CONTENTS

* All sections marked with an asterisk can be omitted without disrupting the continuity.

CHAPTER
ONE

SETS AND PROPOSITIONS

1.1 INTRODUCTION

A major theme of this book is to study discrete objects and relationships among them. The term *discrete objects* is a rather general one. It includes a large variety of items such as people, books, computers, transistors, computer programs, and so on. In our daily lives as well as our technical work we frequently deal with these items, making statements such as, "The people in this room are Computer Science majors in their second year of study," "All the books I bought are detective stories written by A. B. Charles," and "We want to select and buy a computer among those that are suitable for both scientific and business applications at a price not exceeding $200,000." We would like to abstract some of the basic concepts dealing with the many different kinds of discrete objects and establish certain common terminology for dealing with them.

A hint of the possibility of such an abstraction is quite evident when we observe that these three statements all have "something" in common. To be specific, in the first statement we are referring to people who possess the two attributes of being a Computer Science major and of being a sophomore; in the second statement we are referring to books that possess the two attributes of being a detective story and of being written by A. B. Charles; and in the third statement we are referring to computers that possess the three attributes of being suitable for scientific applications, of being suitable for business applications, and of being priced at no more than $200,000. To put it in another way, consider the group of all the Computer Science majors and the group of all the sophomores in the university. In our first statement we are then referring to

those students who belong to both of these groups. Also, consider the collection of all detective stories and the collection of all books written by A. B. Charles. In our second statement we are then referring to those books which belong to both of these collections. Finally, in our third statement we are referring to all computers which belong to the three categories of computers that are suitable for business applications, that are suitable for scientific applications, and that are priced at no more than $200,000.

Our example illustrates the many occasions on which we deal with several classes of objects and wish to refer to those objects that belong to all the classes. Similarly, one would immediately perceive occasions on which we refer to objects that belong to one of several classes of objects, such as in the statement, "I want to interview all the students who speak either German or French," where we refer to those who belong either to the group of German-speaking students or to the group of French-speaking students.

We begin with the introduction of some basic terminology and concepts in elementary set theory.

A *set* is a collection of *distinct* objects. Thus, the group of all sophomores in the university is a set. So is the group of all Computer Science majors in the university, and so is the group of all second-year Computer Science majors. We use the notation $\{a, b, c\}$ to denote the set which is the collection of the objects a, b, and c. The objects in a set are also called the *elements* or the *members* of the set. We usually also give names to sets. For example, we write $S = \{a, b, c\}$ to mean that the set named S is the collection of the objects a, b, and c. Consequently, we can refer to the set S as well as to the set $\{a, b, c\}$. As another example, we may have

Second-year-Computer-Science-majors
$= \{\text{Smith, Jones, Wong, Yamamoto, Vögeli}\}$

(The name of the set {Smith, Jones, Wong, Yamamoto, Vögeli} is Second-year-Computer-Science-majors, which is rather long. The reader probably would want to suggest alternative names such as S or CS. However, there is nothing wrong conceptually with having a "long" name.) We use the notation $a \in S$ to mean that a is an element in the set S. In that case, we also say that S *contains* the element a. We use the notation $d \notin S$ to mean that d is not an element in the set S. In that case, we also say that S does not contain the element d. Thus, in the example above, Jones $\in$ Second-year-Computer-Science-majors, while Kinkaid $\notin$ Second-year-Computer-Science-majors.

Note that a set contains only distinct elements. Thus, $\{a, a, b, c\}$ is a redundant representation of the set $\{a, b, c\}$. Similarly, {The-Midnight-Visitor, The-Midnight-Visitor, The-Missing-Witness, 114-Main-Street} is a redundant representation of the detective stories written by A. B. Charles. One might ask the question: What should we do if our collection of detective stories by A. B. Charles in the library indeed contains two copies of the book The-Midnight-Visitor? In that case, the set {The-Midnight-Visitor, The-Missing-Witness, 114-Main-Street} is a set of distinct titles of detective stories by A. B. Charles in our

library, while the set {The-Midnight-Visitor-1, The-Midnight-Visitor-2, The-Missing-Witness, 114-Main-Street} is the set of detective stories by A. B. Charles in our library where The-Midnight-Visitor-1 is copy 1 of the book The-Midnight-Visitor, and The-Midnight-Visitor-2 is copy 2 of the book. Note that The-Midnight-Visitor-1 and The-Midnight-Visitor-2 are two distinct elements in the latter set.

Note also that the elements in a set are not ordered in any fashion. Thus, $\{a, b, c\}$ and $\{b, a, c\}$ represent the same collection of elements. Later on, we shall introduce the notion of *ordered sets.*

As was introduced above, one way to describe the membership of a set is to list exhaustively all the elements in that set. In many cases, when the elements in a set share some common properties, we can describe the membership of the set by stating the properties that uniquely characterize the elements in the set. For example, let $S = \{2, 4, 6, 8, 10\}$. We can also specify the elements of S by saying that S is the set of all even positive integers that are not larger than 10. Indeed, we can use the notation

$$S = \{x \mid x \text{ is an even positive integer not larger than } 10\}$$

for the set {2, 4, 6, 8, 10}. In general, we use the notation

$$\{x \mid x \text{ possesses certain properties}\}$$

for a set of objects that share some common properties. Thus,

$$S = \{\text{Smith, Jones, Wong, Yamamoto, Vögeli}\}$$

and

$$S = \{x \mid x \text{ is a second year Computer Science major}\}$$

are two different ways to describe the same set of elements.

It should be pointed out that our definition of a set does not preclude the possibility of having a set containing *no* elements. The set that contains no element is known as the *empty set*, and is denoted by $\emptyset$. (Following our established notation of using a pair of braces to enclose all the elements in a set, we could also denote the empty set by { }.) For example, let S denote the set of all detective stories by A. B. Charles that were published in 1924. Clearly, S is the empty set if A. B. Charles was born in 1925.

Let us note that we did not place any restriction on the elements in a set. Thus, S = {Smith, The-Midnight-Visitor, CDC-6600} is a well-defined set. That the elements, Smith (a person), The-Midnight-Visitor (the title of a book), and CDC-6600 (a computer) do not seem to share anything in common does not prohibit them from being elements of the same set. Indeed, we should point out that it is perfectly all right to have sets as members of a set. Thus, for example, the set $\{\{a, b, c\}, d\}$ contains the two elements $\{a, b, c\}$ and d, and the set $\{\{a, b, c\}, a, b, c\}$ contains the four elements $\{a, b, c\}$, a, b, and c. The set of all committees in the U.S. Senate could be represented by $\{\{a, b, c\}, \{a, d, e, f\}, \{b, e, g\}\}$ where each element of the set is a committee which in turn is a set with the senators in the committee as elements. Similarly, $\{a, \{a\}, \{\{a\}\}\}$ is a set with

three *distinct* elements $a, \{a\}, \{\{a\}\}$. Also, the set $\{\emptyset\}$ contains one element—the empty set—and the set $\{\emptyset, \{\emptyset\}\}$ contains two elements—the empty set and a set that contains the empty set as its only element.

Given two sets P and Q, we say that P is a *subset* of Q if every element of P is also an element in Q. We shall use the notation $P \subseteq Q$ to denote that P is a subset of Q. For example, the set $\{a, b\}$ is a subset of the set $\{y, x, b, c, a\}$, but it is not a subset of the set $\{a, c, d, e\}$. The set of all second-year Computer Science majors is a subset of the set of all sophomores. It is also a subset of the set of all Computer Science majors. On the other hand, the set of all Computer Science majors is not a subset of the set of all sophomores nor is the set of all sophomores a subset of the set of all Computer Science majors. Let $A = \{a, b, c\}$ and $B = \{\{a, b, c\}, a, b, c\}$. We note that it is indeed possible to have both $A \in B$ and $A \subseteq B$. As further examples, we ask the reader to check the following statements:

For any set P, P is a subset of P.
The empty set is a subset of any set.
The set $\{\emptyset\}$ is not a subset of the set $\{\{\emptyset\}\}$.

Two sets P and Q are said to be *equal* if they contain the same collection of elements. For example, the two sets

$$P = \{x | x \text{ is an even positive integer not larger than } 10\}$$

$$Q = \{x | x = y + z \text{ where } y \in \{1, 3, 5\},\ z \in \{1, 3, 5\}\}$$

are equal. In a seemingly roundabout way, we can also say that two sets P and Q are equal if P is a subset of Q, and Q is a subset of P. We shall see later that on some occasions, this is a convenient way to define the equality of two sets.

Let P be a subset of Q. We say that P is a *proper* subset of Q if P is not equal to Q, that is, there is at least one element in Q that is not in P. For example, the set $\{a, b\}$ is a proper subset of the set $\{y, x, b, c, a\}$. We use the notation $P \subset Q$ to denote that P is a proper subset of Q.

1.2 COMBINATIONS OF SETS

We show now how sets can be *combined* in various ways to yield new sets. For example, let P be the set of students taking the course Theory of Computation and Q be the set of students taking the course Music Appreciation. If a certain announcement was made in both the Theory of Computation and the Music Appreciation classes, what is the set of students who know about the news announced? Clearly, it is the set of students who are taking either Theory of Computation or Music Appreciation, or both. If both these courses have their final examinations scheduled in the same hours, what is the set of students who will have conflicting final examinations? Clearly, it is the set of students who

are taking both Theory of Computation and Music Appreciation. To formalize these notions, we define the union and the intersection of sets. The *union* of two sets P and Q, denoted $P \cup Q$, is the set whose elements are exactly the elements in either P or Q (or both).† For example,

$$\{a, b\} \cup \{c, d\} = \{a, b, c, d\}$$

$$\{a, b\} \cup \{a, c\} = \{a, b, c\}$$

$$\{a, b\} \cup \emptyset = \{a, b\}$$

$$\{a, b\} \cup \{\{a, b\}\} = \{a, b, \{a, b\}\}$$

The *intersection* of two sets P and Q, denoted $P \cap Q$, is the set whose elements are exactly those elements that are in both P and Q. For example,

$$\{a, b\} \cap \{a, c\} = \{a\}$$

$$\{a, b\} \cap \{c, d\} = \emptyset ‡$$

$$\{a, b\} \cap \emptyset = \emptyset$$

If the elements of P are characterized by a common property and the elements in Q are characterized by another common property, then the union of P and Q is the set of elements possessing at least one of these properties, and the intersection of P and Q is the set of elements possessing both of these properties. According to the definitions, $P \cup Q$ and $Q \cup P$ denote the same set, as do $P \cap Q$ and $Q \cap P$.

In general, the union of k sets $P_1, P_2, \ldots, P_k$, denoted $P_1 \cup P_2 \cup \cdots \cup P_k$, is the set containing exactly the elements in P_1, the elements in $P_2, \ldots$, and the elements in P_k. Similarly, the intersection of k sets $P_1, P_2, \ldots, P_k$, denoted $P_1 \cap P_2 \cap \cdots \cap P_k$, is the set containing exactly the elements that are in P_1 and in $P_2 \ldots$ and in P_k. For example, the set of all undergraduate students in a university is the union of the sets of freshmen, sophomores, juniors, and seniors, and the set of graduating seniors is the intersection of the set of seniors, the set of students who have accumulated 144 or more credit hours, and the set of students who have a C or better grade-point average.

Let P denote the set of students taking Theory of Computation, Q denote the set of students taking Music Appreciation, and R denote the set of students having type AB blood. Suppose an emergency announcement was made in the classes of Theory of Computation and Music Appreciation calling for type AB blood donors. We want to determine the members of the set of potential donors who heard about the emergency call. Since $S = P \cup Q$ is the set of students who heard about the emergency call, $R \cap S$ is the set of potential donors who heard about the emergency call. Instead of using a new name S for the set $P \cup Q$, we can simply write $R \cap (P \cup Q)$, where the parentheses are used as delimiters to

† We do not wish to introduce the notion of algebraic operations until Chap. 8. Thus, at this moment, $P \cup Q$ is simply a name we have chosen for a set.

‡ Two sets are said to be *disjoint* if their intersection is the empty set.

avoid confusion. Note that the set of potential donors who heard about the emergency call is also the set of students with type AB blood in the Theory of Computation class together with the set of students with type AB blood in the Music Appreciation class. That is, the set $(R \cap P) \cup (R \cap Q)$. This example suggests very strongly that for any sets P, Q, R, the two sets $R \cap (P \cup Q)$ and $(R \cap P) \cup (R \cap Q)$ are equal. Indeed, this is the case, as we now show.

We show first that $R \cap (P \cup Q)$ is a subset of $(R \cap P) \cup (R \cap Q)$ by showing that every element in $R \cap (P \cup Q)$ is also in $(R \cap P) \cup (R \cap Q)$. Let x be an element in $R \cap (P \cup Q)$. The element x must be in R and must be either in P or Q. If x is in P, x is in $R \cap P$. If x is in Q, x is in $R \cap Q$. Consequently, x is in $(R \cap P) \cup (R \cap Q)$, and we conclude that $R \cap (P \cup Q)$ is a subset of $(R \cap P) \cup (R \cap Q)$. Second, we show that $(R \cap P) \cup (R \cap Q)$ is a subset of $R \cap (P \cup Q)$. Let x be an element in $(R \cap P) \cup (R \cap Q)$. Thus, x must either be in $R \cap P$ or be in $R \cap Q$. That is, x must either be in both R and P or be in both R and Q. In other words, x must be in R and must be either in P or in Q. Consequently, x is in $R \cap (P \cup Q)$, and we can conclude that $(R \cap P) \cup (P \cap Q)$ is a subset of $R \cap (P \cup Q)$. It follows that the two sets $R \cap (P \cup Q)$ and $(R \cap P) \cup (R \cap Q)$ are equal.

In a similar manner we can show that for any sets P, Q, R, the two sets $R \cup (P \cap Q)$ and $(R \cup P) \cap (R \cup Q)$ are equal. Furthermore, we have

$$R \cap (P_1 \cup P_2 \cup \cdots \cup P_k) = (R \cap P_1) \cup (R \cap P_2) \cup \cdots \cup (R \cap P_k)$$

$$R \cup (P_1 \cap P_2 \cap \cdots \cap P_k) = (R \cup P_1) \cap (R \cup P_2) \cap \cdots \cap (R \cup P_k)$$

We leave the details to the reader.†

The difference of two sets P and Q, denoted $P - Q$, is the set containing exactly those elements in P that are not in Q. For example,

$$\{a, b, c\} - \{a\} = \{b, c\}$$

$$\{a, b, c\} - \{a, d\} = \{b, c\}$$

$$\{a, b, c\} - \{d, e\} = \{a, b, c\}$$

If P is the set of people who have tickets to a ball game and Q is the set of people who are ill on the day of the game, then $P - Q$ is the set of people who will go to the game. Note that Q might contain some or none of the elements of the set P. However, these elements will not appear in $P - Q$ in any case, just as in the example, those people who are ill but do not have tickets to the ball game will not go to the game anyway. Indeed, if the elements in Q are characterized by some common property, then $P - Q$ is the set of elements in P that do not possess this property. If Q is a subset of P, the set $P - Q$ is also called the *complement of Q with respect to P*. For example, let P be the set of all students

† Again, we do not wish to introduce the notions of algebraic operations, associativity, and distributivity until Chap. 8. Note, however, these notions are not needed here because $P \cap Q$, $P \cup Q$, $P_1 \cup P_2 \cup \cdots \cup P_k$ are simply names for sets obtained according to our definitions.

in the course Theory of Computation and Q be the set of those students who have passed the course. Then $P - Q$ is the set of students who failed the course. On many occasions, when the set P is clear from the context, we shall abbreviate the *complement of Q with respect to P* as *the complement of Q*, which will be denoted $\bar{Q}$. For example, let P be the set of all students in the course Theory of Computation. Let Q be the set of Computer Science majors in the course, and R be the set of sophomores in the course. Then the complement of Q refers to the set of students in the course who are not Computer Science majors, and the complement of R refers to the set of those students who are not sophomores, if it is understood that in our discussion we always restrict ourselves to students in the course Theory of Computation. Indeed, when our discussion is always restricted to the subsets of a set P, P is referred to as the *universe*.

The *symmetric difference* of two sets P and Q, denoted $P \oplus Q$, is the set containing exactly all the elements that are in P or in Q but not in both. In other words, $P \oplus Q$ is the set $(P \cup Q) - (P \cap Q)$. For example,

$$\{a, b\} \oplus \{a, c\} = \{b, c\}$$

$$\{a, b\} \oplus \emptyset = \{a, b\}$$

$$\{a, b\} \oplus \{a, b\} = \emptyset$$

If we let P denote the set of cars that have defective steering mechanisms and Q denote the set of cars that have defective transmission systems, then $P \oplus Q$ is the set of cars that have one but not both of these defects. Suppose that a student will get an A in a course if she did well in both quizzes, will get a B if she did well in one of the two quizzes, and will get a C if she did poorly in both quizzes. Let P be the set of students who did well in the first quiz and Q be the set of students who did well in the second quiz. Then $P \cap Q$ is the set of students who will get A's, $P \oplus Q$ is the set of students who will get B's, and $S - (P \cup Q)$ is the set of students who will get C's, where S is the set of all students in the course. We define $P_1 \oplus P_2 \oplus \cdots \oplus P_k$ to be the set of elements that are in an odd number of the sets $P_1, P_2, \ldots, P_k$.

The *power set* of a set A, denoted $\mathcal{P}(A)$, is the set that contains exactly all the subsets of A. Thus $\mathcal{P}(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$, and $\mathcal{P}(\emptyset) = \{\emptyset\}$. For example, let $A = \{\text{novel, published-in-1975, paperback}\}$ be the three attributes concerning the books in the library in which we are interested. Then $\mathcal{P}(A)$ is the set of all possible combinations of these attributes the books might possess, ranging from books that have none of these attributes [the empty set in $\mathcal{P}(A)$] to books that have all three of these attributes [the set A in $\mathcal{P}(A)$].

Sets obtained from combinations of given sets can be represented pictorially. If we let P and Q be the sets represented by the cross-hatched areas in Fig. 1.1*a*, then the cross-hatched areas in Fig. 1.1*b* represent the sets $P \cup Q$, $P \cap Q$, $P - Q$, and $P \oplus Q$, respectively. These diagrams are known as *Venn diagrams*.

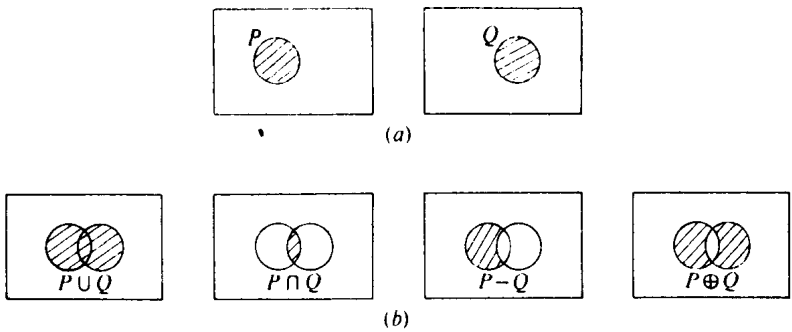


Figure 1.1

1.3 FINITE AND INFINITE SETS

Intuitively, it is quite clear that by the size of a set we mean the number of distinct elements in the set. Thus, there is little doubt when we say the size of the set $\{a, b, c\}$ is 3, the size of the set $\{a, \emptyset, d\}$ is also 3, the size of the set $\{\{a, b\}\}$ is 1, and the size of the set $\emptyset$ is 0. Indeed, we could stop our discussion on the size of sets at this point if we were only interested in the size of "finite" sets. However, a much more intriguing topic is the size of "infinite" sets. At this point, a perceptive reader will probably ask the question, "What is an infinite set in the first place?" An evasive answer such as, "An infinite set is not a finite set," is no answer at all, because if we start to think about it, we should also ask the question, "What is a finite set anyway?"

Let us begin by declaring that we have not yet committed ourselves to the precise definitions of finite sets and infinite sets. As the basis of our discussion, we want to construct an example of an infinite set. For a given set A, we define the *successor* of A, denoted A^+, to be the set $A \cup \{A\}$. Note that $\{A\}$ is a set that contains A as the only element. In other words, A^+ is a set that consists of all the elements of A together with an additional element which is the set A. For example, if $A = \{a, b\}$, then $A^+ = \{a, b\} \cup \{\{a, b\}\} = \{a, b, \{a, b\}\}$; and if $A = \{\{a\}, b\}$, then $A^+ = \{\{a\}, b, \{\{a\}, b\}\}$. Let us now construct a sequence of sets starting with the empty set $\emptyset$. The successor of the empty set is $\{\emptyset\}$, whose successor is $\{\emptyset, \{\emptyset\}\}$, and whose successor, in turn, is $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$. It is clear that we can go on to construct more and more successors. Let us also assign names to these sets. Let

$$0 = \emptyset$$

$$1 = \{\emptyset\}$$

$$2 = \{\emptyset, \{\emptyset\}\}$$

$$3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$$

.....

We have, clearly, $1 = 0^+$, $2 = 1^+$, $3 = 2^+$, and so on.† Let us now define a set N such that

1. N contains the set 0.
2. If the set n is an element in N so is the set n^+.
3. N contains no other sets.

Since for every set in N its successor is also in N, the reader probably would agree that N is indeed an "infinite set." However, let us proceed in a more precise way.

We shall talk about the sizes of sets in a comparative manner. To this end, let us introduce a definition: Given two sets P and Q, we say that there is a *one-to-one correspondence* between the elements in P and the elements in Q if it is possible to pair off the elements in P and Q such that every element in P is paired off with a distinct element in Q.‡ Thus, there is a one-to-one correspondence between the elements in the set $\{a, b\}$ and the elements in the set $\{c, d\}$, because we can pair a with c and b with d, or we can pair a with d and b with c. There is also a one-to-one correspondence between the elements in the set $\{a, b, c\}$ and the elements in the set $\{\emptyset, a, d\}$. On the other hand, there is no one-to-one correspondence between the elements in the set $\{a, b, c\}$ and $\{a, d\}$. The intention of introducing the notion of one-to-one correspondence between the elements of two sets is quite obvious, because we can now compare two sets and say that they are of the same size or that they are of different sizes. The basis of our comparison is indeed the sets we constructed above, namely, 0, 1, 2, 3, ..., and N. We are now ready to introduce some formal definitions. A set is said to be a *finite* set if there is a one-to-one correspondence between the elements in the set and the elements in some set n, where $n \in N$; n is said to be the *cardinality* of the set. Thus, for example, the cardinalities of the sets $\{a, b, c\}$, $\{a, \emptyset, d\}$, $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$ are all equal to 3. Note that it is now precise for us to say that a set is an infinite set if it is not a finite one. We can, however, be more precise about the "size" of infinite sets: A set is said to be a *countably infinite* set (or the cardinality of the set is countably infinite§) if there is a one-to-one correspondence between the elements in the set and the elements in N. We observe first of all that the set of all natural numbers $\{0, 1, 2, 3, \ldots\}$¶ is a countably infinite set. It follows that the set of all nonnegative even integers $\{0, 2, 4, 6, 8, \ldots\}$ is a countably infinite set because there is an obvious one-to-one correspondence between all nonnegative even integers and all natural numbers, namely, the even integer $2i$ corresponds to the natural number i for $i = 0, 1,$

† Using 0, 1, 2, 3, ... as names of sets is just as good as using $A, B, C, D, \ldots$. As will be seen, it is intentional that we choose 0, 1, 2, 3, ... as names.

‡ Such an intuitive definition will be made more formal later on.

§ In the literature, the cardinality of a countably infinite set is also referred to as $\aleph_0$. ($\aleph$ is the first letter in the Hebrew alphabet.)

¶ The notation is perhaps confusing. However, it is intentional, because the set N is *indeed* a precise definition of the set of natural numbers.

2, ... Similarly, the set of all nonnegative multiples of 7 {0, 7, 14, 21, ...} is also a countably infinite set. So is the set of all positive integers {1, 2, 3, ...}. We note that a set is a countably infinite set if starting from a certain element we can sequentially list all the elements in the set one after another, because such a listing will yield a one-to-one correspondence between the elements in the set and the natural numbers. For example, the set of all integers {..., −2, −1, 0, 1, 2, ...} is a countably infinite set, since its elements can be listed sequentially as {0, 1, −1, 2, −2, 3, −3, ...}. This example suggests that the union of two countably infinite sets is also a countably infinite set. It indeed is the case. As a matter of fact, the union of a finite number of countably infinite sets is a countably infinite set and, furthermore, so is the union of a countably infinite number of countably infinite sets (see Prob. 1.19).

Finally, we show that there are infinite sets whose cardinalities are not countably infinite. In particular, we shall show that the set of real numbers between 0 and 1 is not a countably infinite set. Our proof procedure is to assume that the set is a countably infinite set and then show the existence of a contradiction. If the cardinality of the set of real numbers between 0 and 1 is countably infinite, there is a one-to-one correspondence between these real numbers and the natural numbers. Consequently, we can exhaustively list them one after another in decimal form as in the following:†

$$0.a_{11}a_{12}a_{13}a_{14}\cdots$$

$$0.a_{21}a_{22}a_{23}a_{24}\cdots$$

$$0.a_{31}a_{32}a_{33}a_{34}\cdots$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$0.a_{i1}a_{i2}a_{i3}a_{i4}\cdots$$

$$\cdots\cdots\cdots\cdots\cdots$$

where a_{ij} denotes the jth digit of the ith number in the list. Consider the number

$$0.b_1b_2b_3b_4\cdots$$

where

$$b_i = \begin{cases} 1 & \text{if } a_{ii} = 9 \\ 9 - a_{ii} & \text{if } a_{ii} = 0, 1, 2, \ldots, 8 \end{cases}$$

for all i. Clearly, the number $0.b_1b_2b_3b_4\cdots$ is a real number between 0 and 1 that does not have an infinite string of trailing 0's (i.e., $0.34000\cdots$). Moreover, it is different from each of the numbers in the list above because it differs from the first number in the first digit, the second number in the second digit, ..., the ith number in the ith digit, ..., and so on. Consequently, we conclude that the list above is not an exhaustive listing of the set of all real numbers between 0 and 1, contradicting the assumption that this set is a countably infinite set.

† A number such as 0.34 can be written in two different forms, namely, 0.34000 ... or 0.339999 ... We follow an arbitrarily chosen convention of writing it in the latter form.

It is possible to continue in this direction to classify infinite sets so that notions such as some infinite sets are "more infinite" than other infinite sets can be made precise. This, however, will be beyond our scope of discussion.

1.4 MATHEMATICAL INDUCTION

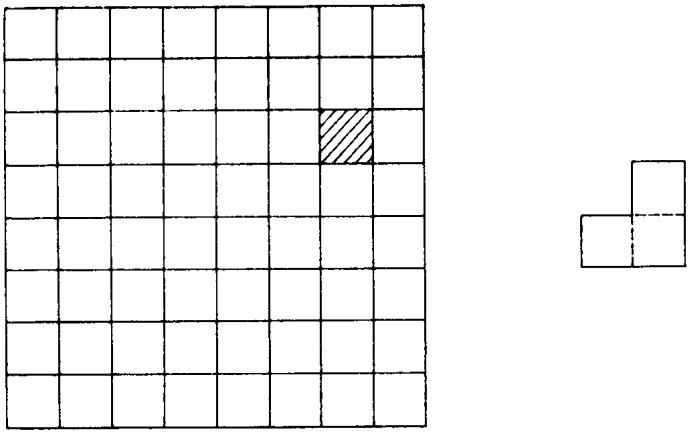
Let us consider some illustrative examples:

Example 1.1 Suppose we have stamps of two different denominations, 3 cents and 5 cents. We want to show that it is possible to make up exactly any postage of 8 cents or more using stamps of these two denominations. Clearly, the approach of showing case by case how to make up postage of 8 cents, 9 cents, 10 cents, ..., using 3-cent and 5-cent stamps will not be a fruitful one, because there is an infinite number of cases to be examined.† Let us consider an alternative approach. We want to show that if it is possible to make up exactly a postage of k cents using 3-cent and 5-cent stamps, then it is also possible to make up exactly a postage of $k + 1$ cents using 3-cent and 5-cent stamps. We examine two cases: Suppose we make up a postage of k cents using at least one 5-cent stamp. Replacing a 5-cent stamp by two 3-cent stamps will yield a way to make up a postage of $k + 1$ cents. On the other hand, suppose we make up a postage of k cents using 3-cent stamps only. Since $k \geq 8$, there must be at least three 3-cent stamps. Replacing three 3-cent stamps by two 5-cent stamps will yield a way to make up a postage of $k + 1$ cents. Since it is obvious how we can make up a postage of 8 cents, we conclude that we can make up a postage of 9 cents, which, in turn, leads us to conclude that we can make up a postage of 10 cents, which, in turn, leads us to conclude that we can make up a postage of 11 cents and so on. □

Example 1.2 Suppose we remove a square from a standard 8×8 chessboard as shown in Fig. 1.2*a*. Given 21 L-shaped triominoes‡ as shown in Fig. 1.2*b*, we want to know whether it is possible to *tile* the 63 remaining squares of the chessboard with the triominoes. (By tiling the remaining squares of the chessboard, we mean covering each of them exactly once without parts of the triominoes extending over the removed square or the edges of the board.) The answer to our question is affirmative, as Fig. 1.3 shows. We can actually prove a more general result as we shall proceed to do.

† See, however, Prob. 1.20.

‡ The word *tri*omino is derived from the word domino. Also, there are *tet*rominoes, *pent*ominoes, *hex*ominoes, ... and, in general *poly*ominoes. For many interesting results in connection with polyominoes, see Golomb[4].



(*a*) (*b*)

Figure 1.2

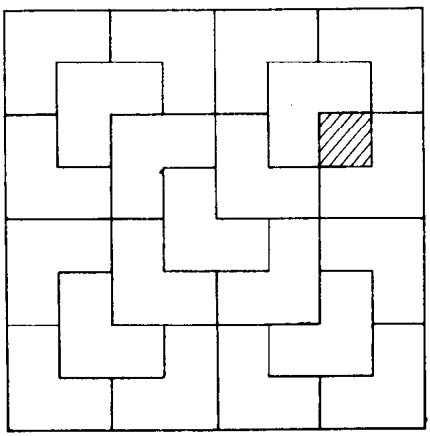


Figure 1.3

A chessboard with one of its squares removed will be referred to as a *defective* chessboard. We want to show that any defective $2^n \times 2^n$ chessboard can be tiled with L-shaped triominoes.† It is trivially obvious that a defective 2×2 chessboard can be tiled with an L-shaped triomino. Let us now assume that any defective $2^k \times 2^k$ chessboard can be tiled with L-shaped triominoes and proceed to show that any defective $2^{k+1} \times 2^{k+1}$ chessboard can also be tiled with L-shaped triominoes. Consider a defective $2^{k+1} \times 2^{k+1}$ chessboard as shown in Fig. 1.4*a*. Let us divide the chessboard into four quadrants, each of which is a $2^k \times 2^k$ chessboard, as shown in Fig. 1.4*b*. One of these $2^k \times 2^k$ chessboards is a defective one.

† One would immediately question whether $2^n \times 2^n - 1$ is always divisible by 3. The answer is affirmative. (See Prob. 1.23.)

Furthermore, by placing an L-shaped triomino at the center of the $2^{k+1} \times 2^{k+1}$ chessboard as shown in Fig. 1.4*c*, we can imagine that the other three quadrants are also defective $2^k \times 2^k$ chessboards. Since we assume that any defective $2^k \times 2^k$ chessboard can be tiled with L-shaped triominoes, we can tile each of the quadrants with L-shaped triominoes, and conclude that any defective $2^{k+1} \times 2^{k+1}$ chessboard can be tiled with L-shaped triominoes. Thus, starting with the tiling of any defective 2×2 chessboard, we have proved that we can tile any $2^n \times 2^n$ defective chessboard. □

These two examples illustrate a very powerful proof technique in mathematics known as the principle of *mathematical induction.* For a given

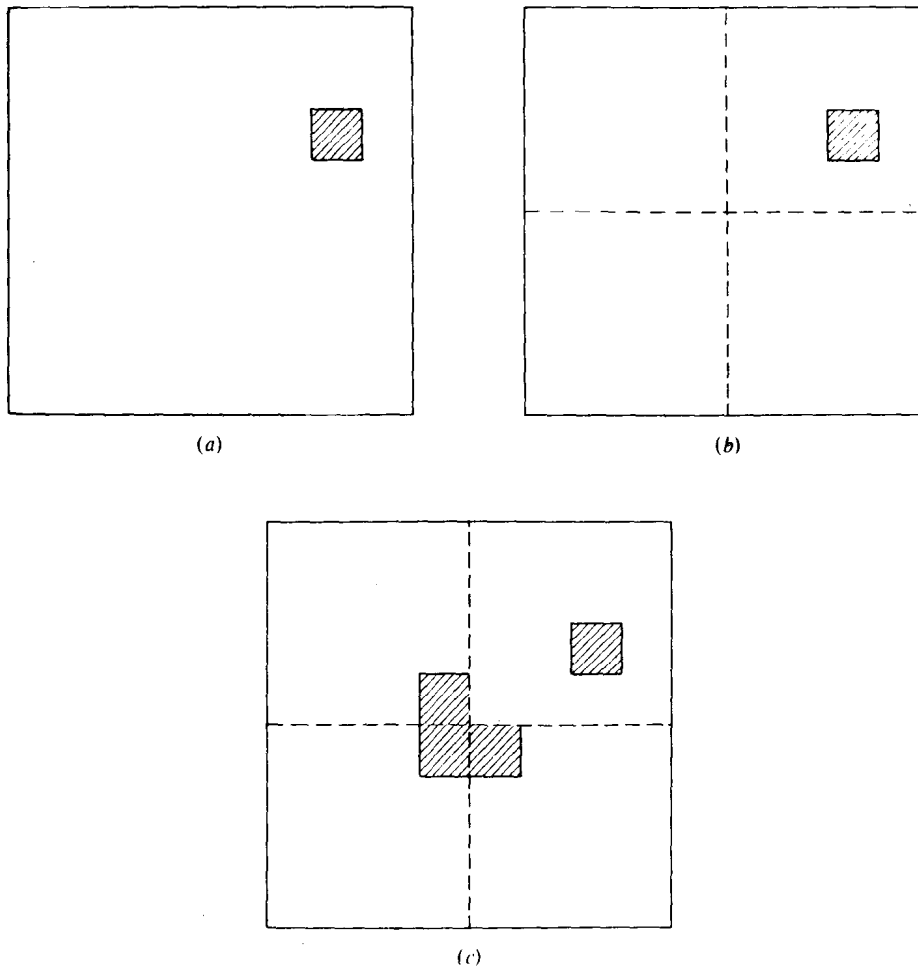


Figure 1.4

statement involving a natural number n, if we can show that:

1. The statement is true for $n = n_0$; and
2. The statement is true for $n = k + 1$, assuming that the statement is true for $n = k$, $(k \geqq n_0)$,

then we can conclude that the statement is true for all natural numbers $n \geqq n_0$. (1) is usually referred to as the *basis of induction*, and (2) is usually referred to as the *induction step*. For example, in the postage-stamp problem, we want to prove the statement, "It is possible to make up exactly any postage of n cents using 3-cent stamps and 5-cent stamps for $n \geqq 8$." In order to prove the statement we show that:

1. *Basis of induction.* It is possible to make up exactly a postage of 8 cents.
2. *Induction step.* It is possible to make up exactly a postage of $k + 1$ cents, assuming it is possible to make up exactly a postage of k cents.

We note that the principle of mathematical induction is a direct consequence of the definition of natural numbers. Consider a set S such that

1. The natural number n_0 is in S.
2. If the natural number k is in S, then the natural number $k + 1$ is also in S.

According to the definition of the set of natural numbers, we can conclude that S contains all the natural numbers larger than or equal to k. However, this is exactly the statement of the principle of mathematical induction when we consider S to be the set of natural numbers for which a given statement is true.

We consider now more examples:

Example 1.3 Show that

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} \qquad n \geqq 1$$

by mathematical induction.

1. *Basis of induction.* For $n = 1$, we have

$$1^2 = \frac{1(1+1)(2+1)}{6}$$

2. *Induction step.* Assume that

$$1^2 + 2^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}$$