two implications in the preceding proof; some people find them easier to understand.

101.5.3 Exercise Write a formula for the inverse of each of these bijections.

a) $x \mapsto x^2 : \mathrm{R}^+ \to \mathrm{R}^+$.
b) $x \mapsto x - 1 : \mathrm{R} \to \mathrm{R}$.
c) $x \mapsto 2x : \mathrm{R} \to \mathrm{R}$.
d) $x \mapsto (1/x) : \mathrm{R} \to \mathrm{R}$.

(Answer on page 248.)

101.5.4 Exercise Prove that a function has a left inverse if and only if it is injective.

101.5.5 Exercise Prove that a function has a right inverse if and only if it is surjective.

101.5.6 Exercise Give a right inverse of the function $\mathrm{GCD}:\mathrm{N}^+ \times \mathrm{N}^+ \to \mathrm{N}^+$. (You are being asked to give the right inverse explicitly, not merely show it exists.)

101.5.7 Exercise Show that $\mathrm{GCD}:\mathrm{N}^+ \times \mathrm{N}^+ \to \mathrm{N}^+$ does not have a left inverse.

101.5.8 Exercise (hard) A function $F: A \to B$ is **left cancellable** if whenever $G: D \to A$ and $H: D \to A$ are functions for which $F \circ G = F \circ H$, then G must be the same as H. **Right cancellable** is defined analogously. Prove that a function is left cancellable if and only if it is injective and right cancellable if and only if it is surjective.

101.5.9 Exercise (hard) Let $F: A \to B$ be a function and suppose A has more than one element. Show that if F has exactly one left inverse then the left inverse is also a right inverse (hence F has an inverse).

101.5.10 Exercise (hard) Let A and B be sets. Let $\beta: \mathrm{Rel}(A,B) \to (\mathcal{P}B)^A$ defined in Problem 94.1.11, page 137. Let $\gamma: (\mathcal{P}B)^A \to \mathrm{Rel}(A,B)$ be the function (defined in Definition 53.3, page 76) that takes a function $F: A \to \mathcal{P}B$ to the relation α_F defined by

$$a \, \alpha_F \, b \text{ if and only if } b \in F(a)$$

Prove that γ is the inverse of β (hence β is the inverse of γ).

102. Notation for sums and products

In this section we introduce a notation for sums and products that may be familiar to you from calculus. This will be used in studying induction in Chapter 103.

102.1 Definition: sum and product of a sequence
Let $a_1, a_2, \ldots, a_n$ be a sequence of numbers (not necessarily integers). The expression $\sum_{i=1}^n a_i$ denotes the sum $a_1 + a_2 + \cdots + a_n$ of the numbers in the sequence and the expression $\prod_{i=1}^n a_i$ denotes the product $a_1 a_2 \cdots a_n$ of the numbers in the sequence.

102.1.1 Example $\sum_{i=1}^4 i = 1+2+3+4 = 10$ and $\prod_{i=1}^4 i = 1 \times 2 \times 3 \times 4 = 24$.

102.1.2 Example $\sum_{k=1}^5 2k-1 = 1+3+5+7+9 = 25$. The sum $\sum_{i=1}^5 2i-1$ also gives 25 — the i is a **dummy variable** just like the x in $\int_3^5 x^2\,dx$, which has the same value as $\int_3^5 t^2\,dt$. On the other hand, $\sum_{i=1}^5 2k-1 = 10k-5$.

102.1.3 Exercise What is $\sum_{k=1}^5 k^2$? What is $\prod_{k=1}^5 k^2$? (Answer on page 248.)

102.1.4 Example For b any fixed number, $\sum_{i=1}^4 b = b+b+b+b = 4b$ and $\prod_{i=1}^4 b = b \cdot b \cdot b \cdot b = b^4$.

102.1.5 Remark The numbering of the sequence does not have to start at 1. Thus a sequence a_3, $a_4, \ldots, a_{12}$ would have sum $\sum_{i=3}^{12} a_i$.

102.1.6 Exercise What are $\sum_{k=1}^{5} 2k$, $\sum_{k=0}^{4} 2(k+1)$ and $\sum_{k=0}^{4} 2k+1$? Two of them are the same. Explain why.

102.1.7 Sums and products in Mathematica To compute

$$\sum_{k=1}^{5} 2k-1$$

in Mathematica, you type the expression `Sum[2 k-1,{k,1,5}]`. Similarly, the expression `Product[k,{k,1,6}]` calculates $1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6$.

The expression `{k,1,5}` is a **range expression**; range expressions are used in many Mathematica commands. The range expression `{x,a,b}` means that the variable `x` ranges from the value of `a` to the value of `b`.

103. Mathematical induction

The positive integers contain some fascinating patterns. For example,

$$\begin{array}{ll} 1 = 1, & 1{+}3{+}5{+}7 = 16, \\ 1{+}3 = 4, & 1{+}3{+}5{+}7{+}9 = 25, \\ 1{+}3{+}5 = 9, & 1{+}3{+}5{+}7{+}9{+}11 = 36, \end{array}$$

In general it appears that the sum of the first n odd positive integers is n^2. This is a statement $Q(n)$ about an infinite number of positive integers n.

The subject of this section is an inference rule allowing the proof of such statements. Before the rule is stated, we will reformulate $Q(n)$ and see why it is true.

Using summation notation, $Q(n)$ is the statement

$$\sum_{k=1}^{n}(2k-1) = n^2$$

(You should check that $2k-1$ is indeed the kth odd positive integer.) Clearly $Q(1)$ is true: it says $1=1$.

I will now prove that for any positive integer n, $Q(n) \Rightarrow Q(n+1)$, using the direct method. The direct method requires us to assume the hypothesis is true, so suppose we knew that $Q(n)$ is true, that is that the sum of the first n odd integers is n^2. Then the sum of the first $n+1$ odd integers is

$$(\text{the sum of the first } n \text{ odd integers}) + (\text{the } n{+}1\text{st odd integer})$$

We know the left term is n^2 because we are assuming $Q(n)$, and the right term is $2n+1$. Hence the sum of the first $n+1$ odd integers is n^2+2n+1. But $n^2+2n+1=(n+1)^2$, in other words the sum of the first $n+1$ odd integers is $(n+1)^2$, so that $Q(n+1)$ is true. This proves that $Q(n) \Rightarrow Q(n+1)$.

Now we know these two things:

a) $Q(1)$.

b) For any n, $Q(n) \Rightarrow Q(n+1)$.

Using these facts, you should be able to convince yourself that $Q(n)$ is true for any positive integer, since $Q(1)$ is true, and the implication $Q(n) \Rightarrow Q(n+1)$ allows you to see that $Q(2)$ is true, $Q(3)$ is true, ..., jacking the proof up, so to speak, until you get to any positive integer. You need to know *both* $Q(1)$ and the implication $Q(n) \Rightarrow Q(n+1)$ for all n to make this work.

This approach is the basis for the following rule of inference:

103.2 Theorem: The principle of mathematical induction
For any statement about the positive integers, this rule of inference is valid:

$$P(1),\ (\forall n{:}\mathrm{N}^+)\big(P(n) \Rightarrow P(n+1)\big) \vdash (\forall n{:}\mathrm{N}^+)P(n)$$

103.2.1 Usage A proof using the principle of mathematical induction is called an **inductive proof**. The proof that $P(1)$ is true is the **basis step** and that $P(n) \Rightarrow P(n+1)$ is the **induction step**.

103.2.2 Remarks

a) The induction step is sometimes stated as $P(n-1) \Rightarrow P(n)$, which must hold for all integers >1, but that is only a change in notation.
b) The proof of the induction step, which is an implication, may be carried out by the direct method as was done above, or by the contrapositive method. If it is carried out by the direct method, one assumes that $P(n)$ is true and deduces $P(n+1)$. In doing this, $P(n)$ is called the **induction hypothesis**.
c) The principle of mathematical induction gives you a *scheme for proving a statement about all positive integers.* You still have to be clever somewhere in the proof. In the example just given, algebraic cleverness was required in the induction step.

103.3 Other starting points for proofs by induction

We have formulated mathematical induction as a scheme for proving a statement about all positive integers. One can similarly prove statements about all nonnegative integers by starting the induction at 0 instead of at 1 (see Example 103.3.1 below). In that case you must prove $P(0)$ and

$$(\forall n{:}\mathrm{N})\big(P(n) \Rightarrow P(n+1)\big)$$

Indeed, a proof by mathematical induction can be started at any integer, positive or negative. For example, if you prove $P(-47)$ and $P(n) \Rightarrow P(n+1)$ for $n \geq -47$, then $P(n)$ is true for all $n \geq -47$.

One could also go down instead of up, but we won't do that in this text.

103.3.1 Example Let's prove that for all nonnegative integers n, $3 \mid n^3 + 2n$.

Basis step: We must show $3 \mid 0^3 + 0$, which is obvious.

Induction step: assume $3 \mid n^3+2n$. (This is the induction hypothesis.) Then $n^3+2n = 3k$ for some integer k. Then

$$\begin{aligned}(n+1)^3+2(n+1) &= n^3+3n^2+3n+1+2n+2 \\ &= n^3+2n+3n^2+3n+3 \qquad (103.1)\\ &= 3(k+n^2+n+1)\end{aligned}$$

so is divisible by 3 as required.

103.3.2 Remark The statement $3 \mid n^3+2n$ is true of negative integers, too. Once you know it for positive integers, the proof for negative integers is easy: substitute $-n$ for n in the statement and do a little algebra. This trick often works for proving things about all integers. However, the principle of mathematical induction by itself is *not a valid method of proof* for proving statements about all integers.

103.3.3 Example A statement about the value of a sum or product can often be proved by induction. Let us prove that

$$\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$$

Proof Basis step: $\sum_{k=1}^{1} k = 1 = \frac{1}{2}\cdot 1 \cdot 2$, as required.
Induction step:

$$\begin{aligned}\sum_{k=1}^{n+1} k &= n+1+\sum_{k=1}^{n} k \\ &= n+1+\tfrac{1}{2}n(n+1) \quad \text{(by the induction hypothesis)}\\ &= (\tfrac{1}{2}n+1)(n+1)\\ &= \tfrac{1}{2}(n+1)(n+2) \quad \text{(by algebra)}\end{aligned}$$

This proof uses a basic trick: separate out the term in the sum (or product) of highest index, in this case $n+1$. Then the rest of the sum can be evaluated using the induction hypothesis.

103.3.4 Remark In all proofs by induction you should label the basis step, the induction step and the induction hypothesis. If you find yourself writing "and so on…" or "continuing in this way…" or anything like that, *you are not doing an inductive proof.*

103.4 Exercise set

Prove the statements in Exercises 103.4.1 through 103.4.8 by induction.

103.4.1 $\sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}$. (Answer on page 248.)

103.4.2

$$\sum_{k=1}^{n}(-1)^k k = \begin{cases} \frac{n}{2} & (n \text{ even}) \\ \frac{-(n+1)}{2} & (n \text{ odd}) \end{cases}$$

(Answer on page 249.)

103.4.3 $\sum_{k=1}^{n} k(k+1) = \frac{1}{3}n(n+1)(n+2)$.

103.4.4 $\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$.

103.4.5 $\sum_{k=1}^{n} 2^k = 2^{n+1} - 2$.

103.4.6 $\sum_{k=1}^{n} k2^k = (n-1)2^{n+1} + 2$.

103.4.7 $\sum_{k=1}^{n} k^3 = \left(\frac{1}{2}n(n+1)\right)^2$.

103.4.8 $\sum_{k=1}^{n} (-1)^k k^2 = \frac{(-1)^n}{2}n(n+1)$.

103.4.9 Exercise Prove the following inequalities by induction.
a) $2^n > 2n+1$ for $n \geq 3$.
b) $2^n \geq n^2$ for $n \geq 4$.

104. Least counterexamples

Proof by induction as described in Chapter 103 is based on a very basic fact about the positive integers that has wider applications. Suppose $P(n)$ is a statement about positive integers, and suppose the statement $(\forall n{:}\mathrm{N}^+)P(n)$ is *false.* Then there is a counterexample m, a positive integer for which $P(m)$ is false. Among all such counterexamples, there is a smallest one:

104.1 Theorem: The Principle of the Least Counterexample
Every false statement of the form $(\forall n{:}\mathrm{N}^+)P(n)$ *about the positive integers has a smallest counterexample.*

104.1.1 Usage This property of the positive integers is often referred to by saying the positive integers are **well-ordered**.

104.1.2 Remark Of course, one can replace the positive integers by the nonnegative integers, or by the set of all integers greater than a particular one, in the statement of Theorem 104.1.

The existence of least counterexamples is characteristic of such sets; for most other data types, least counterexamples need not exist. For example, the statement, "All integers are even" is a false universally quantified statement about the integers which has many counterexamples, but no *smallest* one.

104.2 Least counterexample and induction

The principle of mathematical induction, in other words Theorem 103.2, can be proved using the principle of the least counterexample. The proof is by contradiction.

Suppose that the hypotheses of the theorem are true: $P(1)$, and

$$(\forall n{:}\mathrm{N}^+)\Big(P(n) \Rightarrow P(n+1)\Big)$$

Suppose that $(\forall n{:}\mathrm{N}^+)P(n)$ is *false.* Then there is a least counterexample m, so $P(k)$ is true if $k < m$ but $P(m)$ is false. Now we have two cases.

(i) $m = 1$. Then $P(1)$ is false — but this contradicts one of the hypotheses of the theorem.

(ii) $m > 1$. In this case, $P(m)$ is false, since m is a counterexample to the statement $(\forall n{:}\mathrm{N}^+)P(n)$. Since m is the *least* counterexample, the statement $P(m-1)$ is *true.* It follows from the truth table for implication that the statement $P(m-1) \Rightarrow P(m)$ is false. But that means the hypothesis

$$(\forall n{:}\mathrm{N}^+)\Big(P(n) \Rightarrow P(n+1)\Big)$$

is false since $n = m-1$ provides a counterexample.

So in either case, one of the hypotheses of Theorem 103.2 must be false. Therefore there can be no least counterexample, so by Theorem 104.1 there can be no counterexample. Hence $(\forall n{:}\mathrm{N}^+)P(n)$ is true.

The principle of mathematical induction and the principle of the least counterexample are actually equivalent.

104.2.1 Exercise Use the principle of mathematical induction (Theorem 103.2) to prove Theorem 104.1.

104.3 Strong induction

The principle of the least counterexample is useful in its own right for proving things. For example, it is used in Problems 104.3.3 and 104.4.4 to prove the Fundamental Theorem of Arithmetic.

The principle allows you to assume as a kind of induction hypothesis that $P(k)$ is true for *all* $k < n$, not just for $n-1$. This is stated formally in Exercise 104.3.1 below. It is handy for proving things about factoring integers, since the prime factorization of an integer n has little to do with the factorization of $n-1$. This more general approach is often called **strong induction**, and another statement of it is in Problem 104.3.1.

In this book, proofs using this technique are usually presented as direct applications of the least counterexample principle.

104.3.1 Exercise Use the principle of the least counterexample to prove the following rule of inference for positive integers n. This rule is called the **Principle of Strong Induction**.

$$(\forall n{:}\mathrm{N}^+)\Bigg((\forall m{:}\mathrm{N}^+)\Big(m < n \Rightarrow P(m)\Big) \Rightarrow P(n)\Bigg) \vdash (\forall n{:}\mathrm{N}^+)P(n)$$

104.3.2 Example: Existence of quotient and remainder We will use the Principle of the Least Counterexample to prove the existence half of Theorem 60.2, page 84. That is, we will prove that for given integers m and n with $n \neq 0$, there are integers q and r satisfying

Q.1 $m = qn + r$, and

Q.2 $0 \leq r < |n|$.

That there is only one such pair of integers was proved on page 84.

We will give the proof for $m \geq 0$ and $n > 0$ and leave the other cases to you (Exercise 104.3.4). Let S be the set of all nonnegative integers of the form $m - xn$. S is nonempty (any negative x makes $m - xn$ positive, but there may also be positive x that do so). Let $m - qn$ be the smallest element of S. Let $r = m - qn$. Then $qn + r = qn + m - qn = m$, so Q.1 is true. Since $m - qn \in \mathrm{N}$ by assumption, we know that r, which is $m - qn$, is nonnegative, which is half of Q.2. As for the other half, suppose *for the purpose of proof by contradiction* that $r \geq n$. Then $m - qn \geq n$, that is, $m \geq n + qn = (q+1)n$. But then $m - (q+1)n$ is nonnegative, and it is certainly smaller than $m - qn$, contradicting our choice of $m - qn$ as the least element of S.

104.3.3 Exercise (hard) Show that if m is any integer greater than 1, then there is a finite list of primes $p_1, \ldots, p_k$ and integers $e_1, \ldots, e_k$ for which $m = \prod_{i=1}^{k} p_i^{e_i}$. Use the principle of the least counterexample. *Do not use the Fundamental Theorem of Arithmetic.* Note that if m is prime, then this holds for $k = 1$, $p_1 = m$, and $e_1 = 1$.

104.3.4 Exercise Complete the proof that the quotient (of integers) and remainder exist (see 104.3.2).

104.4 Proof of the Fundamental Theorem of Arithmetic

Exercises 104.4.1 through 104.4.4, together with Exercise 104.3.3, lead up to a proof of the Fundamental Theorem of Arithmetic. Thus the Fundamental Theorem should not be used in the proofs of those problems.

104.4.1 Exercise Show that if p is a prime and m an integer not divisible by p, then $\mathrm{GCD}(p, m) = 1$. (Answer on page 249.)

104.4.2 Exercise Show that if p is a prime and m and n are integers for which $p \mid mn$ but p does not divide m, then $p \mid n$. (Hint: Use Problem 104.4.1 and Bézout's Lemma, page 128.) (Answer on page 249.)

Suppose p is prime, $p \mid mn$, but p does not divide m. Then $\mathrm{GCD}(p, m) = 1$, so there are integers a and b for which $ap + bm = 1$. There is also an integer k for which $mn = pk$. Putting these facts together, $n = anp + bmn = anp + bkp = (an + bk)p$, so n is divisible by p.

104.4.3 Exercise Use Problem 104.4.2 to show that if p is a prime and $m_1, \ldots, m_k$ are positive integers for which $p \mid \prod_{i=1}^{k} m_i$, then for some i, $p \mid m_i$.

104.4.4 Exercise (hard) Show that if $p_1 < p_2 < \ldots < p_k$ in the prime factorization $m = \prod_{i=1}^{k} p_i^{e_i}$ in Exercise 104.3.3, then the factorization is unique. (Hint: Assume m is the least positive integer which has two such factorizations and use Problem 104.4.3 to obtain a prime which occurs in both factorizations. Then divide by that prime to obtain a smaller integer with two factorizations.)

105. Recursive definition of functions

Many functions are defined in such a way that the value at one input is defined in terms of other values of the function. Such a definition is called **recursive**.

105.1.1 Example One way of defining the function $F : \mathrm{N} \to \mathrm{N}$ for which $F(n) = 2^n$ would be to say

$$\begin{cases} F(0) = 1 \\ F(n+1) = 2 \cdot F(n) \end{cases} \tag{105.1}$$

for all $n \in \mathrm{N}$.

Programs 105.1 and 105.2 give Pascal functions which calculate 2^n.

```
FUNCTION TWOREC(N:INTEGER):INTEGER;
  BEGIN
    IF N=0 THEN TWOREC := 1
    ELSE TWOREC := 2*TWOREC(N-1)
  END;
```

Program 105.1: Program for 2^n

Program 105.1 simply copies Definition 105.1. Since the function `TWOREC` calls itself during its execution, this program is also said to be recursive. Program 105.2 is a translation of Definition 105.1 which avoids the function calling itself. Since it is implemented by a loop it is called an **iterative** implementation of the function.

105.1.2 Remark Many common algorithms are easily to define recursively, so the study of recursively-defined functions and how to implement them is a major part of computer science. Very often, the iterative implementation like Program 105.2 is to be preferred to the recursive one, but in complicated situations it is not always easy to transform the recursive definition into an iterative one. In some applications, for example in writing programs to parse expressions, the recursively written program may be the preferred method for writing the first attempt, since the iterative version can be much harder to understand and debug.

```
FUNCTION TWOIT(N:INTEGER):INTEGER; VAR COUNT:INTEGER;
                BEGIN
                  COUNT := 0; POWER := 1;
                  (*POWER is a integer variable declared in
                  the program containing this procedure.*)
                  WHILE COUNT<N DO
                    BEGIN
                      POWER := 2*POWER;
                      COUNT := COUNT+1
                    END
                  TWOIT := POWER
                END;
```

Program 105.2: Another program for 2^n

105.1.3 Exercise Find the values for $n=1$ through 5 of the functions defined as follows:

a) $F(0)=-3,\ F(n+1)=(n+1)F(n)$
b) $F(1)=1,\ F(n)=n^2+F(n-1)$
c) $F(n)=\begin{cases}0 & \text{if } 3\mid n\\ 1+F(n+1) & \text{otherwise}\end{cases}$
d) $F(0)=1,\ F(1)=3,\ F(n)=F(n-1)+F(n-2)$
e) $F(1)=0,\ F(2)=1,\ F(n)=(n-1)\Big(F(n-1)+F(n-2)\Big)$

(Answer on page 249.)

105.1.4 Example For a fixed sequence $\{a_k\}_{k\in\mathrm{N}^+}$,

$$F(n)=\sum_{k=1}^{n} a_k$$

is a function from N^+ to N^+ which has a natural definition by induction:

$$\begin{cases} \sum_{k=1}^{1} a_k = a_1 \\ \sum_{k=1}^{n+1} a_k = a_{n+1} + \sum_{k=1}^{n} a_k \end{cases} \tag{105.2}$$

105.1.5 Example The product has a similar definition:

$$\begin{cases} \prod_{k=1}^{1} a_k = a_1 \\ \prod_{k=1}^{n+1} a_k = a_{n+1} \cdot (\prod_{k=1}^{n} a_k) \end{cases} \tag{105.3}$$

105.1.6 The factorial function A particularly important function which can be defined by induction is the **factorial function**. Its value at n is denoted $n!$ and it is defined this way:

$$\begin{cases} 0! = 1 \\ (n+1)! = (n+1)\cdot n! \end{cases} \tag{105.4}$$

Thus for $n > 0$, $n! = \prod_{k=1}^{n} k$; you can prove this by induction because both $n!$ and the product are defined by induction (Exercise 105.2.1). The factorial function will be used in various combinatorial formulas in later sections.

105.2 Proofs involving inductively defined functions

Defining a function by induction makes it convenient to prove things about it by induction. For example, let us use induction to prove that $n! > 2^n$ for $n > 3$. We start the induction at $n = 4$. Then $4! = 24$ and $2^4 = 16$, so the statement is true for $n = 4$. For the induction step, suppose $n! > 2^n$ and $n \geq 4$. It is necessary to prove that $(n+1)! > 2^{n+1}$. Both these functions are defined by induction, so we can apply their definitions and the induction hypothesis to get

$$(n+1)! = (n+1) \cdot n! > 2 \cdot n! \geq 2 \cdot 2^n = 2^{n+1}$$

as required.

105.2.1 Exercise Prove directly from the inductive definition of $n!$ that $n! = \prod_{k=1}^{n} k$ for all positive integers n. (Answer on page 249.)

105.2.2 Exercise Prove that for all integers $n > 0$, $2^n \leq 2(n!)$.

105.2.3 Exercise Find constants C and D for which for all integers $n > 0$, $3^n \leq C(n!)$ and $4^n \leq D(n!)$. Prove your answers are correct.

106. Inductive and recursive

Definition 105.1 gives the value at n in terms of the value of the function at a smaller integer. In general, a function $F : \mathrm{N} \to \mathrm{N}$ is **defined by induction** if certain initial values $F(0), F(1), \ldots, F(k)$ are defined and for each $n \in \mathrm{N}$, $F(n+1)$ is defined in terms of some or all of the preceding values $F(0), F(1), \ldots, F(n)$. Thus inductive definition is a special case of recursive definition. In a more formal treatment of this subject, the phrase "in terms of" would have to be precisely defined.

Recursive definitions which are not inductive may involve domains other than N which have no natural ordering (so that "in terms of smaller values" makes no sense) or functions on N which involve definition in terms of both larger and smaller values. The general concept of recursion is fundamental to much of theoretical computer science. It is a common theme uniting the different threads in [Hofstadter, 1979].

106.1.1 Example The **ninety-one function** $F : \mathrm{N} \to \mathrm{N}$ is defined by:

$$F(n) = \begin{cases} F\Big(F(n+11)\Big) & (n \leq 100) \\ n - 10 & (n > 100) \end{cases} \tag{106.1}$$

This is a well defined function. It has the property that $F(n) = 91$ if $n \leq 100$ and $F(n) = n - 10$ if $n > 100$.

106.1.2 Example The **Collatz function** $T : \mathrm{N}^+ \to \mathrm{N}^+$ defined by:

$$T(n) = \begin{cases} 1 & (\text{if } n = 1) \\ T(\frac{n}{2}) & (\text{if } n \text{ is even}) \\ T(3n+1) & (\text{if } n \text{ is odd and } n > 1) \end{cases}$$

Looking at the formula, there is no reason to believe that the computation wouldn't loop forever for some value of the input, but no one has ever been able to discover such a value or to prove that such a value does not exist. (Every input that has ever been computed does in fact given an answer, namely 1.) In other words, although we called it "the Collatz function", we don't actually know that it is a function! Note that if you change the '$3n+1$' to '$3n-1$' in the third line, then $T(5)$ is not defined. There is much more about this in [Guy, 1981], Problem E-16, page 120 and in [Lagarias, 1985].

106.1.3 Exercise (hard) Prove that the ninety-one function defined by Equation (106.1) on page 159 satisfies $F(n) = 91$ if $n \leq 100$ and $F(n) = n - 10$ if $n > 100$.

107. Functions with more than one starting point

The Fibonacci function is an example of a function defined in terms of *two* previous values (hence requiring two initial conditions):

107.1 Definition: Fibonacci function

The **Fibonacci function** $F : \mathrm{N} \to \mathrm{N}$ is defined by

$$\begin{cases} F(0) = 0 \\ F(1) = 1 \\ F(n) = F(n-1) + F(n-2) \end{cases} \tag{107.1}$$

107.1.1 Remarks

a) The Fibonacci function is called "Fibonacci" after Leonardo di Pisa, who described it in 1220 AD. He was the son (Fi, short for Figlio) of Bonaccio.

b) The Fibonacci function has traditionally been described as the formula for the number of pairs of rabbits you have after n months under these assumptions: initially you have just one pair of rabbits, and every month each pair of rabbits over one month old have a pair of children, one male and one female. And none of them die.

   Suppose you buy (trap?) the first pair of rabbits at the beginning of month 1. Then $F(0) = 0$ and $F(1) = 1$. At the nth month, F must satisfy the equation

$$F(n) = F(n-1) + F(n-2) \quad (n \geq 2)$$

   since the $F(n-1)$ rabbits you had one month ago are still around and you have a new pair for each of the $F(n-2)$ pairs born two or more months ago.

This explanation bears no relation to reality since rabbits take six months, not one, to mature sexually, and they do not reliably produce one male and one female each gestation period.

107.1.2 Example The **Perrin function** is defined with *three* starting points:

$$\begin{cases} P(0) = 3 \\ P(1) = 0 \\ P(2) = 2 \\ P(n) = P(n-2) + P(n-3) \end{cases} \tag{107.2}$$

For integers larger than 1 up to a fairly large number, this function has the property

$$n \mid P(n) \Leftrightarrow n \text{ is prime.}$$

The smallest integer > 1 for which this is false is apparently $271{,}441$, which is 521^2, but I have not been able to check this.

A number n for which $n \mid P(n)$ is called a **Perrin pseudoprime**.

107.2 Recurrence relations

Since the Fibonacci function has domain N, it is the same as an infinite sequence (see Example 97.2.2). The values $F(0), F(1), F(2), \ldots$ are often called the **Fibonacci numbers**. When expressed in sequence notation, the definition becomes

$$\begin{cases} f_0 = 0 \\ f_1 = 1 \\ f_n = f_{n-1} + f_{n-2} \end{cases} \tag{107.3}$$

Fibonacci function is called a **recurrence relation** or simply a **recurrence**.

Sometimes, but not always, a function defined by a recurrence relation can be given a noninductive definition by a formula. Finding such a "closed form" definition is called **solving the recurrence relation**. We have already solved some recurrence relations. For example, the statement that the sum of the first n odd integers is n^2 can be reworded to say that the solution to the recurrence relation

$$\begin{cases} s_1 = 1 \\ s_{n+1} = 2n + 1 + s_n \end{cases} \tag{107.4}$$

is $s_n = n^2$.

If you can guess a solution to a recurrence relation, you can often prove it is correct by induction. Problem 107.3.11 gives a closed solution to the Fibonacci recurrence. Note that it would generally be better to calculate Fibonacci numbers for small n using the recurrence relation rather that the complicated formula given in Problem 107.3.11.

107.3 Exercise set

Exercises 107.3.2 through 107.3.11 refer to the Fibonacci sequence.

107.3.1 Exercise Prove that for all nonnegative integers n, $f_{n+1}^2 - f_n f_{n+2} = (-1)^n$. (Answer on page 249.)

107.3.2 Exercise Prove that for all nonnegative integers n, f_n is even if and only if $3 \mid n$.

107.3.3 Exercise Prove that for all [positive integers n, $f_{n+1} \operatorname{div} f_n = 1$ and $f_{n+1} \bmod f_n = f_{n-1}$.

107.3.4 Exercise Prove that for all nonnegative integers n,

$$f_n f_{n+3} - f_{n+1} f_{n+2} = (-1)^{n+1}$$

107.3.5 Exercise Prove that for all nonnegative integers n, $\mathrm{GCD}(f_{n+1}, f_n) = 1$. (Hint: You can use Exercise 107.3.3, or you can look at Exercise 107.3.4 and meditate upon Bézout.)

107.3.6 Exercise Prove by induction that

$$\Sigma_{k=1}^n f_k^2 = f_n f_{n+1}$$

107.3.7 Exercise Give a proof by induction on n that for all $n \geq 0$,

$$f_{n+2} \geq (\frac{8}{5})^n$$

(You can also prove this using Problem 107.3.11 below.)

107.3.8 Exercise Show that for all $n \geq 0$, $f_{n+1}^2 - f_n f_{n+1} - f_n^2 = \pm 1$.

107.3.9 Exercise (hard) (Matijasevich) Prove that if x and y are nonnegative integers such that $y^2 - xy - x^2 = \pm 1$, then for some nonnegative integer n, $x = f_n$ and $y = f_{n+1}$. (Be careful: You are *not* being asked to show that $\langle f_n, f_{n+1} \rangle$ is a solution of the equation for each n — that is what the Problem 107.3.8 asks for. You are being asked to show that *no other pair of integers* is a solution.)

107.3.10 Exercise (hard) (Matijasevich) Show that for all nonnegative integers m and n, if $f_m^2 \mid f_n$, then $f_m \mid n$.

107.3.11 Exercise (hard) Prove that for all nonnegative integers n,

$$f_n = (1/\sqrt{5})(r^n - s^n)$$

where r and s are the two roots of the equation $x^2 - x - 1 = 0$ and $r > s$.

107.3.12 Exercise Let a function $F : \mathrm{N} \to \mathrm{N}$ be defined by

$$\begin{cases} F(0) = 0 \\ F(1) = 1 \\ F(n) = 5F(n-1) - 6F(n-2) \quad (n > 1) \end{cases}$$

Prove by induction that for all $n \geq 0$, $F(n) = 3^n - 2^n$.

107.3.13 Exercise Define a function $F : \mathrm{N} \to \mathrm{N}$ by

$$\begin{cases} F(0) = F(1) = 1 \\ F(n) = 2F(n-1) + F(n-2) \qquad (n > 1). \end{cases}$$

Show
a) $F(n)$ is always odd.
b) $F(4k+2)$ is divisible by 3 for any integer $k \geq 0$.

107.3.14 Exercise (hard) (Myerson and van der Poorten [1995]) Define a function $G : \mathrm{N} \to \mathrm{N}$ by $G(1) = G(3) = G(5) = 0$, $G(0) = G(4) = 8$, $G(2) = 9$, and $G(n+6) = 6G(n+4) - 12G(n+2) + 8G(n)$ for $n > 5$. Show that $G(n) = 0$ if n is odd and

$$G(n) = (n-8)^2 \cdot 2^{\frac{n-6}{2}}$$

otherwise.

107.3.15 Exercise (Myerson and van der Poorten [1995]) Define a function $G : \mathrm{N} \to \mathrm{Z}$ by $G(0) = 0$, $G(1) = 1$, $G(2) = -1$, and

$$G(n) = -G(n-1) + G(n-2) + G(n-3)$$

for $n > 2$. Show that

$$G(n) = \begin{cases} -\frac{n}{2} & n \text{ even} \\ \frac{n+1}{2} & n \text{ odd} \end{cases}$$

(Compare Exercise 94.1.4, page 136.)

108. Functions of several variables

Functions $F : \mathrm{N}^2 \to \mathrm{N}$ can be defined by induction, too. One technique is to define a function of two variables for all values of one variable by induction on the other variable.

108.1.1 Example Multiplication in N, which is a function $\mathrm{N}^2 \to \mathrm{N}$, can be defined by

$$\begin{cases} m \cdot 0 = 0 \\ m \cdot (n+1) = m \cdot n + m \end{cases} \tag{108.1}$$

This defines $m \cdot n$ for each $m \in \mathrm{N}$ by induction on n. The definition shows how to define multiplication in terms of adding one.

108.1.2 Exercise The **successor function** $s : \mathrm{N} \to \mathrm{N}$ is the function which takes each natural number to the next one: $s(n) = n + 1$. Show how to define addition inductively in terms of the successor function.

108.1.3 Exercise Show how to define the operation $(m, n) \mapsto m^n$ inductively in terms of the successor function and multiplication (defined inductively in Example 108.1.1).

108.1.4 Example Theorem 65.1, page 92, gives a recursive definition of the GCD function. It translates directly into the Pascal function in Program 108.1.

```
FUNCTION GCD(M,N:INTEGER);
BEGIN
  IF M=0 THEN GCD := N
  ELSE
    IF N=0 THEN GCD := M
    ELSE
      GCD := GCD(N,M MOD N)
END;
```

Program 108.1: Program to compute the GCD

108.1.5 Exercise Define the function $A : \mathrm{N} \times \mathrm{N} \to \mathrm{N}$ by

$$\begin{cases} A(0,y) = 1 & \\ A(1,0) = 2 & \\ A(x,0) = x+2 & \text{for } x \geq 2 \\ A(x,y) = A(A(x-1,y),y-1) & \end{cases}$$

a) Prove by induction that $A(x,1) = 2x$ for all $x \geq 1$.
b) Prove by induction that $A(x,2) = 2^x$ for all $x \geq 0$.
c) Prove by induction that $A(x,3) = 2^{A(x-1,3)}$ for all $x \geq 0$.
d) Calculate $A(4,4)$.

109. Lists

Informally, a list of elements of a set A consists of elements of A arranged from first to last, with order and repetition mattering. We will write them using the same notation that we use for tuples. Thus $\langle 1,4,3,3,2\rangle$ is a list of elements of N. It is not the same list as $\langle 1,4,3,2\rangle$ or as $\langle 4,1,3,3,2\rangle$. A particular list is the **empty list**, denoted $\langle\rangle$.

We could have said that a list of elements of A is just a tuple of elements of A. However, the *specification* for lists is different from that for tuples, so our formal treatment will start from scratch. The definition is recursive.

109.1 Definition: list
For any set A, a **list of elements of** A is either the **empty list** $\langle\rangle$ or a **nonempty list**. A nonempty list of elements of A has a **head**, which is an element of A, and a **tail**, which is a list of elements of A. The list with head a and tail $\langle b_1,\ldots,b_k\rangle$ is denoted $\langle a,b_1,\ldots,b_k\rangle$. The list with head a and empty tail is denoted $\langle a\rangle$. Every list of elements of A is constructed by repeated application of this definition starting with the empty list.

109.1.1 Remark The head of a nonempty list is *not* a list, but the tail is a list. The empty list does not have a head or a tail.

109.1.2 Example $\langle\rangle$, $\langle 5\rangle$, $\langle 2,1,1,-3\rangle$ and $\langle 3,3,3\rangle$ are all lists of elements of Z (lists of integers). The head of $\langle 2,1,1,-3\rangle$ is 2 and the tail is $\langle 1,1,-3\rangle$. The head of $\langle 5\rangle$ is 5 and the tail is $\langle\rangle$.

109.2 Definition: set of lists

The set of all lists of elements of A is denoted A^*. The set of all nonempty lists of elements of A is denoted A^+.

109.2.1 Example Let A be the English alphabet. Then the lists $\langle\rangle$, $\langle a,a,b\rangle$ and $\langle c,a,t,c,h\rangle$ are all elements of A^*. The list $\langle 2,2\rangle$ is an element of N^*, and $\langle c,a,t,c,h,2,2\rangle$ is an element of $(A\cup\mathrm{N})^*$ but not of A^* or of N^*.

109.2.2 Lists in Mathematica A list such as $\langle 1,5,3,6\rangle$ in Mathematica is written $\{1,5,3,6\}$.

109.3 The list constructor

Most concepts connected with lists are defined recursively using Definition 109.1. To make this easy, we introduce the **list constructor function** $\mathrm{cons}: S\times S^*\to S^+$ (note carefully the domain and codomain of this function), which is defined by requiring

$$\mathrm{cons}(a,\langle b_1,b_2,\ldots,b_n\rangle)=\langle a,b_1,b_2,\ldots,b_n\rangle \tag{109.1}$$

Thus $\mathrm{cons}(c,\langle a,t,c,h\rangle)=\langle c,a,t,c,h\rangle$ and $\mathrm{cons}(a,\langle\rangle)=\langle a\rangle$.

109.4 Definition: length of a list

The **length (of a list)** of a list L of elements of S is denoted $|L|$ and is defined by

LL.1 $|\langle\rangle|=0$.

LL.2 $|\mathrm{cons}(a,L)|=1+|L|$.

109.4.1 Example $|\langle c,a,t\rangle|=3$, because, by repeatedly applying Rule (109.1), page 165, and LL.1 and LL.2, we have

$$\begin{aligned}
|\langle c,a,t\rangle| &= |\mathrm{cons}(c,\langle a,t\rangle)|\\
&= |\mathrm{cons}(c,\mathrm{cons}(a,\langle t\rangle))|\\
&= |\mathrm{cons}(c,\mathrm{cons}(a,\mathrm{cons}(t,\langle\rangle)))|\\
&= 1+|\mathrm{cons}(a,\mathrm{cons}(t,\langle\rangle))|\\
&= 1+1+|\mathrm{cons}(t,\langle\rangle)|\\
&= 1+1+1+|\langle\rangle|\\
&= 1+1+1+0 = 3
\end{aligned}$$

109.4.2 Remark It can be proved by induction on the length of a list that a list of length k satisfies the specification for a k-tuple (Definition 36.2, page 50). Nevertheless, the recursive definition of list given above has provides a useful alternative approach to the idea which simplifies much of the theory of lists.

109.5 Concatenation

Informally, the concatenate of two lists is obtained by writing the entries of one and then the other in a single list. Concatenation is denoted by juxtaposition; thus $\langle 1,4,4\rangle\langle 2,3\rangle = \langle 1,4,4,2,3\rangle$ and $\langle 3,2,2\rangle\langle\rangle = \langle 3,2,2\rangle$.

Again, we give a formal definition by induction.

109.6 Definition: concatenate of lists

The concatenate LN of two lists L and N is defined recursively as follows:

CL.1 $\langle\rangle N = N$

CL.2 $\operatorname{cons}(a,L)N = \operatorname{cons}(a,LN)$.

109.6.1 Example

$$\begin{aligned}
\langle c,a,t\rangle\langle c,h\rangle &= \operatorname{cons}(c,\langle a,t\rangle)\langle c,h\rangle \\
&= \operatorname{cons}(c,\langle a,t\rangle\langle c,h\rangle) \\
&= \operatorname{cons}(c,\operatorname{cons}(a,\langle t\rangle)\langle c,h\rangle) \\
&= \operatorname{cons}(c,\operatorname{cons}(a,\langle t\rangle\langle c,h\rangle)) \\
&= \operatorname{cons}(c,\operatorname{cons}(a,\operatorname{cons}(t,\langle\rangle)\langle c,h\rangle)) \\
&= \operatorname{cons}(c,\operatorname{cons}(a,\operatorname{cons}(t,\langle\rangle\langle c,h\rangle))) \\
&= \operatorname{cons}(c,\operatorname{cons}(a,\operatorname{cons}(t,\langle c,h\rangle))) \\
&= \operatorname{cons}(c,\operatorname{cons}(a,\langle t,c,h\rangle)) \\
&= \operatorname{cons}(c,\langle a,t,c,h\rangle) \\
&= \langle c,a,t,c,h\rangle
\end{aligned}$$

109.6.2 Remark Definition 109.6 implies that, for example, $\langle\rangle\langle c,a,t\rangle = \langle c,a,t\rangle$. We would expect that $\langle c,a,t\rangle\langle\rangle = \langle c,a,t\rangle$ as well. This can be proved by induction:

109.7 Theorem

For any list L, $L\langle\rangle = L$.

Proof If L has length 0, that is, if $L=\langle\rangle$, then $L\langle\rangle = \langle\rangle\langle\rangle = \langle\rangle$ by CL.1. Otherwise, assume the theorem is true for lists of length k and let L have length $k+1$. Then $L = \operatorname{cons}(a,L')$ for some element a and list L' of length k, and

$$L\langle\rangle = \operatorname{cons}(a,L')\langle\rangle = \operatorname{cons}(a,L'\langle\rangle) = \operatorname{cons}(a,L') = L$$

by CL.2 and the induction hypothesis.

109.8 Theorem

Concatenation is associative. Precisely, for any lists L, M and N, $(LM)N = L(MN)$.

Proof This is also proved by induction on the length of L. If $L = \langle\rangle$, then $(LM)N = (\langle\rangle M)N = MN = \langle\rangle(MN)$ by CL.1 applied twice. Now assume that $L = \text{cons}(a, L')$ and that $(L'M)N = L'(MN)$. Then

$$\begin{aligned}
(LM)N &= \big(\text{cons}(a, L')M\big)N \\
&= \text{cons}(a, L'M)N && \text{by CL.2} \\
&= \text{cons}(a, (L'M)N) && \text{by CL.2} \\
&= \text{cons}(a, L'(MN)) && \text{induction hypothesis} \\
&= \text{cons}(a, L')(MN) && \text{by CL.2} \\
&= L(MN)
\end{aligned}$$

109.8.1 Exercise Prove by induction that the length of the concatenate of two lists is the sum of the lengths of the lists. Use Definitions 109.4 and 109.6 explicitly.

109.8.2 Exercise Give an inductive definition of the last entry of a list. (Answer on page 249.)

109.8.3 Exercise Give an inductive definition of the maximum of a nonempty list of real numbers. It should satisfy $\max\langle 1, 3, 17, 2\rangle = 17$ and $\max\langle 5\rangle = 5$, for example.

109.8.4 Exercise Give an inductive definition of the sum of the entries of a list of real numbers. It should satisfy $\text{SUM}\langle 3, 4, 2, 3\rangle = 12$ and $\text{SUM}\langle 42\rangle = 42$. The sum of the empty list should be zero.

109.8.5 Exercise (hard) Prove that a list of length k satisfies the specification for a tuple of length k (Definition 36.2, page 50).

110. Strings

110.1 Definition: string
A **string** is a list of characters in some alphabet.

110.1.1 Example $\langle c, a, t\rangle$ is a string in the English alphabet.

110.1.2 Usage It is customary to denote such a string by writing the characters down next to each other and enclosing them in quotes. We will use single quotes. Thus '*cat*' is another notation for the string $\langle c, a, t\rangle$. We specifically regard '*cat*' and $\langle c, a, t\rangle$ as the same mathematical object written using two different notations.

110.1.3 Remarks

a) Note carefully that '*cat*' is a string, "cat" is an English word, and a cat is a mammal! Similarly, '52' is a string and 52 is a number.
b) The alphabet can be any set of characters. For example '0101' is a string in the alphabet of binary digits.

110.2 Concatenation of strings

In string notation, concatenation is simply juxtaposition: to say that the concatenate of '*cat*' and '*ch*' is '*catch*', we write

$$\text{`}cat\text{'`}ch\text{'} = \text{`}catch\text{'}$$

Strings are often denoted by lowercase letters, particularly those late in the alphabet. For example, let $w = $ '*cat*' and $x = $ '*doggie*'. Then $wx = $ '*catdoggie*', $ww = $ '*catcat*' and $xw = $ '*doggiecat*'. It is very important to distinguish w, which here is the name of a string, from 'w' which is a string of length one.

110.3 The empty string

The empty string could be denoted '', but this makes it hard to read, so we will follow common practice and use a symbol to denote the empty string. In this text, the symbol will be Λ. Other texts use ϵ or 0.

110.3.1 Example Λ'*abba*' = '*abba*' = '*abba*'Λ, and $\Lambda\Lambda = \Lambda$.

110.3.2 Remark Note carefully that '*cat*' is a string, but that "Λ" is the *name* of a string.

110.4 Exponential notation for concatenation

To designate a string concatenated with itself several times an exponential notation is used. If w is a string, w^n is the concatenate of the string w with itself n times.

110.4.1 Example Let $w = $ '0110'. Then it follows that

$$w^2 = \text{`01100110'} \quad \text{and} \quad w^3 = \text{`011001100110'}$$

Note in particular that '0'^{3} = '000' and '1'^{2}'0'^{4} = '110000'. We always take $w^1 = w$ and $w^0 = \Lambda$.

110.4.2 Exercise Find the concatenate wx if

a)	$w = $ '011', $x = $ '1010'	d)	$w = x = \Lambda$.
b)	$w = \Lambda$, $x = $ '011'	e)	$w = $ '011', $x = w^2$.
c)	$w = $ '011', $x = \Lambda$.	f)	$x = $ '011', $w = x^2$.

(Answer on page 249.)

110.4.3 Exercise Let $A = \{a, b\}$ and let E be the set of strings in A^* of even length. Give an inductive definition of E. (Answer on page 249.)

110.4.4 Exercise Give an inductive definition of the set of strings in $\{a, b\}$ of odd length.

110.4.5 Exercise Give an inductive definition of the kth entry of a string. It should exist for strings of length k or greater but not for strings of length less than k. Follow the pattern of the answer to Exercise 109.8.2, using cons.

110.4.6 Exercise Give an inductive definition of w^n for an arbitrary string w. The induction should be on n.

111. Formal languages

111.1 Definition: language
A **language** is a set of strings in some finite alphabet A.

111.1.1 Usage

a) In the research literature, this concept of language is often call "formal language".
b) If L is a language consisting of strings in A^* for some finite alphabet A, then one says that L is a "language in A". This is common terminology but may be slightly confusing since in fact the elements of L are not elements of A, they are elements of A^*.

111.1.2 Remark The definition says that a language is a subset of A^*. Note that the language may be infinite although the alphabet is finite.

111.1.3 Example The empty language is the set $\emptyset$. No strings are elements of the empty language.

111.1.4 Example Another example is the language $\{\Lambda\}$ whose only element is the empty string. It is important to distinguish this from the empty language $\emptyset$.

111.1.5 Example Another uninteresting language is the language A^*, containing as elements every string in the alphabet A.

111.1.6 Example The set $\{$'01','011','1'$\}$ is a language in $\{0,1\}$.

111.1.7 Example The set of strings in $\{0,1\}^*$ with 1 in the second place is a language. Note that '0110' is in the language but '1' and '100' are not in the language.

111.1.8 Example If n is a positive integer, then A^n denotes the set of strings in the alphabet A of length n. Thus if $A=\{0,1\}$, then $A^2=\{$'00','01','10','11'$\}$. We take $A^0=\{\Lambda\}$. Note that A^1 is the set of strings of length 1 in A, and so is *not* the same thing as A.

111.1.9 Example The set L of strings in $\{0,1\}^*$ which read the same forward and backward is a language. For example, '0110' $\in L$, but '10010' $\notin L$. Such strings are called **palindromes**.

111.2 Theorem

For any alphabet A,

$$A^* = A^0 \cup A^1 \cup \cdots \cup A^n \cup \cdots \tag{111.1}$$

the union of the infinite sequence of languages $A^0, A^1, \ldots, A^n, \ldots$.

Proof This follows from the fact that every string in A^* has some length n.

111.2.1 Remark An element of A^* is a string of *finite* length. A^* contains as elements no infinite sequences of elements of A, although Equation (111.1) expresses it as the union of an infinite sequence of *sets*. This follows from the definition of "union": to be in A^* according to 111.1, an element has to be in A^n for some integer n, so has to be a string of length n for some n.

111.3 Inductive definition of languages

A language can sometimes be given an inductive definition paralleling the definition of A^* given previously.

111.3.1 Example Let L be the set of strings in $\{0,1\}$ of the form 0^k1^k, for $k = 1$, 2, ... In other words, L consists of Λ, '01', '0011', '000111', '00001111', and so on. Then L can be defined by induction this way:

L.1 The empty string Λ is a string in L.
L.2 If $w \in L$, then '0'w'1' $\in L$.
L.3 Every string in L is given by one of the preceding rules.

111.3.2 Example The set P of palindromes can be defined this way:

111.4 Definition: the set of palindromes

Let A be a set.
PAL.1 The empty string Λ is a string in P.
PAL.2 If $a \in A$, then 'a' is a string in P.
PAL.3 If w is a string in P and $a \in A$, then awa is a string in P.
PAL.4 Every string in P is given by one of the preceding rules.

111.4.1 Remark Thus to show that '$abba$' is a palindrome, we say that Λ is a palindrome by PAL.1, so 'bb' (which is '$b\Lambda b$') is a palindrome by PAL.3, so '$abba$', which is 'a''bb''a', is a palindrome by PAL.3.

111.4.2 Exercise Give inductive definitions of the following languages in the alphabet $\{a,b\}$:

a) The set of strings containing no a's.
b) The set of strings containing exactly one a.
c) The set of strings containing exactly two a's.

112. Families of sets

112.1 Definition: family of sets
A tuple whose coordinates are sets is called a **family of sets**.

112.1.1 Usage A variant of this concept is to consider a *set* whose elements are sets. For some authors, a family of sets is a set of sets instead of a tuple of sets.

112.1.2 Example Let $A_1 = \{1,2,3\}$, $A_2 = \{2,3,4,5\}$ and $A_3 = \{3,4,5,7\}$. Then $\langle A_1, A_2, A_3\rangle$ is a family of sets, and so is $\langle A_1, \{4,5,6\}, \emptyset\rangle$.

112.2 Definition: union and intersection of a family of sets
Let $S = \langle A_i\rangle_{i\in\mathbf{n}}$ be an n-tuple of sets $A_1, A_2, \ldots, A_n$ Then

$$\bigcup_{i=1}^n A_i = \{x \mid \exists i(x \in A_i)\} \tag{112.1}$$

$$\bigcap_{i=1}^n A_i = \{x \mid \forall i(x \in A_i)\} \tag{112.2}$$

112.2.1 Example Let $A_1 = \{1,2,3\}$, $A_2 = \{2,3,4,5\}$ and $A_3 = \{3,4,5,7\}$. Then $\bigcup_{i=1}^3 A_i = \{1,2,3,4,5,7\}$ and $\bigcap_{i=1}^3 A_i = \{3\}$.

112.2.2 Example This notation is frequently used for infinite sets. As an example, recall that $(a\,..\,b)$ denotes the subset $\{r \in \mathrm{R} \mid a < r < b\}$ of the reals. Then if $\mathcal{F} = \{(-n\,..\,n) \mid n \in \mathrm{N}^+\}$, then $\bigcap\mathcal{F} = (-1\,..\,1)$, and, by the Archimedean property, $\bigcup\mathcal{F} = \mathrm{R}$. This is often written in the notation of infinite sequences:

$$\bigcup_{n=1}^{\infty}(-n\,..\,n) = \mathrm{R} \quad\text{and}\quad \bigcap_{n=1}^{\infty}(-n\,..\,n) = (-1\,..\,1)$$

112.2.3 Warning The symbol $\bigcup_{i=1}^3 A_i$ denotes $A_1 \cup A_2 \cup A_3$. In contrast, the symbol $\bigcup_{i=1}^\infty A_i$ denotes the union of all the sets A_i for each positive integer i, *specifically not including anything denoted* A_∞. Since "∞" is not an integer, A_∞ (if such a thing has been defined) is not included in the union.

Thus "$\bigcup_{i=1}^3 A_i$" goes up to 3 and includes 3, but "$\bigcup_{i=1}^\infty A_i$" does not include "∞".

There is notation analogous to that of Definition 112.2 for a set of sets (in contrast to a tuple of sets).

112.3 Definition: union and intersection of a set of sets
If $\mathcal{F}$ is a set whose elements are sets, then

$$\bigcup\mathcal{F} = \{x \mid (\exists A \in \mathcal{F})(x \in A)\} \tag{112.3}$$

and

$$\bigcap\mathcal{F} = \{x \mid (\forall A \in \mathcal{F})(x \in A)\} \tag{112.4}$$

112.3.1 Example Let $\mathcal{F} = \{\{1,2,3\},\{2,3\},\{3,4\}\}$. Then $\bigcup\mathcal{F} = \{1,2,3,4\}$ and $\bigcap\mathcal{F} = \{3\}$.

112.3.2 Exercise Give an explicit description of these sets.

a) $\bigcup_{i=1}^{\infty}(-i\,..\,i+2)$
b) $\bigcup_{i=1}^{\infty}(-1/i\,..\,1/i)$
c) $\bigcap_{i=1}^{\infty}(-1/i\,..\,1+(1/i))$
d) $\bigcap_{i=1}^{\infty}(i-1\,..\,i)$
e) $\bigcap_{i=1}^{\infty}[i-1\,..\,i]$

112.4 Intersection and union over the empty set

If $\mathcal{F}$ is a family of subsets of a set B, then we can reword the definition of the intersection of the sets in $\mathcal{F}$ as follows: it is the set T defined by the property:

$$(\forall S)(S \in \mathcal{F} \Rightarrow x \in S) \Leftrightarrow x \in T$$

If $\mathcal{F}$ is empty, the hypothesis is vacuously true, so $x \in T$ for every $x \in B$; in other words, $T = B$. Thus we define the intersection of the empty set of subsets of a set B to be B itself. Note that this definition is relative to a set containing as subsets all the sets in $\mathcal{F}$, in contrast to the intersection of families of sets in general as defined in the preceding section.

The union U of a family of sets $\mathcal{F}$ of subsets of B can be described by the property:

$$(\exists S)(S \in \mathcal{F} \wedge x \in S) \Leftrightarrow x \in U$$

(note the placement of the parentheses). If $\mathcal{F}$ is empty, then there is no $S \in \mathcal{F}$, so we define the union of an empty family of sets to be the empty set.

112.4.1 Warning In discussing sets of sets, remember that if $\mathcal{F}$ is a set of sets, an *element* of $\mathcal{F}$ is a *set*. It is a mistake to think of the words "element" and "set" as contrasting with each other. An element of a set may or may not be a set itself. Also, *any* set S is an element of some other set, for example of $\{S\}$.

112.4.2 Exercise Give an explicit description of $\bigcup\mathcal{F}$ and $\bigcap\mathcal{F}$ for each of these families of subsets of R:

a) $\mathcal{F} = \{\{2,4\},\{1,3,4\},\{2,5\}\}$.
b) $\mathcal{F} = \{(-3\,..\,3),(-2\,..\,2),(-1\,..\,1)\}$.
c) $\mathcal{F} = \{(-1\,..\,1),(1\,..\,2),(2\,..\,3)\}$.

(Answer on page 249.)

112.4.3 Exercise What are $\bigcup\mathcal{P}A$ and $\bigcap\mathcal{P}A$ for any set A?

113. Finite sets

We begin by giving a mathematical definition of the idea that a set has n elements. No doubt you have no trouble understanding a statement such as "S has 5 elements" without a formal definition; however, giving a formal meaning to such statements allows us to prove theorems about the number of elements of a set that have turned out to have many applications.

In this definition we use the set $\mathbf{n} = \{i \in \mathrm{N} \mid 1 \leq i \leq n\}$ (Definition 36.1, page 50).

113.1 Definition: number of elements of a finite set
Let n be a nonnegative integer. The statement, "A set S has n elements" means there is a bijection $F : \mathbf{n} \to S$.

113.1.1 Example A set has 5 elements if there is a bijection from $\{1,2,3,4,5\}$ to the set. Thus the formal definition captures the usual meaning of number of elements: if a set has 5 elements, the process of counting them — "This is the first element, this is the second element, ..." — in effect constructs a bijection from $\mathbf{n}$ to the set.

113.1.2 Exercise Give an explicit proof that the set of positive divisors of 8 has 4 elements. (Answer on page 249.)

113.2 Definition: finite
A **finite set** is a set with n elements, where n is some nonnegative integer.

113.2.1 Example The empty set is finite, since it has 0 elements, and the set $\{1,3,5,7,9\}$ is finite because it has 5 elements.

113.3 Definition: cardinality
The number of elements of a finite set is the **cardinality** of the set. For any finite set A, the cardinality of A is denoted $|A|$.

113.3.1 Example $|\emptyset| = 0$ and $|\{1,3\}| = 2$.

113.3.2 Exercise Show that if A is a finite set and $\beta : B \to A$ is a bijection then B is finite.

113.3.3 Exercise Show that a subset of a finite set is finite. Make sure you use the definition of finite in your proof.

113.4 Infinite sets

A set which is not finite is **infinite**. Sets such as N, Z and R are infinite. Since "infinite" merely means "not finite", to say that R (for example) is infinite means just that there is no nonnegative integer n for which the statement "R has n elements" is true. This is certainly correct in the case of R, since if you claim (for example) that R has 42 elements, all I have to do is add up the absolute values of those 42 numbers to get a number which is bigger than all of them, so is a 43rd element.

We do not go into the extensive theory of infinite sets in this book, but it is important to understand the difference between "finite" and "infinite" since many theorems, such as the ones in this section, concern only finite sets.

113.4.1 Warning It is tempting when faced with proving a theorem about possibly infinite sets to talk about one set having "more elements than another". Such arguments are often fallacious. For example: "There cannot possibly be an injective function from $\mathrm{N}\times\mathrm{N}$ to N since $\mathrm{N}\times\mathrm{N}$ has more elements than N." But there are such functions: see Exercises 93.1.8 and 113.5.3. Compare the extended hint to Exercise 92.1.8.

113.5 Exercise set

A set S is **countably infinite** if there is a bijection $\beta:\mathrm{N}\to S$. Problems 113.5.1 through 113.5.4 explore this property.

113.5.1 Exercise Show that the set N^+ of positive integers is countably infinite. (Answer on page 249.)

113.5.2 Exercise Show that Z is countably infinite.

113.5.3 Exercise Show that $\mathrm{N}\times\mathrm{N}$ is countably infinite.

113.5.4 Exercise (hard) Show that Q is countably infinite.

114. Multiplication of Choices

The principle of multiplication of choices, stated below, is behind the sort of reasoning illustrated in the following argument: You are at a restaurant whose menu has three columns, A, B and C. To have a complete meal, you order one of the three items in column A, one of the five items in column B, and one of the three items in column C. You can therefore choose $45 = 3\times 5\times 3$ different meals.

114.1 Definition: independent tasks

Suppose that there are k tasks T_1, $T_2,\ldots,T_k$ which must be done in order, and, for each $i = 1,2,\ldots,k$, there are n_i ways of doing task T_i. Suppose furthermore that doing T_i in any particular way does not change the number n_j ways of doing any later task T_j. Then we say that the tasks are **independent** of each other.

114.2 Theorem: The Principle of Multiplication of Choices

Suppose there are k independent tasks T_i $(i = 1, \ldots, k)$ and suppose that for each i there are n_i ways of doing T_i $(i = 1, \ldots, k)$. Then there are $\prod_{i=1}^{k} n_i = n_1 n_2 \cdots n_k$ ways of doing the tasks $T_1, \ldots, T_k$ in order.

Proof We prove Theorem 114.2 by induction on k, starting at 1.

If you have one task T_1 which can be done in n_1 different ways, Theorem 114.2 says you can do T_1 in $\prod_{i=1}^{1} n_i = n_1$ different ways, which of course is true.

Now suppose the theorem is true for k tasks. Assume you have $k+1$ tasks $T_1, \ldots, T_k, T_{k+1}$, and for each i there are n_i ways of doing task T_i. Let m be the total number of ways of doing the tasks $T_1, \ldots, T_k$ in order. Suppose you have done them in one of the m ways. Then you can do T_{k+1} in any of n_{i+1} ways. Thus for each of the m ways of doing the first k tasks, you have n_{i+1} ways of doing the $(k+1)$st; therefore, there are altogether $n_{i+1} + n_{i+1} + \cdots + n_{i+1}$ (sum of m terms) ways of doing the $k+1$ tasks. This means that there are $m \times n_{i+1}$ ways to do $T_1, \ldots, T_{k+1}$ in order.

By induction hypothesis, $m = \prod_{i=1}^{k} n_i$, so the number of ways of doing the tasks $T_1, \ldots, T_{k+1}$ is

$$n_{k+1} \cdot (\prod_{i=1}^{k} n_i)$$

which by 105.1.5 is $\prod_{i=1}^{k+1} n_i$, as required.

114.2.1 Worked Exercise How many three-digit integers (in decimal notation) are there whose second digit is not 5?

Answer Writing such a sequence of digits can be perceived as carrying out three tasks in a row:

T.1 Write any digit except 0.
T.2 Write any digit except 5.
T.3 Write any digit.

There are 9 ways to do T.1, 9 ways to do T.2, and 10 ways to do T.3, so according to Theorem 114.2, there are 810 ways to do T.1, T.2, T.3 in order.

114.2.2 Worked Exercise Find the number of strings of length n in $\{a, b, c\}^*$ that contain exactly one a.

Answer This requires us to look at the problem in a slightly different way from Worked Exercise 114.2.1. To construct a string of length in $\{a, b, c\}^*$ with exactly one a requires us to

a) Choose which of n possible locations to put the one and only a (n ways to do this).
b) For each of the $n-1$ other locations, choose whether to put a b or c there (2 choices for each location, 2^{n-1} choices altogether).

It follows that there are $n \cdot 2^{n-1}$ such strings.

114.2.3 Exercise Find the number of 5-digit integers with '3' in the middle place. (Answer on page 249.)

114.2.4 Exercise Find the number of even 5-digit integers. (Answer on page 249.)

114.3 Exercise set

In exercises 114.3.2 through 114.3.5, $A = \{a, b, c\}$.

114.3.1 Exercise Find the number of strings of length n in $\{a, b, c\}^*$ with *no* a's. (Answer on page 249.)

114.3.2 Exercise Find a formula $F(n)$ for the number of strings in A^* of length n, for each $n \in \mathrm{N}$. (Answer on page 249.)

114.3.3 Exercise Find a formula $G(n)$ for the number of strings in A^* of length n which begin and end with a. (Answer on page 249.)

114.3.4 Exercise Find a formula $H(n)$ for the number of strings in A^* of length n which do not begin or end with c.

114.3.5 Exercise Find a formula for the number of strings in A^* of length $n > 2$ which have a 'a' in the third place.

114.3.6 Exercise In the USA a local telephone number consists of a string of 7 digits, the first two of which cannot be 0 or 1. How many possible local telephone numbers are there?

115. Counting with set operations

Almost every operation associated with set theory has a corresponding combinatorial principle or counting technique applicable to finite sets associated with it. Some of these are obvious, others are more subtle. The first example has to do with inclusion:

115.1 Theorem
If A and B are finite sets and $A \subseteq B$, then $|A| \leq |B|$.

(We *told* you some of the principles were obvious!)

115.1.1 Exercise Show that if A and B are finite then $|A \cap B| \leq |A|$.

There is a principle for powersets, too.

115.2 Theorem
If a set A has n elements then $\mathcal{P}A$ has 2^n elements.

Proof The easiest proof of this theorem uses the Principle of Multiplication of Choices (Theorem 114.2). If A has n elements and you want to describe a subset of A, you may go through the n elements of A one by one and say whether each one is in the subset. There are two choices (yes or no) for each element and n elements, so the Principle of Multiplication of Choices says that you can make 2^n choices altogether.

115.2.1 Remark As is the case with any counting technique based on the Principle of Multiplication of Choices, it is also possible to prove Theorem 115.2 by a direct argument using induction. (Recall that the Principle of Multiplication of Choices was proved by induction.)

115.2.2 Worked Exercise How many subsets with an even number of elements does a set with n elements have? Explain your answer.

Answer A set S with n elements has 2^{n-1} subsets with an even number of elements. Proof: To give a subset A of S, for each element of S *except the last one* you must choose whether that element is in A. That requires 2^{n-1} independent choices. You have no choice concerning the last element: if at that point the subset has an odd number of elements so far, you *must* include the last one, and if it has an even number so far, you *must not* include the last one.

115.2.3 Exercise Let S be an n-element set. How many elements do the following sets have?

a) The set of nonempty subsets of S.
b) The set of singleton subsets of S.
c) The set of subsets of the powerset of S.

(Answer on page 249.)

The following theorem can be proved using Multiplication of Choices.

115.3 Theorem
If A and B are finite, then $|A \times B| = |A|\,|B|$.

115.3.1 Exercise If A has m elements and B has n elements, how many elements do each of these sets have?

a) $A \times A$
b) $\mathcal{P}(A \times A)$
c) $\mathcal{P}(A \times B)$

115.3.2 Exercise Prove Theorem 115.1.

115.3.3 Exercise Prove Theorem 115.3.

115.3.4 Exercise Suppose A has m elements and B has n elements.

a) Prove that $MAX(m,n) \leq |A \cup B| \leq m+n$.
b) Prove that $0 \leq |A \cap B| \leq MIN(m,n)$.
c) Prove that the symbols '$\leq$' in (a) and (b) cannot be replaced by '$<$'.

115.3.5 Exercise Let A be a finite set and $F: A \to B$ a function. Prove that $|\Gamma(F)| = |A|$.

116. The Principle of Inclusion and Exclusion

116.1 Theorem

Let A and B be finite sets. Then

$$|A \cup B| = |A| + |B| - |A \cap B| \tag{116.1}$$

Proof This follows from the fact that the expression $|A| + |B|$ counts the elements which are in *both* sets twice, so to get the correct count for $|A \cup B|$, you have to subtract $|A \cap B|$.

116.1.1 Remark More generally, if C and D are also finite sets, then

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C| \tag{116.2}$$

and

$$\begin{aligned} |A \cup B \cup C \cup D| \quad = \quad & |A| + |B| + |C| + |D| \\ & - |A \cap B| - |A \cap C| - |A \cap D| \\ & - |B \cap C| - |B \cap D| - |C \cap D| \\ & + |A \cap B \cap C| + |A \cap B \cap D| \\ & + |A \cap C \cap D| + |B \cap C \cap D| \\ & - |A \cap B \cap C \cap D| \end{aligned} \tag{116.3}$$

116.1.2 The general principle Equations (116.1)–(116.3) are special cases of a general principle which requires some notation to state properly. Let $\mathcal{F}$ be a family of n distinct finite sets. For each $k = 1, 2, \ldots, n$, let $\mathcal{F}_k$ be the set of k-element subsets of $\mathcal{F}$. For example, if $\mathcal{F} = \{A, B, C, D\}$, then

$$\mathcal{F}_3 = \{\{A,B,C\}, \{A,B,D\}, \{A,C,D\}, \{B,C,D\}\}$$

Then we have:

116.2 Theorem: The Principle of Inclusion and Exclusion

Using the notation of Section 116.1.2,

$$\begin{aligned} |\cup\mathcal{F}| &= \sum_{X\in\mathcal{F}} |X| - \sum_{G\in\mathcal{F}_2} |\cap G| + \sum_{G\in\mathcal{F}_3} |\cap G| - \dots \\ & \quad -(-1)^k \sum_{G\in\mathcal{F}_k} |\cap G| + \dots - (-1)^n |\cap\mathcal{F}| \end{aligned} \tag{116.4}$$

116.2.1 Remarks

a) The first sum is over the elements of $\mathcal{F}$ (which are themselves sets), whereas the others are over intersections of *subfamilies* G of $\mathcal{F}$, with a plus sign for subfamilies with an odd number of elements and a minus sign for those with an even number of elements.
b) You should check that Equations (116.1)–(116.3) are special cases of this Principle.
c) The Principle of Inclusion and Exclusion will not be proved here, but you should be able to see with no trouble why it is true for families of three or four sets.

116.2.2 Example The Principle of Inclusion and Exclusion is stated as an equation, so you can solve for one of its terms if you know all the others.

For example, suppose there was a party with 9 people, including 5 Norwegians. There was only one man at the party who was neither a vegetarian nor a Norwegian. All the vegetarians were Norwegians and two of the women were Norwegians. Exactly one woman was a vegetarian. How many women were at the party?

To solve this, let W be the set of women, N the set of Norwegians, and V the set of vegetarians. The party had 9 people, and only one was not in $W \cup N \cup V$, so $|W \cup N \cup V| = 8$. We are given that $|N| = 5$. Since 2 of the women were Norwegians, $|W \cap N| = 2$, and since one woman was a vegetarian and every vegetarian was a Norwegian, we know $|W \cap V| = |W \cap N \cap V| = 1$ and also $|V| = |N \cap V|$.

Thus in the sum

$$\begin{aligned} |W \cup N \cup V| &= |W| + |N| + |V| - \\ & \quad |W \cap N| - |W \cap V| - |N \cap V| + |W \cap N \cap V| \end{aligned}$$

we have

$$8 = |W| + 5 + |V| - 2 - |W \cap V| - |N \cap V| + |W \cap N \cap V|$$

or since $|V| = |N \cap V|$,

$$8 = |W| + 5 - 2 = |W| + 3$$

so that there were 5 women at the party.

116.2.3 Exercise You have a collection of American pennies. Three of them are zinc pennies and eight of them were minted before 1932. What do you have to know to determine the total number of pennies? Explain your answer! (Answer on page 249.)

116.2.4 Exercise A, B and C are finite sets with the following properties: $A \cup B \cup C$ has 10 elements; B has twice as many elements as A; C has 5 elements; B and C are disjoint; and there is just one element in A that is also in B. Show that A has at least 2 elements.

116.2.5 Exercise Suppose that A, B and C are finite sets with the following properties:
(i) B has one more element than A.
(ii) C has one more element than B.
(iii) $A \cap B$ is twice as big as $A \cap C$.
(iv) B and C have no elements in common.
Prove that $|A \cup B \cup C|$ is divisible by 3.

116.2.6 Exercise Cornwall Computernut has 5 computers with hard disk drives and one without. Of these, several have speech synthesizers, including the one without hard disk. Several have Pascal, including all those with synthesizers. Exactly 3 of the computers with hard disk have Pascal. How many have Pascal?

117. Partitions

117.1 Definition: partition
If C is a set, a family Π of *nonempty* subsets of C is called a **partition** of C if
PAR.1 $C = \cup\Pi$, and
PAR.2 For all $A, B \in \Pi$, $A \neq B \Rightarrow A \cap B = \emptyset$.

117.1.1 Usage The elements of the partition Π are called the **blocks** of Π. If $x \in C$, the block of Π that has x as an element is denoted $[x]_\Pi$, or just $[x]$ if the partition is clear from context.

117.1.2 Fact P.2 says the blocks of Π (remember that these subsets of C) are **pairwise disjoint**: if they are different, they can't overlap.

117.1.3 Fact P.1 and P.2 together are equivalent to saying that *every element of* C *is in exactly one block of* Π.

117.1.4 Example Here are three partitions of the set $\{1, 2, 3, 4, 5\}$:
a) $\Pi_1 = \{\{1, 2\}, \{3, 4\}, \{5\}\}$.
b) $\Pi_2 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$.
c) $\Pi_3 = \{\{1, 2, 3, 4, 5\}\}$.

117.1.5 Example The set $\{\{1,2\},\{3,4\},\{5\},\emptyset\}$ is not a partition of any set because it contains the empty set as an element.

117.1.6 Example Let S be any nonempty set and A any proper nontrivial subset of A. Then $\{S, S-A\}$ is a partition of A with two blocks.

117.1.7 Example The empty set has a unique partition which is also the empty set. It has no blocks.

117.1.8 Exercise Why does Example 117.1.6 have to require that A be a proper nontrivial subset of A?

117.1.9 Worked Exercise Let S be a nonempty finite set with n elements. How many partititions of S with exactly two blocks are there?
Answer There are $2^n - 1$ nonempty subsets of S and, except for S itself, each one induces a two-block partition as in Example 117.1.6. This does *not* mean that there are $2^n - 2$ two-block partitions because that would count each two-block partition twice (a subset and its complement each induce the same two-block partition). So the correct answer is that there are

$$\frac{1}{2}(2^n - 2) = 2^{n-1} - 1$$

two-block partitions.

117.1.10 Warning One of the commonest mistakes made by people just beginning to learn counting is to come up with a seemingly reasonable technique which unfortunately counts some things more than once.

117.1.11 Exercise Find a formula for the number of partitions with exactly three blocks of an n-element set.

117.1.12 Usage A partition with a finite number of blocks (even though the blocks might be infinite sets) is commonly written as a tuple, e.g., $\Pi = \langle A_i \rangle_{i \in \mathbf{n}}$. Even so, if Π' is another partition which is the same as Π except for ordering, they are regarded as the same partition even though they are different tuples. We will follow that practice here.

117.1.13 Exercise Which of the following are partitions of $S = \{1,2,3,4,5\}$? Here, $A = \{1,2\}$, $B = \{3,4,5\}$, $C = \{3\}$, $D = \{4,5\}$.

a)	$\{A, B\}$	e)	$\{S\}$
b)	$\{A, B, C\}$	f)	$\{\{x\} \mid x \in S\}$
c)	$\{A, C, D\}$	g)	$\{C, S - C\}$
d)	$\{A, B, \emptyset\}$	h)	$\{A \cup C, D\}$

(Answer on page 250.)

117.2 Partition of Z by remainders

Any poesitive integer n induces a very important partition of the set Z of integers. This partition is denoted Z/n. The blocks of Z/n are the n sets

$$C_r = \{m \in \mathrm{Z} \mid m \text{ leaves a remainder of } r \text{ when divided by } n\}$$

for $0 \le r < n$. For negative m floored division must be used. (Observe that the notation "C_r" requires you to depend on context to know what n is.) Thus $\mathrm{Z}/n = \{C_r \mid 0 \le r < n\}$.

117.2.1 Remark It is important to understand that Z/n is a *finite set,* even though each block is an infinite set.

117.2.2 Example If $n = 3$, $\mathrm{Z}/3$ has three blocks. One of them is C_1, which is the set of integers which leave a remainder of 1 when divided by 3. Thus 1, -2 and 16 are in C_1. C_0 is the set of integers divisible by 3. Thus $\mathrm{Z}/3 = \{C_0, C_1, C_2\}$.

117.3 Exercise set

In problems 117.3.1 through 117.3.5, provide an example of a partition Π of Z with the given property.

117.3.1 Π has at least one block with exactly three elements. (Answer on page 250.)

117.3.2 $\{1,2\}$ and $\{3\}$ are blocks of Π.

117.3.3 Π has at least one finite block and at least one infinite block.

117.3.4 Π has an infinite number of finite blocks.

117.3.5 Π has an infinite number of infinite blocks.

118. Counting with partitions

P.2 in Definition 117.1 implies that, in the statement of the Principle of Inclusion and Exclusion, the sums over families with more than one element disappear. This gives the following theorem, which is obvious anyway:

118.1 Theorem
If $\Pi = \langle A_i \rangle_{i \in \mathbf{n}}$ is a partition of a finite set C, then $|C| = \Sigma_{i=1}^{n} |A_i|$.

This Theorem together with the phenomenon of Example 117.1.6 gives a method:

118.1.1 Method
To count the number of elements of a subset A of a set S, count the number of elements of S and subtract the number of elements of the complement $S-A$.

118.1.2 Worked Exercise How many strings of length n in $\{a,b,c\}^*$ are there that have more than one a?

Answer We will use Method 118.1.1. We know from Exercises 114.2.2 and 114.3.1 that there are 2^n strings with no a and $n\cdot 2^{n-1}$ strings with one a. Since there are 3^n strings of length n in $\{a,b,c\}^*$, the answer is $3^n-2^n-n\cdot 2^{n-1}$.

118.1.3 Exercise How many strings of length n in $\{a,b\}^*$ are there that have more than one a?

118.1.4 Exercise How many strings of length n in $\{a,b\}^*$ are there that satisfy the following requirement: If it has an a in it, it has at least two.

118.1.5 Exercise How many strings of length n in $\{a,b,c\}^*$ are there that have exactly two different letters in them (so each one is either all a's and b's, all a's and c's, or all b's and c's.)?

118.1.6 Exercise In the USA the identifying name of a radio station consist of strings of letters of length 3 or 4, beginning with K or W. Upper and lower case are not distinguished. How many legal identifying names are there?

119. The class function

119.1 Definition: the class function
If Π is a partition of a set A, then the **class function** $\mathrm{cls}_\Pi : A\to\Pi$ takes an element a of A to the block of Π that has it as an element.

119.1.1 Example If $A=\{1,2,3,4,5\}$ and $\Pi=\{\{1,2\},\{3,4,5\}\}$, then $\mathrm{cls}_\Pi(3)=\{3,4,5\}$.

119.1.2 Usage A common notation for the class function is $[\,]: A\to\Pi$; in Example 119.1.1, one would write $[3]=\{3,4,5\}$.

119.1.3 Example In Example 117.1.4, $[2]_{\Pi_1}=\{1,2\}$.

119.1.4 Warning Note that in Example 119.1.1, $[3]=[4]=[5]$, but $[2]\neq[3]$. *In mathematics, the fact that two different names are used does not mean they name different things.* (This point was made before, in Example 58.1.2.)

119.1.5 Example If Π is the partition $\mathrm{Z}/3$, then $[2]=[5]=[-1]=C_2$, and $[3]=C_0$.

119.1.6 Exercise Prove that for any set S with partition Π, the class function $\mathrm{cls}: S\to\Pi$ is surjective.

120. The quotient of a function

We mentioned the partition $\mathrm{Z}/n = \{C_r \mid 0 \leq r < n\}$ in section 117.2. It is a special case of a construction which works for any function:

120.1 Theorem
Let $F: A \to B$ be a function. Then the family of sets

$$\{F^{-1}(b) \mid b \in \operatorname{Im} F\}$$

is a partition of A.

120.2 Definition: quotient set
The set $\{F^{-1}(b) \mid b \in \operatorname{Im} F\}$ is denoted A/F and is called the **quotient set** of F.

120.2.1 Example Consider the function $F: \{1,2,3\} \to \{2,4,5,6\}$ defined by $F(1) = 4$ and $F(2) = F(3) = 5$. Its quotient set (of a function) is $\{\{1\},\{2,3\}\}$.

120.2.2 Example The quotient set (of a function) of the squaring function $S: \mathrm{R} \to \mathrm{R}$ defined by $S(x) = x^2$ is

$$\mathrm{R}/S = \{\{r, -r\} \mid r \in \mathrm{R}\}$$

Every block of R/S has two elements with the exception of the block $\{0\}$. The notation "$\{\{r,-r\} \mid r \in \mathrm{R}\}$" for R/S lists $\{0\}$ as $\{0,-0\}$, but that is the same set as $\{0\}$. Note that every set except $\{0\}$ is listed twice in the expression "$\{\{r,-r\} \mid r \in \mathrm{R}\}$".

120.2.3 Example Let's look at the remainder function $R_n(k) = k \bmod n$ for a fixed integer n. This function takes an integer k to its remainder when divided by n. (As earlier, we use floored division for negative k). For a particular remainder r, the set of integers which leave a remainder of r when divided by n is the set we called C_r earlier in the section. Thus the quotient set of R_n is the set we called Z/n.

120.3 Proof of Theorem 120.1

We must show that the blocks of A/F are nonempty and that every element of A is in exactly one block of A/F.

That the blocks are nonempty follows the fact that A/F consists of those $F^{-1}(b)$ for which $b \in \operatorname{Im} F$; if $b \in \operatorname{Im} F$, then there is some $a \in A$ with $F(a) = b$, which implies that $a \in F^{-1}(b)$, so that $F^{-1}(b)$ is nonempty. Since $a \in F^{-1}(F(a))$, every element of A is in at *least* one block. If $a \in F^{-1}(b)$ also, then $F(a) = b$ by definition, so $F^{-1}(F(a)) = F^{-1}(b)$, so no element is in more than one block.

120.3.1 Exercise For a function $F: S \to T$, define a condition on the quotient set S/F which is true if and only if F is injective. (Answer on page 250.)

120.3.2 Exercise Give examples of two functions $F:\mathrm{N}\to\mathrm{N}$ and $G:\mathrm{N}\to\mathrm{N}$ with the property that F is surjective, G is not surjective and F and G have the same quotient set. (Thus, in contrast to Exercise 120.3.1, there is no condition on the quotient set of a function that forces the function to be surjective.)

120.4 Exercise set

In Problems 120.4.1 through 120.4.5, provide an example of a function $F:\mathrm{R}\to\mathrm{R}$ for which R/F has the given property.

120.4.1 R/F has at least one block with exactly three elements. (Answer on page 250.)

120.4.2 R/F has exactly three blocks.

120.4.3 R/F is finite.

120.4.4 Every block of R/F is finite.

120.4.5 Every block of R/F has exactly two elements.

120.4.6 Exercise Suppose $F:A\to B$ is a function, and x and y are distinct elements of B. Suppose also that $|A|=7$, $|B|=4$, $\operatorname{Im}F=B-\{y\}$, and that the function $F|\big(A-F^{-1}(x)\big)$ is injective.

a) How many elements does A/F have?
b) How many elements are there in each block of A/F?

120.4.7 Exercise (hard) Let A be a set, Π a partition of A and B a subset of A. Define the set $\Pi|B$ of subsets of B by

$$\Pi|B=\{C\cap B \mid C\in\Pi \text{ and } C\cap B\neq\emptyset\}$$

a) Prove that $\Pi|B$ is a partition of B.
b) Give an example to show that the set $\{C\cap B \mid C\in\Pi\}$ need not be a partition of B.

120.4.8 Exercise (hard) Let A be a set, Π a partition of A, and Φ a partition of Π. For any block $C\in\Phi$, let B_C be the union of all the blocks $B\in\Pi$ for which $B\in C$. Show that $\{B_C \mid C\in\Phi\}$ is a partition of A. (For many people, this exercise will be an excellent example of a common phenomenon in conceptual mathematics: It seems incomprehensible at first, but when you finally figure out what the notation means, you see that it is *obviously* true.)

121. The fundamental bijection theorem

The following theorem forms a theoretical basis for very important constructions in abstract mathematics:

121.1 Theorem: The Fundamental Bijection Theorem for functions
Let $F: A \to B$ be a function, and define β_F to be the function $F^{-1}(b) \mapsto b$. Then β_F is a bijection $\beta_F: A/F \to \operatorname{Im} F$.

121.1.1 Example For the function $F: \{1,2,3\} \to \{2,4,5,6\}$ defined by $F(1) = 4$ and $F(2) = F(3) = 5$, we have $\beta_F(\{1\}) = 4$ and $\beta_F(\{2,3\}) = 5$.

121.1.2 Remark The input to the bijection is a *set*, namely a block of A/F, and the output is an *element of the codomain of* F. The statement that $\beta_F(\{2,3\}) = 5$ means that when you plug $\{2,3\}$ into β_F (*not* when you plug 2 in or 3 in!) you get 5.

121.2 Proof of Theorem 121.1

It is easy to see that β_F really is a bijection. If $b \in \operatorname{Im} F$, then there is some element $a \in A$ for which $F(a) = b$, so $F^{-1}(b)$ is nonempty and hence an element of A/F. Then $\beta_F(F^{-1}(b)) = b$ so β_F is surjective.

Proving injectivity reduces to showing that if $F^{-1}(b) \neq F^{-1}(c)$, then $b \neq c$. If $F^{-1}(b) \neq F^{-1}(c)$, then there is some element $a \in A$ for which $a \in F^{-1}(b)$ but $a \notin F^{-1}(c)$ (or vice versa). The statement $a \in F^{-1}(b)$ means that $F(a) = b$, and the statement $a \notin F^{-1}(c)$ means that $F(a) \neq c$. Thus $b \neq c$, as required.

121.2.1 Exercise Let $A = \{1,2,3,4,5\}$. For each function $F: A \to \mathrm{R}$ given below, write out all the values of the bijection $\beta_F: A/F \to \operatorname{Im} F$ given by Theorem 121.1.

a) $F(1) = F(3) = F(5) = 4$, $F(4) = 6$, $F(2) = 0$.
b) $F(n) = 3$ for all $n \in A$.
c) $F(n) = n$ for all $n \in A$.
d) $F(n) = n^2$ for all $n \in A$.
e) $F(n) = n^3 - 3n^2 + 2n - 5$ for all $n \in A$.

(Answer on page 250.)

122. Elementary facts about finite sets and functions

This chapter contains miscellaneous results, mostly easy, concerning finite sets and functions between them. The facts about finite sets A and B in the following theorem are not difficult to see using examples. We give part of the proof and leave the rest to you.

122.1 Theorem
Let A and B be finite sets. Then:
a) $|A| = |B|$ if and only if there is a bijection $\beta : A \to B$.
b) $|A| \leq |B|$ if and only if there is an injective function $F : A \to B$.
c) If B is nonempty, $|A| \geq |B|$ if and only if there is a surjective function $G : A \to B$.

Proof By Definition 113.1, if A and B both have n elements then there are bijections $\beta : \mathbf{n} \to A$ and $\beta' : \mathbf{n} \to B$. Then, using Theorem 101.5, page 149, Theorem 101.3, page 148 and Exercise 98.2.7 of Chapter 98, $\beta' \circ \beta^{-1} : A \to B$ is a bijection. To finish the proof of (a), we must show that if there is a bijection $\beta : A \to B$ then A and B have the same number of elements. This is left as an exercise.

We also leave (b) as an exercise, and prove half of (c). Suppose A has m elements and B has n elements with $m \geq n > 0$. Then there are bijections $\beta : \mathbf{m} \to A$ and $\beta' : \mathbf{n} \to B$. Let us define a function $F : \mathbf{m} \to \mathbf{n}$ by: $F(k) = k$ if $k < n$, and $F(k) = n$ if $k \geq n$. F is surjective, because if $1 \leq i \leq n$, then $F(i) = i$. Then $\beta' \circ F \circ \beta^{-1} : A \to B$ is the composite of a bijection, a surjection and a bijection, so is a surjection by Exercise 98.2.7 of Chapter 98.

122.1.1 Exercise Complete the proof of Theorem 122.1.

122.1.2 Exercise Use the principles of counting for finite sets that we have introduced to prove that if Π is a partition of a finite set A, then $|\Pi| \leq |A|$.

Here is another useful theorem:

122.2 Theorem
If A and B are finite sets and $|A| = |B|$, then a function $F : A \to B$ is injective if and only if it is surjective.

Proof Let $F : A \to B$ be injective. Then $\operatorname{Im} F$, being a subset of B, has no more than $|B|$ elements by Theorem 115.1. Since F is injective, $\operatorname{Im} F$ has at least $|A|$ elements by Theorem 122.1(a). Since $|A| = |B|$, it follows that $\operatorname{Im} F$ has exactly $|B|$ elements, so $\operatorname{Im} F = B$. Hence F is surjective.

Conversely, if F is *not* injective, then the quotient A/F has fewer elements than A. The fundamental bijection theorem (Theorem 121.1) says that then $\operatorname{Im} F$ has fewer elements than A, so it has fewer elements than B since $|A| = |B|$. That means $\operatorname{Im} F \neq B$, so F is not surjective.

122.2.1 Warning Observe that if $|A| = |B|$, then Theorem 122.1(a) says there is an injection from A to B and Theorem 122.1(b) says that there is a surjection from A to B. But Theorems 122.1(a) and (b) *do not say that the injection and the surjection have to be the same function,* so it would be a fallacy to deduce Theorem 122.2 from those two facts.

122.2.2 Warning Theorem 122.2 allows you to determine whether a function from a finite set to itself is a bijection by testing either injectivity or surjectivity — you don't have to test both. However, *you have to test both for infinite sets.* For example, the **shift function** $n \mapsto n+1 : \mathrm{N} \to \mathrm{N}$ is injective but not surjective (0 is not a value) and $0 \mapsto 0$, $n \mapsto n-1$ for $n > 0$ defines a function $\mathrm{N} \to \mathrm{N}$ which is surjective but not injective, since 0 and 1 both have value 0.

Here is a counting principle for function sets:

122.3 Theorem
If $|A| = n$ and $|B| = m$, then there are m^n functions from A to B. In other words, $\left|B^A\right| = |B|^{|A|}$.

Proof To construct an element of B^A, that is, a function from A to B, you have to say what $F(a)$ is for each element of A. For each a you have m choices for $F(a)$ since $F(a)$ has to be an element of B and B has m elements. There are n elements a of A for each of which you have to make these choices, so by the Principle of Multiplication of Choices there are m^n possibilities altogether.

122.3.1 Exercise How many ways are there of assigning a letter of the alphabet to each decimal digit, allowing the same letter to be assigned to different digits? (Answer on page 250.)

122.3.2 Exercise

a) Show by quoting principles enunciated here that if A and B are finite, $A \subseteq B$ and $A \neq B$, then there is no bijection from A to B.
b) Show that the statement in (a) can be false if A and B are infinite.

122.3.3 Exercise Let $F(n)$ be the number of functions from $\mathcal{P}S$ to S, where S is a set with n elements, and let $G(n)$ be the number of functions from S to its powerset. For which integers n is $F(n) = G(n)$?

123. The Pigeonhole Principle

In its contrapositive form, Theorem 122.1(b) says the following:

123.1 Theorem
For any finite sets A and B, if $|A| > |B|$, then no function from A to B is injective.

123.1.1 Example If you have a set A of pigeons and a set B of pigeonholes, $|A| > |B|$, and you put each pigeon in a pigeonhole (thereby giving a function from A to B), then at least one pigeonhole has to have two pigeons in it (the function is not injective). For this reason, Theorem 123.1 is called the **Pigeonhole Principle**.

123.1.2 Example An obvious example of the use of the Pigeonhole Principle is that in any room containing 367 people, two of them must have the same birthday. Note that the Pigeonhole Principle gives you no way to find out who they are.

123.1.3 Worked Exercise Let $S = \{n : \mathrm{N} \mid 1 \leq n \leq 10\}$. Show that any subset T of S with more than 5 elements contains two numbers that add up to 11.

Answer The following are all the two-element subsets of S whose elements add up to 11: $\{1,10\}$, $\{2,9\}$, $\{3,8\}$, $\{4,7\}$, $\{5,6\}$. They form a partition of S with five blocks. Every element of T is in one of these subsets, and since T has more than five elements, by the Pigeonhole Principle two different elements must be in the same block of the partition.

123.1.4 Exercise Let S be as in Worked Exercise 123.1.3. Show that if $T \subseteq S$ and $|T| \geq 4$ then there are two different elements of T that have the same remainder when divides by 3.

123.1.5 Exercise Let $A = \{n : \mathrm{N} \mid 1 \leq n \leq 12\}$. Find the least integer n so that the following statement is true: If $T \subseteq A$ and $|T| \geq n$, then T contains two distinct elements whose product is 12.

124. Recurrence relations in counting

Many counting formulas can be derived as recurrence relations. In many cases, you can then find a closed formula which evaluates the recurrence relation, but even if you cannot do that, the recurrence relation gives you a way of evaluating the formula for successive values of n.

124.1 Theorem
If A has n elements, then there are $n!$ different permutations of A.

To prove this, it is useful to prove something more general.

124.2 Theorem
The number of bijections between two n-element sets is $n!$.

Proof Let $P(n)$ be the number of bijections between two n-element sets. Then $P(0) = P(1) = 1$. Let A and B be two sets with $n+1$ elements. Let $a \in A$. Then in constructing a bijection from A to B we have $n+1$ choices for the value of the bijection at a. If we choose $b \in B$, then what is left is a bijection from $A - \{a\}$ to $B - \{b\}$. These are both n-element sets, so there are $P(n)$ of these, by definition of $P(n)$. Hence

$$P(n+1) = (n+1) \cdot P(n)$$

This is the recurrence relation which (with $P(0) = 1$) defines $n!$ (see Section 105.1.6, page 158), so $P(n) = n!$.

Here is another example of using recurrences in counting:

124.2.1 Worked Exercise Derive a formula or recurrence relation for the number of strings of length n in $\{0,1\}^*$ with an even number of 1's.
Answer Let $F(n)$ be the number of such strings. Obviously $F(0) = F(1) = 1$. There are $F(n)$ strings of length n with an even number of ones and $2^n - F(n)$ with an odd number of ones. (Note that there is no justification at this point for assuming that the number of strings of length n with an even number of ones and the number with an odd number of ones are the same.) You can adjoin a 0 to a string of the first type and a 1 to a string of the second type to get a string of length $n+1$ with an even number of ones. Thus $F(n+1) = F(n) + 2^n - F(n) = 2^n$. This is a case of a recurrence relation that solves itself!

124.2.2 Exercise Derive a formula or recurrence relation for the number of ways to arrange n people around a circular table. (All that matters is who sits on each person's left and who sits on his or her right.)

124.2.3 Exercise Derive a formula or recurrence relation for the amount of money in a savings account after n years if the interest rate is $i\%$ compounded annually and you start with \$100.

125. The number of subsets of a set

125.1 Definition: binomial coefficient
$C(n,k)$ denotes the number of k-element subsets of an n-element set.

125.1.1 Example $C(4,0) = 1$ (there is exactly one subset with no elements in a set with 4 elements), $C(4,1) = 4$ (there are four singleton subsets of a four-element set) and $C(4,2) = 6$ (count them).

We can deduce some immediate consequences of the definition:

125.2 Theorem

For all $n \geq 0$ and $k \geq 0$,

a) $C(n,0) = 1$.
b) $C(n,n) = 1$.
c) $C(n,k) = C(n,n-k)$.
d) $C(n,k) = 0$ *if* $k > n$.

Proof

a) There is exactly one empty subset of any set, so $C(n,0) = 1$ for any n.
b) An n-element set clearly has exactly one subset with n elements, namely itself.
c) This follows from the fact that for a particular k there is a bijection between k element subsets of an n-element set and their complements, which of course are $(n-k)$-element subsets.
d) Obvious.

$C(n,k)$ is called a **binomial coefficient** because of the formula in the following theorem. $C(n,k)$ is also written $\binom{n}{k}$.

125.3 Theorem

For all real x and y and all nonnegative integers n and k,

$$(x+y)^n = \sum_{k=0}^{n} C(n,k) x^{n-k} y^k \tag{125.1}$$

I won't give a formal proof, but just sketch the idea.

$$(x+y)^n = (x+y)(x+y)\cdots(x+y) \tag{125.2}$$

where $(x+y)$ occurs n times in the expression on the right. In the expanded version of $(x+y)^n$, each term occurs by selecting an x or a y in each factor of the right side of Equation (125.2) and multiplying them together (try this on $(x+y)(x+y)(x+y)$). You get one occurrence of $x^{n-k}y^k$ by choosing a subset of k factors (out of the n that occur) and using y from those factors and x from the $n-k$ other factors. There are $C(n,k)$ ways to do this, so that Equation (125.1) follows.

125.4 Recurrence relation for $C(n,k)$

We can get a recurrence relation for $C(n,k)$ which will allow us to calculate it.

Suppose, for a fixed k, we want to know $C(n+1,k)$, the number of k-element subsets of an $n+1$-element set A. Let $a \in A$. Then we can get each subset of A that has a in it exactly once by adjoining a to a $(k-1)$-element subset of $A - \{a\}$, so there are $C(n,k-1)$ k-element subsets of A that have a as an element. On the other hand, every k-element subset of A that does *not* contain a as an element is a k-element subset of $A - \{a\}$, so there are $C(n,k)$ of them.

Every subset of A either has a as an element or not, so we have the following theorem:

basis step 152
recurrence relation 161
theorem 2

125.5 Theorem

For all $n \geq 0$ *and* $k > 0$,

$$\begin{cases} C(n,0) &= 1 & \\ C(n,k) &= 0 & \text{if } k > n \\ C(n+1,k) &= C(n,k-1)+C(n,k) & \text{otherwise.} \end{cases} \tag{125.3}$$

125.5.1 Example

$$\begin{aligned} C(4,2) &= C(3,1)+C(3,2) \\ &= C(2,0)+C(2,1)+C(2,1)+C(2,2) \\ &= 1+2\cdot C(2,1)+C(2,2) \\ &= 1+2(C(1,0)+C(1,1))+C(1,1)+C(1,2) && (125.4) \\ &= 1+2(1+C(0,0)+C(0,1))+C(0,0)+C(0,1) \\ &= 1+2\cdot 2+1=6 && (125.5) \end{aligned}$$

125.5.2 Example The recurrence relation for $C(n,k)$ can be used to give an inductive proof of Theorem 125.3.

The basis step is to prove that

$$(x+y)^0 = \sum_{k=0}^{0} C(0,k)x^{-k}y^k$$

The sum on the right has only one term, namely $C(0,0)x^0y^0$, which is 1, as is the expression on the left.

Inductive step: Assume

$$(x+y)^n = \sum_{k=0}^{n} C(n,k)x^{n-k}y^k$$

We must prove

$$(x+y)^{n+1} = \sum_{k=0}^{n+1} C(n+1,k)x^{n+1-k}y^k$$

We now make a calculation. In this calculation it is convenient to define $C(n,-1)$

to be 0.

$$
\begin{aligned}
(x+y)^{n+1} &= (x+y)(x+y)^n \\
&= (x+y)\sum_{k=0}^{n} C(n,k)x^{n-k}y^k \quad \text{by induction hypothesis} \\
&= x\sum_{k=0}^{n} C(n,k)x^{n-k}y^k + y\sum_{k=0}^{n} C(n,k)x^{n-k}y^k \\
&= \sum_{k=0}^{n} C(n,k)x^{n+1-k}y^k + \sum_{k=0}^{n} C(n,k)x^{n-k}y^{k+1} \\
&\qquad \text{(now change } k \text{ to } k-1 \text{ in the second term)} \\
&= \sum_{k=0}^{n} C(n,k)x^{n+1-k}y^k + \sum_{k=1}^{n+1} C(n,k-1)x^{n-(k-1)}y^k \\
&= \sum_{k=0}^{n+1} \big(C(n,k)+C(n,k-1)\big)x^{n+1-k}y^k \\
&= \sum_{k=0}^{n+1} C(n+1,k)x^{n+1-k}y^k \quad \text{by Theorem 125.5}
\end{aligned}
$$

Note that I changed the limits on the sum in the next to last line of this proof, using the facts that $C(n,n+1)=0$ and $C(n,-1)=0$.

There is a sense in which this proof forces you to believe Theorem 125.3, but the earlier proof (on page 191) *explains* why it is true. Mathematicians sometimes call a proof like the earlier one a **conceptual proof**.

The following theorem gives an explicit formula for the binomial coefficient.

125.6 Theorem

For $0 \le k \le n$,

$$C(n,k) = \frac{n!}{k!(n-k)!} \tag{125.6}$$

The proof is omitted.

125.6.1 Worked Exercise Find the number of strings of length n in $\{a,b,c\}^*$ that contain exactly two a's.

Answer Now that we have the function $C(n,r)$ we can solve this using the idea of Worked Exercise 114.2.2. To construct such a string, we must choose two locations in the string where the two a's will be. There are $C(n,2)$ ways of doing this. Then there are two choices (b or c) for each of the other locations, so the answer is $C(n,2)\cdot 2^{n-2}$, which by Theorem 125.6 is

$$\frac{n(n-1)}{2}2^{n-2}$$

125.6.2 Proving identities for the binomial coefficient An enormous number of identities are known for the binomial coefficient. We consider one here to illustrate how one goes about proving such identities. The identity is

$$\sum_{k=0}^{n} C(n,k)^2 = C(2n,n) \tag{125.7}$$

This can be proved using the recurrence relation of Theorem 125.5, but the proof is rather tedious. I quail with terror at the idea of using the formula in Theorem 125.6 to prove this theorem.

It is much easier to use Definition 125.1. $C(2n,n)$ is the number of ways of choosing n balls from a set of $2n$ balls. Now suppose that we have $2n$ balls and n of them are red and n of them are green. Then an alternative way of looking at the task of choosing n balls from this set is that we must choose k red balls and $n-k$ green balls for some integer k such that $0 \leq k \leq n$. For a particular k there are $C(n,k)C(n,n-k)$ ways of doing this. By Theorem 125.2(c), this is the same as $C(n,k)^2$. Altogether this alternative method of choosing a n-element subset gives

$$\sum_{k=0}^{n} C(n,k)^2$$

possibilities.

125.6.3 Remark Like most concepts in mathematics, $C(n,k)$ has a *conceptual definition*, namely Definition 125.1, and a *method of calculating it*, in this case two of them: Theorems 125.5 and 125.6. It is generally good advice to *try the conceptual approach first*.

In this case there is a second conceptual description, as coefficients in a polynomial (Formula 125.1), and in fact that formula allows a faily easy second proof of Formula (125.7).

125.6.4 Exercise Prove that $\sum_{k=0}^{n} C(n,k) = 2^n$.

125.6.5 Exercise Prove that $\sum_{k=0}^{n} (-1)^k C(n,k) = 0$ for $n > 0$.

125.6.6 Exercise Prove two ways that for all $n \geq 4$,

$$C(n,3) = \frac{n-2}{3} \cdot C(n,2)$$

a) Prove it by using the definition of $C(n,k)$.
b) Prove it using formula (125.6).

125.6.7 Exercise Prove Theorem 125.6. (It can be done by induction, but is a bit complicated.)

125.6.8 Exercise Prove Formula (125.7) using Formula (125.1).

125.6.9 Exercise Let $F(n,k)$ be the number of strings of length n in $\{a,b,c\}^*$ with exactly k b's. Find a formula or recurrence relation for $F(n,k)$.

125.6.10 Exercise Derive a formula or recurrence relation for the number of strings of length n in $\{a,b\}^*$ with the same number of a's as b's.

125.6.11 Exercise (hard) Find a recurrence relation for the number of partitions of an n-element set that have exactly k blocks.

125.6.12 Exercise (hard) Prove formula (125.6).

126. Composition of relations

126.1 Definition: composition of relations
Let α be a relation from A to B and β be a relation from B to C. The **composite** $\alpha\circ\beta$ is a relation from A to C, defined this way: For all $a\in A$ and $c\in C$,

$$a(\alpha\circ\beta)c \Leftrightarrow \exists b\in B(a\,\alpha\,b\wedge b\,\beta\,c)$$

126.1.1 Example Let $A=\{1,2,3,4,5\}$, $B=\{3,5,7,9\}$ and $C=\{1,2,3,4,5,6\}$, with

$$\alpha=\Big\{\langle 1,3\rangle,\langle 1,5\rangle,\langle 2,7\rangle,\langle 3,5\rangle,\langle 3,9\rangle,\langle 5,7\rangle\Big\}$$

and

$$\beta=\Big\{\langle 3,1\rangle,\langle 3,2\rangle,\langle 3,3\rangle,\langle 7,4\rangle,\langle 9,4\rangle,\langle 9,5\rangle,\langle 9,6\rangle\Big\}$$

Then

$$\alpha\circ\beta=\Big\{\langle 1,1\rangle,\langle 1,2\rangle,\langle 1,3\rangle,\langle 2,4\rangle,\langle 3,4\rangle,\langle 3,5\rangle,\langle 3,6\rangle,\langle 5,4\rangle\Big\}$$

126.1.2 Usage As you can see, although functions are composed from right to left, relations are composed from left to right. It is not hard to see that if $F:A\to B$ and $G:B\to C$ are functions, then

$$\Gamma(G\circ F)=\Gamma(F)\circ\Gamma(G)$$

126.1.3 Exercise Let $A=\{2,3,4,5\}$, $B=\{6,7,8,9\}$, $C=\{a,b,c,d,e\}$, and $\alpha\in\mathrm{Rel}(A,B)$, $\beta\in\mathrm{Rel}(B,C)$ be defined as follows. Give the ordered pairs in $\alpha\circ\beta$.

a) α is "divides", β is $\Big\{\langle 6,a\rangle,\langle 6,c\rangle,\langle 7,b\rangle,\langle 9,d\rangle\Big\}$.

b) α is "divides", β is $\Big\{\langle 7,a\rangle,\langle 7,b\rangle,\langle 7,c\rangle\Big\}$.

c) $\alpha=\Big\{\langle 2,7\rangle,\langle 2,8\rangle,\langle 3,7\rangle,\langle 3,9\rangle,\langle 4,8\rangle,\langle 4,9\rangle\Big\}$ and
$\beta=\Big\{\langle 6,a\rangle,\langle 6,b\rangle,\langle 7,c\rangle,\langle 8,c\rangle,\langle 9,c\rangle,\langle 9,d\rangle,\langle 9,e\rangle\Big\}$

(Answer on page 250.)

126.2 Theorem
Composition of relations is associative: if $\alpha \in \mathrm{Rel}(A,B)$, $\beta \in \mathrm{Rel}(B,C)$, *and* $\gamma \in \mathrm{Rel}(C,D)$, *then*

$$(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma) \in \mathrm{Rel}(A,D)$$

Proof Left as Problem 126.3.4.

126.3 Definition: composition powers
The **composition powers** of a relation α on a set A are $\alpha^0 = \Delta_A$ (the equals relation), $\alpha^1 = \alpha$, $\alpha^2 = \alpha \circ \alpha$, and in general $\alpha^n = \alpha \circ \alpha^{n-1}$.

126.3.1 Exercise For each relation R in Exercise 52.1.3, page 75, determine whether $1\,\mathrm{R}^2\,3$, $1\,\mathrm{R}^3\,3$, and $3\,\mathrm{R}^2\,1$. (Answer on page 250.)

126.3.2 Exercise Prove that if $F : A \to B$ and $G : B \to C$ are functions, then $\Gamma(G \circ F) = \Gamma(F) \circ \Gamma(G)$.

126.3.3 Exercise Let $A = \{1,2,3,4\}$.

a) Construct a nonempty relation α on A for which α^2 is empty.
b) Construct a relation $\alpha \neq A \times A$ on A for which $\alpha^2 = A \times A$.

126.3.4 Exercise Prove that composition of relations is associative.

126.3.5 Exercise Show that the composite of functional relations is a functional relation.

126.3.6 Exercise Let α be a relation on a set A. Prove that α is transitive if and only if $\alpha \circ \alpha \subseteq \alpha$.

126.3.7 Exercise A relation α on a set A is **interpolative** if $\alpha \subseteq \alpha \circ \alpha$. Show that $<$, as a relation on R, is interpolative, but as a relation on Z, it is not interpolative.

127. Closures

Given any relation α on S, and any property P that a relation can have there may be a "smallest" relation with property P containing α as a subset. It may not exist, but if it does, it is called the **P-closure** of α. Here is the formal definition.

127.1 Definition: closure
A relation β on A is the P-closure of α if
C.1 β has property P.
C.2 $\alpha \subseteq \beta$.
C.3 If γ has property P and $\alpha \subseteq \gamma$, then $\beta \subseteq \gamma$.

127.1.1 How to think of closures β is the "smallest" (in the sense of inclusion) relation with property P containing α as a subset.

127.1.2 Fact The reflexive, symmetric, and transitive closures of relations always exist. We will look at each of these in turn. The antisymmetric closure of a relation need not exist (Problem 128.2.5).

127.2 Theorem
The reflexive closure of a relation α is $\alpha \cup \Delta_S$. It is denoted by α^R.

Proof To prove this formally you must show that it fits Definition 127.1; that is, that
RC.1 $\alpha \cup \Delta_S$ is reflexive,
RC.2 $\alpha \subseteq \alpha \cup \Delta_S$, and
RC.3 if γ is a reflexive relation and $\alpha \subseteq \gamma$, then $\alpha \cup \Delta_S \subseteq \gamma$.

RC.1 and RC.2 are obvious. As for RC.3, suppose that $\alpha \subseteq \gamma$ and γ is reflexive. If $x\,(\alpha \cup \Delta_S)\,z$ then either $x\,\alpha\,z$ or $x = z$ (that is, $x\,\Delta_S\,z$). In the first case $x\gamma z$ because $\alpha \subseteq \gamma$, and in the second case, $x\gamma z$ because γ is reflexive. Thus

$$x(\alpha \cup \Delta_S)y \Rightarrow x\gamma y$$

so $\alpha \cup \Delta_S \subseteq \gamma$, as required.

127.2.1 Exercise What is the reflexive closure of the relation "$<$" on R? (Answer on page 250.)

127.3 Theorem
The symmetric closure of a relation α is

$$\alpha^S = \alpha \cup \alpha^{\mathrm{op}}$$

127.3.1 Exercise What is the symmetric closure of "$<$" on R? (Answer on page 250.)

127.3.2 Exercise What is the symmetric closure of "$\leq$" on R?

127.3.3 Exercise Give an example of a relation whose symmetric closure has exactly three elements.

127.3.4 Exercise Show that the symmetric closure of a relation α is $\alpha \cup \alpha^{op}$. (Answer on page 250.)

The most important type of closure in practice is the transitive closure:

127.4 Theorem
Let α be a relation on a set S. The transitive closure α^T of α is $\cup_{k=1}^{\infty} \alpha^k$, where $\alpha^k = \alpha \circ \alpha \circ \ldots \circ \alpha$ (k times), the composition power.

Proof Let $\beta = \cup_{k=1}^{n} \alpha^k$. Any member of a family of sets is enclosed in the union of the family, so $\alpha \subseteq \beta$. This verifies C.2 of Definition 127.1. As for C.3, suppose γ is transitive and $\alpha \subseteq \gamma$. Then $\alpha^k \subseteq \gamma$ (Exercise 127.4.2), so $\beta \subseteq \gamma$ because any ordered pair in β is in at least one of the sets α^k.

Finally, we must show that β is transitive. Suppose $x\beta z$ and $z\beta y$. Then for some integers k and m, $x\alpha^k z$ and $z\alpha^m y$. Then it is easy to see that $x\alpha^{k+m}y$, so $x\beta y$ as required.

127.4.1 Exercise What is the transitive closure of the relation α on Z defined by $x\alpha y$ if and only if $y = x+1$?

127.4.2 Exercise Suppose γ is transitive and $\alpha \subseteq \gamma$. Show that $\alpha^k \subseteq \gamma$ for all positive integers k.

$\alpha \cup \Delta_S$ is the only reflexive closure of α. That is why we could use the notation α^R — it means only one thing. It is always true that if a relation has a P-closure, it has only one:

127.5 Theorem
Let P be a property of relations, and suppose β and β' are P-closures of a relation α on a set S. Then $\beta = \beta'$.

Proof By C.2 of Definition 127.1, $\alpha \subseteq \beta$ and $\alpha \subseteq \beta'$. Then by C.3, $\beta \subseteq \beta'$ and $\beta' \subseteq \beta$. Thus $\beta = \beta'$.

128. Closures as intersections

The following set-theoretic description of P-closures is useful. It does not make the P-closure easy to calculate, but it does give a conceptual description useful for proving properties of closures.

128.1 Definition: intersection-closed

A property P of relations on a set A is **intersection-closed** if:

IC.1 $A \times A$ has property P.

IC.2 For any set $\mathcal{S}$ of relations on A, all of which have property P, the intersection of all the relations in $\mathcal{S}$ also has property P.

128.1.1 Remark The set $A \times A$ can be regarded as the intersection of the *empty* family of relations on A. The reasoning is this: In the case of relations, each relation on A is a subset of $A \times A$, and by Section 112.4 the intersection of the empty family of relations on A is $A \times A$. From this point of view, IC.1 is unnecessary.

128.2 Theorem

Let P be an intersection-closed property of relations. Then for any relation α, the P-closure of α exists and is the intersection of the set of all P-closed relations containing α as a subset.

Proof Let β be the intersection of all the P-closed relations containing α as a subset. We must verify C.1, C.2 and C.3. β has property P because P is intersection-closed. $\alpha \subseteq \beta$ because $\alpha \subseteq A \times A$ and $A \times A$ has property P, and β is the intersection of all the relations with property P that contain α as a subset. Finally, the intersection of a family of sets is included in any member of the family.

128.2.1 Exercise Prove that for any property P, if α has property P then the P-closure of α is α itself.

128.2.2 Exercise Show that the following hold for any relation α:

a) $\alpha^{RS} = \alpha^{SR}$.

b) $\alpha^{RT} = \alpha^{TR}$.

128.2.3 Exercise

a) Prove that for any relation α, $\alpha^{TS} \subseteq \alpha^{ST}$.

b) Give an example of a relation α for which $\alpha^{TS} \neq \alpha^{ST}$.

128.2.4 Exercise Let P be the property of a relation β that either $1\beta 2$ or $2\beta 1$. On the set $S = \{1,2\}$, let $\alpha = \{\langle 1,1\rangle\}$. Let $\beta = \{\langle 1,1\rangle, \langle 1,2\rangle\}$ and $\gamma = \{\langle 1,1\rangle, \langle 2,1\rangle\}$. Then β and γ both include α and both have property P. On the other hand, α does not have property P. Does this contradict Theorem 127.5?

128.2.5 Exercise Show that a relation need not have an “antisymmetric closure”.

129. Equivalence relations

If an object a is like an object b in some specified way, then b is like a in that respect. And surely a is like itself — in *every* respect! Thus if you want to give an abstract definition of a type of relation intended to capture the idea of being alike in some respect, two of the properties you could require are reflexivity and symmetry. Relations with those two properties are studied in the literature (the nearness relation $\mathcal{N}$ in Section 55.1.4 is such a relation), but here we are going to require the additional property of transitivity, which roughly speaking forces the objects to fall into discrete types, making a partition of the set of objects being studied.

129.1 Definition: equivalence relation
An **equivalence relation** on a set S is a reflexive, symmetric, transitive relation on S.

129.1.1 Remark This is an abstract definition — you don't have to have some property or mode of similarity in mind to define an equivalence relation.

129.1.2 Example Let $A = \{1,2,3,4,5,6\}$. Here is an equivalence relation α on the set A:

$$\alpha = \{\langle n,n\rangle \mid n \in A\} \cup \{\langle 2,5\rangle, \langle 5,2\rangle, \langle 3,4\rangle, \langle 4,3\rangle, \langle 3,6\rangle, \langle 6,3\rangle, \langle 4,6\rangle, \langle 6,4\rangle\} \tag{129.1}$$

129.1.3 Example The relation "equals" on any set is an equivalence relation.

129.1.4 Example The relation "has the same parity as" on the set $\mathbf{N}$ of natural numbers is an equivalence relation. Two numbers have the same parity if they are both even or both odd.

129.1.5 Example The relation of being in the same suit on a deck of cards is an equivalence relation.

129.1.6 Example Both the congruence relation and the similarity relation on the set of triangles are equivalence relations.

129.1.7 Example The relation called equivalence on the set of propositions or the set of predicates is an equivalence relation. (This example requires that the set of propositions or predicates be precisely defined, which is done in formal treatments of logic but which has not been done in this text.)

129.2 Exercise set

In questions 129.2.1 through 129.2.9, let E be the relation defined in the question on $\mathbf{Z}$. Is E an equivalence relation? Explain your answer.

129.2.1 $mEn \Leftrightarrow m \leq n$ (Answer on page 250.)

129.2.2 $mEn \Leftrightarrow m^2 = n$ (Answer on page 250.)

129.2.3 $mEn \Leftrightarrow m = n+1 \vee n = m+1$ (Answer on page 250.)

129.2.4 $mEn \Leftrightarrow 2 \mid m-n \vee 3 \mid m-n$ (Answer on page 250.)

129.2.5 $mEn \Leftrightarrow m^2 = n^2$

129.2.6 $mEn \Leftrightarrow m \mid n \wedge n \mid m$

129.2.7 $mEn \Leftrightarrow |m-n| < 6$.

129.2.8 $mEn \Leftrightarrow 12 \mid (m-n+1)$.

129.2.9 $mEn \Leftrightarrow (6 \mid (m-n)$ and $8 \mid (m-n))$.

129.3 Exercise set

In questions 129.3.1 through 129.3.6, let E be the relation defined in the question on R. Is E an equivalence relation?

129.3.1 $rEs \Leftrightarrow r/s = 1$ (Answer on page 250.)

129.3.2 $rEs \Leftrightarrow \text{floor}(r) = \text{floor}(s)$. (Answer on page 250.)

129.3.3 $rEs \Leftrightarrow [r = s \vee (0 \le r \le 1 \wedge 0 \le s \le 1)]$ (Answer on page 250.)

129.3.4 $rEs \Leftrightarrow r + s = 1$.

129.3.5 $rEs \Leftrightarrow r - s \in \mathrm{N}$.

129.3.6 $rEs \Leftrightarrow r - s \in \mathrm{Z}$

129.3.7 Exercise If E and F are equivalence relations on a set S, are $E \cap F$ and $E \cup F$ always equivalence relations?

130. Congruence

130.1 Definition: congruence (mod k)

Let k be a fixed positive integer. Two integers m and n are **congruent (mod k)**, written "$m \equiv n \pmod{k}$", if k divides $m-n$, in other words, if there is an integer q for which $m - n = qk$.

130.1.1 Example $9 \equiv 3 \pmod{6}$, $-5 \equiv 16 \pmod{7}$, $146 \equiv -22 \pmod{12}$.

130.1.2 Usage

a) In the phrase "$m \equiv n \pmod{k}$", k is called the **modulus of congruence**.
b) The syntax for "mod" here is different from that of the operator "MOD" used in Pascal and other languages. In Pascal, "MOD" is a binary operator like "+"; when used between two variables, as in the phrase "M MOD K", it causes the calculation of the remainder when M is divided by K. Thus "5 MOD 3", for example, is an expression (not a statement) having value 2. The phrase "$5 \equiv 2 \pmod{3}$", on the other hand, is a *sentence* that is either true or false.

130.1.3 Exercise List all the positive integers ≤ 100 that are congruent to 3 mod 24. (Answer on page 250.)

130.1.4 Exercise List all the positive integers ≤ 100 that are congruent to -3 mod 24.

130.1.5 Remark Recall that the remainder when m is divided by k is the unique integer r with $0 \leq r < |k|$ for which there is an integer q such that $m = qk + r$. Then we can prove:

130.2 Theorem

Two positive integers m and n are congruent mod k if and only if m and n leave the same remainder when divided by k.

Proof If $m = qk + r$ and $n = q'k + r$ (same r), then $m - n = (q - q')k$, so k divides $m - n$. Then by definition $m \equiv n \pmod{k}$.

Conversely, if $m \equiv n \pmod{k}$, let r be the remainder when m is divided by k and r' the remainder when n is divided by k. Then there are quotients q and q' for which $m = qk + r$ and $n = q'k + r'$. Then $r - r' = (m - qk) - (n - q'k) = m - n + (q' - q)k$. Since $m - n$ is divisible by k, this means $r - r'$ is divisible by k. Since r and r' are both between 0 and k (not including k), this means $r = r'$, as required.

130.3 Theorem

Congruence $\pmod{k}$ is an equivalence relation.

Proof Here is the proof that it is transitive; the rest is left to you. Suppose that $m \equiv n \pmod{k}$ and $n \equiv p \pmod{k}$. Then m leaves the same remainder as n when divided by k, and n leaves the same remainder as p when divided by k. Since remainders are unique, m leaves the same remainder as p when divided by k, so, by Theorem 130.2 $m \equiv p \pmod{k}$.

Congruence has an important special property connected with addition and multiplication that has given it extensive applications in computer science:

130.4 Theorem

If $m \equiv m' \pmod{k}$ and $n \equiv n' \pmod{k}$ then $m + n \equiv m' + n' \pmod{k}$ and $mn \equiv m'n' \pmod{k}$.

Proof The hypothesis translates into the statement

$$k \mid m - m' \text{ and } k \mid n - n'$$

Then $(m + n) - (m' + n') = m - m' + n - n'$ is the sum of two numbers divisible by k, so is divisible by k. Hence $m + n \equiv m' + n' \pmod{k}$. Also $mn - m'n' = mn - mn' + mn' - m'n' = m(n - n') + n'(m - m')$, again the sum of two numbers divisible by k, so that $mn \equiv m'n' \pmod{k}$.

130.4.1 Remark The consequence of Theorem 130.4 is that if you have an expression involving integers, addition and multiplication, you can freely substitute integers congruent to the integers you replace and the expression will evaluate to an integer that, although it may be different, will be congruent $\pmod{k}$ to the original value.

130.4.2 Example As an example, what is 5^8 congruent to $\pmod{16}$? The arithmetic is much simplified if you reduce each time you multiply by 5:

$$\begin{array}{rcl} 5 & \equiv & 5 \pmod{16} \\ 5^2 & \equiv & 25 \equiv 9 \pmod{16} \\ 5^3 & \equiv & 5 \cdot 9 \equiv 45 \equiv 13 \pmod{16} \\ 5^4 & \equiv & 5 \cdot 13 \equiv 65 \equiv 1 \pmod{16} \\ 5^8 & \equiv & (5^4)^2 \equiv 1^2 \equiv 1 \pmod{16} \end{array} \tag{130.1}$$

130.4.3 Remark This ability to compute powers fast is the basis of an important technique in cryptography.

130.4.4 Exercise Compute:
a) $5^{12} \pmod{4}$
b) $5^{12} \pmod{10}$
c) $5^{12} \pmod{16}$
(Answer on page 250.)

130.4.5 Exercise Prove that if $s \mid t$, then

$$ms \equiv ns \pmod{t} \Leftrightarrow m \equiv n \pmod{t/s}$$

131. The kernel equivalence of a function

If $F : A \to B$ is a function, it induces an equivalence relation $K(F)$ on its domain A by identifying elements that go to the same thing in B. Formally:

131.1 Definition: kernel equivalence
If $F : A \to B$ is a function, the **kernel equivalence** of F on A, denoted $K(F)$, is defined by

$$aK(F)a' \Leftrightarrow F(a) = F(a')$$

131.1.1 Fact It is easy to see that the kernel equivalence of a function is an equivalence relation.

131.1.2 Example The congruence relations described in the preceding section are kernel equivalences. Let k be a fixed integer ≥ 2. The **remainder function** $F : \mathrm{Z} \to \mathrm{Z}$ is defined by $F(n) = n \pmod{k}$, the remainder when n is divided by k. Theorem 130.2, reworded, says exactly that the relation of congruence $\pmod{k}$ is the kernel equivalence of the remainder function.

131.1.3 Exercise Give an example of a function $F : \mathrm{N} \to \mathrm{N}$ with the property that $3K(F)5$ but $\neg(3K(F)6$. (Answer on page 250.)

132. Equivalence relations and partitions

132.0.4 Discussion If an equivalence relation E is given on a set S, the elements of S can be collected together into subsets, with two elements in the same subset if they are related by E. This collection of subsets of S is a set denoted S/E, the **quotient set** of S by E. Here is the formal definition of S/E:

132.1 Definition: quotient set of an equivalence relation
Let E be an equivalence relation on a set S. For each $x \in S$, the **equivalence class** of x **mod** E, denoted $[x]_E$, is the subset $\{y \in S \mid yEx\}$ of S. The **quotient set (of an equivalence relation)** S/E of E is the set $\{[x]_E \mid x \in S\}$.

132.1.1 Example The quotient set of the equivalence relation α defined in 129.1 above is $\{\{1\}, \{2,5\}, \{3,4,6\}\}$, which is a partition.

132.1.2 Example The quotient set of congruence $\pmod 6$ is the partition of Z by remainders upon division by 6. The quotient set is *always* a partition:

132.2 Theorem
If S is a set and E is an equivalence relation on S, then the quotient set S/E is a partition of S.

Proof To see why S/E is a partition, we have to see why
a) every element of S is in an equivalence class in S/E,
b) no element of S is in two equivalence classes in S/E, and
c) S/E does not contain the empty set as an element.
(This just spells out the definition of partition.)

Part (a) is easy: if $x \in S$ then, by reflexivity, xEx, so $x \in [x]_E$.

Part (c) is similar: by definition of S/E, an element of S/E is an equivalence class $[x]_E$ for some $x \in S$; since $x \in [x]_E$, $[x]_E$ is not empty.

As for (b), $x \in [x]_E$; if also $x \in [y]_E$ for some $y \in S$, then we have to show that $[y]_E = [x]_E$. To do this, we have to show two things:
(i) $[y]_E \subseteq [x]_E$, and
(ii) $[x]_E \subseteq [y]_E$.
For (i), let $z \in [y]_E$. Then zEy by definition. Since $x \in [y]_E$, xEy. By symmetry and transitivity, zEx, so $z \in [x]_E$. Hence $[y]_E \subseteq [x]_E$.

For (ii), let $z \in [x]_E$. Then zEx. Since $x \in [y]_E$, xEy. So by transitivity, zEy. Hence $z \in [y]_E$, as required.

132.2.1 Fact The equivalence class $[x]_E$ is a block of the partition S/E.

132.2.2 Worked Exercise Let $S = \{1,2,3,4,5\}$. Find S/E if

$$E = \Delta_S \cup \{\langle 1,3\rangle, \langle 3,1\rangle, \langle 3,4\rangle, \langle 4,3\rangle, \langle 1,4\rangle, \langle 4,1\rangle\}$$

Answer

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

132.2.3 Exercise Let $S = \{1,2,3,4,5,6\}$. Find S/E if

$$E = \Delta_S \cup \{\langle 1,3\rangle, \langle 3,1\rangle, \langle 3,4\rangle, \langle 4,3\rangle, \langle 1,4\rangle, \langle 4,1\rangle, \langle 2,5\rangle, \langle 5,2\rangle\}$$

132.2.4 Exercise Let $S = \{1,2,3,4,5\}$. Find two different equivalence relations E and E' with the property that the subset $\{1,2\}$ is an element of both S/E and S/E'. (Answer on page 250.)

132.2.5 Exercise Give an example of an equivalence relation E on the set R with the property that

$$\{x \in \mathrm{R} \mid 0 \leq x \leq 1\}$$

is one of the equivalence classes of E.

132.2.6 Exercise Let $S = \{1,2,3,4,5\}$. Find two different equivalence relations E and E' on S with the property that $S/E \cap S/E' = \{\{1,5\},\{3\}\}$

132.2.7 How to think of equivalence relations If E is an equivalence relation on S, the quotient set S/E is often thought of as obtained by *merging equivalent elements of* S. One often says that one **identifies** equivalent elements. Here, "identify" means "make identical" rather than "discover the identity of". Mathematicians informally will say we glue equivalent elements together.

133. Partitions give equivalence relations

For a partition Π of a set S, we will use the notation $[x]_\Pi$ or just $[x]$ if the context makes clear which partition is being used, to denote the (unique) block of Π that has x as an element. Given a partition Π, you get an equivalence relation E_Π by the definition:

$$x\mathrm{E}_\Pi y \Leftrightarrow (x \in [y]_\Pi) \tag{133.1}$$

133.1 Theorem
If Π is a partition of a set S, then the relation E_Π defined by (133.1) is an equivalence relation.

Proof To see that $x\mathrm{E}_\Pi x$ requires $x \in [x]$, which is true by definition of $[x]$. Hence E_Π is reflexive. If $x\mathrm{E}_\Pi y$ then $x \in [y]$. That means $[x] = [y]$, since by definition of partition an element is in only one block. Since $y \in [y]$ by definition and $[x] = [y]$, we know that $y \in [x]$, so $y\mathrm{E}_\Pi x$. Hence E_Π is symmetric. Note that we now know that $x\mathrm{E}_\Pi y$ if and only if x and y are in the same block of Π. To prove transitivity, suppose $x\mathrm{E}_\Pi y$ and $y\mathrm{E}_\Pi z$. Then x and y are in the same block, and y and z are in the same block, so $[x]=[y]=[z]$. This means $x\mathrm{E}_\Pi z$, so E_Π is transitive.

133.2 The fundamental theorem on equivalence relations

We gave two constructions in the preceding sections. Given an equivalence relation E, in Definition 132.1 we constructed a partition S/E, and given a partition Π, in Section 133 we constructed an equivalence relation E_Π.

If we let πS denote the set of partitions of S (this is standard notation) and $E(S)$ denote the set of equivalence relations on S (there is no standard notation for this), we now have functions $E \mapsto S/E : E(S) \to \pi S$ and $\Pi \mapsto \mathrm{E}_\Pi : \pi S \to E(S)$, where E_Π is defined in formula (133.1) above. The basic fact about these constructions is that *these two functions are bijections and each is the inverse of the other.* This fact is the "fundamental theorem on equivalence relations."

In other words, if you have an equivalence relation E, construct the quotient set S/E, which is a partition, and then construct the equivalence relation $E_{S/E}$ corresponding to that partition, you get the equivalence relation E you started with. And if you have a partition Π of S, construct the corresponding equivalence relation E_Π, and then construct the quotient set S/E_Π of E, you get the partition Π back again. The proof of the fundamental theorem involves the same sort of arguments given earlier, and is left as a problem.

133.2.1 Exercise Prove the fundamental theorem on equivalence relations.

133.2.2 Exercise Prove that any partition of a set A is the quotient of some function with domain A.

134. Orderings

An ordering is a special sort of relation that is the mathematical formulation of the concept of comparison or priority. It includes as special cases the relation "$\leq$" between numbers and the relation of inclusion between subsets of a set. Here is the formal definition:

> **134.1 Definition: ordering**
> A relation α on a set A is an **ordering** if it is antisymmetric and transitive. If it is also reflexive, it is a **weak ordering**, and if it is also irreflexive, it is a **strict ordering**.

134.1.1 Example The relation "$\leq$" on a set of numbers is a weak ordering, and "$<$" is a strict ordering.

134.1.2 Example An example of an ordering α on a set S that is neither weak nor strict is the relation

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

on the set $\{1,2,3\}$. It is not reflexive because 2 is not related to itself, but it is not irreflexive because 1 *is* related to itself.

134.1.3 Remark Essentially all the orderings considered in this text are either weak orderings or strict orderings, but the more general concept is occasionally useful.

134.2 Definition: ordered set

If α is an ordering on A, then (A,α) is an **ordered set**. If α is a weak ordering, (A,α) is a **poset**.

134.2.1 Example $(\mathrm{R},\leq)$ and $(\mathrm{R},\geq)$ are posets, and so is $(\mathcal{P}A,\subseteq)$ for any set A. The set of all relations on a set S is ordered by inclusion; it is the poset $(\mathcal{P}(S\times S),\subseteq)$.

134.2.2 Usage In many texts, a weak ordering is called a **partial ordering**, and "poset" is short for "partially ordered set".

134.2.3 Example Not only are "$\leq$" and "$<$" orderings on R, but so are "$\geq$" and "$>$".

134.2.4 Example The relation $m\,|\,n$ on N is a weak ordering; thus $(\mathrm{N},|)$ is a poset. Reflexivity is the obvious fact that $n\,|\,n$ for any $n\in\mathrm{N}$, transitivity requires proving that if $m\,|\,n$ and $n\,|\,p$ then $m\,|\,p$, and antisymmetry is the almost obvious fact that if $m\,|\,n$ and $n\,|\,m$ then $m=n$.

I will prove antisymmetry and leave the others to you. By definition, $m\,|\,n$ means that $n=hm$ for some positive integer h. Likewise $n\,|\,m$ means that $m=kn$ for some positive integer k. Thus $m=kn=khm$. If $m\neq 0$ you can cancel m and get $kh=1$. Since k and h are positive integers, that means $k=h=1$. Hence $m=n$. As for the case $m=0$, the fact that $n=hm$ means $n=0$, so $m=n$ again.

134.2.5 Example If you have a collection $\mathcal{T}$ of tasks, there is a natural ordering of $\mathcal{T}$ defined this way: $t\,\alpha\,u$ if task t must be done before task u can be started. This is obviously transitive. If α were not antisymmetric, that would say there are two *different* tasks t and u, each of which had to be done before the other, so that it is in fact impossible to perform the set of tasks. Thus for any *reasonable* collection $\mathcal{T}$ of tasks, $(\mathcal{T},\alpha)$ is antisymmetric as well as transitive and therefore an ordering.

134.3 Theorem

Let α be an ordering. Then α^{op} (see Section 54.2, page 77) is also an ordering. Moreover, α^{op} is strict if α is strict and weak if α is weak.

134.3.1 How to think of orderings If α is an ordering on a set S and $a\,\alpha\,b$, one says that "a is smaller than b". This phraseology has to be used with caution — one would not use it, for example, for the relation "$\geq$" on R. More subtle problems with this terminology arise with other orderings. For example, in the poset $(\mathrm{N},|)$, 3 is smaller than 6 but 3 is not smaller than 5. Nor, for that matter, is 5 smaller than 3. You have to be very clear that "smaller" here is not the *usual* relation "$\leq$" on N.

The following Theorem, whose proof is left to you, shows that a relationship analogous to that between "<" and "≤" holds for all orderings.

134.4 Theorem
For any ordering α on a set S, $\alpha - \Delta_S$ is a strict ordering of S and the reflexive closure α^R is a weak ordering.

135. Total orderings

135.1 Definition: total ordering
An ordering α on a set A with the property that for any pair of elements $a, b \in A$, either $a\,\alpha\,b$ or $b\,\alpha\,a$, is a **total ordering**.

135.1.1 Usage A total ordering is also called a **linear ordering**.

135.1.2 Example The relations "≤" and "≥" are total orderings on R, as well as other sets of numbers.

135.1.3 Example The ordered set $(\mathrm{N}, |)$ is not totally ordered: as we observed previously, 3 and 5 are not related to (do not divide) each other.

135.1.4 Example If A has more than one element, then $(\mathcal{P}A, \subseteq)$ is not a totally ordered set.

135.2 Theorem
A total ordering is reflexive, in other words is a weak ordering.

135.2.1 Exercise Prove Theorem 135.2.

135.2.2 Usage In most writing in pure mathematics, a total ordering is a type of strict ordering, defined axiomatically in Definition 135.3 below. We call it "strict total ordering" here.

135.3 Definition: strict total ordering
A relation α on a set S is a **strict total ordering** if it is transitive and satisfies **trichotomy**: For all $a, b \in S$, *exactly one* of the following statements hold:
(i) $a\,\alpha\,b$
(ii) $b\,\alpha\,a$
(iii) $a = b$.

135.3.1 Remark This definition has the consequence that a strict total ordering is not a total ordering in the sense of Definition 135.1. However, it is straightforward to prove that if α is a strict total ordering then α^R is a total ordering in the sense of Definition 135.1.

The relation "divides" on Z is not an ordering because it is not antisymmetric. For example, $6\,|\,{-6}$ and $-6\,|\,6$ but $6 \neq -6$. "Divides" is, however, reflexive and transitive on Z.

135.3.2 Exercise Let α be a relation on a set A. Prove that if α is a strict total ordering in the sense of Definition 135.3, then α is a strict ordering. (Answer on page 250.)

135.3.3 Exercise Let α be a relation on a set A.

a) Assume that α is a strict total ordering in the sense of Definition 135.3. Prove that α^R is a total ordering in the sense of Definition 135.1.
b) Prove that if α is a total ordering then $\alpha - \Delta_A$ is a strict total ordering.

135.3.4 Exercise How many total orderings of an n-element set are there? Prove your answer correct.

135.3.5 Exercise For any natural number n, let $D(n)$ denote the set of positive divisors of N. Thus $D(6) = \{1,2,3,6\}$. Show that $(D(n),|)$ is totally ordered if and only if n is a power of a prime.

136. Preorders

> **136.1 Definition: preordering**
> A reflexive, transitive relation α on a set A is called a **preorder** or **preordering** on A, and (A,α) is a **preordered set**.

136.1.1 Usage Sometimes "quasi-ordering" is used for "preordering", but that word is used with other meanings, too.

136.1.2 Remark Every preorder can be converted into a partial order by a process resembling the construction of the quotient of a function. This process is explored in exercises below.

136.1.3 Exercise (hard) Let α be a preorder on a set S.

a) Prove that the relation E defined by

$$xEy \Leftrightarrow (x\alpha y \wedge\ y\alpha x)$$

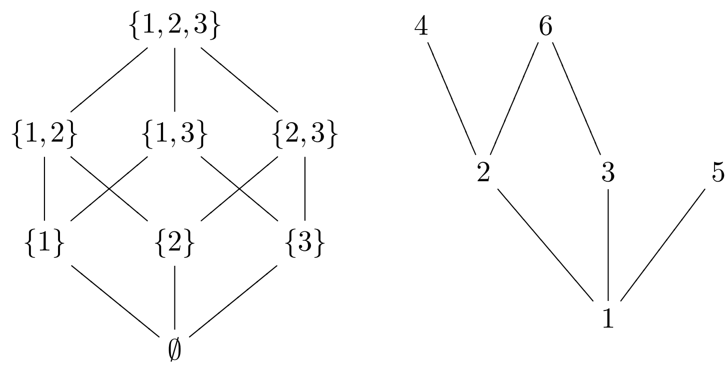
is an equivalence relation.
b) Define a relation λ on S/E by

$$[x]\lambda[y] \Leftrightarrow x\alpha y$$

Prove that λ is well-defined, that is, that if $[x] = [x']$, $[y] = [y']$, and $[x]\lambda[y]$, then $[x']\lambda[y']$.
c) Prove that λ is an ordering

137. Hasse diagrams

Exhibiting an ordering using a digraph as in Section 51.2 tends to be messy-looking because transitivity causes lots of arrows to exist. Orderings are normally illustrated using a different sort of picture called a **Hasse diagram**. The elements of the set are represented as dots, as before, and the diagram is drawn so that when there is a rising line from a to b, then $a \,\alpha\, b$. ("Rising" means toward the top of the page.) The rising line from a to b does not have to go directly from a to b, but may pass through other nodes; this makes use of the fact that the relation is transitive. Note that the diagram does not show whether a node is related to itself. In this text, Hasse diagrams are used only for weak orderings.



(137.1)

137.1.1 Example The two Hasse diagrams in Figure 137.1 show the inclusion relation on the set of subsets of $\{1,2,3\}$ and the relation of division on the set $\{1,2,3,4,5,6\}$.

137.1.2 Remark Note that b can be higher on the page than a without it being true that $a \,\alpha\, b$ — there must be a rising line from a to b to make $a \,\alpha\, b$. For example, in the right diagram, 5 is not less than 6.

137.1.3 Exercise Draw the Hasse diagram of the indicated poset (A,α):

a) $A = \{1,2,3,4,5\}$,

$$\alpha = \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle, \langle 4,4\rangle, \langle 5,5\rangle, \langle 1,2\rangle, \langle 2,3\rangle, \langle 1,3\rangle, \langle 5,4\rangle, \langle 4,3\rangle, \langle 5,3\rangle\}$$

b) $A = \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{2,3\}\}$, α is inclusion.

c) $A =$set of positive divisors of 20, α is divisibility.

d) $A =$set of positive divisors of 25, α is divisibility.

(Answer on page 250.)

137.1.4 Exercise Which of the posets in Exercise 137.1.3 are total orderings? (Answer on page 251.)

137.1.5 Exercise Draw the Hasse diagram for the relation "divides" on:

1. The set of positive divisors of 12.

2. The set $\{n \in \mathrm{N} \mid 1 \leq n \leq 12\}$.

138. Lexical ordering

A finite totally ordered set A used as an alphabet induces a total order on the strings in A^* called the **lexical order** on A^*. When A is the English alphabet, the result is the familiar alphabetical ordering of strings.

To define lexical ordering, we need a preliminary idea.

138.1 Definition: initial segment
A string u is an **initial segment** of a string w if $w = ux$ for some string x in A^*.

138.1.1 Example 'ab' is an initial segment of 'abbac'.

138.1.2 Example Any string is an initial segment of itself (since $\Lambda \in A^*$).

138.1.3 Example Λ is an initial segment of any string.

138.2 Definition: lexical order
Let (A, α) be a finite totally ordered set. Then the **lexical order** or **lexical ordering** λ on A^* is defined as follows: $w \lambda x$ if either
LE.1 w is an initial segment of x, or
LE.2 If i is the first position where w and x differ, then $w_i \, \alpha \, x_i$.

138.2.1 Example If A is the English alphabet with the usual ordering, 'car' comes before 'card' in alphabetical ordering because 'car' is an initial segment of 'card', and 'car' comes before 'cat' because the first place where 'car' and 'cat' differ is the third place, and 'r' comes before 't'.

138.2.2 Example If A is nonempty, A^* is an infinite set. Consider the lexical ordering on $\{0,1\}^*$, where $\{0,1\}$ is ordered so that 0 comes first. The first few elements of $\{0,1\}^*$ are Λ, '0', '00', '000', '0000', '00000', ... Thus if you go through the strings in order, there are strings such as '1' that you can't get to in a finite amount of time: there are an infinite number of strings in $\{0,1\}^*$ before '1'.

138.2.3 Exercise Prove that the lexical ordering on $\{0,1\}^*$ (with $0 < 1$) is a total ordering.

139. Canonical ordering

The canonical ordering, defined below, is often used on infinite sets of strings to remedy the problem described in Example 138.2.2. It is the most commonly used ordering on $\{0,1\}^*$.

139.1 Definition: canonical ordering
The **canonical ordering** on $\{0,1\}^*$, usually denoted "$\leq$", is defined this way: $w \leq x$ if
a) w is shorter than x ($|w| < |x|$) or
b) $|w| = |x|$ and the integer represented by w in binary notation is less than or equal to the integer represented by x in binary notation.

139.1.1 Example 1110 comes before 00001 because it is shorter, and 0011 comes before 0101 because 0011 is 3 in binary and 0101 is 5.

139.1.2 Example In the canonical ordering of $\{0,1\}^*$, the first few strings are Λ, 0, 00, 01, 10, 11, 000, 001, 010, 011, 100, ...

139.1.3 Fact The canonical ordering is linear and, unlike the lexical ordering, there are only a finite number of strings between any two strings.

139.1.4 Remark This idea can obviously be extended to strings in the alphabet $\{0,1,\ldots,n\}$ where n is a small integer (use base $n+1$).

139.1.5 Exercise List the elements of the set

$$A = \{00, 01, 110, 111, 0101, 0111, 10101, 10111, 01111\}$$

in the lexical ordering and in the canonical ordering. (Answer on page 251.)

139.1.6 Exercise Prove that the canonical ordering on $\{0,1\}^*$ is a total ordering, and that there are only a finite number of strings between any two given strings.

140. Upper and lower bounds

140.1 Definition: upper bound
If (A,α) is a poset and $B \subseteq A$, an element $a \in A$ is an **upper bound** of B in (A,α) if $b\,\alpha\,a$ for every $b \in B$.

140.1.1 Remark Note that the upper bound a of Definition 140.1 need not be in B.

140.1.2 Example In the right poset in Figure 137.1, 6 is an upper bound (in fact the only one) of $\{1,2,3\}$ and the set $\{1,2,3,4\}$ has no upper bound.

140.1.3 Example $\{1,2,3,4\}$ has many upper bounds in the poset $(\mathrm{N},|)$, for example 12, 24 and 144.

140.1.4 Remark A **lower bound** of a subset is defined in the analogous way: a is a lower bound of B if $a \,\alpha\, b$ for all $b \in B$.

140.2 Definition: maximum

Let A be a poset and B a subset of A. The **maximum** of B (plural "maxima") is an element m of B with the property that for all $b \in B$, $b \,\alpha\, m$.

140.2.1 Fact The maximum of B, if it exists, is clearly an upper bound of B; unlike an upper bound, however, it must actually be in B. More is true:

140.3 Theorem

The maximum of a subset B of a poset A, if it exists, is unique.

Proof If m and m' were both maxima of B, then both would be elements of B and so it would have to be the case that $m \,\alpha\, m'$ and $m' \,\alpha\, m$. Then antisymmetry forces $m = m'$.

140.3.1 Remark The **minimum** of B is an element n of B with $n \,\alpha\, b$ for all $b \in B$. A similar proof shows that a subset B has at most one minimum. Note that the minimum of B in A is the minimum of B in the opposite poset of A.

140.3.2 Exercise Find all the maxima and minima of the posets in Exercise 137.1.3 of Chapter 134. (Answer on page 251.)

140.3.3 Exercise What are the maxima and minima, if any, of $(\mathrm{N},|)$? Of $(\mathrm{N}-\{0\},|)$? Of $(\mathrm{N}-\{0,1\},|)$? (Answer on page 251.)

141. Suprema

The two ideas of upper bound and minimum combine to form a concept that is more important than either of them.

141.1 Definition: supremum

Let A be a poset with subset B. An element $m \in A$ is a **supremum** of B, or **least upper bound** of B, if it is the minimum of the set of upper bounds of B.

141.1.1 Fact The supremum m must be unique if it exists, and it may or may not be in B. Because of its uniqueness, we denote the supremum of B as $\sup B$.

141.1.2 Reformulation of the definition It is worth spelling out the definition of supremum: If $B \subseteq A$ and $m \in A$, then m is the supremum of B if m is an upper bound of B and $m \,\alpha\, a$ for every other upper bound a of B. This gives rise to a rule of inference.

141.2 Theorem

If (A,α) is a poset and $B \subseteq A$, then

$$(\forall b{:}B)(b \mathrel{\alpha} m),\ (\forall a{:}A)\big((\forall b{:}B)(b \mathrel{\alpha} a) \Rightarrow m \mathrel{\alpha} a\big) \vdash\ m = \sup B$$

141.2.1 Fact Note that m is the "least" upper bound in the sense of the ordering α: if a is an upper bound of B, then $m \mathrel{\alpha} a$. Specifically, *no upper bound can be unrelated to m.*

141.2.2 Example The supremum of $\{\{1\},\{1,2\},\{3\}\}$ in the set of all subsets of $\{1,2,3\}$ is $\{1,2,3\}$ itself (See Figure 137.1).

141.2.3 Example The supremum in $(\mathrm{R},\leq)$ of the open interval $(0\,..\,1)$ is 1, which is also the supremum of the closed interval $[0\,..\,1]$.

141.2.4 Example The set

$$S = \{x \in \mathrm{Q} \mid 0 \leq x \text{ and } x^2 \leq 2\} = \{x \in \mathrm{Q} \mid 0 \leq x \leq \sqrt{2}\}$$

has no supremum in $(\mathrm{Q},\leq)$. That is because if it had a supremum $m \in \mathrm{Q}$, m would have to be its supremum in R, too, but the supremum in R is $\sqrt{2}$, which is not in Q.

> ### 141.3 Definition: infimum
> The **infimum** of B, or $\inf B$, if it exists, is the unique element n for which
> a) $n \mathrel{\alpha} b$ for all $b \in B$, and
> b) if $a \mathrel{\alpha} b$ for all $b \in B$, then $a \mathrel{\alpha} n$.

141.3.1 Example In the set $\{1,2,3,4,5,6\}$ ordered by division, the supremum of the subset $\{2,5\}$ does not exist, and the infimum is 1.

141.3.2 Exercise Find the suprema and infima, if they exist, of the subset S of the poset (T,α):

a) $S = \{3,4,5\}$, $T = \mathrm{N}$, α is "$\leq$".
b) $S = \{3,4,5\}$, $T = \mathrm{N}$, α is "divides".
c) S is the set of all positive primes, $T = \mathrm{N}$, and α is "$\leq$".
d) S is the set of all positive primes, $T = \mathrm{N}$, α is "divides".
e) $S = \{\{1,2\},\{2,3\}\}$, $T = \mathcal{P}\{1,2,3\}$, α is inclusion.

(Answer on page 251.)

141.3.3 Least upper bounds of two elements There is a special notation for suprema and infima of subsets of two elements. If (A,α) is a poset and $a,b \in A$, then the supremum of $\{a,b\}$ is denoted $a \vee b$ and called the **join** of a and b, and the infimum is denoted $a \wedge b$ and called the **meet** of a and b. Using this notation, Rule (141.2) then gives this rule of inference:

$$a \mathrel{\alpha} c,\ b \mathrel{\alpha} c,\ ((\forall d)(a \mathrel{\alpha} d \text{ and } b \mathrel{\alpha} d) \Rightarrow c \mathrel{\alpha} d) \vdash\ c = a \vee b$$

There is a similar rule for $a \wedge b$.

141.3.4 Exercise (hard) Let (T,α) be a poset, and suppose $A \subseteq S \subseteq T$.

a) Show that if m is the supremum of A in S and n is the supremum of A in T, then $n \le m$.
b) Show that if n is the supremum of A in T and $n \in S$, then n is the supremum of A in S.
c) Give an example where the situation in (a) holds and $m \ne n$.

141.3.5 Exercise (hard) Show that if a and b are real numbers and

$$J = \{t \in \mathrm{Q} \mid a \le t \le b\}$$

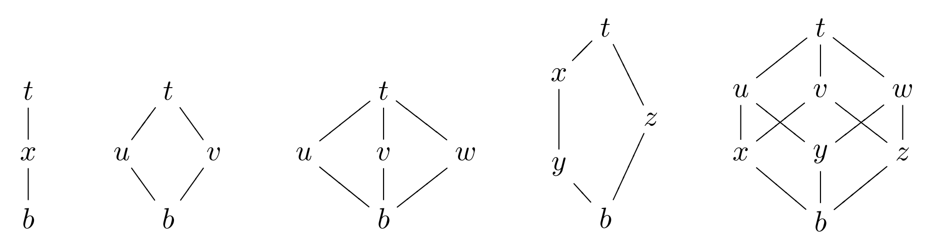
then the supremum of J in Q, if it exists, is b, so that b is rational. (Hint: Let n be the supremum of J in Q. Use Problem 141.3.4 to show that $b \le n$. Now assume $b < n$ and use the Archimedean property to get an integer k for which $1/(n-b) < k$, so that $b < n-(1/k) < n$ and $n-(1/k)$ is rational.)

142. Lattices

> **142.1 Definition: lattice**
> A poset (A,α) with the property that for any two elements a and b, $a \wedge b$ and $a \vee b$ always exist, is called a **lattice**. If $a \wedge b$ always exists, but not necessarily $a \vee b$, then (A,α) is called a **lower semilattice**, and if $a \vee b$ always exists but not necessarily $a \wedge b$, it is an **upper semilattice**.

142.1.1 Remark Some texts require that a lattice have a minimum and a minimum, as well.

142.1.2 Example The following are Hasse diagrams of lattices. Note that, for example, in (d), $x \wedge z = b$, $x \vee z = t$, and $x \vee y = x$.



(a) (b) (c) (d) (e)

(142.1)

142.1.3 Example In the unit interval $\mathrm{I} = \{r \in \mathrm{R} \mid 0 \le r \le 1\}$, the meet $r \wedge s$ and the join $r \vee s$ with respect to the usual weak ordering $\le$ always exist, and in fact $r \wedge s = \min(r,s)$ and $r \vee s = \max(r,s)$. Thus $(\mathrm{I},\le)$ is a lattice. More generally, any total ordering is a lattice (Exercise 142.1.11).

142.1.4 Example Let A be a set and B and C subsets of A. Then in $(\mathcal{P}A,\subseteq)$, $B \wedge C$ and $B \vee C$ always exist and moreover $B \wedge C = B \cap C$ and $B \vee C = B \cup C$. Thus $(\mathcal{P}A,\subseteq)$ is a lattice. (See Exercise 142.1.7.)

142.1.5 Example Let m and n be natural numbers. Then in $(\mathrm{N}, |)$, $m \wedge n$ and $m \vee n$ always exist, and moreover $m \wedge n = \mathrm{GCD}(m,n)$ and $m \vee n = \mathrm{LCM}(m,n)$. Thus $(\mathrm{N}, |)$ is a lattice. This follows immediately from Corollary 64.2, page 90.

142.1.6 Exercise Which of these posets are lattices?

a) $(\mathrm{N}, \leq)$.
b) $(\mathrm{Z}, \leq)$.
c) $(\mathrm{R}, \leq)$.
d) $(A, |)$, where A is the set of positive divisors of 25.
e) $(A, |)$, where A is the set of positive divisors of 30.
f) $(A, |)$, where $A = \{1,2,3,4,5,6\}$.

(Answer on page 251.)

142.1.7 Exercise Prove that for any set A, $(\mathcal{P}A, \subseteq)$ is a lattice. (Answer on page 251.)

142.1.8 Exercise Give an example of a lattice in which for some elements a, b and c, $a \wedge (b \vee c) \neq (a \wedge b) \vee (a \wedge c)$.

142.1.9 Exercise Show that in the lattice $(\mathrm{N} - \{0\}, |)$, every subset has an infimum and every finite subset has a supremum, but not every subset has a supremum.

142.1.10 Exercise Let n be a positive integer. Show that the set of positive divisors of n with "divides" as the relation is a lattice.

142.1.11 Exercise Prove that if (L, α) is a lattice, then α is a total ordering if and only if $x \vee y$ is the minimum of x and y and $x \wedge y$ is the minimum of x and y.

143. Algebraic properties of lattices

The following theorem gives algebraic properties of meet and join.

143.1 Theorem
If (A, α) is an upper semilattice, then for all $a, b, c \in A$,

a) $a \vee a = a$ *(idempotence).*
b) $a \vee b = b \vee a$ *(commutativity).*
c) $a \vee (b \vee c) = (a \vee b) \vee c$ *(associativity).*

Similarly, if (A, α) is a lower semilattice, then for all $a, b, c \in A$,

a) $a \wedge a = a$.
b) $a \wedge b = b \wedge a$
c) $a \wedge (b \wedge c) = (a \wedge b) \wedge c$.

Proof We will prove the associativity of $\wedge$ and leave the rest as an exercise. This proof involves applying the definition of infimum repeatedly to prove that each side of the equation is the infimum of the set $\{a,b,c\}$, and using the uniqueness of the infimum. I will show that $a \wedge (b \wedge c) = \inf\{a,b,c\}$ and leave the other side to you. The definition of infimum tells us that all the following are true:

(1) $b \wedge c \,\alpha\, b$

(2) $b \wedge c \,\alpha\, c$

(3) $a \wedge (b \wedge c) \,\alpha\, a$

(4) $a \wedge (b \wedge c) \,\alpha\, b \wedge c$.

Putting (1), (2) and (4) together and using transitivity gives that

(5) $a \wedge (b \wedge c) \,\alpha\, b$

(6) $a \wedge (b \wedge c) \,\alpha\, c$

(3), (5) and (6) tells us that

(7) $a \wedge (b \wedge c) \,\alpha\, \inf\{a, b, c\}$.

On the other hand, by definition

(8) $\inf\{a, b, c\} \,\alpha\, b$

(9) $\inf\{a, b, c\} \,\alpha\, c$

so

(10) $\inf\{a, b, c\} \,\alpha\, b \wedge c$.

Also

(11) $\inf\{a, b, c\} \,\alpha\, a$

so by (10) and (11),

(12) $\inf\{a, b, c\} \,\alpha\, a \wedge (b \wedge c)$.

Now (7), (12) and antisymmetry give us the desired result.

143.1.1 Exercise Complete the proof of Theorem 143.1.

143.1.2 Exercise Prove that in a lattice, $x \,\alpha\, y \Leftrightarrow x = x \wedge y \Leftrightarrow y = x \vee y$.

143.2 The Axiomatic Method

The proof that $\wedge$ and $\vee$ are associative is rather long, although conceptually not difficult. The value is that having done it once, we know it is true for every situation in which $\wedge$ and $\vee$ occur.

143.2.1 Example We now know immediately, by examples 142.1.3 through 142.1.5, that max and min, intersection, union, and GCD and LCM are all idempotent, commutative and associative. It is not hard to prove these directly (although the proof for GCD and LCM is not trivial), but once we know Theorem 143.1 and the corresponding fact for sups, the associativity doesn't need proof.

143.2.2 The idea is that *we have extracted salient properties of union, intersection,* GCD *and* LCM *and made them into axioms*; then any theorem derived from those axioms is true in all the cases all at once. This is an example of the **axiomatic method** in mathematics. The axiomatic method is largely responsible for the power of modern mathematics.

144. Directed graphs

144.1 About graphs in general

A graph is a mathematical construction that is used to encode information about connections between things. There are two main types of graphs, the kind called “undirected graph” in which only the connection between two things matters, and the kind called “directed graph” or “digraph” in which the direction of the connection matters. Each of these main types occurs in numerous subvarieties, only some of which are commonly used in computer science.

The terminology for different kinds of graphs in the literature is notoriously varied; it is probably true that if two graph theory books by different authors use the same terminology, one of the authors was the graduate student of the other one. The terminology in this text is similar to the usage in many (but not all) computer science books, but is quite different from that in books written by combinatorialists or graph theorists.

In this book, “graph” means undirected graph and “digraph” means directed graph. All graphs here are finite; although the definitions work for infinite graphs, many of the theorems are not true as stated for the infinite case.

144.1.1 Digraphs Informally, a digraph is a bunch of dots called nodes with arrows going from some nodes to others. Here are two examples.

$$
\begin{tikzcd}
x \arrow[r, shift left, "a"] \arrow[r, shift right, "b"'] \arrow[d, shift left, "d"] & y \arrow[loop above, "u"] \\
z \arrow[u, shift left, "c"] \arrow[ur, "e"'] & w
\end{tikzcd}
\qquad
\begin{tikzcd}
x \arrow[r, shift right, "b"'] \arrow[d, "c"'] & y \arrow[l, shift right, "a"'] \\
z &
\end{tikzcd}
\tag{144.1}
$$

Here is a more precise definition:

> **144.2 Definition: directed graph**
> A **directed graph** or **digraph** G consists of two finite sets G_0 and G_1 and two functions $\text{source}: G_1 \to G_0$ and $\text{target}: G_1 \to G_0$.
> The elements of G_0 are called the **nodes** or **vertices** (singular: vertex) of G and the elements of G_1 are the **arrows** or **directed edges** of G.
> If an arrow a has source x and target y we write $a : x \to y$ in the same way we write functions.

144.2.1 Drawing digraphs A digraph $\langle G_0, G_1, s, t\rangle$ is conventionally drawn using dots or labels for the nodes, and an (actual) arrow going from node x to node y for each arrow a (element of G_1) with source x and target y.

144.2.2 Exercise Draw the following digraphs:

a) The graph with nodes $\{A, B, C, D\}$ and exactly one arrow from each node to A.

b) $G = (G_0, G_1, s, t)$ where $G_0 = \{1, 2, 3\}$, $G_1 = \{a, b, c, d, e\}$, $s(a) = s(e) = 1$, $s(b) = s(c) = s(d) = 2$, $t(a) = 2$, $t(b) = t(c) = 1$, and $t(d) = t(e) = 3$.

(Answer on page 251.)

144.2.3 Exercise Draw the graph $G_0 = \{2,3,4,5,6,7,8,9,10\}$, with n arrows going from r to s if and only if $r^n \mid s$ and r^{n+1} does not divide s.

144.3 Definition: abstract description

The information about a digraph given by the definition, that is the sets G_0, G_1 and the source and target functions, is called the **abstract description** of the digraph.

144.3.1 Remark We will frequently encode the abstract description for a digraph as an ordered quadruple: thus "G is the digraph $\langle G_0, G_1, s, t\rangle$" means G_0 is the set of nodes, G_1 the set of arrows, and s and t are the source and target functions.

144.3.2 Example The abstract description of the digraph on the left of Figure (144.1) has $G_0 = \{x,y,z,w\}$, $G_1 = \{a,b,c,d,e,u\}$,

$$\text{source}(a) = \text{source}(b) = \text{source}(d) = \text{target}(c) = x$$

$$\text{target}(a) = \text{target}(b) = \text{target}(e) = \text{source}(u) = \text{target}(u) = y$$

and $\text{source}(c) = \text{source}(e) = z$.

144.4 Graphs and abstraction

A digraph is defined here in an abstract way, not as a picture. The interplay between the abstract definitions and the pictures is analogous to that between the formula of a function such as $f(x) = x^2 + 1$ and its graph (a parabola) in analytic geometry. The pictures are more suggestive and comprehensible than the abstract definition, but it is difficult to prove things using pictures because it is hard to be sure you have the most general case. It may also be difficult or wasteful (or both) to store pictures directly in the computer. The abstract treatment is both more rigorous and more amenable to computation.

144.5 Digraphs in applications

144.5.1 Example Digraphs provide a natural way to encode data about certain kinds of complex systems. The flow chart of a program, for example, is a digraph. The commutative diagrams of sets and functions in Chapter 98 are examples of labeled digraphs. However, the information concerning the composites of the functions is additional information not encoded by the description of the diagrams as a digraph.

144.5.2 Example Digraphs are the natural way to model the sequencing of a collection of tasks that must be performed to accomplish a goal. Each node is a task and there is an arrow from task a to task b if task a must be completed before task b can be started. For example, the task of computing $\log(x^2 + y^3)$ can be modeled this way:

$$\begin{tikzcd}
 & \text{calculate } x^2 \arrow[dr] & & \\
\text{start} \arrow[ur] \arrow[dr] & & \text{add} \arrow[r] & \text{calculate log} \\
 & \text{calculate } y^3 \arrow[ur] & &
\end{tikzcd} \tag{144.2}$$

This graph shows, for example, that if you had two people or two processors to perform the squaring you could speed up the computation. Digraphs arising in this way often have a weight function on the arrows.

144.5.3 Exercise Draw the digraph modeling the computation of the truth value of the equation

$$x^2 + xy^2 = x^2 - y$$

145. Miscellaneous topics about digraphs

> **145.1 Definition: loop**
> An arrow a from a node to itself, in other words $a : x \to x$ for some node x, is called a **loop**.

145.1.1 Example u is a loop in the left digraph in Figure (144.1).

> **145.2 Definition: indegree and outdegree**
> The number of arrows that have a node as source is called the **outdegree** of the node, and the number of arrows that have the node as target is the **indegree**.

145.2.1 Example The node y in the left graph of Figure (144.1) has indegree 4 and outdegree 1.

> **145.3 Definition: opposite of a graph**
> The **opposite** of a digraph G is the digraph with the same nodes and all the arrows reversed. It is called G^{op}. Thus if $G = \langle G_0, G_1, s, t\rangle$, then $G^{\mathrm{op}} = \langle G_0, G_1, t, s\rangle$.

145.3.1 Example The digraphs below are opposites of each other.

$$\begin{tikzcd}
A \arrow[r] \arrow[d] \arrow[dr] & B \arrow[d, shift right] \\
C & D \arrow[u, shift right]
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
A & B \arrow[l] \arrow[d, shift right] \\
C \arrow[u] & D \arrow[u, shift right] \arrow[ul]
\end{tikzcd}$$

145.4 Labeling

A **labeling** of the nodes of a digraph G is a function $L : G_0 \to S$, where S is a set. If x is a node, its label is $L(x)$. Similarly a function $L : G_1 \to S$ would label the arrows. As an example, the digraph below shows the cost of traveling by rail in a (mythical) mountainous country between three cities A, B, and C. (The fare for going to a higher elevation is more than for going to a lower one.)

$$\begin{tikzcd}
& A \arrow[dl, shift left, "100"] \arrow[dr, shift left, "120"] & \\
B \arrow[ur, shift left, "130"] \arrow[rr, shift left, "90"] & & C \arrow[ul, shift left, "90"] \arrow[ll, shift left, "90"]
\end{tikzcd} \tag{145.1}$$

The nodes are labeled by $\{A, B, C\}$ and the arrows are labeled by integers representing cost. Here the labeling is a function $F : G_1 \to \mathrm{Z}$. A function labeling arrows by integers or real numbers is commonly called a **weight function** on the arrows. You can see that the labeling of the nodes is injective but the labeling of the arrows is not. When the labeling of the nodes is injective, there is usually no harm in taking the attitude that the labels are actually the nodes; a similar remark applies to an injective labeling of the arrows.

146. Simple digraphs

> **146.1 Definition: simple digraph**
> A digraph is **simple** if for two distinct arrows a and b, either $\mathrm{source}(a) \neq \mathrm{source}(b)$ or $\mathrm{target}(a) \neq \mathrm{target}(b)$. In other words, only one arrow can go from a node to another node. (However, one arrow *is* allowed each way.)

146.1.1 Example The left graph in Figure (144.1), page 218, is not a simple digraph, whereas the right one is.

146.1.2 Exercise What is the largest number of arrows a simple digraph with n nodes can have?

146.1.3 Variation in terminology In many books the word "digraph" is used only for simple digraphs; those that allow more than one arrow from a node to a node are called "multigraphs" or "multidigraphs".

A simple digraph can be given a much simpler (!) abstract description (of a graph). Since there can be at most one arrow from a node to another one, all you have to do to describe the digraph is to give the set G_0 of nodes and the subset A of $G_0 \times G_0$ of ordered pairs of those nodes that have an arrow going from the first node to the second one. This is summed up in the following definition.

146.2 Definition: relational description

The **relational description** of a simple digraph G is (G_0, A), where $A \subseteq G_0 \times G_0$ is the set of ordered pairs

$$\{\langle m,n\rangle \mid \text{There is an arrow from } m \text{ to } n\}$$

146.2.1 Remark We saw this correspondence between simple digraphs and relations from the opposite point of view in 51.2.

146.2.2 Example In the case of the right graph in Figure (144.1), which is simple, G_0 is $\{x,y,z\}$ and A is $\{\langle x,y\rangle, \langle y,x\rangle, (x,z\rangle\}$.

146.2.3 Exercise Which of the digraphs in Exercise 144.2.2 are simple? Give the relational description of each one that is. (Answer on page 251.)

146.2.4 Exercise Give the relational description of the graph (147.1), page 223.

146.2.5 Fact The relational description can be converted to the original definition of digraph by calling a pair $\langle x,y\rangle$ in A an arrow from x to y; thus the source is the first coordinate and the target is the second.

To sum up:

(i) If $\langle G_0, G_1, s, t\rangle$ is the abstract description (of a graph) of a simple digraph, you get the relational description $\langle G, A\rangle$ of the same graph by taking $G = G_0$ and

$$A = \{\langle x,y\rangle \in G_0 \times G_0 \mid (\exists s)(s : x \to y) \text{ in } G_1\}$$

(ii) If $\langle G, A\rangle$ is the relational description of a simple digraph, the abstract description (of a graph) of the same graph is defined to be $\langle G_0, G_1, s, t\rangle$, where $G_0 = G$, $G_1 = A$, $s = p_1$ (the first coordinate function) and $t = p_2$.

147. Isomorphisms

The two digraphs below are abstractly identical in a sense that can be made precise. The idea is that node a in the left digraph plays the same role as node 2 in the right digraph, and similarly b and 1 match up and c and 3 match up. "Playing the same role" means precisely that if you match node x in one digraph to node m in another, and similarly node y to n, then the arrows from x to y must match up with the arrows from m to n. (You should check these two digraphs to see that this happens).

$$\begin{tikzcd} a \arrow[rr, shift left, "f"] \arrow[dr, "h"'] & & b \arrow[ll, shift left, "g"] \arrow[dl, "k"] \\ & c & \end{tikzcd} \qquad \begin{tikzcd} 1 \arrow[r, bend left, "u"] \arrow[rr, bend right, "x"'] & 2 \arrow[l, "v"] \arrow[r, "w"] & 3 \end{tikzcd} \tag{147.1}$$

This is made precise this way:

147.1 Definition: isomorphism
Let $G = \langle G_0, G_1, s, t\rangle$ and $G' = \langle G'_0, G'_1, s', t'\rangle$ be digraphs. An **isomorphism** from G to G' is a pair of bijections $\beta_0 : G_0 \to G'_0$ and $\beta_1 : G_1 \to G'_1$ with the property that $a : x \to y$ in G if and only if $\beta_1(a) : \beta_0(x) \to \beta_0(y)$ in G'.

147.1.1 Remark Since there is rarely any problem with ambiguity, the subscripts may be omitted from β_0 and β_1.

147.1.2 Example In Figure 147.1 there is an isomorphism β from the left figure to the right figure defined by

$$\begin{array}{ll} \beta(a) = 2 & \beta(f) = v \\ \beta(b) = 1 & \beta(g) = u \\ \beta(c) = 3 & \beta(h) = w \\ & \beta(k) = x \end{array}$$

The inverse of this isomorphism (meaning $\langle \beta_0^{-1}, \beta_1^{-1}\rangle$) is also an isomorphism; in fact the inverse of any digraph isomorphism is also an isomorphism.

147.1.3 Remark It is easily possible for two digraphs to be isomorphic *in more than one way.* This happens in Figure 147.1, for example.

147.1.4 Exercise (hard) Show that two digraphs are isomorphic if and only if there is an ordering of their nodes for which their adjacency matrices are identical.

147.1.5 Exercise Draw both (nonisomorphic) simple digraphs that have only one node, and all ten (nonisomorphic) simple digraphs that have two nodes.

147.1.6 Exercise Let $G = \langle G_0, G_1, s, t\rangle$ and $G' = \langle G'_0, G'_1, s', t'\rangle$ be digraphs. Prove that $\beta_0 : G_0 \to G'_0$ and $\beta_1 : G_1 \to G'_1$ constitute an isomorphism if and only if β and β' are bijections and $s' \circ \beta_1 = \beta_0 \circ s$ and $t' \circ \beta_1 = \beta_0 \circ t$.

147.1.7 Exercise Let $\beta = \langle \beta_0 : G_0 \to G'_0,\ \beta_1 : G_1 \to G'_1 \rangle$ be a digraph isomorphism from $G = \langle G_0, G_1, s, t \rangle$ to $G' = \langle G'_0, G'_1, s', t' \rangle$. Show that β^{-1}, i.e., $\langle \beta_0^{-1}, \beta_1^{-1} \rangle$, is a digraph isomorphism from G' to G.

147.2 Definition: automorphism
An isomorphism $\beta : G \to G$ of a digraph with itself is called an **automorphism**.

147.2.1 Example For any digraph, the identity function is an automorphism. The digraphs in Figure 147.1 each have two automorphisms, the identity and one other.

147.2.2 Exercise Find the automorphisms of the digraphs in exercise 144.2.2. (Answer on page 251.)

147.2.3 Exercise (hard) Let G be a digraph with exactly n automorphisms, and let G' be a digraph isomorphic to G. Show that there are exactly n isomorphisms from G to G'.

147.2.4 Exercise (hard) For any positive integer n, show how to construct a digraph with exactly n automorphisms.

148. The adjacency matrix of a digraph

A convenient way for representing a digraph G in a computer program is by means of its adjacency matrix.

148.1 Definition: adjacency matrix
The **adjacency matrix** of a digraph G is a matrix of nonnegative integers whose entries are indexed by $G_0 \times G_0$ and whose entry in the location indexed by the pair of nodes $\langle x, y \rangle$ is the number of arrows from x to y.

148.1.1 Example For the left digraph in Figure 144.1 the adjacency matrix is

$$\begin{array}{ccccc} & x & y & z & w \\ x & 0 & 2 & 1 & 0 \\ y & 0 & 1 & 0 & 0 \\ z & 1 & 1 & 0 & 0 \\ w & 0 & 0 & 0 & 0 \end{array}$$

148.1.2 Remark The adjacency matrix depends on the way the nodes are ordered; thus if you permute the nodes you get a different adjacency matrix for the same graph. Note that the adjacency matrix does not contain the information concerning the names of the arrows.

148.1.3 Exercise Draw the graph with this adjacency matrix:

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 1 & 1 \\ 2 & 0 & 0 & 1 & 1 \\ 3 & 0 & 1 & 0 & 1 \\ 4 & 1 & 0 & 0 & 0 \end{array}$$

(Answer on page 251.)

148.1.4 Exercise Give the relational description of the digraph in Exercise 148.1.3. (Answer on page 251.)

148.1.5 Uses of the adjacency matrix You can use the adjacency matrix of a graph to determine properties of the graph:

(i) It is simple if no entry in the adjacency matrix is greater than 1.

(ii) It has no loops if the entries down the main diagonal (the one from upper left to lower right) are all 0.

(iii) The outdegree of a node is the sum over its row and the indegree is the sum over its column.

The adjacency matrix will be used in the next section to calculate which nodes can be reached from a given node.

148.1.6 Exercise Give the adjacency matrices of the digraphs in Figure 147.1. (Answer on page 251.)

148.1.7 Exercise Draw this digraph and give its adjacency matrix: The nodes are the numbers 1,2,3,4,6,12 and there is an arrow from a to b if and only if a and b have the same prime factors (in other words, for all primes p, $p \mid a \Leftrightarrow p \mid b$).

149. Paths and circuits

149.1 Definition: directed walk

A **directed walk** of length k from a node p to a node q in a digraph is a tuple $\langle a_1, \ldots, a_k \rangle$ of arrows for which

P.1 $\text{source}(a_1) = p$;

P.2 $\text{target}(a_k) = q$; and

P.3 if $k > 1$, then for each $i = 1, \ldots, k-1$, $\text{source}(a_{i+1}) = \text{target}(a_i)$.

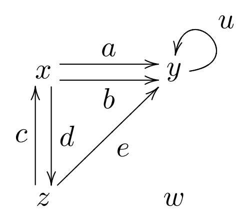
149.1.1 Remarks

a) By definition, the length of a directed walk is the number of *arrows* it goes through. If it goes through an arrow twice, the arrow is counted twice. A directed walk of length n will thus make $n+1$ visits to nodes, counting the start and finish nodes, and the same node may be visited more than once.

b) We allow the empty walk $\langle\rangle$ from any node to itself.

149.1.2 Example All these refer to digraph (149.1) below.

a) The walk $\langle u \rangle$ on the left digraph is of length one and touches the node y twice.
b) The empty walk $\langle \rangle$ from y to y is also a walk (of length 0); it is not the same as $\langle u \rangle$.
c) The walk $\langle c,d,e \rangle$ goes from z to y and touches z twice.
d) The walk $\langle c,d,c,d \rangle$ goes from z to z and touches each of x and z twice.
e) $\langle e,a,d \rangle$ is not a directed walk because an arrow goes the wrong way.



(149.1)

149.2 Definition: directed path
A **directed path** is a directed walk in which the arrows $a_1, \ldots, a_k$ are all different.

149.2.1 Example In the digraph (149.1):

a) $\langle c,a,u \rangle$ is a directed path of length 3 from z to y.
b) $\langle d,c,a \rangle$ is a directed path of length 3 from x to y.
c) $\langle e \rangle$ is a directed path of length 1 from z to y.
d) $\langle d,c,d,e \rangle$ is a directed walk that is not a directed path.

149.3 Definition: directed circuit
A **directed circuit** is a directed path from a node to itself.

149.3.1 Remark A directed circuit must be a path, not merely a walk.

149.3.2 Example In the digraph (149.1), the only directed circuits are the three empty paths, $\langle c,d \rangle$, $\langle d,c \rangle$ and $\langle u \rangle$. (Thus a loop is a directed circuit.)

149.4 Definition: simple directed path
A **simple directed path** is a directed path not containing any directed circuits, so that you never hit a node twice.

149.4.1 Example The only simple directed paths from z to y in the digraph (149.1) are $\langle c,a \rangle$, $\langle c,b \rangle$, and $\langle e \rangle$.

149.4.2 Example Programs in many languages such as Pascal are made up of procedures or functions that call on each other. It is often useful to draw a digraph in which the nodes are the procedures and functions and there is an arrow from P to Q if Q is called when P is run. A loop in such a digraph indicates a procedure or function that calls itself recursively. Larger circuits indicate indirect recursion.

149.4.3 Exercise Find all the simple directed paths from 1 to 3 in the digraph $G=(G_0,G_1,s,t)$, where $G_0=\{1,2,3\}$, $G_1=\{a,b,c,d,e\}$, $s(a)=s(e)=1$, $s(b)=s(c)=s(d)=2$, $t(a)=2$, $t(b)=t(c)=1$, and $t(d)=t(e)=3$. (This is the same as the digraph in Exercise 144.2.2(b).) (Answer on page 251.)

149.4.4 Exercise A digraph is **transitive** if whenever there are arrows $x\to y$ and $y\to z$, there must be an arrow $x\to z$. Show that a digraph is transitive if and only if whenever there is a walk from x to y there is an arrow $x\to y$.

150. Matrix addition and multiplication

The adjacency matrix of a digraph can be used to compute directed walks from one node to another. This involves the concepts of matrix addition and multiplication, which are described briefly here.

150.1 Definition: scalar product
Let V and W be two n-tuples of real numbers. The **scalar product** $V\cdot W$ is the sum $\Sigma_{i=1}^{n} V_iW_i$.

150.1.1 Example $\langle 3,5,-1,0\rangle\cdot\langle 1,2,3,4\rangle = 10$.

150.1.2 Usage The scalar product is also called the "dot" product. You may be familiar with its geometrical meaning when the tuples represent vectors.

150.1.3 Remark The scalar product is only defined for two tuples of the same length. For each positive integer n, it is a function $\mathrm{R}^n\times\mathrm{R}^n\to\mathrm{R}$.

150.2 Definition: product of matrices
Let A be a $k\times m$ matrix with real entries, and B an $m\times n$ matrix with real entries; specifically, A has the same number of *columns* as B has *rows.* Then the **product** AB of the matrices is the $k\times n$ matrix whose $\langle i,j\rangle$th entry is the scalar product of the ith row of A and the jth column of B. In other words,

$$(AB)_{ij}=\Sigma_{k=1}^{m}A_{ik}B_{kj} \tag{150.1}$$

150.2.1 Example

$$\begin{pmatrix} 1 & 3 & 0\\ 2 & 2 & 2\end{pmatrix}\cdot\begin{pmatrix} 2 & 1 & 0 & 5\\ 3 & -2 & 1 & -1\\ 5 & 1 & 1 & 0\end{pmatrix}=\begin{pmatrix} 11 & -5 & 3 & 2\\ 20 & 0 & 4 & 8\end{pmatrix} \tag{150.2}$$

150.2.2 Fact Matrix multiplication is associative, when it is defined; in other words, for a $k\times m$ matrix A, an $m\times n$ matrix B and an $n\times p$ matrix C, $(AB)C=A(BC)$. Matrix multiplication is not, however, commutative. There are $n\times n$ matrices A and B for which $AB\neq BA$. (Note that if AB and BA are both defined, then A and B must be square matrices.)

150.2.3 Exercise Give examples of 2×2 matrices showing that matrix multiplication is not commutative.

150.2.4 Exercise Show that matrix multiplication is associative when it is defined.

150.3 Definition: sum of matrices

Let M and N be $m \times n$ matrices. Then the sum $M + N$ is defined by requiring that $(M + N)_{ij} = M_{ij} + N_{ij}$.

150.3.1 Remark Two matrices can be added if and only if they have the same dimensions.

150.3.2 Example

$$\begin{pmatrix} 2 & 5 \\ 3 & -3 \end{pmatrix} + \begin{pmatrix} 7 & -1 \\ 5 & 5 \end{pmatrix} = \begin{pmatrix} 9 & 4 \\ 8 & 2 \end{pmatrix} \tag{150.3}$$

150.4 Powers of matrices

In the following, we will use powers of square matrices with integer coefficients. If M is a square $m \times m$ matrix, M^n denotes M multiplied by itself $n-1$ times. This is best defined by induction: $M^0 = I$, $M^1 = M$, and $M^n = M^{n-1} \cdot M$. It follows from this and Definition 150.2 that

$$(M^n)_{ij} = \Sigma_{k=1}^{m} (M^{n-1})_{ik} M_{kj} \tag{150.4}$$

151. Directed walks and matrices

151.1 Theorem

If $G = (G_0, G_1, s, t)$ is a digraph with adjacency matrix M, then the number of directed walks of length k from node p to node q is the $\langle p, q \rangle$ th entry of M^k.

Proof This fact can be proved by induction on k. It is clear for $k = 1$, since a directed walk of length 1 is just an arrow, and the $\langle p, q \rangle$ th entry in $M^1 = M$ is the number of arrows from p to q by definition.

Suppose it is true that for all nodes p and q, the $\langle p, q \rangle$ th entry of M^k is the number of directed walks of length k from p to q. A directed walk of length $k + 1$ from p to q is a directed walk of length k from p to some node r followed by an arrow (directed walk of length 1) from r to q. By the induction hypothesis, there are $(M^k)_{pr}$ directed walks of length k from p to r, and there are M_{rq} arrows from r to q. Hence the number of directed walks of length $k + 1$ from p to q that consist of a directed walk of length k from p to r followed by an arrow from r to q is $(M^k)_{pr} \times M_{rq}$. The total number of directed walks of length $k + 1$ from p to q must be obtained by adding up this number $(M^k)_{pr} \times M_{rq}$ for each node r of the

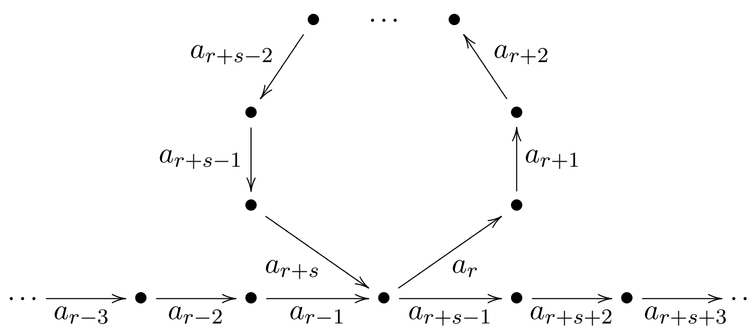


Figure 151.1: Walk with loop.

digraph; in other words, if there are n nodes in the digraph, the total number of walks is

$$\Sigma_{r=1}^{n}\Big((M^k)_{pr} \times M_{rq}\Big) \tag{151.1}$$

That sum, by formula (150.1), is the $\langle p,q\rangle$th entry of M^{k+1}, which is $M^k \times M$, and that is what we had to prove.

151.2 Reachability

Let p and q be nodes of a digraph G. One says that q is **reachable** from p if there is at least one directed walk of some length (possibly zero) from p to q.

Since a directed walk of length k touches $k+1$ nodes, it follows from the pigeonhole principle that a directed walk of length n or more in a digraph G with n nodes must touch some node twice. Suppose such a walk $\langle a_1,\ldots,a_k\rangle$ touches a node x twice; say arrow a_r has source x and arrow a_{r+s} (with $s \geq 0$) has target x. Then the directed walk $\langle a_r,\ldots,a_{r+s}\rangle$ can be eliminated from the walk, as in Figure 151.1, giving

$$\langle a_1,\ldots,a_{r-1},a_{r+s+1},\ldots,a_k\rangle \tag{151.2}$$

from p to q. (Note: if $r=1$ or $r+s=k$, the walk (151.2) has to be modified in an obvious way.)

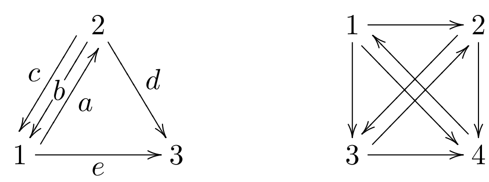
Clearly, by successively eliminating circuits, one can replace the walk by a path (not just a walk) of length$< n$. This leads to:

151.3 Corollary

Let G be a digraph as in Theorem 151.1 with n nodes and matrix M. Then q is reachable from p if and only if the $\langle p,q\rangle$th entry of the matrix $K = I + M + M^2 + \ldots + M^{n-1}$ is nonzero.

Proof If there is a directed walk from p to q, then the argument before the statement of the Corollary shows that there must be one of length $n-1$ or less. This means that one of the matrices M, M^2, $\ldots$, M^{n-1} has a nonzero $\langle p,q\rangle$th entry. Since all the entries in these matrices are nonnegative, this means that the $\langle p,q\rangle$th entry of K is nonzero. Conversely, if that entry is nonzero it must be because the $\langle p,q\rangle$th entry in M^i for some i is nonzero.

151.3.1 Exercise Use matrix multiplication to find all the directed walks of length 1, 2, 3 and 4 that go from 1 to 3 in these digraphs:



(151.3)

(a) (b)

(Answer on page 251.)

> **151.4 Definition: reachability matrix**
> The matrix
> $$K = I + M + M^2 + \ldots + M^{n-1}$$
> is called the **reachability matrix** for the digraph G.

151.4.1 Exercise Calculate the reachability matrices for the digraphs in Figure 144.1, page 218. (Answer on page 251.)

151.4.2 Exercise Let G be the digraph whose set of nodes is $\{1,2,3,4\}$, with an arrow from a to b if and only if a is even and b is 2 or 3. Find the reachability matrix of G by counting paths and by direct addition and multiplication of matrices. (You may use Mathematica for the latter.)

151.4.3 Exercise Let D be a digraph with adjacency matrix M. Show that D is transitive (as defined in the preceding problem) if and only if

$$(M^2)_{ij} \neq 0 \Rightarrow M_{ij} \neq 0$$

for all pairs $\langle i,j\rangle$.

152. Undirected graphs

In Chapters 144 through 151, we considered digraphs that consisted of nodes and arrows between some of the nodes. The graphs considered in this section have nodes with edges between them, but the edges have no direction assigned to them.

> **152.1 Definition: graph**
> A **graph** G consists of two finite sets G_0 and G_1 together with a function Γ from G_1 to the set of two-element subsets of G_0. The elements of G_0 are called **nodes** or x**vertices** of G and the elements of G_1 are called **edges**.

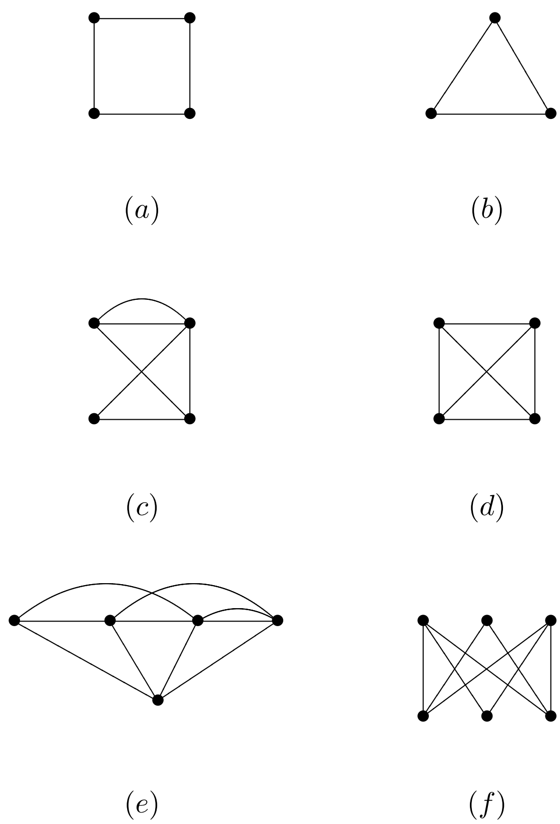


Table 152.1: Some graphs

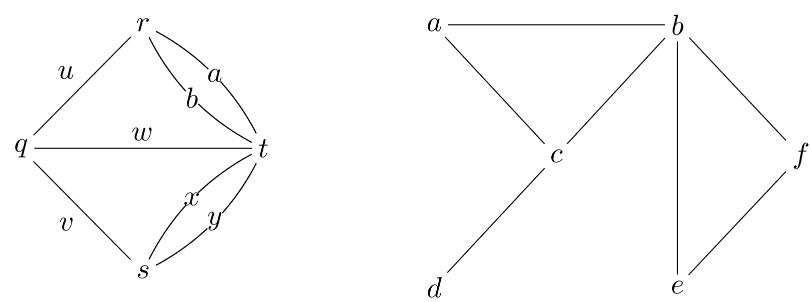
152.2 Definition: simple graph
G is a **simple graph** if Γ is injective, so that there is no more than one edge connecting two nodes.

152.2.1 Exercise Which of the graphs in Table (152.1) are simple? (Answer on page 251.)

152.2.2 Remark We will sometimes use the word "multigraph" to emphasize that we are talking about a graph that is not necessarily simple.

152.2.3 Drawing graphs One draws a graph by using dots for the nodes, and drawing a line between nodes p and q for each edge e for which $\Gamma(e) = \{p, q\}$. In common with most of the literature on the subject, our graphs do not have loops: the requirement that Γ have values in the set of *two*-element subsets rules out the possibility of loops.

152.2.4 Example The figure below shows two graphs; the one on the right is simple.



(152.1)

In the left graph the set of nodes is $\{q,r,s,t\}$, the set of edges is $\{a,b,u,v,x,w,y\}$, and, for example, $\Gamma(a)=\{r,t\}$.

152.3 Definition: incidence

If e is an edge in a graph and $\Gamma(e)=\{p,q\}$ then e is said to be **incident** on p (and on q). Two nodes connected by an edge in a simple graph are **adjacent**. If n edges connect two nodes the nodes are said to be **adjacent with multiplicity** n.

152.4 Definition: adjacency matrix

The **adjacency matrix** of a graph is the square matrix A whose rows and columns are indexed by the set of nodes, with $A(p,q) =$ the number of edges between p and q.

152.4.1 Fact It follows from the definition that for any (multi)graph with adjacency matrix A,

(i) for any node p, $A(p,p)=0$;
(ii) for any nodes p and q, $A(p,q)=A(q,p)$ (this says A is **symmetric**); and
(iii) if the graph is simple, A has only 0's and 1's as entries.

152.4.2 Remark Because of 152.4.1(i) and (ii), all the information about the graph is contained in the triangular matrix consisting of the entries $A(p,q)$ with $p<q$.

152.4.3 Example The adjacency matrix of the left graph in Figure (152.1) is

	r	q	t	s
r	0	1	2	0
q	1	0	1	1
t	2	1	0	2
s	0	1	2	0

152.5 Definition: degree
The **degree** of node is the number of edges incident on that node.

152.5.1 Example The degree of the node c in the right graph in Figure (152.1), page 232, is 3, and the degree of d is 1.

152.5.2 Fact The degree of a node is the sum over the row (and also over the column) of the adjacency matrix corresponding to that node.

152.5.3 Exercise Show that the sum of the degrees of the nodes of a graph is twice the number of edges.

153. Special types of graphs

Two special kinds of graphs that will be referred to later are given in the following definitions.

153.1 Definition: complete graph on n nodes
A **complete graph on n nodes** is a simple graph with n nodes, each pair of which are adjacent. Such a graph is denoted K_n.

153.1.1 Example K_4 is shown in diagram (153.1) below.

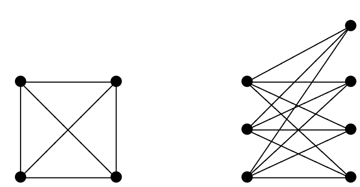
153.1.2 Exercise Give a formula for the number of edges of K_n for $n > 0$.

153.2 Definition: bipartite graph
A **bipartite graph** G is a graph whose nodes are the union of two disjoint nonempty subsets A and B, called its **moieties**, with the property that every edge of G connects a node of A to a node of B.

153.2.1 Fact It follows from Definition 153.2 that no two nodes of A are adjacent, and similarly for B.

153.3 Definition: complete bipartite graph
A bipartite graph G with moieties A and B is a **complete bipartite graph** if *every* node of A is connected to *every* node of B. A complete bipartite graph for which A has m elements and B has n elements with $m \leq n$ is denoted $K_{m,n}$.

153.3.1 Example The right graph in the following figure is $K_{3,4}$.



(153.1)

K_4 $\qquad$ $K_{3,4}$

153.3.2 Exercise Which of the graphs in Table (152.1), page 231 are complete graphs? (Answer on page 251.)

153.3.3 Exercise Which of the graphs in Table (152.1), page 231 are bipartite graphs? Which are complete bipartite graphs? (Answer on page 251.)

153.3.4 Exercise Give a formula for the number of edges of the complete bipartite graph $K_{m,n}$.

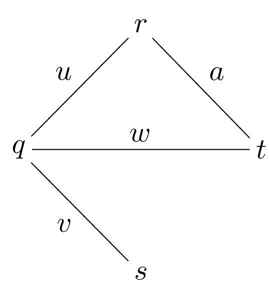
154. Subgraphs

154.1 Definition: subgraph
A **subgraph** of a graph G is a graph G' whose nodes G'_0 are a subset of the nodes G_0 of G, and for which every edge of G' is an edge of G between nodes of G'. If *every* edge of G that connects nodes of G' is an edge of G', then G' is a **full** subgraph of G.

154.1.1 Usage For some authors, "subgraph" means what we call a full subgraph.

154.1.2 Fact If G' is a subgraph of G, the edge function Γ' for G' is the restriction to G'_0 of the edge function Γ of G.

154.1.3 Example The following graph is a non-full subgraph of the left graph in Figure (152.1), page 232.



(154.1)

154.1.4 Exercise Show that if K_n is a subgraph of a simple graph G, then it is a full subgraph. Is the same true of $K_{m,n}$?

155. Isomorphisms

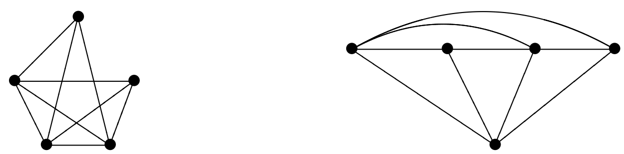
155.0.5 Remark Isomorphism of graphs is analogous to isomorphism of digraphs: it captures the idea that two graphs are the same in their connectivity — there is a way of matching up the nodes so that the edges match up too.

155.1 Definition: isomorphism

Let G and H be simple graphs. A function $\beta : G_0 \to H_0$ is an **isomorphism from G to H** if it is a bijection with the property that p and q are adjacent in G if and only if $\beta(p)$ and $\beta(q)$ are adjacent in H. G and H are **isomorphic** if there is an isomorphism from G to H.

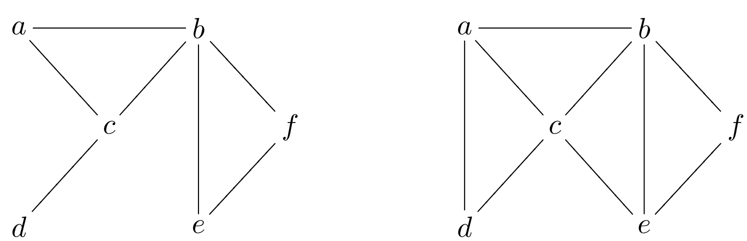
155.1.1 Usage In electrical engineering, isomorphic graphs are said to have the "same topology".

155.1.2 Example In general there may be more than one isomorphism between G and H. The graphs below are isomorphic. Altogether, there are 12 isomorphisms between them.



(155.1)

155.1.3 Example The left graph below is *not* isomorphic to the right graph. The identity map is a bijection on the nodes, and if nodes are adjacent in the left graph, they are adjacent in the right graph, but there are nodes in the right graph that are adjacent there but not in the left graph. The definition of isomorphism requires that p and q be adjacent *if and only if* $\beta(p)$ and $\beta(q)$ are adjacent.



(155.2)

155.1.4 Exercise Group the graphs in Table (152.1), page 231 according to which are isomorphic to each other. (Answer on page 251.)

155.1.5 Exercise In Table (152.1), page 231, show that (b) is isomorphic to a full subgraph of (c), and to a nonfull subgraph of (c). (Answer on page 251.)

155.1.6 Exercise

a) Prove that any two complete graphs on n nodes are isomorphic.
b) Prove that if $n \leq m$, then a complete graph on n nodes is isomorphic to a full subgraph of a complete graph on m nodes.
c) Prove that for fixed integers m and n, two complete bipartite graphs, each of which has one moiety with m nodes and the other moiety with n nodes, are isomorphic.

155.1.7 Exercise

a) Give a definition of isomorphism for multigraphs.
b) Prove that a graph isomorphic to a simple graph (using your definition of isomorphic) is simple.
c) Prove that for simple graphs your definition of isomorphism is the same as Definition 147.1.

156. Connectivity in graphs

We talk about walks, paths and circuits in graphs in much the same way as for digraphs.

156.1 Definition: walk
A **walk** from node p to node q in a graph is a sequence

$$\langle n_0, e_1, n_1, e_2, \ldots, n_{k-1}, e_k, n_k \rangle$$

of alternating nodes and edges for which $n_0 = p$, $n_k = q$, and e_i is incident on n_{i-1} and n_i for $i = 1, 2, \ldots, k$. The **length** of such a walk is k, which is the number of edges occurring in the list (counting repetitions), or one less than the number of nodes ocurring in the list.

156.2 Definition: path
A **path** in a graph is a walk in which no edges are repeated. A **simple path** is a path in which no nodes are repeated.

156.3 Definition: circuit
A **circuit** is a path (not a walk) from a node to itself, and a **cycle** is a circuit in which no nodes are repeated except that the beginning and end are the same.

156.3.1 Fact It is easy to see (eliminate circuits) that if there is a walk between two nodes then there is a simple path between them.

156.4 Definition: connected
A graph is **connected** if there is a path (hence a simple path) between any two nodes. If p is a node in a graph, let $C(p)$ denote the set consisting of p and of all nodes q for which there is a path between p and q. The sets $C(p)$ are called the **connected components** of the graph G.

156.4.1 Fact Part (a) of the theorem below implies that two nodes in a graph are joined by a path if and only if they are in the same connected component. A graph is therefore connected if and only if it has just one connected component.

156.5 Theorem
Let G be a graph.
a) Let p be a node in G. For any two nodes q and r in $C(p)$ there is a path from q to r.
b) If $q \in C(p)$ then $C(p) = C(q)$.
c) The set $\{C(p) \mid p \in G_0\}$ is a partition of G.

Proof For (a), if $p = q$ or $p = r$ there is a path from q to r by definition of $C(p)$. Otherwise, just connect the path from p to q to the path from p to r. The result might only be a walk, but by eliminating circuits, you get a path. That proves (a).

If $q \in C(p)$, (a) implies there is a path from p to r if and only if there is a path from q to r, so (b) follows. Finally, any node p is an element of $C(p)$; this and (b) implies that every node is in exactly one set $C(p)$, so the sets $C(p)$ form a partition of the nodes. That proves (c).

156.6 Definition: distance
The **distance** between two nodes p and q in a connected graph is the length of the shortest simple path between p and q.

156.6.1 Example In the right graph of Figure (152.1), the distance between nodes d and f is 3. There are of course simple paths of length 4 and 5 between nodes d and f, but the shortest one has length 3.

156.7 Definition: diameter
The **diameter** of a connected graph is the maximum distance between any two nodes in the graph.

156.7.1 Example The diameter of the graph just mentioned is 3.

157. Special types of circuits

157.1 Definition: Eulerian circuit
An **Eulerian circuit** is a circuit in a graph which contains each edge exactly once. It need not be a cycle; in other words, nodes may be repeated, but not edges.

A graph need not have an Eulerian circuit. For example, the graph in Figure (152.1) has no Eulerian circuit. There is a simple criterion for whether a graph has an Eulerian circuit:

157.2 Theorem
A connected graph G has an Eulerian circuit if and only if the degree of every node is even.

Proof Suppose G has an Eulerian circuit. As you go around the circuit, you have to hit every edge exactly once. Every time you go through a node, you must therefore leave by a different edge from the one you entered. So for each node p, you can divide the edges incident to p into two groups: those you enter p on and those you leave p on. Since you enter and leave p the same number of times, these two groups of edges must have the same number of elements. Thus the number of edges incident on p is even.

Now for the converse: suppose every node of G has even degree. To construct an Eulerian circuit, pick a node p. If that is the only node in G you are finished. Otherwise, there is an edge on p. Travel along that edge to some node q and mark the edge so you won't use it again. Because there are an even number of edges incident on q, there is an unmarked edge. Leave on the edge and repeat the process until you arrive at p again.

This process will produce a circuit containing p. No edge can be repeated because you are marking the ones you use, and because of finiteness you have to return to p sometime. However, the circuit may not pass over every edge. If it does not, there is an unmarked edge e incident on some node q already in your circuit, because G is connected. Start with that node and that edge and repeat the process, continuing until you return to q. This will give another circuit containing q. Note that the second circuit may hit nodes of the first circuit, but there will always be an unmarked edge to leave on because each node in the first circuit has even degree and an even number of marked edges. You now can put these two circuits together into a big circuit — go around the first circuit starting at p until you get to q, go around the second circuit until you return to q, and then continue around the first circuit until you get back to p. If you still don't hit all the edges, you can repeat this process a second time, and so on until all the edges are used up. The result will be an Eulerian circuit.

This problem was first solved by Leonhard Euler, who was asked whether it was possible to walk around the city of Königsberg (then in Prussia, now in Russia and called Kaliningrad) in such a way that you could traverse each of its seven bridges exactly once. The arrangement of bridges in Euler's time is represented by the left graph in Figure (152.1), page 232 (each edge represents a bridge), which clearly has no Eulerian circuit since in fact none of its nodes has even degree.

157.2.1 Exercise For which integers n does K_n have an Eulerian circuit?

157.2.2 Exercise For which integers m and n does $K_{m,n}$ have Eulerian circuit?

157.3 Definition: Hamiltonian circuit
A **Hamiltonian circuit** in a graph is a circuit which hits each *node* exactly once.

157.3.1 Fact Such a graph must be connected (why?).

157.3.2 Remark Our main purpose in mentioning Hamiltonian circuits is to contrast their theory with that of Eulerian circuits: there is no known simple criterion to determine whether a graph has a Hamiltonian circuit or not. The problem is computationally difficult in general, although for special classes of graphs the question can be answered more easily (Problems 157.4.5 and 157.3.3).

157.3.3 Exercise For which integers m and n does $K_{m,n}$ have a Hamiltonian circuit?

157.4 Exercise set

Exercises 157.4.1 through 157.4.3 concern the graphs in Table 152.1, page 231.

157.4.1 Give the diameter of each graph. (Answer on page 251.)

157.4.2 Which of the graphs has an Eulerian circuit? (Answer on page 252.)

157.4.3 Which of the graphs has a Hamiltonian circuit? (Answer on page 252.)

157.4.4 Give examples of:

a) A graph which has an Eulerian circuit but not a Hamiltonian circuit.
b) A graph which has a Hamiltonian circuit but not an Eulerian circuit.

157.4.5 For which integers n does K_n have a Hamiltonian circuit?

158. Planar graphs

158.1 Definition: Planar
A graph is **embedded in the plane** if it is drawn in such a way that no two edges cross. It is **planar** if can can be embedded in the plane.

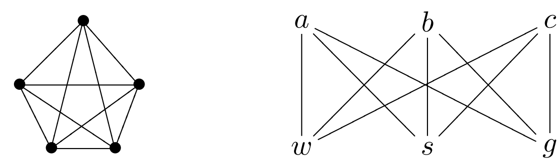
158.1.1 Example Graphs can be used to represent electric circuits. It is desirable in a printed circuit that no two lines (edges of the graph) cross each other. This is exactly the statement that the graph is embedded in the plane.

158.1.2 Example The left graph in Figure (155.1), page 235, can be embedded in the plane as the right graph in the same figure.

158.1.3 Warning *The fact that a graph is drawn with edges crossing does not mean it is not planar.* For example, K_4 is planar, in spite of the way it is drawn in Figure (153.1), page 233.

158.1.4 Exercise Which graphs on page 231, are planar? (Answer on page 252.)

158.1.5 Example Not all graphs can be embedded in the plane. For example, the complete graph on 5 vertices (left graph below) cannot be embedded in the plane. Another such graph is the **utility graph**, the right graph below (which is the complete bipartite graph $K_{3,3}$). It arises if you have three houses a, b and c that must each be connected to the water, sewer and gas plants (w, s and g). If it is drawn in the plane, edges must cross.



K_5 $K_{3,3}$

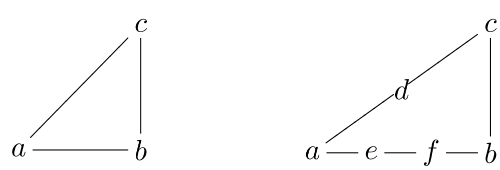
(158.1)

There is an easy-to-use criterion to determine whether a graph is planar. It requires a new concept:

158.2 Definition: subdivision

A **subdivision** of a graph is obtained by repeatedly applying the following process zero or more times: take an edge e connecting two nodes x and y and replace it by a new node z and two edges e' and e'' with e' connecting x and z and e'' connecting y and z.

158.2.1 Example The graph H below is a subdivision of G; it is obtained by subdividing three times. Note that a graph is always a subdivision of itself.



(G) (H)

(158.2)

158.3 Theorem: Kuratowski's Theorem

A graph is not planar if and only if it contains as a subgraph either a subdivision of K_5 or a subdivision of the utility graph

.

158.3.1 Remark This theorem has a fairly technical proof that will not be given here. Note that it turns a property that it would appear difficult to verify into one that is fairly easy to verify.

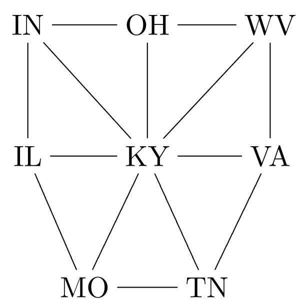
159. Graph coloring

Some very difficult questions arise concerning labeling the node of a simple graph.

> **159.1 Definition: coloring**
> A **coloring** of a simple graph G is a labeling $L : G_0 \to S$ (where S is some finite set) with the property that if nodes p and q are adjacent, then $L(p) \neq L(q)$. In this context the elements of S are called **color**s.

159.1.1 Remark This terminology arises from the problem of coloring a map of countries in such a way that countries with a common border are colored with different colors. In the (very large) literature on coloring problems, two states or countries that have only a point on their borders in common, such as Arizona and Colorado in the U.S.A., are regarded as not bordering each other. The common border must have a nonzero length.

159.1.2 Example The state of Kentucky in the U.S.A., and the seven states bordering it, require four colors to color them in such a way that adjoining states do not have the same color. This is turned into a problem of graph theory by drawing a graph with one node for each state and an edge between two nodes if the corresponding states border each other:



(159.1)

> **159.2 Definition: chromatic number**
> The *smallest* number of colors needed to color a simple graph G is called the **chromatic number** of G, denoted $\chi(G)$.

Warning: Note that we have already used χ for the characteristic function of a subset of a set.

159.2.1 Example The chromatic number of the graph in Figure (159.1) is 4, and the chromatic number of the right graph in Figure (152.1) is 3.

159.2.2 Exercise Show that a graph with at least one edge is bipartite if and only if its chromatic number is 2.

159.2.3 Exercise Show that a graph has chromatic number 2 if and only if it has no cycles of odd length.

159.2.4 Remark It is in general a nontrivial question to determine the chromatic number of a graph. However, some things can be said.

a) The complete graph on n nodes has chromatic number n, since every node is adjacent to every other one.
b) A bipartite graph has chromatic number 2 (if it has any edges): since none of the nodes in one of the moieties are adjacent to each other, they can all be colored the same color, and the nodes in the other moiety can be colored another color.
c) It is known that any *planar* graph has chromatic number ≤ 4. This fact is called the **Four Color Theorem** and is difficult to prove.

159.2.5 Example As an indication of the problems involved in proving the Four Color Theorem, observe that the graph of states in Figure (159.1) has chromatic number 4, although it does not contain the complete graph K_4 as a subgraph. In other words, although there is no four-element subset of the states involved in Figure (159.1) that all border each other (thus turning into a copy of K_4 in (159.1)), it nevertheless takes four colors to color the whole graph. It follows that you can't use Kuratowski's Theorem to prove the Four Color Theorem: the fact that no planar graph contains K_5 as a subgraph does not rule out the possibility that a planar graph needs five colors to color it.

159.2.6 Exercise Give an example of a graph with chromatic number 3 that does not contain a subgraph isomorphic to K_3.

159.2.7 Exercise Find a place in the world with four political subdivisions that all border each other. (There are no four states in the U.S.A. like this, although you will observe that North Carolina, South Carolina, Georgia and the Atlantic Ocean all "border" each other.)

159.2.8 Exercise A **Kempe graph** is a graph with $n+1$ nodes, consisting of n nodes in a cycle and another node connected to each node in the cycle, and no other edges. Figure (159.1), page 241, is a Kempe graph.

a) Show that a Kempe graph is planar.
b) Find the chromatic number of a Kempe graph. (It will depend on n.)

159.2.9 Garbage routes The effort to prove the Four Color Theorem resulted in the discovery of fast coloring algorithms and of a lot of detailed information about graph coloring. This has other applications besides coloring maps. For example, consider the following problem: A city is divided into a number of garbage pickup routes. Some of the routes overlap, because businesses must be picked up more often than residences and therefore are assigned to two or more routes. What is the best way to distribute the routes among the five working days of the week, with each route traveled once a week?

If each route is regarded as a node, with two routes adjacent if they overlap, the result is a graph. A scheduling of the routes that avoids scheduling overlapping routes on the same day is a five-coloring of this graph. An efficient way of coloring the graph would be a start towards finding a good schedule. Note that this problem has nothing to do with planarity or the Four Color Theorem.

Answers to Selected Exercises

3.1.5 Yes, because $-(-3) = 3$ and $3 > 0$, so by Definition 2.2, $-(-3)$ is positive.

4.1.2 Yes, because $52 = 4 \cdot 13$.

4.1.10 -2, -1, 1, 2.

5.5.1 $333 = 9 \times 37$ and 9 is an integer, so $37 \mid 333$ by Definition 4.1.

5.5.2 Suppose $0 \le k < n$ and suppose k is divisible by n. By Definition 4.1, there is an integer q for which $k = qn$. Since k and n are nonnegative, so is q. Since $k = qn < n$, dividing through the inequality by n (which is positive) gives $q < 1$. Since q is nonnegative, it must be 0. Since $k = qn$, $k = 0$ as well.

6.1.5 $91 = 7 \times 13$; $98 = 2 \times 7^2$; $108 = 2^2 \times 3^3$; $111 = 3 \times 37$; 211 is prime

7.5.1 No. For example,

$$\frac{1}{4} + \frac{1}{4} = \frac{2}{4} \text{ and } \frac{2}{3}\frac{3}{4} = \frac{2}{4}$$

9.2.4 Only the pair in (c) are equal.

10.1.2 $5.\overline{1} = 46/9$; $4.\overline{36} = 48/11$; $4.1\overline{36} = 91/22$.

12.2.6 $x^2 - \frac{6}{x} + 4x > 2x$.

12.4.1 $m = 2$ makes it true and $m = 8$ makes it false.

12.4.2 Any m makes it true. No value of m makes it false.

12.5.2 $Q(-1)$ is $1 < 4$ and $Q(x-1)$ is $(x-1)^2 < 4$.

12.5.3 a. $2 < 5$. b. $3 < 4$. c. $x^2 < x + y + 1$. d. $x(x+y) < x + y + z + 1$.

13.2.7 (a) and (b) are true, and the others are false. It is wrong to say that (c) is "sometimes true" or "usually true". The statement that $3 \cdot 0 > 0$ is false, so the statement $(\forall x{:}\mathrm{N})(3x > x)$ is simply false, with no qualification.

14.2.3

	2	6	7
a	T	T	T
b	T	T	F
c	T	T	T

14.2.4 a) True: $n = 5$. False: Any n other than 5.
b) True: $n = 8$, for example, or $n = 0$. False: $n = 4, 5, 6, 7$ are the only ones.
c) True: Impossible. False: any n.
d) True: Any n.

14.2.5 Only (d).

17.1.4 3.

18.1.5 a) 2. b) 3. c) 2. d) 0. (For (b), see Remark 8.1.3.)

18.1.16 You must show that $P(a)$ is false.

19.2.5 (a) and (c) are true and (b) is false.

19.2.6 $-13, -7, -5, -4, -3, -2, 0, 1, 2, 3, 5, 11$. b) $1, 4, 9, 16, 36, 144$. c) Same as (b).

20.1.3 (a) and (c) are the same, and so are (b) and (d).

22.1.6 Only (d) is the empty set.

23.1.5 d is the empty set and b, c and g are singletons.

23.1.6 (a) D_1 is the only singleton. (n) 1 is the only integer which is an element of D_n for every positive integer n.

25.1.4 Item (a) is true for all integers m but (b) and (c) are false. For example, (b) is false for $m = 6$ (then the hypothesis is true and the conclusion is false, and that is the line in the truth table that makes the implication false), and (c) is false for $m = -2$.

26.1.5 a) True: $n = 6$, for example (this is vacuously true). False: $n = 8$.
b) True: any n. False: not possible.
c) True: $n = 10$. False: $n = 8$.
d) True: any n. False: not possible.
e) True: any n (always vacuously true). False: not possible.
f) True: Any n except 1. False: $n = 1$.

27.2.1 (a), (c), (d) and (e) say the same thing, and (b) and (f) say the same thing.

30.4.5 The contrapositive is "If n is not prime, then 3 does not divide n", which is not true for some integers n. The converse is "If n is prime, then $3 \mid n$", which is also false for some n.

31.4.2

	$\in$	$\subseteq$	$=$
a)	N	Y	N
b)	Y	N	N
c)	N	Y	Y
d)	N	N	N
e)	N	N	N

31.5.3 You must show that there is an element $x \in S$ that is not an element of T. This is because of Definition 31.1, which defines $A \subseteq B$ to mean the implication $x \in A \Rightarrow x \in B$, and the only way that implication can be false is for the hypothesis to be true and the conclusion false.

32.1.6 a: 4. b: 0. c: 1. d: 2.

32.1.7 $\{\emptyset, \{5\}, \{6\}, \{7\}, \{5,6\}, \{6,7\}, \{5,7\}, \{5,6,7\}\}$

32.1.8

	a	b	c	d	e	f	g
a	Y	Y	Y	Y	N	N	N
b	Y	Y	Y	N	N	N	N
c	N	Y	N	N	N	N	N
d	N	N	N	N	Y	N	N
e	N	N	N	N	N	N	N
f	N	N	N	N	Y	N	N
g	N	N	N	N	Y	N	Y

33.2.2 $\{1,2,3\} \cup \{2,3,4,5\} = \{1,2,3,4,5\}$ and $\{1,2,3\} \cap \{2,3,4,5\} = \{2,3\}$.

33.2.3 $\mathrm{N} \cup \mathrm{Z} = \mathrm{Z}$ and $\mathrm{N} \cap \mathrm{Z} = \mathrm{N}$.

33.2.7 By Definition 31.1, we must show that if $x \in A \cap B$, then $x \in A \cup B$. By Definition 33.2 (of intersection), $x \in A \cap B$ implies that $x \in A$ and $x \in B$. By Definition 33.1 (of union), if $x \in A$, then $x \in A \cup B$.

33.3.1 There are of course an infinite number of answers. Some correct answers are: The set of all negative integers, the set of all negative even integers, $\{-1,-2,-3\}$, $\{-42\}$, and the empty set (which is disjoint from *every* set).

34.2.2 $\mathrm{Z} - \mathrm{N}$ is the set of all negative integers. $\mathrm{N} - \mathrm{Z} = \emptyset$.

34.2.5 (a) 1,2,3,4,5; (b) 2,3; (c) 1,2,3,4,5,7,8; (d) none; (e) 1; (f) 2,3,4,5; (g) 1,2,3; (h) 1,2,3,4,5; (i) 2,3,4,5.

34.2.6 1) 1 and 2. 2) 1. 3) 1, 3 and 5. 4) 5. 5) 6 and 7. 6) None. 7) 6 and 7.

35.1.3 The pairs in (a) are different; the pairs in (b) and (c) are equal.

36.1.2 $\mathbf{m} \cap \mathbf{n}$ is $\mathbf{k}$, where k is the minimum of m and n, and $\mathbf{m} \cup \mathbf{n}$ is l, where l is the maximum of m and n.

36.3.1 None of them are equal.

36.4.1 1. a) $3,4$. b) $2, \langle 1,5 \rangle$. c) $2, \langle 5, \langle 2,1 \rangle \rangle$. d) $2,9$. e) $2, \{1,2\}$. f) $4, \mathrm{Z}$.

37.1.2 $\langle 1,a \rangle, \langle 1,b \rangle, \langle 2,a \rangle, \langle 2,b \rangle$.

37.6.1 This is false for any *nonempty* set A because the elements of $A \times A$ are pairs of elements of A, and an ordered pair is distinct from its coordinates (see 35.1). (The last statement implies that in fact for nonempty A, $A \times A$ and A have no elements in common.) The statement $A \times A = A$ is true if $A = \emptyset$.

37.7.1 "For all sets A and B and all nonempty sets $C, \ldots$"

37.9.1

(a) Λ

(b) $1,2$

(c) $\langle 1,1 \rangle, \langle 1,2 \rangle, \langle 2,1 \rangle, \langle 2,2 \rangle$

(d) $\langle 1,1,1 \rangle, \langle 1,1,2 \rangle, \langle 1,2,1 \rangle, \langle 1,2,2 \rangle,$
$\langle 2,1,1 \rangle, \langle 2,1,2 \rangle, \langle 2,2,1 \rangle, \langle 2,2,2 \rangle$

(e) $\langle 1,3 \rangle, \langle 1,4 \rangle, \langle 1,5 \rangle, \langle 2,3 \rangle, \langle 2,4 \rangle, \langle 2,5 \rangle$

(f) $\langle 3,1 \rangle, \langle 3,2 \rangle, \langle 4,1 \rangle, \langle 4,2 \rangle, \langle 5,1 \rangle, \langle 5,2 \rangle$

(g) $\langle 1,1,3 \rangle, \langle 1,1,4 \rangle, \langle 1,1,5 \rangle, \langle 1,2,3 \rangle, \langle 1,2,4 \rangle, \langle 1,2,5 \rangle,$
$\langle 2,1,3 \rangle, \langle 2,1,4 \rangle, \langle 2,1,5 \rangle, \langle 2,2,3 \rangle, \langle 2,2,4 \rangle, \langle 2,2,5 \rangle$

(h) $\langle 1, \langle 1,3 \rangle \rangle, \langle 1, \langle 1,4 \rangle \rangle, \langle 1, \langle 1,5 \rangle \rangle, \langle 1, \langle 2,3 \rangle \rangle,$
$\langle 1, \langle 2,4 \rangle \rangle, \langle 1, \langle 2,5 \rangle \rangle, \langle 2, \langle 1,3 \rangle \rangle, \langle 2, \langle 1,4 \rangle \rangle,$
$\langle 2, \langle 1,5 \rangle \rangle, \langle 2, \langle 2,3 \rangle \rangle, \langle 2, \langle 2,4 \rangle \rangle, \langle 2, \langle 2,5 \rangle \rangle$

(i) $\langle 1,3 \rangle, \langle 1,4 \rangle, \langle 1,5 \rangle, \langle 2,3 \rangle, \langle 2,4 \rangle, \langle 2,5 \rangle, 1, 2$

(j) $\emptyset$

37.9.2

	1	2	3	4	5	6	7
1	N	N	Y	N	N	N	Y
2	Y	N	Y	Y	N	N	Y
3	N	N	N	N	N	N	N
4	N	N	N	N	Y	N	N
5	N	N	N	N	N	Y	N
6	N	N	N	N	Y	N	N
7	N	Y	N	N	N	N	N

38.2.1 $\Big\{\langle x,n \rangle \mid x > n\Big\} \subseteq \mathrm{R} \times \mathrm{N}$.

38.2.2 $\Big\{\langle x,y \rangle \mid x \in \mathrm{R}, y = 1\Big\} = \Big\{\langle x,1 \rangle \mid x \in \mathrm{R}\Big\} \subseteq \mathrm{R} \times \mathrm{R}$

38.2.3 $\{1\} \subseteq \mathrm{R}$

38.2.4 $\Big\{\langle x,y,z,w\rangle \mid x+y=z\Big\} \subseteq \mathrm{R}\times\mathrm{R}\times\mathrm{R}\times\mathrm{R}$.

39.3.7 $F(1) = \{\{1\},\{1,2\},\{1,3\},\{1,2,3\}\}$ and $F(2) = \{\{2\},\{1,2\},\{2,3\},\{1,2,3\}\}$.

40.2.6 (a) and (d) only.

41.1.8

	$F(2)$	$F(4)$
a)	2	4
b)	42	42
c)	2	4

41.1.9 a) $\langle 2,2\rangle, \langle 3,3\rangle$
b) $\langle 2,2\rangle, \langle 3,3\rangle$
c) $\langle 2,2\rangle, \langle 3,3\rangle$
d) $\langle 1,3\rangle, \langle 2,3\rangle, \langle 3,3\rangle$
e) $\langle\langle 1,2\rangle,1\rangle\rangle, \langle\langle 1,3\rangle,1\rangle, \langle\langle 2,2\rangle,2\rangle\rangle,$
$\langle\langle 2,3\rangle,2\rangle, \langle\langle 3,2\rangle,3\rangle\rangle, \langle\langle 3,3\rangle,3\rangle$

42.2.3 a) $\lambda x.x^3$; $x \mapsto x^3 : \mathrm{R}\to\mathrm{R}$ b) $\lambda\langle a,b\rangle.a$; $\langle a,b\rangle \mapsto a : A\times B \to A$. c) $\lambda\langle a,b\rangle.a+b$; $\langle a,b\rangle \mapsto a+b : \mathrm{R}\times\mathrm{R}\to\mathrm{R}$

43.1.4 a) $\langle 1,\mathrm{FALSE}\rangle, \langle 2,\mathrm{TRUE}\rangle, \langle 3,\mathrm{TRUE}\rangle$
b) $\langle 1,\mathrm{TRUE}\rangle, \langle 2,\mathrm{FALSE}\rangle, \langle 3,\mathrm{TRUE}\rangle$
c) $\langle\langle 2,2\rangle,4\rangle, \langle\langle 2,3\rangle,5\rangle, \langle\langle 3,2\rangle,5\rangle, \langle\langle 3,3\rangle,6\rangle$

44.1.5 a) (1) only. b) (2) only. c) (3) only. d) (1) only. Note that (4) is not an answer to (d) because the function is given as having codomain R. Of course there is a function $x \mapsto x^2 : \mathrm{R}\to\mathrm{R}^+$ with the same graph but it is technically a different function. For many purposes, this is merely a technicality, but there are places in mathematics where the distinction is quite important.

46.4.3 35 22 $+$6 5 $+\ast$.

48.1.5 We must show, for all subsets A, B and C of S, that $A\cup(B\cup C) = (A\cup B)\cup C$. We will do this using Method 21.2.1, page 32. Suppose that $x \in A\cup(B\cup C)$. Then by (33.1), page 47, either $x \in A$ or $x \in B\cup C$. If $x\in A$, then $x \in A\cup B$, so $x \in (A\cup B)\cup C$ by using the definition of union twice. If $x \in B\cup C$, then either $x\in B$ or $x \in C$. If $x\in B$, then $x\in A\cup B$, so $x\in(A\cup B)\cup C$. If $x\in C$, then again by definition of union, $x\in(A\cup B)\cup C$. So we have verified that in every case,

$$x \in A\cup(B\cup C) \Rightarrow x\in(A\cup B)\cup C$$

so that by Definition 31.1, page 43, $A\cup(B\cup C)\subseteq(A\cup B)\cup C$. A similarly tedious argument shows that $(A\cup B)\cup C \subseteq A\cup(B\cup C)$. Therefore by Method 21.2.1, $A\cup(B\cup C) = (A\cup B)\cup C$.

50.1.4 (1) is associative, not commutative, and does not have an identity. (2) is not associative (because $(a\,\Delta\, b)\,\Delta\, c = a$ but $a\,\Delta\,(b\,\Delta\, c) = b$), is commutative, and does not have an identity.

50.1.7 The empty set, since for any subset A of S, $A\cup\emptyset = \emptyset\cup A = A$.

51.1.5
a) $\langle 1,3\rangle, \langle 1,5\rangle, \langle 2,1\rangle, \langle 2,3\rangle, \langle 2,5\rangle, \langle 3,1\rangle, \langle 3,5\rangle$
b) $\langle 2,2\rangle, \langle 2,4\rangle, \langle 2,6\rangle, \langle 2,8\rangle, \langle 2,10\rangle, \langle 3,3\rangle,$
$\langle 3,6\rangle, \langle 3,9\rangle, \langle 5,5\rangle, \langle 5,10\rangle, \langle 7,7\rangle$
c) $\langle 1,1\rangle, \langle 1,2\rangle, \langle 1,3\rangle, \langle 2,2\rangle, \langle 3,3\rangle$

52.1.3
a) $\langle 1,2\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \langle 2,3\rangle, \langle 2,4\rangle, \langle 3,4\rangle$
b) $\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle, \langle 4,4\rangle$. (This is Δ_A.)
c) $\langle 1,3\rangle, \langle 2,3\rangle, \langle 3,3\rangle, \langle 4,3\rangle$.
d) $\langle 1,1\rangle, \langle 3,3\rangle, \langle 1,3\rangle, \langle 3,1\rangle$.

53.1.2 (a), (c) and (e) are functional relations.

53.2.3 $1\mapsto\{3,5\}$, $2\mapsto\{1,3,5\}$, $3\mapsto\{1,5\}$.

53.3.3 $\{\langle 1,3\rangle, \langle 1,4\rangle, \langle 2,1\rangle, \langle 2,3\rangle, \langle 2,4\rangle, \langle -666,0\rangle\}$

55.1.9 (b) is not reflexive, the others are reflexive.

56.1.4 (b) and (c) are symmetric, (a) and (d) are not.

57.1.9 (a), (b) and (c) are antisymmetric; (d) is not. Note that (c) is vacuously antisymmetric.

59.1.3

	ref	sym	ant	trs	irr
a	Y	N	Y	Y	N
b	N	N	Y	N	Y
c	N	Y	N	N	Y
d	N	N	Y	Y	N
e	Y	Y	Y	Y	N
f	N	N	N	N	Y

59.1.4

	ref	sym	ant	trs	irr	Note
a	N	Y	N	N	Y	
b	Y	Y	N	Y	N	
c	N	Y	N	N	N	
d	N	N	N	N	N	
e	Y	N	Y	Y	N	
f	Y	Y	N	Y	N	
g	N	N	N	N	N	

concerning (d): $2\le 3^2$, $3\le 2^2$, $8\le 3^2$.

60.1.2 a: $q=0$, $r=2$. b: $q=0$, $r=0$. c: $q=2$, $r=0$. d: $q=3$, $r=1$.

60.1.4 Suppose $a = qm + r$ and $b = q'm + r$. Then $a - b = qm - q'm = (q - q')m$ so it is divisible by m.

60.2.3 Since $m \operatorname{div} n = a$, $m = an + r$ for some integer r such that $0 \le r < n$. We are given that $m = an + n + b + 2$, so $r = n + b + 2$. Hence $n + b + 2 < n$, so that $b + 2 < 0$, so $b < 0$.

60.2.4 Since $n \mid s$, $s = qn$ for some integer q. q is not less than 0 since n and s are nonnegative. It is not greater than 0 since then $qn \ge n > s$ but we are given $s = qn$. So q must be 0, so that s is 0 too.

60.5.2 By Definition 60.1, we must show that $37 = 7 \cdot 5 + 2$ and that $0 \le 2 < 5$. Both are simple arithmetic. It follows from Theorem 60.2 that the quotient is 7 and the remainder 2 as claimed. (Yes, you knew this in fourth grade. The point here is that it follows from the definitions and theorems we have.)

60.5.4 4, because $m = 36q + 40 = 36(q + 1) + 4$ and $0 \le 4 < 36$.

61.1.3 $n \le r < n + 1 \vdash n = \operatorname{floor}(r)$, where n is of type integer.

61.2.3 a: $\operatorname{trunc}(7/5) = \operatorname{floor}(7/5) = 1$.
b: $\operatorname{trunc}(-7/5) = -1$; $\operatorname{floor}(-7/5) = -2$.
c: $\operatorname{trunc}(-7) = \operatorname{floor}(-7) = -7$.
d: $\operatorname{trunc}(-6.7) = -6$; $\operatorname{floor}(-6.7) = -7$.

62.2.2 $30 = 2^1 \times 3^1 \times 5^1$, $35 = 5^1 \times 7^1$, $36 = 2^2 \times 3^2$, $37 = 37^1$, $38 = 2^1 \times 19^1$.

62.3.2

prime	98	99	100	111	1332	1369
3	0	2	0	1	2	0
7	2	0	0	0	0	0
37	0	0	0	1	1	2

62.5.1

$$\begin{aligned}
90 &= 2^1 \times 3^2 \times 5^1\\
91 &= 7^1 \times 13^1\\
92 &= 2^2 \times 23^1\\
93 &= 3^1 \times 31^1\\
94 &= 2^1 \times 47^1\\
95 &= 5^1 \times 19^1\\
96 &= 2^5 \times 3^1\\
97 &= 97^1\\
98 &= 2^1 \times 7^2\\
99 &= 3^2 \times 11^1
\end{aligned}$$

63.2.2

PAIR	GCD	LCM
12, 12	12	12
12, 13	1	156
12, 14	2	84
12, 24	12	24

63.2.4 False: for example $\operatorname{GCD}(4,2) = \operatorname{GCD}(2,2) = 2$. If you said "TRUE" you may have fallen into the trap of saying "the GCD of m and n is the product of the primes that m and n have in common," which is incorrect.

63.2.5 $\langle 1,1\rangle, \langle 1,2\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \langle 2,1\rangle, \langle 2,3\rangle, \langle 3,1\rangle, \langle 3,2\rangle, \langle 3,4\rangle, \langle 4,1\rangle, \langle 4,3\rangle$

63.3.2 If d divides both n and $n + 1$ it must divide their difference, which is 1. Hence the largest integer dividing n and $n + 1$ is 1.

64.2.2 Suppose $e \mid m$ and $e \mid n$. Let p be any prime. Then $\mathrm{e}_p(e)$ must be less than or equal to $\mathrm{e}_p(m)$ and also less than or equal to $\mathrm{e}_p(n)$. Thus it is less than or equal to $\mathrm{e}_p(d)$, which by Theorem 64.1 is the minimum of $\mathrm{e}_p(m)$ and $\mathrm{e}_p(n)$. This is true for every prime p, so in the prime factorization of e, every prime occurs no more often than it does in d, so by Theorem 62.4, $e \mid d$.

64.2.4 Let p be any prime. By Theorem 64.1, $\mathrm{e}_p(d) = \min(\mathrm{e}_p(m), \mathrm{e}_p(n))$. Observe that $\mathrm{e}_p(m/d) = \mathrm{e}_p(m) - \mathrm{e}_p(d)$ and $\mathrm{e}_p(n/d) = \mathrm{e}_p(n) - \mathrm{e}_p(d)$. We know that $\mathrm{e}_p(d) = \min(\mathrm{e}_p(m), \mathrm{e}_p(n))$, so one of the numbers $\mathrm{e}_p(m) - \mathrm{e}_p(d)$ and $\mathrm{e}_p(n) - \mathrm{e}_p(d)$ is zero. That means p does not divide both m and n. Since p was assumed to be *any* prime, this means *no* prime divides both m and n. Therefore, $\operatorname{GCD}(m/d, n/d) = 1$, as required.

66.6.3 By Definition 66.4,

$$n = d_m b^m + \cdots + d_1 b^1 + d_0 b^0$$

so

$$bn = d_m b^{m+1} + \cdots + d_1 b^2 + d_0 b^1 + 0 b^0$$

which means that bn is represented by $d_m d_{m-1} \cdots d_1 0$.

67.2.3 a) 1100000. b) 11010010. c) 110001111. d) 1010111100.

67.2.4 a) 1525. b) b00. c) 10c9a.

68.4.1

DEC	OCT	HEX	BASE 36	BINARY
100	144	64	2s	1100100
111	157	6f	33	1101111
127	177	7f	3j	1111111
128	200	80	3k	10000000

69.3.1 $(x \geq 10) \vee (x \leq 12)$. Of course, this is true of all real numbers.

69.3.2 $(x \geq 10) \vee (x \geq 12)$. Of course, this is the same as saying $x \geq 10$.

71.2.5 Here are the truth tables:

P	Q	$P \vee Q$	$\neg P$	$\neg Q$	$\neg P \wedge \neg Q$	$\neg(\neg P \wedge \neg Q)$
T	T	T	F	F	F	T
T	F	T	F	T	F	T
F	T	T	T	F	F	T
F	F	F	T	T	T	F

The third and seventh columns are the same.

71.2.9

P	Q	$\neg P$	$\neg P \vee Q$	$P \Rightarrow Q$	$\neg Q$	$P \wedge \neg Q$	$\neg(P \wedge \neg Q)$
T	T	F	T	T	F	F	T
T	F	F	F	F	T	T	F
F	T	T	T	T	F	F	T
F	F	T	T	T	T	F	T

The fourth, fifth and eighth columns are the same.

74.2.1 Valid.

74.2.2 Valid.

74.2.3 Invalid.

74.2.7 Let P be $3 > 5$ and Q be $4 > 6$. Then $P \Rightarrow Q$ is true because both hypothesis and conclusion are false; on the other hand, Q is false. Since the hypothesis of $(P \Rightarrow Q) \Rightarrow Q$ is therefore true and the conclusion false, the statement is false.

75.3.4 a: True. Witness: 2. b: False. Counterexample: 9. c: True. Witness: 2. d: False. Counterexample: 3.

75.3.5 (a) True. (b) True. (c) True. (d) False; a counterexample is given by taking P to be $x > 7$ and Q to be $x < 7$.

76.1.4 There are no counterexamples to $(\forall y)P(14, y)$ since it is the statement

$$(\forall y)\,((14 = y) \vee (14 > 5))$$

which is true because "$14 > 5$" is true.

The number 3 and any number greater than 5 is a witness to $(\exists x)P(x,3)$.

77.2.1 (a) means that for every real number the statement $(\exists y)(x > y)$ is true. A witness for that statement is $x - 1$, so the statement is true. (b) means that there is a real number greater than any real number, which is false. (c) is true. Witness: Let $x = y = 3$. Then the statement becomes $((3 > 3) \Rightarrow (3 = 3))$, which is (vacuously) true.

82.2.1 valid: direct method.

82.2.2 invalid: fallacy of affirming the hypothesis.

82.2.3 invalid: fallacy of denying the consequence.

82.2.4 valid with false hypothesis.

82.2.5 invalid: fallacy of denying the consequence.

85.1.3 This follows from Rule (85.1), page 124, going from top to bottom. To use it, we must verify the two hypotheses of the rule with $r = m - qn$. The first is $qn + r = qn + (m - qn) = m$, as required. The other, $0 \leq r < n$, is immediate. Therefore the conclusion, part of which states that $q = m \operatorname{div} n$, must be true.

86.2.4 This is a proof by contradiction. Suppose $p > 2$ and p is not odd. Then p is even, so it is divisible by 2. Therefore p is divisible by a number other that p and 1 (namely 2, which is not p because $p > 2$). This contradiction to the definition of prime (Definition 6.1, page 10) shows that the claim is correct.

88.3.1

a	b	c	
2	12	16	Impossible, since $GCD(12,16) = 4$.
4	12	16	$4 = 16 - 12$.
2	26	30	$2 = 7 \times 26 - 6 \times 30$.
4	26	30	$4 = 14 \times 26 - 12 \times 30$
-2	26	30	$-2 = 6 \times 30 - 7 \times 26$.
1	51	100	$1 = 25 \times 100 - 49 \times 51$.

88.3.3 If m and n are relatively prime, then by Theorem 87.2 there are integers a'' and b'' for which $a''m + b''n = 1$. Then $(a + a'')m + (b + b'')n = am + bn + a''m + b''n = e + 1$. Note: If you reasoned as follows: "Because a and b are relatively prime and $am + bn = e$, it follows that $e = 1$ by Theorem 87.2," then you are guilty of the fallacy of affirming the hypothesis (page 121).

89.1.7 The set of positive integers.

90.1.5 $F(\{2,3\} = \{5\}$ and $F(\{3\})$ is also $\{5\}$.

93.1.4

	inj?	surj?	image
a)	N	Y	B
b)	N	N	$\{2,3\}$
c)	Y	Y	A
d)	Y	N	B
e)	Y	N	B
f)	N	N	$\{3\}$
g)	N	Y	$\{\text{TRUE}, \text{FALSE}\}$
h)	N	Y	A
i)	N	N	$\{4,5,6,7,8\}$
j)	N	Y	$\{\text{TRUE}, \text{FALSE}\}$

93.1.5

	inj?	surj?	image
a)	Y	Y	R
b)	Y	Y	R
c)	N	N	$\{r \in \mathrm{R} \mid r \geq 1\}$
d)	N	N	$\{r \in \mathrm{R} \mid r \leq 2\}$

93.1.7 You must show that there are two different elements a and a' of A for which $F(a) = F(a')$. That is because the definition of injective is the implication

$$a \neq a' \Rightarrow F(a) \neq F(a')$$

and the negation of that implication is the statement

$$a \neq a' \wedge \neg(F(a) \neq F(a'))$$

in other words

$$a \neq a' \wedge F(a) = F(a')$$

96.2.5

domain	R		R^+	
	inj?	surj?	inj?	surj?
a)	N	N	Y	N
b)	Y	Y	Y	N
c)	Y	Y	Y	N

If the answers in the last column puzzle you, remember that the codomain of the restriction of a function is the same as the codomain of the function.

97.1.3 a) Domain: $\{1,2,3,4,5\}$.
Graph: $\Big\{\langle 1,2\rangle, \langle 2,5\rangle, \langle 3,-1\rangle, \langle 4,3\rangle, \langle 5,6\rangle\Big\}$.
b) Domain: $\{1,2,3,4\}$.
Graph: $\Big\{\langle 1,\pi\rangle, \langle 2,5\rangle, \langle 3,\pi-1\rangle,\ \langle 4,\sqrt{2}\rangle\Big\}$.
c) Domain: $\{1,2,3\}$.
Graph: $\Big\{\langle 1,\langle 3,5\rangle\rangle, \langle 2,\langle 8,-7\rangle\rangle, \langle 3,\langle 5,5\rangle\rangle\Big\}$.

97.3.1 a) $\langle 5,5,3,17,-1\rangle$. b) $\langle 2\pi,3\pi,4\pi,5\pi,6\pi\rangle$. c) $\langle 1,4,9,16,25,36\rangle$.

98.2.6

a) $G \circ F : \{1,2,3,4\} \to \{1,3,5,7,9\}$, graph $\Big\{\langle 1,1\rangle, \langle 2,7\rangle, \langle 3,3\rangle, \langle 4,7\rangle\Big\}$.
b) $G \circ F : \mathrm{R} \to \mathrm{R}$, $(G \circ F)(x) = 2x^3$.
c) $G \circ F : \mathrm{R} \to \mathrm{R}$, $(G \circ F)(x) = 8x^3$.
d) $n \mapsto (n/2) : \mathrm{N} \to \mathrm{R}$.
e) $\langle x,y\rangle \mapsto \langle 3,x\rangle : \mathrm{R} \times \mathrm{R} \to \mathrm{R} \times \mathrm{R}$.

99.1.5 $1 \mapsto 1,\ 2 \mapsto 2,\ 3 \mapsto 2$.

100.1.5

(a)
$$\begin{tikzcd} \mathrm{R}^+ \arrow[r, "x \mapsto \sqrt{x}"] \arrow[dr, "\mathrm{id}"'] & \mathrm{R}^+ \arrow[d, "x \mapsto x^2"] \\ & \mathrm{R}^+ \end{tikzcd}$$

(a)
$$\begin{tikzcd} x\ \mathrm{R} \arrow[r, "x \mapsto x^2"] \arrow[dr, "x \mapsto |x|"'] & \mathrm{R} \arrow[d, "x \mapsto \sqrt{x}"] \\ & \mathrm{R} \end{tikzcd}$$

101.2.3 Only (a) and (f) have inverses. For (a) the inverse is $F^{-1} : \{3,4,5,6\} \to \{1,2,3,4\}$ with graph $\Big\{\langle 3,1\rangle, \langle 4,2\rangle, \langle 6,3\rangle, \langle 5,4\rangle\Big\}$. For (f) it is $n \mapsto n-1 : \mathrm{Z} \to \mathrm{Z}$.

101.2.4 All except (c) and (h) have left inverses. (a), (f) and (h) have right inverses.

101.2.5 If L is a left inverse of $G : A \to B$, then for any x in the domain of G, $L = L \circ \mathrm{id}_B = L \circ (G \circ F) = (L \circ G) \circ F = \mathrm{id}_A \circ F = F$.

101.5.3

a) $x \mapsto \sqrt{x}$.
b) $x \mapsto x + 1$.
c) $x \mapsto x/2$.
d) This one is its own inverse.

102.1.3 $\sum_{k=1}^{5} k^2 = 55$ and $\prod_{k=1}^{5} k^2 = 14{,}400$.

103.4.1 Basis: $\sum_{k=1}^{1} \frac{1}{k(k+1)} = \frac{1}{2}$. Induction step:

$$\begin{aligned}
\sum_{k=1}^{n+1} \frac{1}{k(k+1)} &= \frac{1}{(n+1)(n+2)} + \sum_{k=1}^{n} \frac{1}{k(k+1)} \\
&= \frac{1}{(n+1)(n+2)} + \frac{n}{n+1} \\
&= \frac{1+n(n+2)}{(n+1)(n+2)} \\
&= \frac{n^2+2n+1}{(n+1)(n+2)} \\
&= \frac{(n+1)^2}{(n+1)(n+2)} \\
&= \frac{n+1}{n+2}
\end{aligned}$$

as required.

103.4.2 Induction step: If n is even,

$$\begin{aligned}
\sum_{k=1}^{n+1} (-1)^k k &= -(n+1) + \sum_{k=1}^{n} (-1)^k k \\
&= -(n+1) + \frac{n}{2} \\
&= \frac{1}{2}(n - 2n - 2) \\
&= \frac{-n-2}{2} \\
&= \frac{-(n+1+1)}{2}
\end{aligned}$$

as required, and if n is odd,

$$\begin{aligned}
\sum_{k=1}^{n+1} (-1)^k k &= (n+1) + \sum_{k=1}^{n} (-1)^k k \\
&= n+1 - \frac{n+1}{2} \\
&= \frac{n+1}{2}
\end{aligned}$$

again as required.

104.4.1 Suppose d is a positive integer and $d \mid p$ and $d \mid m$. The only divisors of p are 1 and p. If $d = p$, then p does not divide m. So the only possibility is that $d = 1$. Thus 1 is the largest divisor of p and m, so $\mathrm{GCD}(p,m) = 1$.

104.4.2

105.1.3

	1	2	3	4	5
a)	-3	-6	-18	-72	-360
b)	1	5	14	30	55
c)	2	1	0	2	1
d)	3	4	7	11	18
e)	0	1	2	9	44

105.2.1 $1! = 1$, and $(n+1)! = (n+1)n!$ which by the induction hypothesis is

$$(n+1)\Pi_{k=1}^{n} k = \Pi_{k=1}^{n+1} k$$

as required.

107.3.1 For $n = 1$, this is $1^2 - 0 = (-1)^2$. For the induction step, suppose $f_n^2 - f_{n-1}f_{n+1} = (-1)^{n+1}$. Then

$$\begin{aligned}
f_{n+1}^2 - f_n f_{n+2} &= f_{n+1}^2 - f_n(f_n + f_{n+1}) \\
&= f_{n+1}^2 - f_n^2 - f_n f_{n+1} \\
&= f_{n+1}^2 - f_n f_{n+1} - f_{n-1} f_{n+1} \\
&\qquad - f_n^2 + f_{n-1} f_{n+1}
\end{aligned}$$

The first three terms are $f_{n+1}(f_{n+1} - f_n - f_{n-1})$, which is 0 by definition of the Fibonacci recurrence. By the induction hypothesis, the last two terms are $(-1)(-1)^{n+1} = (-1)^{n+2}$ as required.

109.8.2 The last entry of $\langle a \rangle$ is a, and the last entry of $\mathrm{cons}(a, L)$ is the last entry of L.

110.4.2 (a) '0111010'. (b) '011'. (c) '011'. (d) Λ. (e) '011011011'. (f) '011011011'.

110.4.3

EV.1 The empty string Λ is a string in E.

EV.2 If w is a string in E then 'awa', 'awb', 'bwa' and 'bwb' are all strings in E.

EV.3 Every string in E is given by one of the preceding rules.

112.4.2

a) $\bigcup \mathcal{F} = \{1,2,3,4,5\}$, $\bigcap F = \emptyset$.

b) $\bigcup \mathcal{F} = (-3 \mathinner{..} 3)$, $\bigcap F = (-1 \mathinner{..} 1)$.

c) $\bigcup \mathcal{F} = (-1 \mathinner{..} 3) - \{1,2\}$, $\bigcap F = \emptyset$.

113.1.2 The set of positive divisors of 8 is $\{1,2,4,8\}$. Let the bijection β required by Definition 113.1 be defined by: $\beta(1) = 1$, $\beta(2) = 2$, $\beta(3) = 4$, and $\beta(4) = 8$.

113.5.1 $x \mapsto x+1 : \mathrm{N} \to \mathrm{N}^+$ is a bijection.

114.2.3 $9 \cdot 10 \cdot 10 \cdot 10 = 9000$.

114.2.4 $9 \cdot 10 \cdot 10 \cdot 10 \cdot 5 = 45{,}000$.

114.3.1 $2^n - 1$.

114.3.2 $F(n) = 3^n$.

114.3.3 $G(0) = 0$, $G(1) = 1$, and for $n \geq 2$, $G(n) = 3^{n-2}$.

115.2.3 (a) $2^n - 1$. (b) n. (c) $2^{(2^n)}$.

116.2.3 Let Z the set of zinc pennies, B the set of pennies minted before 1932, and A the set of pennies that are neither zinc nor minted before 1932. Let P be your whole collection. Then

$$\begin{aligned} |P| &= |Z| + |B| + |A| - |Z \cap B| - |Z \cap A| \\ &\quad - |A \cap B| + |Z \cap A \cap B| \end{aligned}$$

Since

$$|Z \cap A| = |A \cap B| = |A \cap B \cap Z| = 0$$

we have

$$|P| = 3 + 8 + |A| - |Z \cap B|$$

so you need to know the number of pennies that are neither zinc nor minted before 1932 and the number of zinc pennies minted before 1932. (In fact, all zinc pennies were minted in 1943.)

117.1.13 All are partitions except (b) and (d). Even though every element of S is an element of exactly one set in (d), (d) is not a partition because it contains the empty set as an element.

117.3.1 Let $A = \{1, 3, 5\}$ and let $\Pi = \{A, \mathrm{Z} - A\}$.

120.3.1 Every block of S/F must be a singleton.

120.4.1 Let $F(1) = F(2) = F(\pi) = 42$ and $F(x) = 41$ for all other real numbers x.

121.2.1

a) $\beta_F(\{1,3,5\}) = 4$; $\beta_F(\{4\}) = 6$; $\beta_F(\{2\}) = 0$.
b) $\beta_F(A) = 3$.
c) $\beta_F(\{n\}) = n$ for $n \in A$.
d) $\beta_F(\{n\}) = n^2$ for $n \in A$. (Observe that for (c) and (d), A/F is the same set.)
e) $\beta_F(\{1,2\}) = -5$; $\beta_F(\{3\}) = 1$; $\beta_F(\{4\}) = 21$; $\beta_F(\{5\}) = 55$.

122.3.1 26^{10}.

126.1.3

a) $\Big\{\langle 2,a\rangle, \langle 2,c\rangle, \langle 3,a\rangle, \langle 3,c\rangle, \langle 3,d\rangle\Big\}$
b) $\emptyset$
c) $\Big\{\langle 2,c\rangle, \langle 3,c\rangle, \langle 3,d\rangle, \langle 3,e\rangle, \langle 4,c\rangle, \langle 4,d\rangle, \langle 4,e\rangle\Big\}$.

126.3.1

	$1\,\mathrm{R}^2\,3$?	$1\,\mathrm{R}^3\,3$?	$3\,\mathrm{R}^2\,1$?
a	Y	N	N
b	N	N	N
c	Y	Y	N
d	Y	Y	Y

127.2.1 "$\leq$".

127.3.1 $\mathrm{R} \times \mathrm{R} - \Delta$: any two *different* real numbers are related.

127.3.4 By Definition 127.1, we must show that

C.1 $\alpha \cup \alpha^{op}$ is symmetric.
C.2 $\alpha \subseteq \alpha \cup \alpha^{op}$.
C.3 If γ is symmetric and $\alpha \subseteq \gamma$, then $\alpha \cup \alpha^{op} \subseteq \gamma$.

To prove C.1, suppose $x(\alpha \cup \alpha^{op})y$. Then $x\alpha y$ and $x\alpha^{op}y$, so $y\alpha^{op}x$ and $y(\alpha^{op})^{op}x$, that is, $y\alpha x$. So $y(\alpha \cup \alpha^{op})x$. Hence $\alpha \cup \alpha^{op}$ is symmetric.

C.2 follows because for any sets S and T, $S \subseteq S \cup T$. As for C.3, suppose γ is symmetric and $\alpha \subseteq \gamma$. Suppose $x\alpha^{op}y$. Then $y\alpha x$, so $y\gamma x$ because $\alpha \subseteq \gamma$. Since γ is symmetric, $x\gamma y$. Thus $\alpha^{op} \subseteq \gamma$. We already know that $\alpha \subseteq \gamma$, so it follows that $\alpha \cup \alpha^{op} \subseteq \gamma$ as required.

129.2.1 No; not symmetric.

129.2.2 No; not symmetric or transitive.

129.2.3 No. Not reflexive or transitive.

129.2.4 No. Not transitive.

129.3.1 No, not symmetric or transitive.

129.3.2 Yes. $[0]_E = [0..1)$ and $[3]_E = [3..4)$.

129.3.3 Yes. $[0]_E = [0..1]$ and $[3]_E = \{3\}$.

130.1.3 $3, 27, 51, 75, 99$.

130.4.4 a) 1. b) 5. c) 1.

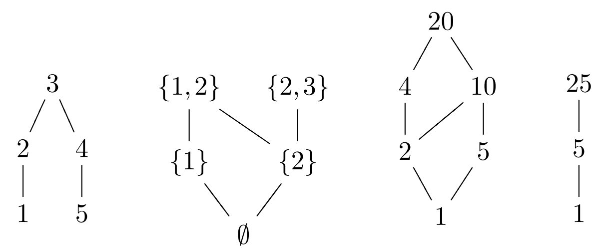
131.1.3 $F(6) = 0$, $F(n) = 1$ otherwise. (There are many answers.)

132.2.4 Here are all the possible values of E and E':

E	S/E
$\Delta_S \cup \{\langle 1,2\rangle, \langle 2,1\rangle\}$	$\{\{1,2\},\{3\},\{4\},\{5\}\}$
$\Delta_S \cup \{\langle 1,2\rangle, \langle 2,1\rangle, \langle 3,4\rangle, \langle 4,3\rangle\}$	$\{\{1,2\},\{3,4\},\{5\}\}$
$\Delta_S \cup \{\langle 1,2\rangle, \langle 2,1\rangle, \langle 3,5\rangle, \langle 5,3\rangle\}$	$\{\{1,2\},\{3,5\},\{4\}\}$
$\Delta_S \cup \{\langle 1,2\rangle, \langle 2,1\rangle, \langle 4,5\rangle, \langle 5,4\rangle\}$	$\{\{1,2\},\{3\},\{4,5\}\}$
$\Delta_S \cup \{\langle 1,2\rangle, \langle 2,1\rangle, \langle 3,4\rangle, \langle 3,5\rangle,$ $\langle 5,4\rangle, \langle 5,3\rangle, \langle 4,5\rangle, \langle 5,4\rangle\}$	$\{\{1,2\},\{3,4,5\}\}$

135.3.2 We must show that α is antisymmetric, transitive, and irreflexive. If $a\,\alpha\,b$ and $b\,\alpha\,a$, this contradicts the requirement that exactly one of the statements in 135.3 holds *unless* $a = b$. Thus $a\,\alpha\,b$ and $b\,\alpha\,a$ imply $a = b$, so α is antisymmetric. α is transitive by assumption. Finally, for any $a \in A$, $a = a$, so that rules out $a\,\alpha\,a$, so α is irreflexive.

137.1.3



(a) (b) (c) (d)

137.1.4 Only (d).

139.1.5 Lexical ordering: 00, 01, 0101, 0111, 01111, 10101, 10111, 110, 111.
Canonical ordering: 00, 01, 110, 111, 0101, 0111, 01111, 10101, 10111.

140.3.2 (a) max$=3$, no min.
(b) no max, min$=\emptyset$.
(c) max$=20$, min$=1$.
(d) max$=25$, min$=1$.

140.3.3 (a) max$=0$, min$=1$ (b) no max, min$=1$ (c) no max, no min.

141.3.2 (a) sup$=5$, ninf$=3$
(b) sup$=60$, inf$=1$
(c) no sup, inf$=2$
(d) sup$=0$, inf$=1$
(e) sup$=\{1,2,3\}$, inf$=\{2\}$

142.1.6 All except (f).

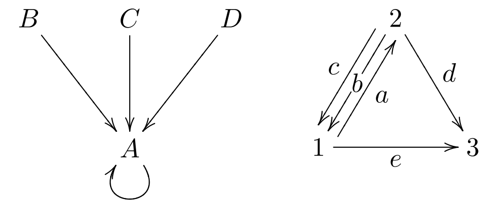
142.1.7 We will show that the infimum of any two elements is the intersection. The proof for the supremum is similar. By Theorem 141.2, we must show

(i) If $B \subseteq A$ and $C \subseteq A$, then $B \cap C \subseteq B$ and $B \cap C \subseteq C$.

(ii) If $B \subseteq A$, $C \subseteq A$, $D \subseteq B$ and $D \subseteq C$, then $D \subseteq B \cap C$.

To see (i), suppose $x \in B \cap C$. By Definition 33.2, page 47, $x \in B$ and $x \in C$. Then by Definition 31.1, $B \cap C \subseteq B$ and $B \cap C \subseteq C$. For (ii), suppose $x \in D$. Then by assumption, $x \in B$ and $x \in C$. Then by Definition 33.2, $x \in B \cap C$. Hence $D \subseteq B \cap C$.

144.2.2



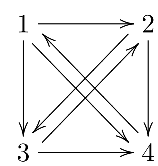
(a) (b)

146.2.3 (a) is simple. The relational definition of (a) is:

$$G_0 = \{A, B, C, D\}$$
$$G_1 = \{\langle A, A\rangle, \langle B, A\rangle, \langle C, A\rangle, \langle D, A\rangle\}$$

147.2.2 There are six automorphisms of (a), representing every possible way of permuting the set $\{B, C, D\}$. There are two automorphisms of (b) (the identity and the one that switches b and c.

148.1.3



(0.2)

148.1.4

$G_0 = \{1, 2, 3, 4\}$
$G_1 = \{\langle 1,2\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \langle 2,3\rangle, \langle 2,4\rangle, \langle 3,2\rangle, \langle 3,4\rangle, \langle 4,1\rangle\}$

148.1.6

(left)	a	b	c
a	0	1	1
b	1	0	1
c	0	0	0

(right)	1	2	3
1	0	1	1
2	1	0	1
3	0	0	0

149.4.3 $\langle e\rangle$ and $\langle a, d\rangle$. Note that a path of length n or more in a digraph with n nodes cannot be simple.

151.3.1

(a) 1 of length 1, 1 of length 2, 2 of length 3, 2 of length 4.

(b) 1 each of length 1 and 2, 2 of length 3 and 4 of length 4.

151.4.1 (a)

	x	y	z	w
x	2	10	2	0
y	0	4	0	0
z	2	8	2	0
w	0	0	0	1

(b)

2	1	1
1	2	1
0	0	1

152.2.1 All but c and e.

153.3.2 b and d.

153.3.3 a and f are bipartite, a is complete bipartite.

155.1.4 No pair of the graphs are isomorphic.

155.1.5 Map (b) to the triangle with horizontal bottom edge (full) and to one of the triangles with horizontal top edge (nonfull).

157.4.1 b and d have diameter 1, f has diameter 3, the others have diameter 2.

157.4.2 a and b.

157.4.3 All of them.

158.1.4 All of them!

Bibliography

At the end of each entry, the pages on which that entry is cited are listed in parentheses.

Bagchi, A. and C. Wells (1998). "On the communication of mathematical reasoning". *PRIMUS*, volume **8**, pages 15–27. Also available by web browser from `URL: http://www.cwru.edu/artsci/math/wells/pub/papers.html`. (117)

Bagchi, A. and C. Wells (1998). "Varieties of mathematical prose". *PRIMUS*, volume **8**, pages 116–136. Also available by web browser from `URL: http://www.cwru.edu/artsci/math/wells/pub/papers.html`. (117)

Ebbinghaus, H. D., J. Flum, and W. Thomas (1984). *Mathematical Logic*. Springer-Verlag.

Graham, R. L., D. E. Knuth, and O. Patashnik (1989). *Concrete Mathematics*. Addison-Wesley.

Guy, R. (1981). *Unsolved Problems in Number Theory*. Springer-Verlag. (160)

Hofstadter, D. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books. (vi, 159)

Knuth, D. E. (1971). *The Art of Computer Programming, Volume 2*. Addison-Wesley.

Lagarias, J. (1985). "The 3x +1 problem and its generalizations". *American Mathematical Monthly*, volume **92**. (160)

Myerson, G. and A. J. van der Poorten (1995). "Some problems concerning recurrence sequences". *American Mathematical Monthly*, volume **102**. (163)

Raymond, E. S. (1991). *The New Hacker's Dictionary*. The MIT Press.

Riesel, H. (1985). *Prime Numbers and Computer Methods for Factorization*. Birkhauser.

Rosen, K. H. (1992). *Elementary Number Theory and its Applications, Third Edition*. Addison-Wesley.

Skiena, S. (1990). *Implementing Discrete Mathematics*. Addison-Wesley.

Wells, C. (1995). "Communicating mathematics: Useful ideas from computer science". *American Mathematical Monthly*, volume **102**, pages 397–408. Also available by web browser from `URL: http://www.cwru.edu/artsci/math/wells/pub/papers.html`. (117)

Wells, C. (1998). "Handbook of mathematical discourse". `URL: http://www.cwru.edu/artsci/math/wells/pub/papers.html`. (117)

Wilder, R. L. (1965). *Introduction to the Foundations of Mathematics. Second Edition*. John Wiley and Sons. (35)

Wilf, H. (1990). *Generatingfunctionology*. Academic Press.

Index

The page number(s) in boldface indicate where the definition or basic explanation of the word is found. The other page numbers refer to examples and further information about the word.

Index of Symbols