CHAPTER 11

Relations

In mathematics there are endless ways that two entities can be related to each other. Consider the following mathematical statements.

$$\begin{array}{ccccccc} 5<10 & 5\le 5 & 6=\frac{30}{5} & 5\,|\,80 & 7>4 & x\neq y & 8\nmid 3 \\ a\equiv b\,(\bmod n) & 6\in\mathbb{Z} & X\subseteq Y & \pi\approx 3.14 & 0\ge -1 & \sqrt{2}\notin\mathbb{Z} & \mathbb{Z}\not\subseteq\mathbb{N} \end{array}$$

In each case two entities appear on either side of a symbol, and we interpret the symbol as expressing some relationship between the two entities. Symbols such as $<, \le, =, |, \nmid, \ge, >$, $\in$ and $\subset$, etc., are called *relations* because they convey relationships among things.

Relations are significant. In fact, you would have to admit that there would be precious little left of mathematics if we took away all the relations. Therefore it is important to have a firm understanding of relations, and this chapter is intended to develop that understanding.

Rather than focusing on each relation individually (an impossible task anyway since there are infinitely many different relations) we will develop a general theory that encompasses *all* relations. Understanding this general theory will give us the conceptual framework and language needed to understand and discuss any specific relation.

Before stating the theoretical definition of a relation, let's look at a motivational example. This example will lead us naturally to our definition.

Consider the set $A = \{1,2,3,4,5\}$. (There's nothing special about this particular set; any set of numbers would do for this example.) Elements of A can be compared to each other by the symbol "$<$." For example, $1<4$, $2<3$, $2<4$, and so on. You have no trouble understanding this because the notion of numeric order is so ingrained. But imagine you had to explain it to an idiot savant, one with an obsession for detail but absolutely no understanding of the meaning of (or relationships between) integers. You might consider writing down for your student the following set:

$$R = \{(1,2),(1,3),(1,4),(1,5),(2,3),(2,4),(2,5),(3,4),(3,5),(4,5)\}$$

The set R encodes the meaning of the $<$ relation for elements in A. An ordered pair (a,b) appears in the set if and only if $a<b$. If asked whether or not it is true that $3<4$, your student could look through R until he found the ordered pair $(3,4)$; then he would know $3<4$ is true. If asked about $5<2$, he would see that $(5,2)$ *does not* appear in R, so $5 \not< 2$. The set R, which is a subset of $A \times A$, completely describes the relation $<$ for A.

Though it may seem simple-minded at first, this is exactly the idea we will use for our main definition. This definition is general enough to describe not just the relation $<$ for the set $A=\{1,2,3,4,5\}$, but *any* relation for *any* set A.

Definition 11.1 A **relation** on a set A is a subset $R \subseteq A \times A$. We often abbreviate the statement $(x,y) \in R$ as xRy. The statement $(x,y) \notin R$ is abbreviated as $x \not R y$.

Notice that a relation is a set, so we can use what we know about sets to understand and explore relations. But before getting deeper into the theory of relations, let's look at some examples of Definition 11.1.

Example 11.1 Let $A=\{1,2,3,4\}$, and consider the following set:

$$R=\{(1,1),(2,1),(2,2),(3,3),(3,2),(3,1),(4,4),(4,3),(4,2),(4,1)\} \subseteq A \times A$$

The set R is a relation on A, by Definition 11.1. Since $(1,1) \in R$, we have $1R1$. Similarly $2R1$ and $2R2$, and so on. However notice that (for example) $(3,4) \notin R$, so $3 \not R 4$. Observe that R is the familiar relation $\geq$ for the set A.

Chapter 1 proclaimed that all of mathematics can be described with sets. Just look at how successful this program has been! The greater-than relation is now a set R. (We might even express this in the rather cryptic form $\geq = R$.)

Example 11.2 Let $A=\{1,2,3,4\}$, and consider the following set:

$$S=\{(1,1),(1,3),(3,1),(3,3),(2,2),(2,4),(4,2),(4,4)\} \subseteq A \times A$$

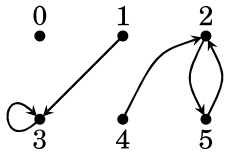
Here we have $1S1$, $1S3$, $4S2$, etc., but $3 \not S 4$ and $2 \not S 1$. What does S mean? Think of it as meaning "has the same parity as." Thus $1S1$ reads "1 has the same parity as 1," and $4S2$ reads "4 has the same parity as 2."

Example 11.3 Let $B = \{0,1,2,3,4,5\}$, and consider the following set:

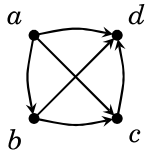
$$U = \{(1,3),(3,3),(5,2),(2,5),(4,2)\} \subseteq B \times B$$

Then U is a relation on B because $U \subseteq B \times B$. You may be hard-pressed to invent any "meaning" for this particular relation. A relation does not have to have any meaning. Any random subset of $B \times B$ is a relation on B, whether or not it describes anything familiar.

Some relations can be described with pictures. For example, we can depict the above relation U on B by drawing points labeled by elements of B. The statement $(x,y) \in U$ is then represented by an arrow pointing from x to y, a graphic symbol meaning *"x relates to y."* Here's a picture of U:



The next picture illustrates the relation R on the set $A = \{a,b,c,d\}$, where xRy means x comes before y in the alphabet. According to Definition 11.1, as a set this relation is $R = \{(a,b),(a,c),(a,d),(b,c),(b,d),(c,d)\}$. You may feel that the picture conveys the relation better than the set does. They are two different ways of expressing the same thing. In some instances pictures are more convenient than sets for discussing relations.



Although pictures can help us visualize relations, they do have their limitations. If A and R were infinite, then the diagram would be impossible to draw, but the set R might be easily expressed in set-builder notation. For instance here is a relation that is too big to be described by a picture.

Example 11.4 Consider the set $R = \{(x,y) \in \mathbb{Z} \times \mathbb{Z} : x - y \in \mathbb{N}\} \subseteq \mathbb{Z} \times \mathbb{Z}$. This is the $>$ relation on the set $A = \mathbb{Z}$. It is infinite because there are infinitely many ways to have $x > y$ where x and y are integers.

Example 11.5 The set $R = \{(x,x) : x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$ is the relation $=$ on the set $\mathbb{R}$, because xRy means the same thing as $x = y$. Thus R is a set that expresses the notion of equality of real numbers.

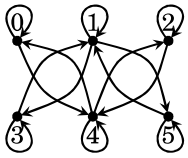
Exercises for Section 11.0

1. Let $A = \{0,1,2,3,4,5\}$. Write out the relation R that expresses $>$ on A. Then illustrate it with a diagram.

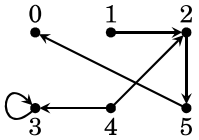
2. Let $A = \{1,2,3,4,5,6\}$. Write out the relation R that expresses $|$ (divides) on A. Then illustrate it with a diagram.

3. Let $A = \{0,1,2,3,4,5\}$. Write out the relation R that expresses $\geq$ on A. Then illustrate it with a diagram.

4. The following diagram represents a relation R on a set A. Write the sets A and R.



5. The following diagram represents a relation R on a set A. Write the sets A and R.



6. Congruence modulo 5 is a relation on the set $A = \mathbb{Z}$. In this relation xRy means $x \equiv y \pmod 5$. Write out the set R in set-builder notation.

7. Write the relation $<$ on the set $A = \mathbb{Z}$ as a subset R of $\mathbb{Z} \times \mathbb{Z}$. This is an infinite set, so you will have to use set-builder notation.

8. Let $A = \{1,2,3,4,5,6\}$. Observe that $\emptyset \subseteq A \times A$, so $R = \emptyset$ is a relation on A. Draw a diagram for this relation.

9. Let $A = \{1,2,3,4,5,6\}$. How many different relations are there on the set A?

11.1 Properties of Relations

Some relations have special properties that other relations don't have. For example, the relation $\leq$ on $\mathbb{Z}$ has the property that $x \leq x$ for every $x \in \mathbb{Z}$, but the relation $<$ on $\mathbb{Z}$ does not have this property, for $x < x$ is never true. The next definition lays out three particularly significant properties that relations may have.

Definition 11.2 Suppose R is a relation on a set A.

1. Relation R is **reflexive** if xRx for every $x \in A$.
   (That is, R is reflexive if $\forall x \in A, xRx$.)
2. Relation R is **symmetric** if xRy implies yRx for all $x, y \in A$
   (That is, R is symmetric if $\forall x, y \in A,\ xRy \Rightarrow yRx$.)
3. Relation R is **transitive** if whenever xRy and yRz, then also xRz.
   (That is, R is transitive if $\forall x, y, z \in A, \big((xRy) \wedge (yRz)\big) \Rightarrow xRz$.)

To illustrate this, let's consider the set $A = \mathbb{Z}$. Examples of reflexive relations on $\mathbb{Z}$ include $\leq$, $=$, and $|$, for $x \leq x$, $x = x$ and $x | x$ are all true for any $x \in \mathbb{Z}$. On the other hand, $>$, $<$, $\neq$ and $\nmid$ are not reflexive for none of the statements $x < x$, $x > x$, $x \neq x$ and $x \nmid x$ is true.

The relation $\neq$ **is** symmetric, for if $x \neq y$, then surely $y \neq x$ also. Also, the relation $=$ is symmetric because $x = y$ always implies $y = x$.

The relation $\leq$ is **not** symmetric, as $x \leq y$ does not necessarily imply $y \leq x$. For instance $5 \leq 6$ is true but $6 \leq 5$ is false. Notice that $(x \leq y) \Rightarrow (y \leq x)$ *is* true for some x and y (for example, it is true when $x = 2$ and $y = 2$) but still $\leq$ is not symmetric because it is not the case that $(x \leq y) \Rightarrow (y \leq x)$ is true for *all* integers x and y.

The relation $\leq$ is transitive because whenever $x \leq y$ and $y \leq z$, it also is true that $x \leq z$. Likewise $<, \geq, >$ and $=$ are all transitive. Examine the following table and be sure you understand why it is labeled as it is.

Relations on $\mathbb{Z}$:	$<$	$\leq$	$=$	$\mid$	$\nmid$	$\neq$
Reflexive	no	yes	yes	yes	no	no
Symmetric	no	no	yes	no	no	yes
Transitive	yes	yes	yes	yes	no	no

Example 11.6 Here $A = \{b, c, d, e\}$ and R is the following relation on A: $R = \{(b,b),(b,c),(c,b),(c,c),(d,d),(b,d),(d,b),(c,d),(d,c)\}$.

Relation R is **not** reflexive, for although bRb, cRc and dRd, it is **not** true that eRe. For a relation to be reflexive, xRx must be true for *all* $x \in A$.

The relation R **is** symmetric, because whenever we have xRy, it follows that yRx too. Observe that bRc and cRb; bRd and dRb; dRc and cRd. If we took away the ordered pair (c, b) from R, then R would no longer be symmetric.

The relation R is transitive, but it takes some work to check it. We must check that the statement $(xRy \wedge yRz) \Rightarrow xRz$ is true for all $x, y, z \in A$. In other words, we must check that whenever xRy and yRz, then also xRz.

Notice that bRc and cRd and also bRd, so the statement $(bRc \wedge cRd) \Rightarrow bRd$ is true. Likewise, bRd, dRc and also bRc, so $(bRd \wedge dRc) \Rightarrow bRc$ is true, and so on. Moreover, note that $(bRc \wedge cRb) \Rightarrow bRb$ fits the pattern $(xRy \wedge yRz) \Rightarrow xRz$, where $x = b$, $y = c$ and $z = b$; and $(bRc \wedge cRb) \Rightarrow bRb$ is true because $(bRc \wedge cRb)$ and bRb are both true. We emphasize that for R to be transitive, it is necessary that $(xRy \wedge yRz) \Rightarrow xRz$ is true for **all** choices of x, y, z from A. Even if we took $x = b$, $y = e$ and $z = c$, then $(bRe \wedge eRc)$ is false and bRc is true, but the statement $(bRe \wedge eRc) \Rightarrow bRc$ is true. It's not much fun, but going through all the combinations, you can verify that $(xRy \wedge yRz) \Rightarrow xRz$ is true for all choices $x, y, z \in A$. (You should try at least a few of them.)

The relation R from Example 11.6 has a meaning. You can think of xRy as meaning that x and y are both consonants. Thus bRc because b and c are both consonants; but $b \not R e$ because it's not true that b and e are both consonants. Once we look at it this way, it's immediately clear that R has to be transitive. If x and y are both consonants and y and z are both consonants, then surely x and z are both consonants. This illustrates a point that we will see again later in this section: Knowing the meaning of a relation can help us understand it and prove things about it.

Here is a picture of the relation from Example 11.6. Notice that we can immediately spot several properties of R that may not have been so clear from its set description. For instance, we see that R is not reflexive because it lacks a loop at e, hence $e \not R e$.

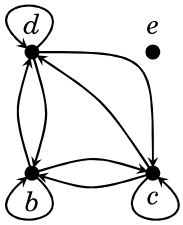
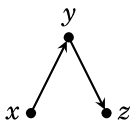
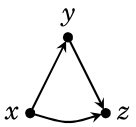


Figure 11.1. The relation R from Example 11.6

In what follows, we summarize how to spot the various properties of a relation from its diagram. Compare these with Figure 11.1.

1. A relation is **reflexive** if for each point x ...  ...there is a loop at x. 

2. A relation is **symmetric** if whenever there is an arrow from x to y ...  ...there is also an arrow from y back to x: 

3. A relation is **transitive** if whenever there are arrows from x to y and y to z ...  ...there is also an arrow from x to z: 

   (This also means that whenever there is an arrow from x to y and from y to x ...  ...there is also a loop from x back to x.) 

Although these visual aids can be illuminating, their use is limited because many relations are too large and complex to be adequately described as diagrams. For example, it would be impossible to draw a diagram for the relation $\equiv \pmod{n}$, where $n \in \mathbb{N}$. Such a relation would best be explained in a more theoretical (and less visual) way.

In the next example we prove that $\equiv \pmod{n}$ is reflexive, symmetric and transitive. Obviously we will not glean this from a drawing. Instead we will prove it from the properties of $\equiv \pmod{n}$ and Definition 11.2. Pay special attention to this example. It illustrates how to **prove** things about relations.

Example 11.7 Prove the following proposition.

Proposition Let $n \in \mathbb{N}$. The relation $\equiv \pmod{n}$ on the set $\mathbb{Z}$ is reflexive, symmetric and transitive.

Proof. First we will show that $\equiv \pmod{n}$ is reflexive. Take any integer $x \in \mathbb{Z}$ and observe that $n \mid 0$, so $n \mid (x-x)$. By definition of congruence modulo n, we have $x \equiv x \pmod{n}$. This shows $x \equiv x \pmod{n}$ for every $x \in \mathbb{Z}$, so $\equiv \pmod{n}$ is reflexive.

Next, we will show that $\equiv \pmod{n}$ is symmetric. For this, we must show that for all $x, y \in \mathbb{Z}$, the condition $x \equiv y \pmod{n}$ implies that $y \equiv x \pmod{n}$. We will use direct proof. Suppose $x \equiv y \pmod{n}$. Thus $n \mid (x-y)$ by definition of congruence modulo n. Then $x - y = na$ for some $a \in \mathbb{Z}$ by definition of divisibility. Multiplying both sides by -1 gives $y - x = n(-a)$. Therefore $n \mid (y-x)$, and this means $y \equiv x \pmod{n}$. We've shown that $x \equiv y \pmod{n}$ implies that $y \equiv x \pmod{n}$, and this means $\equiv \pmod{n}$ is symmetric.

Finally we will show that $\equiv \pmod{n}$ is transitive. For this we must show that if $x \equiv y \pmod{n}$ and $y \equiv z \pmod{n}$, then $x \equiv z \pmod{n}$. Again we use direct proof. Suppose $x \equiv y \pmod{n}$ and $y \equiv z \pmod{n}$. This means $n \mid (x-y)$ and $n \mid (y-z)$. Therefore there are integers a and b for which $x - y = na$ and $y - z = nb$. Adding these two equations, we obtain $x - z = na + nb$. Consequently, $x - z = n(a+b)$, so $n \mid (x-z)$, hence $x \equiv z \pmod{n}$. This completes the proof that $\equiv \pmod{n}$ is transitive.

The past three paragraphs have shown that $\equiv \pmod{n}$ is reflexive, symmetric and transitive, so the proof is complete. ■

As you continue your mathematical education, you will find that the reflexive, symmetric and transitive properties take on special significance in a variety of settings. In preparation for this, the next section explores further consequences of these relations. But first you should work some of the following exercises.

Exercises for Section 11.1

1. Consider the relation $R = \{(a,a),(b,b),(c,c),(d,d),(a,b),(b,a)\}$ on set $A = \{a,b,c,d\}$. Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why.

2. Consider the relation $R = \{(a,b),(a,c),(c,c),(b,b),(c,b),(b,c)\}$ on the set $A = \{a,b,c\}$. Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why.

3. Consider the relation $R = \{(a,b),(a,c),(c,b),(b,c)\}$ on the set $A = \{a,b,c\}$. Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why.

4. Let $A = \{a,b,c,d\}$. Suppose R is the relation

$$R = \{(a,a),(b,b),(c,c),(d,d),(a,b),(b,a),(a,c),(c,a),\\ (a,d),(d,a),(b,c),(c,b),(b,d),(d,b),(c,d),(d,c)\}.$$

Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why.

5. Consider the relation $R = \{(0,0),(\sqrt{2},0),(0,\sqrt{2}),(\sqrt{2},\sqrt{2})\}$ on $\mathbb{R}$. Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why.

6. Consider the relation $R = \{(x,x) : x \in \mathbb{Z}\}$ on $\mathbb{Z}$. Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why. What familiar relation is this?

7. There are 16 possible different relations R on the set $A = \{a,b\}$. Describe all of them. (A picture for each one will suffice, but don't forget to label the nodes.)

8. Define a relation on $\mathbb{Z}$ as xRy if and only if $|x-y| < 1$. Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why. What familiar relation is this?

9. Define a relation on $\mathbb{Z}$ by declaring xRy if and only if x and y have the same parity. Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why. What familiar relation is this?

10. Suppose $A \neq \emptyset$. Since $\emptyset \subseteq A \times A$, the set $R = \emptyset$ is a relation on A. Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why.

11. Suppose $A = \{a,b,c,d\}$ and $R = \{(a,a),(b,b),(c,c),(d,d)\}$. Say whether R is reflexive, symmetric and transitive. If a property does not hold, say why.

12. Prove that the relation $|$ (divides) on the set $\mathbb{Z}$ is reflexive and transitive. (Use Example 11.7 as a guide if you are unsure of how to proceed.)

13. Consider the relation $R = \{(x,y) \in \mathbb{R} \times \mathbb{R} : x - y \in \mathbb{Z}\}$ on $\mathbb{R}$. Prove that this relation is reflexive, symmetric and transitive.

14. Suppose R is a symmetric and transitive relation on a set A, and there is an element $a \in A$ for which aRx for every $x \in A$. Prove that R is reflexive.

15. Prove or disprove: If a relation is symmetric and transitive, then it is also reflexive.

16. Define a relation R on $\mathbb{Z}$ by declaring that xRy if and only if $x^2 \equiv y^2 \pmod 4$. Prove that R is reflexive, symmetric and transitive.

11.2 Equivalence Relations

The relation $=$ on the set $\mathbb{Z}$ (or on any set A) is reflexive, symmetric and transitive. There are many other relations that are also reflexive, symmetric and transitive. Relations which have all three of these properties occur very frequently in mathematics and often play quite significant roles. (For instance, this is certainly true of the relation $=$.) Such relations are given a special name. They are called *equivalence relations*.

Definition 11.3 A relation R on a set A is an **equivalence relation** if it is reflexive, symmetric and transitive.

As an example, Figure 11.2 illustrates four different equivalence relations R_1, R_2, R_3 and R_4 on the set $A = \{-1, 1, 2, 3, 4\}$. Each one has its own meaning, as labeled. For example, in the second row the relation R_2 literally means *"has the same parity as."* So $1 R_2 3$ means "1 *has the same parity as* 3," etc.

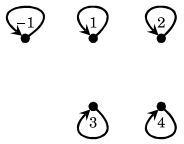
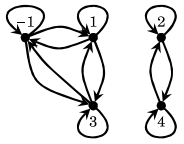
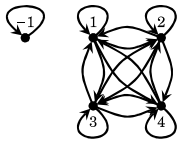
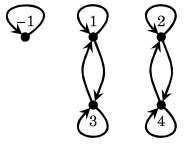
Relation R	Diagram	Equivalence classes (see next page)
"is equal to" (=) $R_1 = \{(-1,-1),(1,1),(2,2),(3,3),(4,4)\}$		$\{-1\}$, $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$
"has same parity as" $R_2 = \{(-1,-1),(1,1),(2,2),(3,3),(4,4),$ $(-1,1),(1,-1),(-1,3),(3,-1),$ $(1,3),(3,1),(2,4),(4,2)\}$		$\{-1,1,3\}$, $\{2,4\}$
"has same sign as" $R_3 = \{(-1,-1),(1,1),(2,2),(3,3),(4,4),$ $(1,2),(2,1),(1,3),(3,1),(1,4),(4,1),$ $(2,3),(3,2),(2,4),(4,2),(1,3),(3,1)\}$		$\{-1\}$, $\{1,2,3,4\}$
"has same parity and sign as" $R_4 = \{(-1,-1),(1,1),(2,2),(3,3),(4,4),$ $(1,3),(3,1),(2,4),(4,2)\}$		$\{-1\}$, $\{1,3\}$, $\{2,4\}$

Figure 11.2. Examples of equivalence relations on the set $A = \{-1, 1, 2, 3, 4\}$

The diagrams in Figure 11.2 make it easy to check that each relation is reflexive, symmetric and transitive, i.e. that each is an equivalence relation. As you can see from these examples, equivalence relations on a set tend to express some measure of "sameness" among the elements of the set, whether it is true equality or something weaker (like having the same parity).

It's time to introduce an important definition. Whenever you have an equivalence relation R on a set A, it divides A into subsets called *equivalence classes*. Here is the definition.

Definition 11.4 Suppose R is an equivalence relation on a set A. Given any element $a \in A$, the **equivalence class containing** a is the subset $\{x \in A : xRa\}$ of A consisting of all the elements of A that relate to a. This set is denoted as $[a]$. Thus the equivalence class containing a is the set $[a] = \{x \in A : xRa\}$.

Example 11.8 Consider the relation R_1 in Figure 11.2. The equivalence class containing 2 is the set $[2] = \{x \in A : xR_1 2\}$. Since in this relation 2 relates to itself and nothing else, we have $[2] = \{2\}$. Other equivalence classes for R_1 are $[-1] = \{-1\}$, $[1] = \{1\}$, $[3] = \{3\}$ and $[4] = \{4\}$. Thus this relation has five separate equivalence classes.

Example 11.9 Consider the relation R_2 in Figure 11.2. The equivalence class containing 2 is the set $[2] = \{x \in A : xR_2 2\}$. Since 2 relates only to itself and 4, we have $[2] = \{2,4\}$. Observe that we also have $[4] = \{x \in A : xR_2 4\} = \{2,4\}$, so $[2] = [4]$. Another equivalence class for R_2 is $[1] = \{x \in A : xR_2 1\} = \{-1,1,3\}$. In addition, note that $[1] = [-1] = [3] = \{-1,1,3\}$. Thus this relation has just two equivalence classes.

Example 11.10 The relation R_4 in Figure 11.2 has three equivalence classes. They are $[-1] = \{-1\}$, and $[1] = [3] = \{1,3\}$, and $[2] = [4] = \{2,4\}$.

Don't be misled by Figure 11.2. It's important to realize that not every equivalence relation can be drawn as a diagram involving nodes and arrows. Even the simple relation $R = \{(x,x) : x \in \mathbb{R}\}$ which expresses equality in the set $\mathbb{R}$ is too big to be drawn. Its picture would involve a point for every real number, and a loop at each point. Clearly that's too many points and loops to draw.

We close this section with several other examples of equivalence relations on infinite sets.

Example 11.11 Let P be the set of all polynomials. Define a relation R on P as follows. Given two polynomials $f(x), g(x) \in P$, let $f(x) R\, g(x)$ mean that $f(x)$ and $g(x)$ have the same degree. Thus, for example $(x^2+3x-4) R\, (3x^2-2)$, and $(x^3+3x^2-4) \not R\, (3x^2-2)$. It takes just a quick mental check to see that R is an equivalence relation. (Do it.) It's easy to describe the equivalence classes of R. For example $[3x^2+2]$ is the set of all polynomials that have the same degree as $3x^2+2$, that is the set of all polynomials of degree 2. We can write this as $[3x^2+2] = \{ax^2+bx+c : a,b,c \in \mathbb{R}, a \neq 0\}$.

Recall that in Example 11.7 we proved that for a given $n \in \mathbb{N}$ the relation $\equiv \pmod{n}$ is reflexive, symmetric and transitive. Thus, in our new parlance, $\equiv \pmod{n}$ is an equivalence relation on $\mathbb{Z}$. Consider the case $n = 3$. Let's find the equivalence classes of the equivalence relation $\equiv \pmod{3}$. The equivalence class containing 0 seems like a reasonable place to start. Observe that

$$\begin{aligned}[0] &= \{x \in \mathbb{Z} : x \equiv 0 \pmod{3}\} \\ &= \{x \in \mathbb{Z} : 3 \,|\, (x-0)\} = \{x \in \mathbb{Z} : 3 \,|\, x\} = \{\ldots, -3, 0, 3, 6, 9, \ldots\}.\end{aligned}$$

Thus the class $[0]$ consists of all the multiples of 3. (Or, said differently, $[0]$ consists of all integers that have a remainder of 0 when divided by 3). Note that $[0] = [3] = [6] = [9]$, etc. The number 1 does not show up in the set $[0]$ so let's next look at the equivalence class $[1]$:

$$[1] = \{x \in \mathbb{Z} : x \equiv 1 \pmod{3}\} = \{x \in \mathbb{Z} : 3 \,|\, (x-1)\} = \{\ldots, -5, -2, 1, 4, 7, 10, \ldots\}$$

The equivalence class $[1]$ consists of all integers that give a remainder of 1 when divided by 3. The number 2 is in neither of the sets $[0]$ or $[1]$, so we next look at the equivalence class $[2]$.

$$[2] = \{x \in \mathbb{Z} : x \equiv 2 \pmod{3}\} = \{x \in \mathbb{Z} : 3 \,|\, (x-2)\} = \{\ldots, -4, -1, 2, 5, 8, 11, \ldots\}$$

The equivalence class $[2]$ consists of all integers that give a remainder of 2 when divided by 3. Observe that any integer is in one of the sets $[0]$, $[1]$ or $[2]$, so we have listed all of the equivalence classes. Thus $\equiv \pmod{3}$ has exactly three equivalence classes, as described above.

Similarly, you can show that the equivalence relation $\equiv \pmod{n}$ has n equivalence classes $[0], [1], [2], \ldots, [n-1]$.

Exercises for Section 11.2

1. Let $A = \{1, 2, 3, 4, 5, 6\}$, and consider the following equivalence relation on A:

$$\begin{aligned}R = \{&(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),(2,3), \\ &(3,2),(4,5),(5,4),(4,6),(6,4),(5,6),(6,5)\}.\end{aligned}$$

List the equivalence classes of R.

2. Let $A = \{a, b, c, d, e\}$. Suppose R is an equivalence relation on A. Suppose R has two equivalence classes. Also aRd, bRc and eRd. Write out R as a set.

3. Let $A = \{a,b,c,d,e\}$. Suppose R is an equivalence relation on A. Suppose R has three equivalence classes. Also aRd and bRc. Write out R as a set.

4. Let $A = \{a,b,c,d,e\}$. Suppose R is an equivalence relation on A. Suppose also that aRd and bRc, eRa and cRe. How many equivalence classes does R have?

5. There are two different equivalence relations on the set $A = \{a,b\}$. Describe them. Diagrams will suffice.

6. There are five different equivalence relations on the set $A = \{a,b,c\}$. Describe them all. Diagrams will suffice.

7. Define a relation R on $\mathbb{Z}$ as xRy if and only if $3x-5y$ is even. Prove R is an equivalence relation. Describe its equivalence classes.

8. Define a relation R on $\mathbb{Z}$ as xRy if and only if x^2+y^2 is even. Prove R is an equivalence relation. Describe its equivalence classes.

9. Define a relation R on $\mathbb{Z}$ as xRy if and only if $4 \mid (x+3y)$. Prove R is an equivalence relation. Describe its equivalence classes.

10. Suppose R and S are two equivalence relations on a set A. Prove that $R \cap S$ is also an equivalence relation. (For an example of this, look at Figure 11.2. Observe that for the equivalence relations R_2,R_3 and R_4, we have $R_2 \cap R_3 = R_4$.)

11. Prove or disprove: If R is an equivalence relation on an infinite set A, then R has infinitely many equivalence classes.

12. Prove or disprove: If R and S are two equivalence relations on a set A, then $R \cup S$ is also an equivalence relation on A.

11.3 Equivalence Classes and Partitions

This section collects several properties of equivalence classes.

Our first result proves that $[a] = [b]$ if and only if aRb. This is useful because it assures us that whenever we are in a situation where $[a] = [b]$, we also have aRb, and vice versa. Being able to switch back and forth between these two pieces of information can be helpful in a variety of situations, and you may find yourself using this result a lot. Be sure to notice that the proof uses all three properties (reflexive, symmetric and transitive) of equivalence relations. Notice also that we have to use some Chapter 8 techniques in dealing with the sets $[a]$ and $[b]$.

Theorem 11.1 Suppose R is an equivalence relation on a set A. Suppose also that $a,b \in A$. Then $[a] = [b]$ if and only if aRb.

Proof. Suppose $[a] = [b]$. Note that $a \in [a]$, because $[a] = \{x \in A : xRa\}$ and aRa by the reflexive property of R. But since $[a] = [b]$, we also have

$a \in [b] = \{x \in A : xRb\}$. Then since a belongs to the set $\{x \in A : xRb\}$, it follows that aRb. This completes the first part of the if-and-only-if proof.

Conversely, suppose aRb. We need to show $[a] = [b]$. This will be accomplished by showing $[a] \subseteq [b]$ and $[b] \subseteq [a]$.

First we show $[a] \subseteq [b]$. Suppose $c \in [a]$. By definition of $[a]$ (recall $[a] = \{x \in A : xRa\}$), it follows that cRa. Now we have cRa and aRb, and since R is transitive we obtain cRb. But cRb implies $c \in \{x \in A : xRb\} = [b]$. This demonstrates that $c \in [a]$ implies $c \in [b]$, so $[a] \subseteq [b]$.

Next we show $[b] \subseteq [a]$. Suppose $c \in [b]$. By definition of $[b]$ (recall $[b] = \{x \in A : xRb\}$), it follows that cRb. Remember that we are assuming aRb, and since R is symmetric, it follows that bRa. Now we have cRb and bRa, and since R is transitive we obtain cRa. But cRa implies $c \in \{x \in A : xRa\} = [a]$. This demonstrates that $c \in [b]$ implies $c \in [a]$, so $[b] \subseteq [a]$.

The previous two paragraphs imply that $[a] = [b]$. ■

To illustrate Theorem 11.1, recall how we worked out the equivalence classes of $\equiv \pmod 3$ at the end of Section 11.2. We observed that $[-3] = [9] = \{\ldots, -3, 0, 3, 6, 9, \ldots\}$. Note that $[-3] = [9]$ and $-3 \equiv 9 \pmod 3$, just as Theorem 11.1 predicts. The theorem assures us that this will work for any equivalence relation. In this course and beyond you may find yourself using the result of Theorem 11.1 quite often. Over time it may become natural and familiar, and you will use it automatically, without even thinking of it as a theorem.

Our next topic addresses the fact that an equivalence relation on a set A divides A into various equivalence classes. There is a special word for this kind of situation. We address it now, as you are likely to encounter it in subsequent mathematics classes.

Definition 11.5 A **partition** of a set A is a set of subsets of A, such that the union of all the subsets equals A, and the intersection of any two different subsets is $\emptyset$.

Example 11.12 Let $A = \{a, b, c, d\}$. One partition of A is $\{\{a, b\}, \{c\}, \{d\}\}$. This is a set of three subsets $\{a, b\}$, $\{c\}$ and $\{d\}$ of A. The union of the three subsets equals A; the intersection of any two subsets is $\emptyset$.

Other partitions of A are

$$\{\{a, b\}, \{c, d\}\}, \quad \{\{a, c\}, \{b\}, \{d\}\}, \quad \{\{a\}, \{b\}, \{c\}\{d\}\}, \quad \text{and} \quad \{\{a, b, c, d\}\},$$

to name a few. Intuitively, a partition is just a dividing up of A into pieces.

Example 11.13 Consider the equivalence relations in Figure 11.2. Each of these is a relation on the set $A = \{-1, 1, 2, 3, 4\}$. The equivalence classes of each relation are listed on the right side of the figure. Observe that, in each case, the set of equivalence classes forms a partition of A. For example, the relation R_1 yields the partition $\{\{-1\}, \{1\}, \{2\}, \{3\}, \{4\}\}$ of A. Likewise the equivalence classes of R_2 form the partition $\{\{-1, 1, 3\}, \{2, 4\}\}$.

Example 11.14 Recall that we worked out the equivalence classes of the equivalence relation $\equiv \pmod 3$ on the set $\mathbb{Z}$. These equivalence classes give the following partition of $\mathbb{Z}$: $\{\{\ldots, -3, 0, 3, 6, 9, \ldots\}, \{\ldots, -2, 1, 4, 7, 10, \ldots\}, \{\ldots, -1, 2, 5, 8, 11, \ldots\}\}$. We can write it more compactly as $\{[0], [1], [2]\}$.

Our examples and experience suggest that the equivalence classes of an equivalence relation on a set form a partition of that set. This is indeed the case, and we now prove it.

Theorem 11.2 Suppose R is an equivalence relation on a set A. Then the set $\{[a] : a \in A\}$ of equivalence classes of R forms a partition of A.

Proof. To show that $\{[a] : a \in A\}$ is a partition of A we need to show two things: We need to show that the union of all the sets $[a]$ equals A, and we need to show that if $[a] \neq [b]$, then $[a] \cap [b] = \emptyset$.

Notationally, the union of all the sets $[a]$ is $\bigcup_{a \in A} [a]$, so we need to prove $\bigcup_{a \in A} [a] = A$. Suppose $x \in \bigcup_{a \in A} [a]$. This means $x \in [a]$ for some $a \in A$. Since $[a] \subseteq A$, it then follows that $x \in A$. Thus $\bigcup_{a \in A} [a] \subseteq A$. On the other hand, if $x \in A$, then $x \in [x]$, so $x \in [a]$ for some $a \in A$. Therefore $x \in \bigcup_{a \in A} [a]$, and this shows $A \subseteq \bigcup_{a \in A} [a]$. Since $\bigcup_{a \in A} [a] \subseteq A$ and $A \subseteq \bigcup_{a \in A} [a]$, it follows that $\bigcup_{a \in A} [a] = A$.

Next we need to show that if $[a] \neq [b]$ then $[a] \cap [b] = \emptyset$. Let's use contrapositive proof. Suppose it's not the case that $[a] \cap [b] = \emptyset$, so there is some element c with $c \in [a] \cap [b]$. Thus $c \in [a]$ and $c \in [b]$. Now, $c \in [a]$ means cRa, and then aRc since R is symmetric. Also $c \in [b]$ means cRb. Now we have aRc and cRb, so aRb (because R is transitive). Theorem 11.1 now implies $[a] = [b]$, so $[a] \neq [b]$ is not true. ■

Theorem 11.2 says the equivalence classes of any equivalence relation on a set A form a partition of A. Conversely, any partition of A describes an equivalence relation R where xRy if and only if x and y belong to the same set in the partition. Thus equivalence relations and partitions are really just two different ways of looking at the same thing.

Exercises for Section 11.3

1. List all the partitions of the set $A = \{a, b\}$. Compare your answer to the answer to Exercise 5 of Section 11.2.

2. List all the partitions of the set $A = \{a, b, c\}$. Compare your answer to the answer to Exercise 6 of Section 11.2.

3. Describe the partition of $\mathbb{Z}$ resulting from the equivalence relation $\equiv \pmod 4$.

11.4 The Integers Modulo n

Example 11.7 proved that for a given $n \in \mathbb{N}$, the relation $\equiv \pmod n$ is reflexive, symmetric and transitive, so it is an equivalence relation. This is a particularly significant equivalence relation in mathematics, and in the present section we deduce some of its properties.

To make matters simpler, let's pick a concrete n, say $n = 5$. Let's begin by looking at the equivalence classes of the relation $\equiv \pmod 5$. There are five equivalence classes, as follows.

$$\begin{array}{rcccccl}
[0] &=& \{x \in \mathbb{Z} : x \equiv 0 \pmod 5)\} &=& \{x \in \mathbb{Z} : 5 | (x-0)\} &=& \{\ldots, -10, -5, 0, 5, 10, 15, \ldots\} \\
[1] &=& \{x \in \mathbb{Z} : x \equiv 1 \pmod 5)\} &=& \{x \in \mathbb{Z} : 5 | (x-1)\} &=& \{\ldots, -9, -4, 1, 6, 11, 16, \ldots\} \\
[2] &=& \{x \in \mathbb{Z} : x \equiv 2 \pmod 5)\} &=& \{x \in \mathbb{Z} : 5 | (x-2)\} &=& \{\ldots, -8, -3, 2, 7, 12, 17, \ldots\} \\
[3] &=& \{x \in \mathbb{Z} : x \equiv 3 \pmod 5)\} &=& \{x \in \mathbb{Z} : 5 | (x-3)\} &=& \{\ldots, -7, -2, 3, 8, 13, 18, \ldots\} \\
[4] &=& \{x \in \mathbb{Z} : x \equiv 4 \pmod 5)\} &=& \{x \in \mathbb{Z} : 5 | (x-4)\} &=& \{\ldots, -6, -1, 4, 9, 14, 19, \ldots\}
\end{array}$$

Notice how these equivalence classes form a partition of the set $\mathbb{Z}$. We label the five equivalence classes as $[0], [1], [2], [3]$, and $[4]$, but you know of course that there are other ways to label them. For example, $[0] = [5] = [10] = [15]$, and so on; and $[1] = [6] = [-4]$, etc. Still, for this discussion we denote the five classes as $[0], [1], [2], [3]$, and $[4]$.

These five classes form a set, which we shall denote as $\mathbb{Z}_5$. Thus

$$\mathbb{Z}_5 = \{[0], [1], [2], [3], [4]\}$$

is a set of five sets. The interesting thing about $\mathbb{Z}_5$ is that even though its elements are sets (and not numbers) it is possible to add and multiply them. In fact, we can define the following rules that tell how elements of

$\mathbb{Z}_5$ can be added and multiplied.

$$[a]+[b] = [a+b]$$

$$[a]\cdot[b] = [a\cdot b]$$

For example, $[2]+[1]=[2+1]=[3]$, and $[2]\cdot[2]=[2\cdot 2]=[4]$. We stress that in doing this we are adding and multiplying *sets* (more precisely equivalence classes), not numbers. We added (or multiplied) two elements of $\mathbb{Z}_5$ and obtained another element of $\mathbb{Z}_5$.

Here is a trickier example. Observe that $[2]+[3]=[5]$. This time we added elements $[2],[3]\in\mathbb{Z}_5$, and got the element $[5]\in\mathbb{Z}_5$. That was easy, except where is our answer $[5]$ in the set $\mathbb{Z}_5=\{[0],[1],[2],[3],[4]\}$? Since $[5]=[0]$, it is more appropriate to write $[2]+[3]=[0]$.

In a similar vein, $[2]\cdot[3]=[6]$ would be written as $[2]\cdot[3]=[1]$ because $[6]=[1]$. Test your skill with this by verifying the following addition and multiplication tables for $\mathbb{Z}_5$.

+	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[1]	[2]	[3]	[4]
[1]	[1]	[2]	[3]	[4]	[0]
[2]	[2]	[3]	[4]	[0]	[1]
[3]	[3]	[4]	[0]	[1]	[2]
[4]	[4]	[0]	[1]	[2]	[3]

·	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]
[2]	[0]	[2]	[4]	[1]	[3]
[3]	[0]	[3]	[1]	[4]	[2]
[4]	[0]	[4]	[3]	[2]	[1]

We call the set $\mathbb{Z}_5=\{[0],[1],[2],[3],[4]\}$ the **integers modulo 5**. As our tables suggest, $\mathbb{Z}_5$ is more than just a set: It is a little number system with its own addition and multiplication. In this way it is like the familiar set $\mathbb{Z}$ which also comes equipped with an addition and a multiplication.

Of course, there is nothing special about the number 5. We can also define $\mathbb{Z}_n$ for any natural number n. Here is the definition.

Definition 11.6 Let $n\in\mathbb{N}$. The equivalence classes of the equivalence relation $\equiv \pmod n$ are $[0],[1],[2],\ldots,[n-1]$. The **integers modulo n** is the set $\mathbb{Z}_n=\{[0],[1],[2],\ldots,[n-1]\}$. Elements of $\mathbb{Z}_n$ can be added by the rule $[a]+[b]=[a+b]$ and multiplied by the rule $[a]\cdot[b]=[ab]$.

Given a natural number n, the set $\mathbb{Z}_n$ is a number system containing n elements. It has many of the algebraic properties that $\mathbb{Z},\mathbb{R}$ and $\mathbb{Q}$ possess.

For example, it is probably obvious to you already that elements of $\mathbb{Z}_n$ obey the commutative laws $[a]+[b]=[b]+[a]$ and $[a]\cdot[b]=[b]\cdot[a]$. You can also verify that the distributive law $[a]\cdot([b]+[c])=[a]\cdot[b]+[a]\cdot[c]$ holds, as follows.

$$\begin{aligned}
[a]\cdot([b]+[c]) &= [a]\cdot[b+c] \\
&= [a(b+c)] \\
&= [ab+ac] \\
&= [ab]+[ac] \\
&= [a]\cdot[b]+[a]\cdot[c]
\end{aligned}$$

The integers modulo n are significant because they more closely fit certain applications than do other number systems such as $\mathbb{Z}$ or $\mathbb{R}$. If you go on to take a course in abstract algebra, then you will work extensively with $\mathbb{Z}_n$ as well as other, more exotic, number systems. (In such a course you will also use all of the proof techniques that we have discussed, as well as the ideas of equivalence relations.)

To close this section we take up an issue that may have bothered you earlier. It has to do with our definitions of addition $[a]+[b]=[a+b]$ and multiplication $[a]\cdot[b]=[ab]$. These definitions define addition and multiplication of equivalence classes in terms of representatives a and b in the equivalence classes. Since there are many different ways to choose such representatives, we may well wonder if addition and multiplication are consistently defined. For example, suppose two people, Alice and Bob, want to multiply the elements [2] and [3] in $\mathbb{Z}_5$. Alice does the calculation as $[2]\cdot[3]=[6]=[1]$, so her final answer is [1]. Bob does it differently. Since $[2]=[7]$ and $[3]=[8]$, he works out $[2]\cdot[3]$ as $[7]\cdot[8]=[56]$. Since $56\equiv 1 \pmod 5$, Bob's answer is $[56]=[1]$, and that agrees with Alice's answer. Will their answers always agree or did they just get lucky (with the arithmetic)?

The fact is that no matter how they do the multiplication in $\mathbb{Z}_n$, their answers will agree. To see why, suppose Alice and Bob want to multiply the elements $[a],[b]\in\mathbb{Z}_n$, and suppose $[a]=[a']$ and $[b]=[b']$. Alice and Bob do the multiplication as follows.

$$\begin{aligned}
&\text{Alice:} \quad [a]\cdot[b]=[ab] \\
&\text{Bob:} \quad [a']\cdot[b']=[a'b']
\end{aligned}$$

We need to show that their answers agree, that is we need to show $[ab]=[a'b']$. Since $[a]=[a']$, we know by Theorem 11.1 that $a\equiv a' \pmod n$.

Thus $n|(a-a')$, so $a-a'=nk$ for some integer k. Likewise, as $[b]=[b']$, we know $b\equiv b' \pmod{n}$, or $n|(b-b')$, so $b-b'=n\ell$ for some integer ℓ. Thus we get $a=a'+nk$ and $b=b'+n\ell$. Therefore:

$$\begin{aligned} ab &= (a'+nk)(b'+n\ell) \\ ab &= a'b'+a'n\ell+nkb'+n^2k\ell \\ ab-a'b' &= n(a'\ell+kb'+nk\ell) \end{aligned}$$

This shows $n|(ab-a'b')$, so $ab\equiv a'b' \pmod{n}$, and from that we conclude $[ab]=[a'b']$. Consequently Alice and Bob really do get the same answer, so we can be assured that the definition of multiplication in $\mathbb{Z}_n$ is consistent.

In one of the exercises, you will be asked to show that addition in $\mathbb{Z}_n$ is similarly consistent.

Exercises for Section 11.4

1. Write the addition and multiplication tables for $\mathbb{Z}_2$.

2. Write the addition and multiplication tables for $\mathbb{Z}_3$.

3. Write the addition and multiplication tables for $\mathbb{Z}_4$.

4. Write the addition and multiplication tables for $\mathbb{Z}_6$.

5. Suppose $[a],[b]\in\mathbb{Z}_5$ and $[a]\cdot[b]=[0]$. Is it necessarily true that either $[a]=[0]$ or $[b]=[0]$?

6. Suppose $[a],[b]\in\mathbb{Z}_6$ and $[a]\cdot[b]=[0]$. Is it necessarily true that either $[a]=[0]$ or $[b]=[0]$?

7. Do the following calculations in $\mathbb{Z}_9$, in each case expressing your answer as $[a]$ with $0\le a\le 8$.
(a) $[8]+[8]$ **(b)** $[24]+[11]$ **(c)** $[21]\cdot[15]$ **(d)** $[8]\cdot[8]$

8. Suppose $[a],[b]\in\mathbb{Z}_n$, and $[a]=[a']$ and $[b]=[b']$. Alice adds $[a]$ and $[b]$ as $[a]+[b]=[a+b]$. Bob adds them as $[a']+[b']=[a'+b']$. Show that their answers $[a+b]$ and $[a'+b']$ are the same.

11.5 Relations Between Sets

In the beginning of this chapter, we defined a relation on a set A to be a subset $R\subseteq A\times A$. This created a framework that could model any situation in which elements of A are compared to themselves. In this setting, the statement xRy has elements x and y from A on either side of the R because R compares elements from A. But there are other

relational symbols that don't work this way. Consider $\in$. The statement $5 \in \mathbb{Z}$ expresses a relationship between 5 and $\mathbb{Z}$ (namely that the element 5 is in the set $\mathbb{Z}$) but 5 and $\mathbb{Z}$ are not in any way naturally regarded as both elements of some set A. To overcome this difficulty, we generalize the idea of a relation on A to a *relation from A to B*.

Definition 11.7 A **relation** from a set A to a set B is a subset $R \subseteq A \times B$. We often abbreviate the statement $(x,y) \in R$ as xRy. The statement $(x,y) \notin R$ is abbreviated as $x \not R y$.

Example 11.15 Suppose $A = \{1,2\}$ and $B = \mathscr{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$. Then $R = \{(1,\{1\}),(2,\{2\}),(1,\{1,2\}),(2,\{1,2\})\} \subseteq A \times B$ is a relation from A to B. Notice that we have $1R\{1\}$, $2R\{2\}$, $1R\{1,2\}$ and $2R\{1,2\}$. The relation R is the familiar relation $\in$ for the set A, that is xRX means exactly the same thing as $x \in X$.

Diagrams for relations from A to B differ from diagrams for relations on A. Since there are two sets A and B in a relation from A to B, we have to draw labeled nodes for each of the two sets. Then we draw arrows from x to y whenever xRy. The following figure illustrates this for Example 11.15.

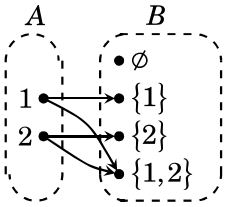


Figure 11.3. A relation from A to B

The ideas from this chapter show that any relation (whether it is a familiar one like $\geq$, $\leq$, $=$, $|$, $\in$ or $\subseteq$, or a more exotic one) is really just a set. Therefore the theory of relations is a part of set theory. In the next chapter we will see that this idea touches on another important mathematical construction, namely functions. We will define a function to be a special kind of relation from one set to another, and in this context we will see that any function is really just a set.

CHAPTER 12

Functions

You know from calculus that functions play a fundamental role in mathematics. You likely view a function as a kind of formula that describes a relationship between two (or more) quantities. You certainly understand and appreciate the fact that relationships between quantities are important in all scientific disciplines, so you do not need to be convinced that functions are important. Still, you may not be aware of the full significance of functions. Functions are more than merely descriptions of numeric relationships. In a more general sense, functions can compare and relate different kinds of mathematical structures. You will see this as your understanding of mathematics deepens. In preparation of this deepening, we will now explore a more general and versatile view of functions.

The concept of a relation between sets (Definition 11.7) plays a big role here, so you may want to quickly review it.

12.1 Functions

Let's start on familiar ground. Consider the function $f(x) = x^2$ from $\mathbb{R}$ to $\mathbb{R}$. Its graph is the set of points $R = \{(x, x^2) : x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$.

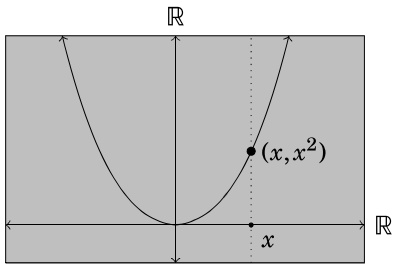


Figure 12.1. A familiar function

Having read Chapter 11, you may see f in a new light. Its graph $R \subseteq \mathbb{R} \times \mathbb{R}$ is a relation on the set $\mathbb{R}$. In fact, as we shall see, functions are just special kinds of relations. Before stating the exact definition, we

look at another example. Consider the function $f(n) = |n| + 2$ that converts integers n into natural numbers $|n| + 2$. Its graph is $R = \{(n, |n| + 2) : n \in \mathbb{Z}\} \subseteq \mathbb{Z} \times \mathbb{N}$.

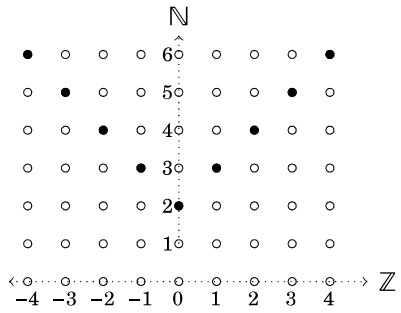


Figure 12.2. The function $f : \mathbb{Z} \to \mathbb{N}$, where $f(n) = |n| + 2$

Figure 12.2 shows the graph R as darkened dots in the grid of points $\mathbb{Z} \times \mathbb{N}$. Notice that in this example R is not a relation on a single set. The set of input values $\mathbb{Z}$ is different from the set $\mathbb{N}$ of output values, so the graph $R \subseteq \mathbb{Z} \times \mathbb{N}$ is *a relation from* $\mathbb{Z}$ *to* $\mathbb{N}$.

This example illustrates three things. First, a function can be viewed as sending elements from one set A to another set B. (In the case of f, $A = \mathbb{Z}$ and $B = \mathbb{N}$.) Second, such a function can be regarded as a relation from A to B. Third, for every input value n, there is *exactly one* output value $f(n)$. In your high school algebra course, this was expressed by the "vertical line test": Any vertical line intersects a function's graph at most once. It means that for any input value x, the graph contains exactly one point of form $(x, f(x))$. Our main definition, given below, incorporates all of these ideas.

Definition 12.1 Suppose A and B are sets. A **function** f **from** A **to** B (denoted as $f : A \to B$) is a relation $f \subseteq A \times B$ from A to B, satisfying the property that for each $a \in A$ the relation f contains exactly one ordered pair of form (a, b). The statement $(a, b) \in f$ is abbreviated $f(a) = b$.

Example 12.1 Consider the function f graphed in Figure 12.2. According to Definition 12.1, we regard f as the set of points in its graph, that is $f = \{(n, |n| + 2) : n \in \mathbb{Z}\} \subseteq \mathbb{Z} \times \mathbb{N}$. This is a relation from $\mathbb{Z}$ to $\mathbb{N}$, and indeed given any $a \in \mathbb{Z}$ the set f contains exactly one ordered pair $(a, |a| + 2)$ whose first coordinate is a. Since $(1, 3) \in f$, we write $f(1) = 3$; and since $(-3, 5) \in f$ we write $f(-3) = 5$, etc. In general, $(a, b) \in f$ means that f sends the input

value a to the output value b, and we express this as $f(a) = b$. This function can be expressed by a formula: for each input value n, the output value is $|n| + 2$, so we may write $f(n) = |n| + 2$. All this agrees with the way we thought of functions in algebra and calculus; the only difference is that now we also think of a function as a relation.

Definition 12.2 For a function $f : A \to B$, the set A is called the **domain** of f. (Think of the domain as the set of possible "input values" for f.) The set B is called the **codomain** of f. The **range** of f is the set $\{f(a) : a \in A\} = \{b : (a,b) \in f\}$. (Think of the range as the set of all possible "output values" for f. Think of the codomain as a sort of "target" for the outputs.)

Continuing Example 12.1, the domain of f is $\mathbb{Z}$ and its codomain is $\mathbb{N}$. Its range is $\{f(a) : a \in \mathbb{Z}\} = \{|a| + 2 : a \in \mathbb{Z}\} = \{2,3,4,5,\ldots\}$. Notice that the range is a subset of the codomain, but it does not (in this case) equal the codomain.

In our examples so far, the domains and codomains are sets of numbers, but this needn't be the case in general, as the next example indicates.

Example 12.2 Let $A = \{p,q,r,s\}$ and $B = \{0,1,2\}$, and

$$f = \{(p,0),(q,1),(r,2),(s,2)\} \subseteq A \times B.$$

This is a function $f : A \to B$ because each element of A occurs exactly once as a first coordinate of an ordered pair in f. We have $f(p) = 0$, $f(q) = 1$, $f(r) = 2$ and $f(s) = 2$. The domain of f is $\{p,q,r,s\}$, and the codomain and range are both $\{0,1,2\}$.

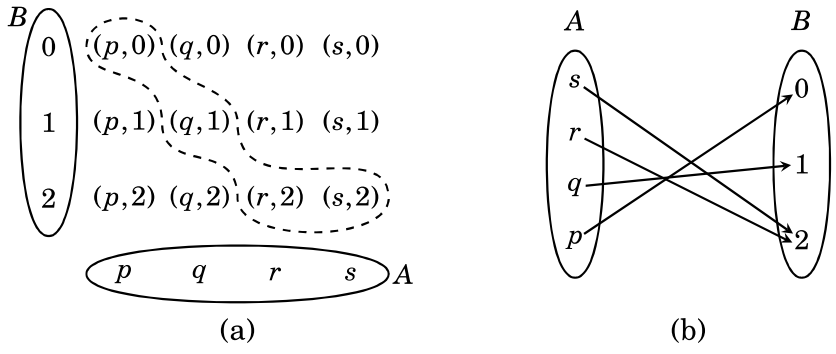


Figure 12.3. Two ways of drawing the function $f = \{(p,0),(q,1),(r,2),(s,2)\}$

If A and B are not both sets of numbers it can be difficult to draw a graph of $f : A \to B$ in the traditional sense. Figure 12.3(a) shows an attempt at a graph of f from Example 12.2. The sets A and B are aligned roughly as x- and y-axes, and the Cartesian product $A \times B$ is filled in accordingly. The subset $f \subseteq A \times B$ is indicated with dashed lines, and this can be regarded as a "graph" of f. A more natural visual description of f is shown in 12.3(b). The sets A and B are drawn side-by-side, and arrows point from a to b whenever $f(a) = b$.

In general, if $f : A \to B$ is the kind of function you may have encountered in algebra or calculus, then conventional graphing techniques offer the best visual description of it. On the other hand, if A and B are finite or if we are thinking of them as generic sets, then describing f with arrows is often a more appropriate way of visualizing it.

We emphasize that, according to Definition 12.1, a function is really just a special kind of set. Any function $f : A \to B$ is a subset of $A \times B$. By contrast, your calculus text probably defined a function as a certain kind of "rule." While that intuitive outlook is adequate for the first few semesters of calculus, it does not hold up well to the rigorous mathematical standards necessary for further progress. The problem is that words like "rule" are too nebulous. Phrasing the definition of a function in the language of sets removes the ambiguity.

Still, in practice we tend to think of functions as rules. Given $f : \mathbb{Z} \to \mathbb{N}$ where $f(x) = |x| + 2$, we think of this as a rule that associates any number $n \in \mathbb{Z}$ to the number $|n| + 2$ in $\mathbb{N}$, rather than a set containing ordered pairs $(n, |n| + 2)$. It is only when we have to understand or interpret the theoretical nature of functions (as we do in this text) that Definition 12.1 comes to bear. The definition is a foundation that gives us license to think about functions in a more informal way.

The next example brings up a point about notation. Consider a function such as $f : \mathbb{Z}^2 \to \mathbb{Z}$ whose domain is a Cartesian product. This function takes as input an ordered pair $(m, n) \in \mathbb{Z}^2$ and sends it to a number $f((m, n)) \in \mathbb{Z}$. To simplify the notation, it is common to write $f(m, n)$ instead of $f((n, m))$, even though this is like writing fx instead of $f(x)$. We also remark that although we've been using the letters f, g and h to denote functions, any other reasonable symbol could be used. Greek letters such as φ and θ are common.

Example 12.3 Say a function $\varphi : \mathbb{Z}^2 \to \mathbb{Z}$ is defined as $\varphi(m, n) = 6m - 9n$. Note that as a set, this function is $\varphi = \{((m, n), 6m - 9n) : (m, n) \in \mathbb{Z}^2\} \subseteq \mathbb{Z}^2 \times \mathbb{Z}$. Find the range of φ.

To answer this, first observe that for any $(m,n) \in \mathbb{Z}^2$, the value $f(m,n) = 6m - 9n = 3(2m - 3n)$ is a multiple of 3. Thus every number in the range is a multiple of 3, so the range is a *subset* of the set of all multiples of 3. On the other hand if $b = 3k$ is a multiple of 3 we have $\varphi(-k,-k) = 6(-k) - 9(-k) = 3k = b$, which means any multiple of 3 is in the range of φ. Therefore the range of φ is the set $\{3k : k \in \mathbb{Z}\}$ of all multiples of 3.

To conclude this section, let's use Definition 12.1 to help us understand what it means for two functions $f : A \to B$ and $g : C \to D$ to be equal. According to our definition, $f \subseteq A \times B$ and $g \subseteq C \times D$. If the two functions are to be equal, we require that the following sets be equal: $f = g$, $A = C$ and $B = D$.

Suppose for example, that $A = \{1,2,3\}$ and $B = \{a,b\}$. The two functions $f = \{(1,a),(2,a),(3,b)\}$ and $g = \{(3,b),(2,a),(1,a)\}$ from A to B are equal because the sets f and g are equal. Observe that the equality $f = g$ means $f(x) = g(x)$ for every $x \in A$. We repackage this idea in the following definition.

Definition 12.3 Two functions $f : A \to B$ and $g : C \to D$ are **equal** if $A = C$, $B = D$ and $f(x) = g(x)$ for every $x \in A$.

According to the definition, to show functions f and g are equal, we just need to confirm that their domains are equal, their codomains are equal, and $f(x) = g(x)$ for every x in the domain. There is a shade of meaning to watch out for here. Consider the functions $f : \mathbb{Z} \to \mathbb{N}$ and $g : \mathbb{Z} \to \mathbb{Z}$ defined as $f(x) = |x| + 2$ and $g(x) = |x| + 2$. Even though their domains are the same and $f(x) = g(x)$ for every x in the domain, they are technically not equal because their codomains differ.

Exercises for Section 12.1

1. Suppose $A = \{0,1,2,3,4\}$, $B = \{2,3,4,5\}$ and $f = \{(0,3),(1,3),(2,4),(3,2),(4,2)\}$. State the domain and range of f. Find $f(2)$ and $f(1)$.

2. Suppose $A = \{a,b,c,d\}$, $B = \{2,3,4,5,6\}$ and $f = \{(a,2),(b,3),(c,4),(d,5)\}$. State the domain and range of f. Find $f(b)$ and $f(d)$.

3. There are four different functions $f : \{a,b\} \to \{0,1\}$. List them all. Diagrams will suffice.

4. There are eight different functions $f : \{a,b,c\} \to \{0,1\}$. List them all. Diagrams will suffice.

5. Give an example of a relation from $\{a,b,c,d\}$ to $\{d,e\}$ that is not a function.

6. Suppose $f : \mathbb{Z} \to \mathbb{Z}$ is defined as $f = \{(x, 4x+5) : x \in \mathbb{Z}\}$. State the domain, codomain and range of f. Find $f(10)$.

7. Consider the set $f = \{(x, y) \in \mathbb{Z} \times \mathbb{Z} : 3x + y = 4\}$. Is this a function from $\mathbb{Z}$ to $\mathbb{Z}$? Explain.

8. Consider the set $f = \{(x, y) \in \mathbb{Z} \times \mathbb{Z} : x + 3y = 4\}$. Is this a function from $\mathbb{Z}$ to $\mathbb{Z}$? Explain.

9. Consider the set $f = \{(x^2, x) : x \in \mathbb{R}\}$. Is this a function from $\mathbb{R}$ to $\mathbb{R}$? Explain.

10. Consider the set $f = \{(x^3, x) : x \in \mathbb{R}\}$. Is this a function from $\mathbb{R}$ to $\mathbb{R}$? Explain.

11. Is the set $\theta = \{(X, |X|) : X \subseteq \mathbb{Z}_5\}$ a function? If so, what is its domain and range?

12. Is the set $\theta = \{((x, y), (3y, 2x, x + y)) : x, y \in \mathbb{R}\}$ a function? If so, what is its domain, codomain and range?

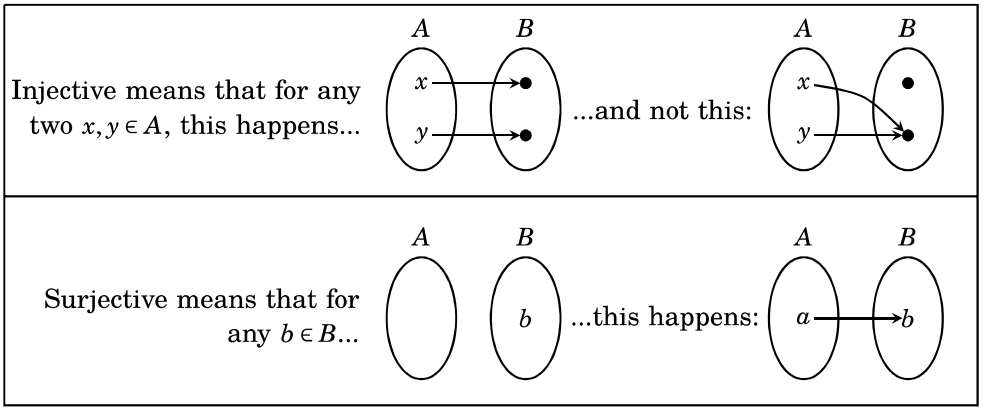
12.2 Injective and Surjective Functions

You may recall from algebra and calculus that a function may be *one-to-one* and *onto*, and these properties are related to whether or not the function is invertible. We now review these important ideas. In advanced mathematics, the word *injective* is often used instead of *one-to-one*, and *surjective* is used instead of *onto*. Here are the exact definitions.

Definition 12.4 A function $f : A \to B$ is:

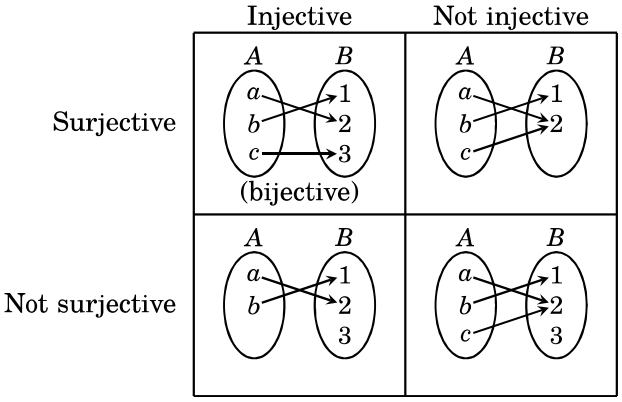
1. **injective** (or one-to-one) if for every $x, y \in A$, $x \neq y$ implies $f(x) \neq f(y)$;
2. **surjective** (or onto) if for every $b \in B$ there is an $a \in A$ with $f(a) = b$;
3. **bijective** if f is both injective and surjective.

Below is a visual description of Definition 12.4. In essence, injective means that unequal elements in A always get sent to unequal elements in B. Surjective means that every element of B has an arrow pointing to it, that is, it equals $f(a)$ for some a in the domain of f.



For more concrete examples, consider the following functions from $\mathbb{R}$ to $\mathbb{R}$. The function $f(x) = x^2$ is not injective because $-2 \neq 2$, but $f(-2) = f(2)$. Nor is it surjective, for if $b = -1$ (or if b is any negative number), then there is no $a \in \mathbb{R}$ with $f(a) = b$. On the other hand, $g(x) = x^3$ is both injective and surjective, so it is also bijective.

There are four possible injective/surjective combinations that a function may posses. This is illustrated in the following figure showing four functions from A to B. Functions in the first column are injective, those in the second column are not injective. Functions in the first row are surjective, those in the second row are not.



We note in passing that, according to the definitions, a function is surjective if and only if its codomain equals its range.

Often it is necessary to prove that a particular function $f : A \to B$ is injective. For this we must prove that for any two elements $x, y \in A$, the conditional statement $(x \neq y) \Rightarrow \big(f(x) \neq f(y)\big)$ is true. The two main approaches for this are summarized below.

How to show a function $f : A \to B$ is injective:

Direct approach:	**Contrapositive approach:**
Suppose $x, y \in A$ and $x \neq y$.	Suppose $x, y \in A$ and $f(x) = f(y)$.
$\vdots$	$\vdots$
Therefore $f(x) \neq f(y)$.	Therefore $x = y$.

Of these two approaches, the contrapositive is often the easiest to use, especially if f is defined by an algebraic formula. This is because the contrapositive approach starts with the *equation* $f(x) = f(y)$ and proceeds

to the *equation* $x = y$. In algebra, as you know, it is usually easier to work with equations than inequalities.

To prove that a function is *not* injective, you must *disprove* the statement $(x \neq y) \Rightarrow \big(f(x) \neq f(y)\big)$. For this it suffices to find example of two elements $x, y \in A$ for which $x \neq y$ and $f(x) = f(y)$.

Next we examine how to prove that $f : A \to B$ is *surjective*. According to Definition 12.4, we must prove the statement $\forall\, b \in B, \exists\, a \in A, f(a) = b$. In words, we must show that for any $b \in B$, there is at least one $a \in A$ (which may depend on b) having the property that $f(a) = b$. Here is an outline.

How to show a function $f : A \to B$ is surjective:

Suppose $b \in B$.
[Prove there exists $a \in A$ for which $f(a) = b$.]

In the second step, we have to prove the existence of an a for which $f(a) = b$. For this, just finding an example of such an a would suffice. (How to find such an example depends on how f is defined. If f is given as a formula, we may be able to find a by solving the equation $f(a) = b$ for a. Sometimes you can find a by just plain common sense.) To show f is *not* surjective, we must prove the negation of $\forall\, b \in B, \exists\, a \in A, f(a) = b$, that is we must prove $\exists\, b \in B, \forall\, a \in A, f(a) \neq b$.

The following examples illustrate these ideas. (For the first example, note that the set $\mathbb{R} - \{0\}$ is $\mathbb{R}$ with the number 0 removed.)

Example 12.4 Show that the function $f : \mathbb{R} - \{0\} \to \mathbb{R}$ defined as $f(x) = \frac{1}{x} + 1$ is injective but not surjective.

We will use the contrapositive approach to show that f is injective. Suppose $x, y \in \mathbb{R} - \{0\}$ and $f(x) = f(y)$. This means $\frac{1}{x} + 1 = \frac{1}{y} + 1$. Subtracting 1 from both sides and inverting produces $x = y$. Therefore f is injective.

Function f is not surjective because there exists an element $b = 1 \in \mathbb{R}$ for which $f(x) = \frac{1}{x} + 1 \neq 1$ for every $x \in \mathbb{R} - \{0\}$.

Example 12.5 Show that the function $g : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ defined by the formula $g(m, n) = (m + n, m + 2n)$, is both injective and surjective.

We will use the contrapositive approach to show that g is injective. Thus we need to show that $g(m, n) = g(k, \ell)$ implies $(m, n) = (k, \ell)$. Suppose $(m, n), (k, \ell) \in \mathbb{Z} \times \mathbb{Z}$ and $g(m, n) = g(k, \ell)$. Then $(m+n, m+2n) = (k+\ell, k+2\ell)$. It follows that $m + n = k + \ell$ and $m + 2n = k + 2\ell$. Subtracting the first equation from the second gives $n = \ell$. Next, subtract $n = \ell$ from $m + n = k + \ell$ to get $m = k$. Since $m = k$ and $n = \ell$, it follows that $(m, n) = (k, \ell)$. Therefore g is injective.

To see that g is surjective, consider an arbitrary element $(b,c) \in \mathbb{Z} \times \mathbb{Z}$. We need to show that there is some $(x,y) \in \mathbb{Z} \times \mathbb{Z}$ for which $g(x,y) = (b,c)$. To find (x,y), note that $g(x,y) = (b,c)$ means $(x+y, x+2y) = (b,c)$. This leads to the following system of equations.

$$\begin{array}{rcrcl} x & + & y & = & b \\ x & + & 2y & = & c \end{array}$$

Solving gives $x = 2b - c$ and $y = c - b$. Then $(x,y) = (2b-c, c-b)$. We now have $g(2b-c, c-b) = (b,c)$, and it follows that g is surjective.

Example 12.6 Consider function $h : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Q}$ defined as $h(m,n) = \dfrac{m}{|n|+1}$. Determine whether this is injective and whether it is surjective.

This function is *not* injective because of the unequal elements $(1,2)$ and $(1,-2)$ in $\mathbb{Z} \times \mathbb{Z}$ for which $h(1,2) = h(1,-2) = \frac{1}{3}$. However, h is surjective: Take any element $b \in \mathbb{Q}$. Then $b = \frac{c}{d}$ for some $c, d \in \mathbb{Z}$. Notice we may assume d is positive by making c negative, if necessary. Then $h(c, d-1) = \frac{c}{|d-1|+1} = \frac{c}{d} = b$.

Exercises for Section 12.2

1. Let $A = \{1,2,3,4\}$ and $B = \{a,b,c\}$. Give an example of a function $f : A \to B$ that is neither injective nor surjective.

2. Consider the logarithm function $\ln : (0,\infty) \to \mathbb{R}$. Decide whether this function is injective and whether it is surjective.

3. Consider the cosine function $\cos : \mathbb{R} \to \mathbb{R}$. Decide whether this function is injective and whether it is surjective. What if it had been defined as $\cos : \mathbb{R} \to [-1,1]$?

4. A function $f : \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ is defined as $f(n) = (2n, n+3)$. Verify whether this function is injective and whether it is surjective.

5. A function $f : \mathbb{Z} \to \mathbb{Z}$ is defined as $f(n) = 2n + 1$. Verify whether this function is injective and whether it is surjective.

6. A function $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ is defined as $f(m,n) = 3n - 4m$. Verify whether this function is injective and whether it is surjective.

7. A function $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ is defined as $f(m,n) = 2n - 4m$. Verify whether this function is injective and whether it is surjective.

8. A function $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ is defined as $f(m,n) = (m+n, 2m+n)$. Verify whether this function is injective and whether it is surjective.

9. Prove that the function $f : \mathbb{R} - \{2\} \to \mathbb{R} - \{5\}$ defined by $f(x) = \dfrac{5x+1}{x-2}$ is bijective.

10. Prove the function $f : \mathbb{R} - \{1\} \to \mathbb{R} - \{1\}$ defined by $f(x) = \left(\dfrac{x+1}{x-1}\right)^3$ is bijective.

11. Consider the function $\theta : \{0,1\} \times \mathbb{N} \to \mathbb{Z}$ defined as $\theta(a,b) = (-1)^a b$. Is θ injective? Is it surjective? Explain.

12. Consider the function $\theta : \{0,1\} \times \mathbb{N} \to \mathbb{Z}$ defined as $\theta(a,b) = a - 2ab + b$. Is θ injective? Is it surjective? Explain.

13. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by the formula $f(x,y) = (xy, x^3)$. Is f injective? Is it surjective?

14. Consider the function $\theta : \mathscr{P}(\mathbb{Z}) \to \mathscr{P}(\mathbb{Z})$ defined as $\theta(X) = \overline{X}$. Is θ injective? Is it surjective?

15. This question concerns functions $f : \{A,B,C,D,E,F,G\} \to \{1,2,3,4,5,6,7\}$. How many such functions are there? How many of these functions are injective? How many are surjective? How many are bijective?

16. This question concerns functions $f : \{A,B,C,D,E\} \to \{1,2,3,4,5,6,7\}$. How many such functions are there? How many of these functions are injective? How many are surjective? How many are bijective?

17. This question concerns functions $f : \{A,B,C,D,E,F,G\} \to \{1,2\}$. How many such functions are there? How many of these functions are injective? How many are surjective? How many are bijective?

12.3 The Pigeonhole Principle

Here is a simple but useful idea. Imagine there is a set A of pigeons and a set B of pigeon holes, and all the pigeons fly into the pigeon holes. You can think of this as describing a function $f : A \to B$, where Pigeon X flies into Pigeon hole $f(X)$. Figure 12.4 illustrates this.

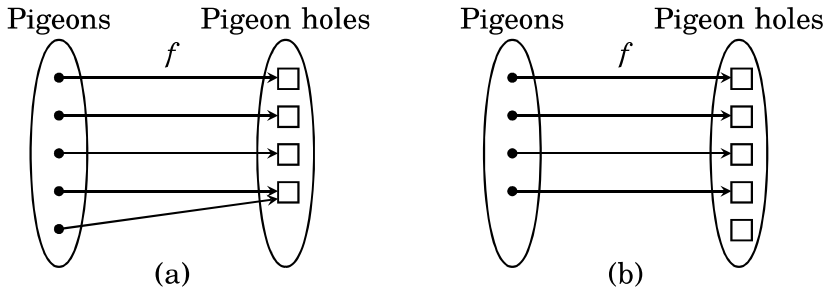


Figure 12.4. The Pigeonhole Principle

In Figure 12.4(a) there are more pigeons than pigeon holes, and it is obvious that in such a case at least two pigeons have to fly into the

same pigeon hole, meaning that f is not injective. In Figure 12.4(b) there are fewer pigeons than pigeon holes, so clearly at least one pigeon hole remains empty, meaning that f is not surjective,

Although the underlying idea expressed by these figures has little to do with pigeons, it is nonetheless called the *Pigeonhole Principle:*

The Pigeonhole Principle Suppose A and B are finite sets and $f : A \to B$ is any function. Then:

1. If $|A| > |B|$, then f is not injective.
2. If $|A| < |B|$, then f is not surjective.

Though the Pigeonhole Principle is obvious, it can be used to prove some things that are not so obvious.

Example 12.7 Prove the following proposition.

Proposition If A is any set of 10 integers between 1 and 100, then there exist subsets $X \subseteq A$ and $Y \subseteq A$ for which the sum of elements in X equals the sum of elements in Y.

To illustrate what this proposition is saying, consider the random set

$$A = \{5, 7, 12, 11, 17, 50, 51, 80, 90, 100\}$$

of 10 integers between 1 and 100. Notice that A has subsets $X = \{5, 80\}$ and $Y = \{7, 11, 17, 50\}$ for which the sum of the elements in X equals the sum of the elements in Y. If we tried to "mess up" A by changing the 5 to a 6, we get

$$A = \{6, 7, 12, 11, 17, 50, 51, 80, 90, 100\}$$

which has subsets $X = \{7, 12, 17, 50\}$ and $Y = \{6, 80\}$ both of whose elements add up to the same number (86). The proposition asserts that this is always possible, no matter what A is. Here is a proof.

Proof. Suppose $A \subseteq \{1, 2, 3, 4, \ldots, 99, 100\}$ and $|A| = 10$, as stated. Notice that if $X \subseteq A$, then X has no more than 10 elements, each between 1 and 100, and therefore the sum of all the elements of X is less than $100 \cdot 10 = 1000$. Consider the function

$$f : \mathscr{P}(A) \to \{0, 1, 2, 3, 4, \ldots, 1000\}$$

where $f(X)$ is the sum of the elements in X. (Examples: $f(\{3, 7, 50\}) = 60$; $f(\{1, 70, 80, 95\}) = 246$.) As $|\mathscr{P}(A)| = 2^{10} = 1024 > 1001 = |\{0, 1, 2, 3, \ldots, 1000\}|$,

it follows from the Pigeonhole Principle that f is not injective. Therefore there are two unequal sets $X, Y \in \mathscr{P}(A)$ for which $f(X) = f(Y)$. In other words, there are subsets $X \subseteq A$ and $Y \subseteq A$ for which the sum of elements in X equals the sum of elements in Y. ■

Example 12.8 Prove the following proposition.

Proposition There are at least two Texans with the same number of hairs on their heads.

Proof. We will use two facts. First, the population of Texas is more than twenty million. Second, it is a biological fact that every human head has fewer than one million hairs. Let A be the set of all Texans, and let $B = \{0, 1, 2, 3, 4, \ldots, 1000000\}$. Let $f : A \to B$ be the function for which $f(x)$ equals the number of hairs on the head of x. Since $|A| > |B|$, the Pigeonhole Principle asserts that f is not injective. Thus there are two Texans x and y for whom $f(x) = f(y)$, meaning that they have the same number of hairs on their heads. ■

Exercises for Section 12.3

1. Prove that if six numbers are chosen at random, then at least two of them will have the same remainder when divided by 5.
2. If a is a natural number, then there exist two unequal natural numbers k and ℓ for which $a^k - a^\ell$ is divisible by 10.
3. Prove that if six natural numbers are chosen at random, then the sum or difference of two of them is divisible by 9.
4. Consider a square whose side-length is one unit. Select any five points from inside this square. Prove that at least two of these points are within $\frac{\sqrt{2}}{2}$ units of each other.
5. Prove that any set of seven distinct natural numbers contains a pair of numbers whose sum or difference is divisible by 10.
6. Given a sphere S, a *great circle* of S is the intersection of S with a plane through its center. Every great circle divides S into two parts. A *hemisphere* is the union of the great circle and one of these two parts. Prove that if five points are placed arbitrarily on S, then there is a hemisphere that contains four of them.

12.4 Composition

You should be familiar with the notion of function composition from algebra and calculus. Still, it is worthwhile to revisit it now with our more sophisticated ideas about functions.

Definition 12.5 Suppose $f : A \to B$ and $g : B \to C$ are functions with the property that the codomain of f equals the domain of g. The **composition of f with g** is another function, denoted as $g \circ f$ and defined as follows: If $x \in A$, then $g \circ f(x) = g(f(x))$. Therefore $g \circ f$ sends elements of A to elements of C, so $g \circ f : A \to C$.

The following figure illustrates the definition. Here $f : A \to B$, $g : B \to C$, and $g \circ f : A \to C$. We have, for example, $g \circ f(0) = g(f(0)) = g(2) = 4$. Be very careful with the order of the symbols. Even though g comes first in the symbol $g \circ f$, we work out $g \circ f(x)$ as $g(f(x))$, with f acting on x first, followed by g acting on $f(x)$.

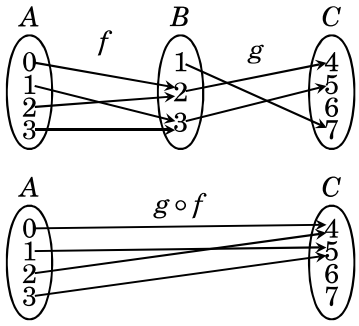


Figure 12.5. Composition of two functions

Notice that the composition $g \circ f$ also makes sense if the range of f is a *subset* of the domain of g. You should take note of this fact, but to keep matters simple we will continue to emphasize situations where the codomain of f equals the domain of g.

Example 12.9 Suppose $A = \{a, b, c\}$, $B = \{0, 1\}$, $C = \{1, 2, 3\}$. Let $f : A \to B$ be the function $f = \{(a, 0), (b, 1), (c, 0)\}$, and let $g : B \to C$ be the function $g = \{(0, 3), (1, 1)\}$. Then $g \circ f = \{(a, 3), (b, 1), (c, 3)\}$.

Example 12.10 Suppose $A = \{a, b, c\}$, $B = \{0, 1\}$, $C = \{1, 2, 3\}$. Let $f : A \to B$ be the function $f = \{(a, 0), (b, 1), (c, 0)\}$, and let $g : C \to B$ be the function $g = \{(1, 0), (2, 1), (3, 1)\}$. In this situation the composition $g \circ f$ is not defined because the codomain B of f is not the same set as the domain C of g.

Remember: In order for $g \circ f$ to make sense, the codomain of f must equal the domain of g. (Or at least be a subset of it.)

Example 12.11 Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = x^2 + x$, and $g : \mathbb{R} \to \mathbb{R}$ be defined as $g(x) = x + 1$. Then $g \circ f : \mathbb{R} \to \mathbb{R}$ is the function defined by the formula $g \circ f(x) = g(f(x)) = g(x^2 + x) = x^2 + x + 1$.

Since the domains and codomains of g and f are the same, we can in this case do the composition in the other order. Note that $f \circ g : \mathbb{R} \to \mathbb{R}$ is the function defined as $f \circ g(x) = f(g(x)) = f(x+1) = (x+1)^2 + (x+1) = x^2 + 3x + 2$.

This example illustrates that even when $g \circ f$ and $f \circ g$ are both defined, they are not necessarily equal. We can express this fact by saying *function composition is not commutative.*

We close this section by proving several facts about function composition that you are likely to encounter in your future study of mathematics. First, we note that, although it is not commutative, function composition *is* associative.

Theorem 12.1 Composition of functions is associative. That is if $f : A \to B$, $g : B \to C$ and $h : C \to D$, then $(h \circ g) \circ f = h \circ (g \circ f)$.

Proof. Suppose f, g, h are as stated. It follows from Definition 12.5 that both $(h \circ g) \circ f$ and $h \circ (g \circ f)$ are functions from A to D. To show that they are equal, we just need to show

$$\Big((h \circ g) \circ f\Big)(x) = \Big(h \circ (g \circ f)\Big)(x)$$

for every $x \in A$. Note that Definition 12.5 yields

$$\Big((h \circ g) \circ f\Big)(x) = (h \circ g)(f(x)) = h(g(f(x)).$$

Also

$$\Big(h \circ (g \circ f)\Big)(x) = h(g \circ f(x)) = h(g(f(x))).$$

Thus

$$\Big((h \circ g) \circ f\Big)(x) = \Big(h \circ (g \circ f)\Big)(x),$$

as both sides equal $h(g(f(x)))$. ■

Theorem 12.2 Suppose $f : A \to B$ and $g : B \to C$. If both f and g are injective, then $g \circ f$ is injective. If both f and g are surjective, then $g \circ f$ is surjective.

Proof. First suppose both f and g are injective. To see that $g \circ f$ is injective, we must show that $g \circ f(x) = g \circ f(y)$ implies $x = y$. Suppose $g \circ f(x) = g \circ f(y)$. This means $g(f(x)) = g(f(y))$. It follows that $f(x) = f(y)$. (For otherwise g wouldn't be injective.) But since $f(x) = f(y)$ and f is injective, it must be that $x = y$. Therefore $g \circ f$ is injective.

Next suppose both f and g are surjective. To see that $g \circ f$ is surjective, we must show that for any element $c \in C$, there is a corresponding element $a \in A$ for which $g \circ f(a) = c$. Thus consider an arbitrary $c \in C$. Because g is surjective, there is an element $b \in B$ for which $g(b) = c$. And because f is surjective, there is an element $a \in A$ for which $f(a) = b$. Therefore $g(f(a)) = g(b) = c$, which means $g \circ f(a) = c$. Thus $g \circ f$ is surjective. ■

Exercises for Section 12.4

1. Suppose $A = \{5,6,8\}$, $B = \{0,1\}$, $C = \{1,2,3\}$. Let $f : A \to B$ be the function $f = \{(5,1),(6,0),(8,1)\}$, and $g : B \to C$ be $g = \{(0,1),(1,1)\}$. Find $g \circ f$.

2. Suppose $A = \{1,2,3,4\}$, $B = \{0,1,2\}$, $C = \{1,2,3\}$. Let $f : A \to B$ be

$$f = \{(1,0),(2,1),(3,2),(4,0)\},$$

and $g : B \to C$ be $g = \{(0,1),(1,1),(2,3)\}$. Find $g \circ f$.

3. Suppose $A = \{1,2,3\}$. Let $f : A \to A$ be the function $f = \{(1,2),(2,2),(3,1)\}$, and let $g : A \to A$ be the function $g = \{(1,3),(2,1),(3,2)\}$. Find $g \circ f$ and $f \circ g$.

4. Suppose $A = \{a,b,c\}$. Let $f : A \to A$ be the function $f = \{(a,c),(b,c),(c,c)\}$, and let $g : A \to A$ be the function $g = \{(a,a),(b,b),(c,a)\}$. Find $g \circ f$ and $f \circ g$.

5. Consider the functions $f, g : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = \sqrt[3]{x+1}$ and $g(x) = x^3$. Find the formulas for $g \circ f$ and $f \circ g$.

6. Consider the functions $f, g : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = \frac{1}{x^2+1}$ and $g(x) = 3x + 2$. Find the formulas for $g \circ f$ and $f \circ g$.

7. Consider the functions $f, g : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ defined as $f(m,n) = (mn, m^2)$ and $g(m,n) = (m+1, m+n)$. Find the formulas for $g \circ f$ and $f \circ g$.

8. Consider the functions $f, g : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ defined as $f(m,n) = (3m - 4n, 2m + n)$ and $g(m,n) = (5m + n, m)$. Find the formulas for $g \circ f$ and $f \circ g$.

9. Consider the functions $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ defined as $f(m,n) = m + n$ and $g : \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ defined as $g(m) = (m,m)$. Find the formulas for $g \circ f$ and $f \circ g$.

10. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by the formula $f(x,y) = (xy, x^3)$. Find a formula for $f \circ f$.

12.5 Inverse Functions

You may recall from calculus that if a function f is injective and surjective, then it has an inverse function f^{-1} that "undoes" the effect of f in the sense that $f^{-1}(f(x)) = x$ for every x in the domain. (For example, if $f(x) = x^3$, then $f^{-1}(x) = \sqrt[3]{x}$.) We now review these ideas. Our approach uses two ingredients, outlined in the following definitions.

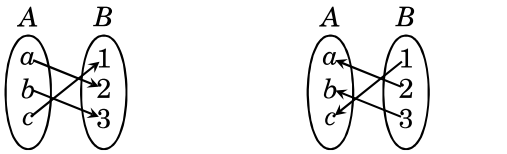
Definition 12.6 Given a set A, the **identity function on** A is the function $i_A : A \to A$ defined as $i_A(x) = x$ for every $x \in A$.

For example if $A = \{1,2,3\}$, then $i_A = \{(1,1),(2,2),(3,3)\}$. Also $i_{\mathbb{Z}} = \{(n,n) : n \in \mathbb{Z}\}$. The identity function on a set is the function that sends any element of the set to itself.

Notice that for any set A, the identity function i_A is bijective: It is injective because $i_A(x) = i_A(y)$ immediately reduces to $x = y$. It is surjective because if we take any element b in the codomain A, then b is also in the domain A, and $i_A(b) = b$.

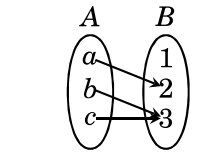
Definition 12.7 Given a relation R from A to B, the **inverse relation of** R is the relation from B to A defined as $R^{-1} = \{(y,x) : (x,y) \in R\}$. In other words, the inverse of R is the relation R^{-1} obtained by interchanging the elements in every ordered pair in R.

For example, let $A = \{a,b,c\}$ and $B = \{1,2,3\}$, and suppose f is the relation $f = \{(a,2),(b,3),(c,1)\}$ from A to B. Then $f^{-1} = \{(2,a),(3,b),(1,c)\}$ and this is a relation from B to A. Notice that f is actually a function from A to B, and f^{-1} is a function from B to A. These two relations are drawn below. Notice the drawing for relation f^{-1} is just the drawing for f with arrows reversed.

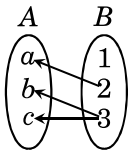


$f = \{(a,2),(b,3),(c,1)\}$ $\quad$ $f^{-1} = \{(2,a),(3,b),(1,c)\}$

For another example, let A and B be the same sets as above, but consider the relation $g = \{(a,2),(b,3),(c,3)\}$ from A to B. Then $g^{-1} = \{(2,a),(3,b),(3,c)\}$ is a relation from B to A. These two relations are sketched below.



$g = \{(a,2),(b,3),(c,3)\}$



$g^{-1} = \{(2,a),(3,b),(3,c)\}$

This time, even though the relation g is a function, its inverse g^{-1} is not a function because the element 3 occurs twice as a first coordinate of an ordered pair in g^{-1}.

In the above examples, relations f and g are both functions, and f^{-1} is a function and g^{-1} is not. This begs a question: What properties does f have and g lack that makes f^{-1} a function and g^{-1} not a function? The answer is not hard to see. Function g is not injective because $g(b) = g(c) = 3$, and thus $(b,3)$ and $(c,3)$ are both in g. This causes a problem with g^{-1} because it means $(3,b)$ and $(3,c)$ are both in g^{-1}, so g^{-1} can't be a function. Thus, in order for g^{-1} to be a function, it would be necessary that g be injective.

But that is not enough. Function g also fails to be surjective because no element of A is sent to the element $1 \in B$. This means g^{-1} contains no ordered pair whose first coordinate is 1, so it can't be a function from B to A. If g^{-1} were to be a function it would be necessary that g be surjective.

The previous two paragraphs suggest that if g is a function, then it must be bijective in order for its inverse relation g^{-1} to be a function. Indeed, this is easy to verify. Conversely, if a function is bijective, then its inverse relation is easily seen to be a function. We summarize this in the following theorem.

Theorem 12.3 Let $f : A \to B$ be a function. Then f is bijective if and only if the inverse relation f^{-1} is a function from B to A.

Suppose $f : A \to B$ is bijective, so according to the theorem f^{-1} is a function. Observe that the relation f contains all the pairs $(x, f(x))$ for $x \in A$, so f^{-1} contains all the pairs $(f(x),x)$. But $(f(x),x) \in f^{-1}$ means $f^{-1}(f(x)) = x$. Therefore $f^{-1} \circ f(x) = x$ for every $x \in A$. From this we get $f^{-1} \circ f = i_A$. Similar reasoning produces $f \circ f^{-1} = i_B$. This leads to the following definitions.

Definition 12.8 If $f : A \to B$ is bijective then its **inverse** is the function $f^{-1} : B \to A$. Functions f and f^{-1} obey the equations $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$.

You probably recall from algebra and calculus at least one technique for computing an inverse: Given f, to find f^{-1}, start with the equation $y = f(x)$. Then interchange variables to get $x = f(y)$. Solving this equation for y (if possible) produces $y = f^{-1}(x)$. The next two examples illustrate this.

Example 12.12 The function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x^3 + 1$ is bijective. Find its inverse.

We begin by writing $y = x^3 + 1$. Now interchange variables to obtain $x = y^3 + 1$. Solving for y produces $y = \sqrt[3]{x-1}$. Thus

$$f^{-1}(x) = \sqrt[3]{x-1}.$$

(You can check your answer by computing

$$f^{-1}(f(x)) = \sqrt[3]{f(x)-1} = \sqrt[3]{x^3+1-1} = x.$$

Therefore $f^{-1}(f(x)) = x$. Any answer other than x indicates a mistake.)

We close with one final example. Exercise 12.5 showed that the function $g : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ defined by the formula $g(m,n) = (m+n, m+2n)$ is bijective. Let's find its inverse. The approach outlined above should work, but we need to be careful to keep track of coordinates in $\mathbb{Z} \times \mathbb{Z}$. We begin by writing $(x,y) = g(m,n)$, then interchanging the variables (x,y) and (m,n) to get $(m,n) = g(x,y)$. This gives

$$(m,n) = (x+y, x+2y),$$

from which we get the following system of equations.

$$\begin{array}{rcrcl} x & + & y & = & m \\ x & + & 2y & = & n \end{array}$$

Solving this system using techniques from algebra with which you are familiar, we get

$$\begin{array}{rcl} x & = & 2m-n \\ y & = & n-m \end{array}$$

Then $(x,y) = (2m-n, n-m)$, so $\boxed{g^{-1}(m,n) = (2m-n, n-m).}$

We can check our work by confirming that $g^{-1}(g(m,n)) = (m,n)$. Doing the math,

$$\begin{aligned} g^{-1}(g(m,n)) &= g^{-1}(m+n, m+2n) \\ &= \big(2(m+n)-(m+2n), (m+2n)-(m+n)\big) \\ &= (m,n). \end{aligned}$$

Exercises for Section 12.5

1. Check that the function $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = 6 - n$ is bijective. Then compute f^{-1}.

2. In Exercise 9 of Section 12.2 you proved that $f : \mathbb{R} - \{2\} \to \mathbb{R} - \{5\}$ defined by $f(x) = \dfrac{5x+1}{x-2}$ is bijective. Now find its inverse.

3. Let $B = \{2^n : n \in \mathbb{Z}\} = \{\ldots, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, \ldots\}$. Show that the function $f : \mathbb{Z} \to B$ defined as $f(n) = 2^n$ is bijective. Then find f^{-1}.

4. The function $f : \mathbb{R} \to (0, \infty)$ defined as $f(x) = e^{x^3+1}$ is bijective. Find its inverse.

5. The function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = \pi x - e$ is bijective. Find its inverse.

6. The function $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ defined by the formula $f(m,n) = (5m+4n, 4m+3n)$ is bijective. Find its inverse.

7. Show that the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by the formula $f(x,y) = ((x^2+1)y, x^3)$ is bijective. Then find its inverse.

8. Is the function $\theta : \mathscr{P}(\mathbb{Z}) \to \mathscr{P}(\mathbb{Z})$ defined as $\theta(X) = \overline{X}$ bijective? If so, what is its inverse?

9. Consider the function $f : \mathbb{R} \times \mathbb{N} \to \mathbb{N} \times \mathbb{R}$ defined as $f(x,y) = (y, 3xy)$. Check that this is bijective; find its inverse.

12.6 Image and Preimage

It is time to take up a matter of notation that you will encounter in future mathematics classes. Suppose we have a function $f : A \to B$. If $X \subseteq A$, the expression $f(X)$ has a special meaning. It stands for the set $\{f(x) : x \in X\}$. Similarly, if $Y \subseteq B$ then $f^{-1}(Y)$ has a meaning *even if f is not invertible.* The expression $f^{-1}(Y)$ stands for the set $\{x \in A : f(x) \in Y\}$. Here are the precise definitions.

Definition 12.9 Suppose $f : A \to B$ is a function.

1. If $X \subseteq A$, the **image of** X is the set $f(X) = \{f(x) : x \in X\} \subseteq B$.
2. If $Y \subseteq B$, the **preimage of** Y is the set $f^{-1}(Y) = \{x \in A : f(x) \in Y\} \subseteq A$.

In words, the image $f(X)$ of X is the set of all things in B that f sends elements of X to. (Roughly speaking, you might think of $f(X)$ as a kind of distorted "copy" or "image" of X in B.) The preimage $f^{-1}(Y)$ of Y is the set of all things in A that f sends into Y.

Example 12.13 Consider $f:\{s,t,u,v,w,x,y,z\}\to\{0,1,2,3,4,5,6,7,8,9\}$, where

$$f=\{(s,4),(t,8),(u,8),(v,1),(w,2),(x,4),(y,6),(z,4)\}.$$

Notice that f is neither injective nor surjective, so it certainly is not invertible. Be sure you understand the following statements.

a. $f(\{s,t,u,z\})=\{8,4\}$
b. $f(\{s,x,z\})=\{4\}$
c. $f(\{s,v,w,y\})=\{1,2,4,6\}$
d. $f^{-1}(\{4\})=\{s,x,z\}$
e. $f^{-1}(\{4,9\})=\{s,x,z\}$
f. $f^{-1}(\{9\})=\emptyset$
g. $f^{-1}(\{1,4,8\})=\{s,t,u,v,x,z\}$

It is important to realize that the X and Y in Definition 12.9 are subsets (not elements!) of A and B. Thus, in the above example we had $f^{-1}(\{4\})=\{s,x,z\}$, though $f^{-1}(4)$ has absolutely no meaning because the inverse function f^{-1} does not exist. Likewise, there is a subtle difference between $f(\{a\})=\{4\}$ and $f(a)=4$. Be careful.

Example 12.14 Consider the function $f:\mathbb{R}\to\mathbb{R}$ defined as $f(x)=x^2$. Observe that $f(\{0,1,2\})=\{0,1,4\}$ and $f^{-1}(\{0,1,4\})=\{-2,-1,0,1,2\}$. This shows $f^{-1}(f(X))\neq X$ in general.

Now check your understanding of the following statements: $f([-2,3])=[0,9]$, and $f^{-1}([0,9])=[-3,3]$. Also $f(\mathbb{R})=[0,\infty)$ and $f^{-1}([-2,-1])=\emptyset$.

If you continue your mathematical studies, you are likely to encounter the following result in the future. For now, you are asked to prove it in the exercises

Theorem 12.4 Suppose $f:A\to B$ is a function. Let $W,X\subseteq A$, and $Y,Z\subseteq B$. Then:

1. $f(W\cap X)\subseteq f(W)\cap f(X)$
2. $f(W\cup X)=f(W)\cup f(X)$
3. $f^{-1}(Y\cap Z)=f^{-1}(Y)\cap f^{-1}(Z)$
4. $f^{-1}(Y\cup Z)=f^{-1}(Y)\cup f^{-1}(Z)$

Exercises for Section 12.6

1. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x^2 + 3$. Find $f([-3,5])$ and $f^{-1}([12,19])$.

2. Consider the function $f : \{1,2,3,4,5,6,7\} \to \{0,1,2,3,4,5,6,7,8,9\}$ given as

$$f = \{(1,3),(2,8),(3,3),(4,1),(5,2),(6,4),(7,6)\}.$$

Find: $f(\{1,2,3\})$, $f(\{4,5,6,7\})$, $f(\emptyset)$, $f^{-1}(\{0,5,9\})$ and $f^{-1}(\{0,3,5,9\})$.

3. This problem concerns functions $f : \{1,2,3,4,5,6,7\} \to \{0,1,2,3,4\}$. How many such functions have the property that $|f^{-1}(\{3\})| = 3$?

4. This problem concerns functions $f : \{1,2,3,4,5,6,7,8\} \to \{0,1,2,3,4,5,6\}$. How many such functions have the property that $|f^{-1}(\{2\})| = 4$?

5. Consider a function $f : A \to B$ and a subset $X \subseteq A$. We observed in Section 12.6 that $f^{-1}(f(X)) \neq X$ in general. However $X \subseteq f^{-1}(f(X))$ is always true. Prove this.

6. Given a function $f : A \to B$ and any subset $Y \subseteq B$, is it always true that $f(f^{-1}(Y)) = Y$? Prove or give a counterexample.

7. Given a function $f : A \to B$ and subsets $W, X \subseteq A$, prove $f(W \cap X) \subseteq f(W) \cap f(X)$.

8. Given a function $f : A \to B$ and subsets $W, X \subseteq A$, then $f(W \cap X) = f(W) \cap f(X)$ is **false** in general. Produce a counterexample.

9. Given a function $f : A \to B$ and subsets $W, X \subseteq A$, prove $f(W \cup X) = f(W) \cup f(X)$

10. Given $f : A \to B$ and subsets $Y, Z \subseteq B$, prove $f^{-1}(Y \cap Z) = f^{-1}(Y) \cap f^{-1}(Z)$.

11. Given $f : A \to B$ and subsets $Y, Z \subseteq B$, prove $f^{-1}(Y \cup Z) = f^{-1}(Y) \cup f^{-1}(Z)$.