5

Functions

5.1. Functions

Suppose P is the set of all people, and let $H = \{(p, n) \in P \times \mathbb{N} \mid \text{the person } p \text{ has } n \text{ children}\}$. Then H is a relation from P to $\mathbb{N}$, and it has the following important property. For every $p \in P$, there is *exactly one* $n \in \mathbb{N}$ such that $(p, n) \in H$. Mathematicians express this by saying that H is a *function* from P to $\mathbb{N}$.

Definition 5.1.1. Suppose F is a relation from A to B. Then F is called a *function from A to B* if for every $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in F$. In other words, to say that F is a function from A to B means:

$$\forall a \in A \exists! b \in B((a, b) \in F).$$

To indicate that F is a function from A to B, we will write $F : A \to B$.

Example 5.1.2.

1. Let $A = \{1, 2, 3\}$, $B = \{4, 5, 6\}$, and $F = \{(1, 5), (2, 4), (3, 5)\}$. Is F a function from A to B?
2. Let $A = \{1, 2, 3\}$, $B = \{4, 5, 6\}$, and $G = \{(1, 5), (2, 4), (1, 6)\}$. Is G a function from A to B?
3. Let C be the set of all cities and N the set of all countries, and let $L = \{(c, n) \in C \times N \mid \text{the city } c \text{ is in the country } n\}$. Is L a function from C to N?
4. Let P be the set of all people, and let $C = \{(p, q) \in P \times P \mid \text{the person } p \text{ is a parent of the person } q\}$. Is C a function from P to P?
5. Let P be the set of all people, and let $D = \{(p, x) \in P \times \mathscr{P}(P) \mid x = \text{the set of all children of } p\}$. Is D a function from P to $\mathscr{P}(P)$?

6. Let A be any set. Recall that $i_A = \{(a, a) \mid a \in A\}$ is called the identity relation on A. Is it a function from A to A?
7. Let $f = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y = x^2\}$. Is f a function from $\mathbb{R}$ to $\mathbb{R}$?

Solutions

1. Yes. Note that 1 is paired with 5 in the relation F, but it is not paired with any other element of B. Similarly, 2 is paired only with 4, and 3 with 5. In other words, each element of A appears as the first coordinate of exactly one ordered pair in F. Therefore F is a function from A to B. Note that the definition of function does *not* require that each element of B be paired with exactly one element of A. Thus, it doesn't matter that 5 occurs as the second coordinate of two different pairs in F and that 6 doesn't occur in any ordered pairs at all.
2. No. G fails to be a function from A to B for two reasons. First of all, 3 isn't paired with any element of B in the relation G, which violates the requirement that every element of A must be paired with some element of B. Second, 1 is paired with two different elements of B, 5 and 6, which violates the requirement that each element of A be paired with *only one* element of B.
3. If we make the reasonable assumption that every city is in exactly one country, then L is a function from C to N.
4. Because some people have no children and some people have more than one child, C is not a function from P to P.
5. Yes, D is a function from P to $\mathscr{P}(P)$. Each person p is paired with exactly one set $x \subseteq P$, namely the set of all children of p. Note that in the relation D, a person p is paired with the set consisting of all of p's children, *not* with the children themselves. Even if p does not have exactly one child, it is still true that there is exactly one set that contains precisely the children of p and nothing else.
6. Yes. Each $a \in A$ is paired in the relation i_A with exactly one element of A, namely a itself. In other words, $(a, a) \in i_A$, but for every $a' \neq a$, $(a, a') \notin i_A$. Thus, we can call i_A the identity *function* on A.
7. Yes. For each real number x there is exactly one value of y, namely $y = x^2$, such that $(x, y) \in f$.

Suppose $f : A \to B$. If $a \in A$, then we know that there is exactly one $b \in B$ such that $(a, b) \in f$. This unique b is called "the value of f at a," or "the image of a under f," or "the result of applying f to a," or just "f of a," and it is written $f(a)$. In other words, for every $a \in A$ and $b \in B$, $b = f(a)$

iff $(a, b) \in f$. For example, for the function $F = \{(1,5),(2,4),(3,5)\}$ in part 1 of Example 5.1.2, we could say that $F(1) = 5$, since $(1,5) \in F$. Similarly, $F(2) = 4$ and $F(3) = 5$. If L is the function in part 3 and c is any city, then $L(c)$ would be the unique country n such that $(c, n) \in L$. In other words, $L(c) =$ the country in which c is located. For example, $L(\text{Paris}) = \text{France}$. For the function D in part 5, we could say that for any person p, $D(p) =$ the set of all children of p. If A is any set and $a \in A$, then $(a, a) \in i_A$, so $i_A(a) = a$. And if f is the function in part 7, then for every real number x, $f(x) = x^2$.

A function f from a set A to another set B is often specified by giving a rule that can be used to determine $f(a)$ for any $a \in A$. For example, if A is the set of all people and $B = \mathbb{R}^+$, then we could define a function f from A to B by the rule that for every $a \in A$, $f(a) = a$'s height in inches. Although this definition doesn't say explicitly which ordered pairs are elements of f, we can determine this by using our rule that for all $a \in A$ and $b \in B$, $(a, b) \in f$ iff $b = f(a)$. Thus,

$$\begin{aligned} f &= \{(a, b) \in A \times B \mid b = f(a)\} \\ &= \{(a, b) \in A \times B \mid b = a\text{'s height in inches}\}. \end{aligned}$$

For example, if Joe Smith is 68 inches tall, then $(\text{Joe Smith}, 68) \in f$ and $f(\text{Joe Smith}) = 68$.

It is often useful to think of a function f from A to B as representing a rule that associates, with each $a \in A$, some corresponding object $b = f(a) \in B$. However, it is important to remember that although a function can be defined by giving such a rule, it need not be defined in this way. Any subset of $A \times B$ that satisfies the requirements given in Definition 5.1.1 is a function from A to B.

Example 5.1.3. Here are some more examples of functions defined by rules.

1. Suppose every student is assigned an academic advisor who is a professor. Let S be the set of students and P the set of professors. Then we can define a function f from S to P by the rule that for every student s, $f(s) =$ the advisor of s. In other words,

$$\begin{aligned} f &= \{(s, p) \in S \times P \mid p = f(s)\} \\ &= \{(s, p) \in S \times P \mid \text{the professor } p \text{ is the academic advisor of} \\ &\qquad \text{the student } s\}. \end{aligned}$$

2. We can define a function g from $\mathbb{Z}$ to $\mathbb{R}$ by the rule that for every $x \in \mathbb{Z}$, $g(x) = 2x + 3$. Then

$$
\begin{aligned}
g &= \{(x, y) \in \mathbb{Z} \times \mathbb{R} \mid y = g(x)\} \\
&= \{(x, y) \in \mathbb{Z} \times \mathbb{R} \mid y = 2x + 3\} \\
&= \{\ldots, (-2, -1), (-1, 1), (0, 3), (1, 5), (2, 7), \ldots\}.
\end{aligned}
$$

3. Let h be the function from $\mathbb{R}$ to $\mathbb{R}$ defined by the rule that for every $x \in \mathbb{R}$, $h(x) = 2x + 3$. Note that the formula for $h(x)$ is the same as the formula for $g(x)$ in part 2. However, h and g are *not* the same function. You can see this by noting that, for example, $(\pi, 2\pi + 3) \in h$ but $(\pi, 2\pi + 3) \notin g$, since $\pi \notin \mathbb{Z}$. (For more on the relationship between g and h, see exercise 7(c).)

Notice that when a function f from A to B is specified by giving a rule for finding $f(a)$, the rule must determine the value of $f(a)$ for *every* $a \in A$. Sometimes when mathematicians are stating such a rule they don't say explicitly that the rule applies to all $a \in A$. For example, a mathematician might say "let f be the function from $\mathbb{R}$ to $\mathbb{R}$ defined by the formula $f(x) = x^2 + 7$." It is understood in this case that the equation $f(x) = x^2 + 7$ applies to all $x \in \mathbb{R}$ even though it hasn't been said explicitly. This means that you can plug in any real number for x in this equation, and the resulting equation will be true. For example, you can conclude that $f(3) = 3^2 + 7 = 16$. Similarly, if w is a real number, then you can write $f(w) = w^2 + 7$, or even $f(2w - 3) = (2w - 3)^2 + 7 = 4w^2 - 12w + 16$.

Because a function f from A to B is completely determined by the rule for finding $f(a)$, two functions that are defined by equivalent rules must be equal. More precisely, we have the following theorem:

Theorem 5.1.4. *Suppose f and g are functions from A to B. If $\forall a \in A(f(a) = g(a))$, then $f = g$.*

Proof. Suppose $\forall a \in A(f(a) = g(a))$, and let (a, b) be an arbitrary element of f. Then $b = f(a)$. But by our assumption $f(a) = g(a)$, so $b = g(a)$ and therefore $(a, b) \in g$. Thus, $f \subseteq g$. A similar argument shows $g \subseteq f$, so $f = g$. □

Commentary. Because f and g are sets, we prove $f = g$ by proving $f \subseteq g$ and $g \subseteq f$. Each of these goals is proven by showing that an arbitrary element of one set must be an element of the other. Note that, now that we have proven Theorem 5.1.4, we have another method for proving that two functions f and g from a set A to another set B are equal. In the future, to prove $f = g$ we will usually prove $\forall a \in A(f(a) = g(a))$ and then apply Theorem 5.1.4.

Because functions are just relations of a special kind, the concepts introduced in Chapter 4 for relations can be applied to functions as well. For example, suppose $f : A \to B$. Then f is a relation from A to B, so it makes sense to talk about the domain of f, which is a subset of A, and the range of f, which is a subset of B. According to the definition of function, every element of A must appear as the first coordinate of some (in fact, exactly one) ordered pair in f, so the domain of f must actually be all of A. But the range of f need not be all of B. The elements of the range of f will be the second coordinates of all the ordered pairs in f, and the second coordinate of an ordered pair in f is what we have called the image of its first coordinate. Thus, the range of f could also be described as the set of all images of elements of A under f:

$$\mathrm{Ran}(f) = \{f(a) \mid a \in A\}.$$

For example, for the function f defined in part 1 of Example 5.1.3, $\mathrm{Ran}(f) = \{f(s) \mid s \in S\} =$ the set of all advisors of students.

We can draw diagrams of functions in exactly the same way we drew diagrams for relations in Chapter 4. If $f : A \to B$, then as before, every ordered pair $(a, b) \in f$ would be represented in the diagram by an edge connecting a to b. By the definition of function, every $a \in A$ occurs as the first coordinate of exactly one ordered pair in f, and the second coordinate of this ordered pair is $f(a)$. Thus, for every $a \in A$ there will be exactly one edge coming from a, and it will connect a to $f(a)$. For example, Figure 5.1 shows what the diagram for the function L defined in part 3 of Example 5.1.2 would look like.

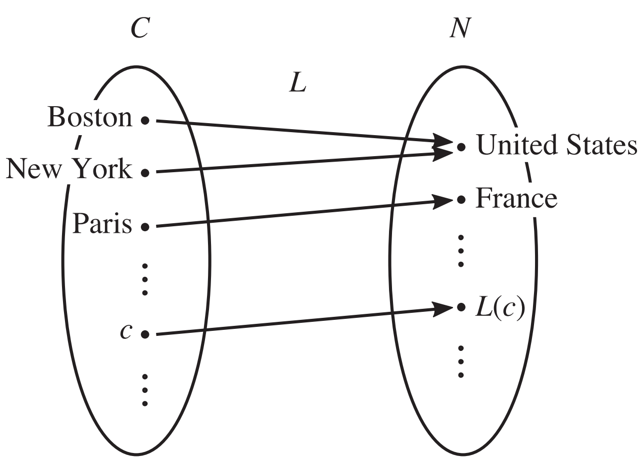


Figure 5.1.

The definition of composition of relations can also be applied to functions. If $f : A \to B$ and $g : B \to C$, then f is a relation from A to B and g is a relation from B to C, so $g \circ f$ will be a relation from A to C. In fact, it turns out that $g \circ f$ is a function from A to C, as the next theorem shows.

Theorem 5.1.5. *Suppose $f : A \to B$ and $g : B \to C$. Then $g \circ f : A \to C$, and for any $a \in A$, the value of $g \circ f$ at a is given by the formula $(g \circ f)(a) = g(f(a))$.*

Scratch work

Before proving this theorem, it might be helpful to discuss the scratch work for the proof. According to the definition of function, to show that $g \circ f : A \to C$ we must prove that $\forall a \in A \exists! c \in C((a, c) \in g \circ f)$, so we will start out by letting a be an arbitrary element of A and then try to prove that $\exists! c \in C((a, c) \in g \circ f)$. As we saw in Section 3.6, we can prove this statement by proving existence and uniqueness separately. To prove existence, we should try to find a $c \in C$ such that $(a, c) \in g \circ f$. For uniqueness, we should assume that $(a, c_1) \in g \circ f$ and $(a, c_2) \in g \circ f$, and then try to prove that $c_1 = c_2$.

Proof. Let a be an arbitrary element of A. We must show that there is a unique $c \in C$ such that $(a, c) \in g \circ f$.

Existence: Let $b = f(a) \in B$. Let $c = g(b) \in C$. Then $(a, b) \in f$ and $(b, c) \in g$, so by the definition of composition of relations, $(a, c) \in g \circ f$. Thus, $\exists c \in C((a, c) \in g \circ f)$.

Uniqueness: Suppose $(a, c_1) \in g \circ f$ and $(a, c_2) \in g \circ f$. Then by the definition of composition, we can choose $b_1 \in B$ such that $(a, b_1) \in f$ and $(b_1, c_1) \in g$, and we can also choose $b_2 \in B$ such that $(a, b_2) \in f$ and $(b_2, c_2) \in g$. Since f is a function, there can be only one $b \in B$ such that $(a, b) \in f$. Thus, since (a, b_1) and (a, b_2) are both elements of f, it follows that $b_1 = b_2$. But now applying the same reasoning to g, since $(b_1, c_1) \in g$ and $(b_1, c_2) = (b_2, c_2) \in g$, it follows that $c_1 = c_2$, as required.

This completes the proof that $g \circ f$ is a function from A to C. Finally, to derive the formula for $(g \circ f)(a)$, note that we showed in the existence half of the proof that for any $a \in A$, if we let $b = f(a)$ and $c = g(b)$, then $(a, c) \in g \circ f$. Thus,

$$(g \circ f)(a) = c = g(b) = g(f(a)). \qquad \square$$

When we first introduced the idea of the composition of two relations in Chapter 4, we pointed out that the notation was somewhat peculiar and promised to explain the reason for the notation in this chapter. We can now provide this explanation. The reason for the notation we've used for composition of relations is that it leads to the convenient formula $(g \circ f)(x) = g(f(x))$ derived in Theorem 5.1.5. Note that because functions are just relations of a special kind, everything we have proven about composition of relations applies to composition of functions. In particular, by Theorem 4.2.5, we know that composition of functions is associative.

Example 5.1.6. Here are some examples of compositions of functions.

1. Let C and N be the sets of all cities and countries, respectively, and let $L : C \to N$ be the function defined in part 3 of Example 5.1.2. Thus, for every city c, $L(c) =$ the country in which c is located. Let B be the set of all buildings located in cities, and define $F : B \to C$ by the formula $F(b) =$ the city in which the building b is located. Then $L \circ F : B \to N$. For example, $F(\text{Eiffel Tower}) =$ Paris, so according to the formula derived in Theorem 5.1.5,

$$\begin{aligned}(L \circ F)(\text{Eiffel Tower}) &= L(F(\text{Eiffel Tower}))\\ &= L(\text{Paris}) = \text{France}.\end{aligned}$$

   In general for every building $b \in B$,

$$\begin{aligned}(L \circ F)(b) = L(F(b)) &= L(\text{the city in which } b \text{ is located})\\ &= \text{the country in which } b \text{ is located.}\end{aligned}$$

   A diagram of this function is shown in Figure 5.2.

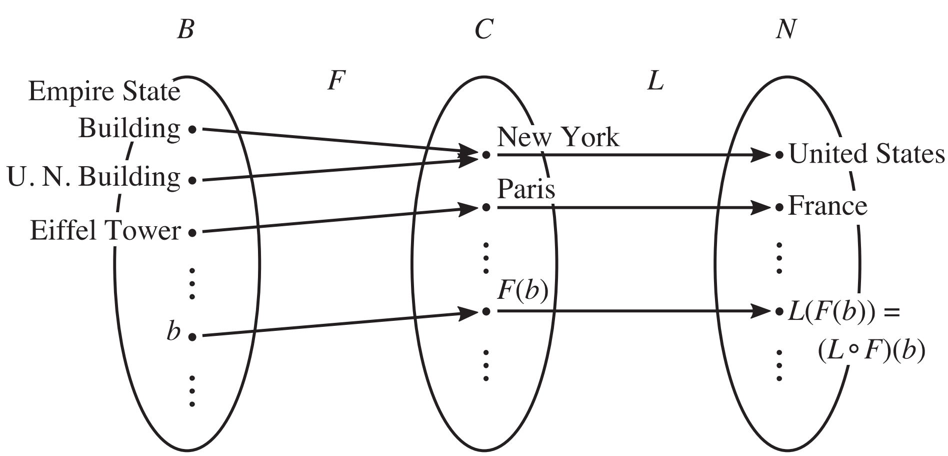


Figure 5.2.

2. Let $g : \mathbb{Z} \to \mathbb{R}$ be the function from part 2 of Example 5.1.3, which was defined by the formula $g(x) = 2x + 3$. Let $f : \mathbb{Z} \to \mathbb{Z}$ be defined by the formula $f(n) = n^2 - 3n + 1$. Then $g \circ f : \mathbb{Z} \to \mathbb{R}$. For example, $f(2) = 2^2 - 3 \cdot 2 + 1 = -1$, so $(g \circ f)(2) = g(f(2)) = g(-1) = 1$. In general for every $n \in \mathbb{Z}$,

$$\begin{aligned}(g \circ f)(n) &= g(f(n)) = g(n^2 - 3n + 1) = 2(n^2 - 3n + 1) + 3\\ &= 2n^2 - 6n + 5.\end{aligned}$$

Exercises

*1. (a) Let $A = \{1, 2, 3\}$, $B = \{4\}$, and $f = \{(1, 4), (2, 4), (3, 4)\}$. Is f a function from A to B?

(b) Let $A = \{1\}$, $B = \{2, 3, 4\}$, and $f = \{(1, 2), (1, 3), (1, 4)\}$. Is f a function from A to B?

(c) Let C be the set of all cars registered in your state, and let S be the set of all finite sequences of letters and digits. Let $L = \{(c, s) \in C \times S \mid$ the license plate number of the car c is $s\}$. Is L a function from C to S?

2. (a) Let f be the relation represented by the graph in Figure 5.3. Is f a function from A to B?

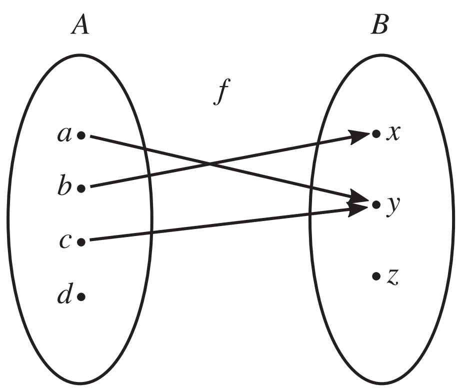


Figure 5.3.

(b) Let W be the set of all words of English, and let A be the set of all letters of the alphabet. Let $f = \{(w, a) \in W \times A \mid$ the letter a occurs in the word $w\}$, and let $g = \{(w, a) \in W \times A \mid$ the letter a is the first letter of the word $w\}$. Is f a function from W to A? How about g?

(c) John, Mary, Susan, and Fred go out to dinner and sit at a round table. Let $P = \{$John, Mary, Susan, Fred$\}$, and let $R = \{(p, q) \in P \times P \mid$ the person p is sitting immediately to the right of the person $q\}$. Is R a function from P to P?

*3. (a) Let $A = \{a, b, c\}$, $B = \{a, b\}$, and $f = \{(a, b), (b, b), (c, a)\}$. Then $f : A \to B$. What are $f(a)$, $f(b)$, and $f(c)$?

(b) Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by the formula $f(x) = x^2 - 2x$. What is $f(2)$?

(c) Let $f = \{(x, n) \in \mathbb{R} \times \mathbb{Z} \mid n \leq x < n + 1\}$. Then $f : \mathbb{R} \to \mathbb{Z}$. What is $f(\pi)$? What is $f(-\pi)$?

4. (a) Let N be the set of all countries and C the set of all cities. Let $H : N \to C$ be the function defined by the rule that for every country n, $H(n) =$ the capital of the country n. What is $H($Italy$)$?

(b) Let $A = \{1, 2, 3\}$ and $B = \mathscr{P}(A)$. Let $F : B \to B$ be the function defined by the formula $F(X) = A \setminus X$. What is $F(\{1, 3\})$?

(c) Let $f : \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ be the function defined by the formula $f(x) = (x + 1, x - 1)$. What is $f(2)$?

*5. Let L be the function defined in part 3 of Example 5.1.2 and let H be the function defined in exercise 4(a). Describe $L \circ H$ and $H \circ L$.

6. Let f and g be functions from $\mathbb{R}$ to $\mathbb{R}$ defined by the following formulas:

$$f(x) = \frac{1}{x^2 + 2}, \qquad g(x) = 2x - 1.$$

Find formulas for $(f \circ g)(x)$ and $(g \circ f)(x)$.

*7. Suppose $f : A \to B$ and $C \subseteq A$. The set $f \cap (C \times B)$, which is a relation from C to B, is called the *restriction* of f to C, and is sometimes denoted $f \restriction C$. In other words,

$$f \restriction C = f \cap (C \times B).$$

(a) Prove that $f \restriction C$ is a function from C to B and that for all $c \in C$, $f(c) = (f \restriction C)(c)$.

(b) Suppose $g : C \to B$. Prove that $g = f \restriction C$ iff $g \subseteq f$.

(c) Let g and h be the functions defined in parts 2 and 3 of Example 5.1.3. Show that $g = h \restriction \mathbb{Z}$.

8. Suppose $f : A \to B$ and $g \subseteq f$. Prove that there is a set $A^{\pm} \subseteq A$ such that $g : A^{\pm} \to B$.

9. Suppose $f : A \to B$, $B \neq \varnothing$, and $A \subseteq A^{\pm}$. Prove that there is a function $g : A^{\pm} \to B$ such that $f \subseteq g$.

*10. Suppose that f and g are functions from A to B and $f \neq g$. Show that $f\ ^3\ g$ is not a function.

11. Suppose A is a set. Show that i_A is the only relation on A that is both an equivalence relation on A and also a function from A to A.
12. Suppose $f : A \to C$ and $g : B \to C$.
    (a) Prove that if A and B are disjoint, then $f \cup g : A \cup B \to C$.
    (b) Prove that $f \cup g : A \cup B \to C$ iff $f \restriction (A \cap B) = g \restriction (A \cap B)$. (See exercise 7 for the meaning of the notation used here.)

*13. Suppose R is a relation from A to B, S is a relation from B to C, $\text{Ran}(R) = \text{Dom}(S) = B$, and $S \circ R : A \to C$.
    (a) Prove that $S : B \to C$.
    (b) Give an example to show that it need not be the case that $R : A \to B$.

14. Suppose $f : A \to B$ and S is a relation on B. Define a relation R on A as follows:

$$R = \{(x, y) \in A \times A \mid (f(x), f(y)) \in S\}.$$

    (a) Prove that if S is reflexive, then so is R.
    (b) Prove that if S is symmetric, then so is R.
    (c) Prove that if S is transitive, then so is R.

*15. Suppose $f : A \to B$ and R is a relation on A. Define a relation S on B as follows:

$$S = \{(x, y) \in B \times B \mid \exists u \in A\, \exists v \in A(f(u) = x \wedge f(v) = y \wedge (u, v) \in R)\}.$$

Justify your answers to the following questions with either proofs or counterexamples.
    (a) If R is reflexive, must it be the case that S is reflexive?
    (b) If R is symmetric, must it be the case that S is symmetric?
    (c) If R is transitive, must it be the case that S is transitive?

16. Suppose A and B are sets, and let $\mathcal{F} = \{f \mid f : A \to B\}$. Also, suppose R is a relation on B, and define a relation S on $\mathcal{F}$ as follows:

$$S = \{(f, g) \in \mathcal{F} \times \mathcal{F} \mid \forall x \in A((f(x), g(x)) \in R)\}.$$

Justify your answers to the following questions with either proofs or counterexamples.
    (a) If R is reflexive, must it be the case that S is reflexive?
    (b) If R is symmetric, must it be the case that S is symmetric?
    (c) If R is transitive, must it be the case that S is transitive?

17. Suppose A is a nonempty set and $f : A \to A$.
    (a) Suppose there is some $a \in A$ such that $\forall x \in A(f(x) = a)$. (In this case, f is called a *constant* function.) Prove that for all $g : A \to A$, $f \circ g = f$.

(b) Suppose that for all $g : A \to A$, $f \circ g = f$. Prove that f is a constant function. (Hint: What happens if g is a constant function?)

18. Let $\mathcal{F} = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$. Let $R = \{(f, g) \in \mathcal{F} \times \mathcal{F} \mid \exists a \in \mathbb{R} \forall x > a(f(x) = g(x))\}$.
    (a) Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be the functions defined by the formulas $f(x) = |x|$ and $g(x) = x$. Show that $(f, g) \in R$.
    (b) Prove that R is an equivalence relation.

*19. Let $\mathcal{F} = \{f \mid f : \mathbb{Z}^+ \to \mathbb{R}\}$. For $g \in \mathcal{F}$, we define the set $O(g)$ as follows:

$$O(g) = \{f \in \mathcal{F} \mid \exists a \in \mathbb{Z}^+ \exists c \in \mathbb{R}^+ \forall x > a(|f(x)| \leq c|g(x)|)\}.$$

(If $f \in O(g)$, then mathematicians say that "f is big-oh of g.")
    (a) Let $f : \mathbb{Z}^+ \to \mathbb{R}$ and $g : \mathbb{Z}^+ \to \mathbb{R}$ be defined by the formulas $f(x) = 7x + 3$ and $g(x) = x^2$. Prove that $f \in O(g)$, but $g \notin O(f)$.
    (b) Let $S = \{(f, g) \in \mathcal{F} \times \mathcal{F} \mid f \in O(g)\}$. Prove that S is a preorder, but not a partial order. (See exercise 25 of Section 4.5 for the definition of *preorder*.)
    (c) Suppose $f_1 \in O(g)$ and $f_2 \in O(g)$, and s and t are real numbers. Define a function $f : \mathbb{Z}^+ \to \mathbb{R}$ by the formula $f(x) = sf_1(x) + tf_2(x)$. Prove that $f \in O(g)$. (Hint: You may find the triangle inequality helpful. See exercise 13(c) of Section 3.5.)

20. (a) Suppose $g : A \to B$ and let $R = \{(x, y) \in A \times A \mid g(x) = g(y)\}$. Show that R is an equivalence relation on A.
    (b) Suppose R is an equivalence relation on A and let $g : A \to A/R$ be the function defined by the formula $g(x) = [x]_R$. Show that $R = \{(x, y) \in A \times A \mid g(x) = g(y)\}$.

*21. Suppose $f : A \to B$ and R is an equivalence relation on A. We will say that f is *compatible* with R if $\forall x \in A \forall y \in A(xRy \to f(x) = f(y))$. (You might want to compare this exercise to exercise 24 of Section 4.5.)
    (a) Suppose f is compatible with R. Prove that there is a unique function $h : A/R \to B$ such that for all $x \in A$, $h([x]_R) = f(x)$.
    (b) Suppose $h : A/R \to B$ and for all $x \in A$, $h([x]_R) = f(x)$. Prove that f is compatible with R.

22. Let $R = \{(x, y) \in \mathbb{N} \times \mathbb{N} \mid x \equiv y \pmod 5\}$. Note that by Theorem 4.5.10 and exercise 14 in Section 4.5, R is an equivalence relation on $\mathbb{N}$.
    (a) Show that there is a unique function $h : \mathbb{N}/R \to \mathbb{N}/R$ such that for every natural number x, $h([x]_R) = [x^2]_R$. (Hint: Use exercise 21.)
    (b) Show that there is no function $h : \mathbb{N}/R \to \mathbb{N}/R$ such that for every natural number x, $h([x]_R) = [2^x]_R$.

5.2. One-to-One and Onto

In the last section we saw that the composition of two functions is again a function. What about inverses of functions? If $f : A \to B$, then f is a relation from A to B, so f^{-1} is a relation from B to A. Is it a function from B to A? We'll answer this question in the next section. As we will see, the answer hinges on the following two properties of functions.

Definition 5.2.1. Suppose $f : A \to B$. We will say that f is *one-to-one* if

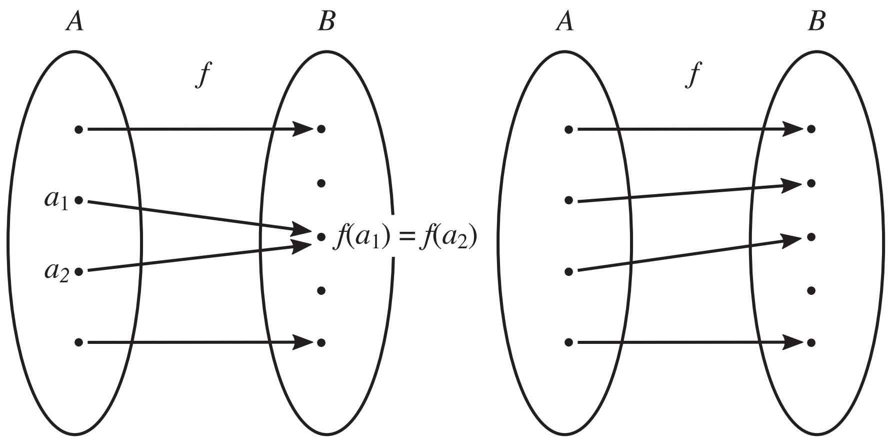
$$\neg\exists a_1 \in A\, \exists a_2 \in A(f(a_1) = f(a_2) \wedge a_1 \neq a_2).$$

We say that f *maps onto* B (or just *is onto* if B is clear from context) if

$$\forall b \in B\, \exists a \in A(f(a) = b).$$

One-to-one functions are sometimes also called *injections*, and onto functions are sometimes called *surjections*.

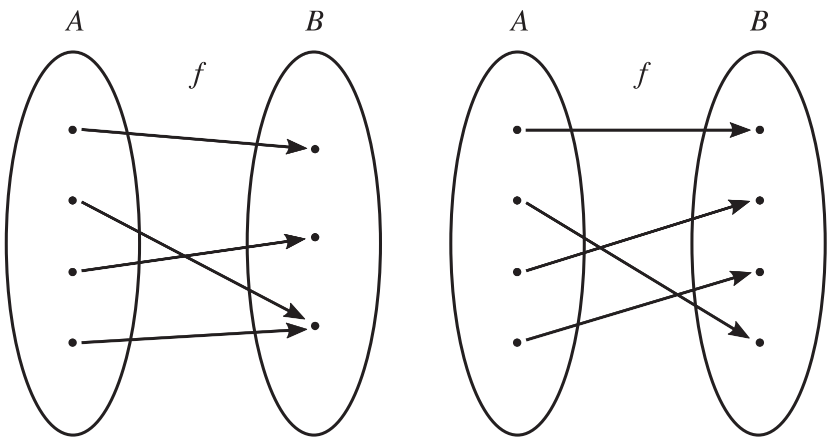
Note that our definition of one-to-one starts with the negation symbol $\neg$. In other words, to say that f is one-to-one means that a certain situation does *not* occur. The situation that must not occur is that there are two different elements of the domain of f, a_1 and a_2, such that $f(a_1) = f(a_2)$. This situation is illustrated in Figure 5.4(a). Thus, the function in Figure 5.4(a) is not one-to-one. Figure 5.4(b) shows a function that is one-to-one.



(a) f is not one-to-one.   (b) f is one-to-one.

Figure 5.4.

If $f : A \to B$, then to say that f is onto means that every element of B is the image under f of some element of A. In other words, in the diagram of f, every element of B has an edge pointing to it. Neither of the functions in Figure 5.4 is onto, because in both cases there are elements of B without edges pointing to them. Figure 5.5 shows two functions that are onto.



(a) f is onto but not one-to-one. (b) f is both one-to-one and onto.

Figure 5.5.

Example 5.2.2. Are the following functions one-to-one? Are they onto?

1. The function F from part 1 of Example 5.1.2.
2. The function L from part 3 of Example 5.1.2.
3. The identity function i_A, for any set A.
4. The function g from part 2 of Example 5.1.3.
5. The function h from part 3 of Example 5.1.3.

Solutions

1. F is not one-to-one because $F(1) = 5 = F(3)$. It is also not onto, because $6 \in B$ but there is no $a \in A$ such that $F(a) = 6$.
2. L is not one-to-one because there are many pairs of different cities c_1 and c_2 for which $L(c_1) = L(c_2)$. For example, $L(\text{Chicago}) = \text{United States} = L(\text{Seattle})$. To say that L is onto means that $\forall n \in N\, \exists c \in C(L(c) = n)$, or in other words, for every country n there is a city c such that the city c is located in the country n. This is probably true, since it is unlikely that there is a country that contains no cities at all. Thus, L is probably onto.

3. To decide whether i_A is one-to-one we must determine whether there are two elements a_1 and a_2 of A such that $i_A(a_1) = i_A(a_2)$ and $a_1 \neq a_2$. But as we saw in Section 5.1, for every $a \in A$, $i_A(a) = a$, so $i_A(a_1) = i_A(a_2)$ means $a_1 = a_2$. Thus, there cannot be elements a_1 and a_2 of A such that $i_A(a_1) = i_A(a_2)$ and $a_1 \neq a_2$, so i_A is one-to-one.

   To say that i_A is onto means that for every $a \in A$, $a = i_A(b)$ for some $b \in A$. This is clearly true because, in fact, $a = i_A(a)$. Thus i_A is also onto.
4. As in solution 3, to decide whether g is one-to-one, we must determine whether there are integers n_1 and n_2 such that $g(n_1) = g(n_2)$ and $n_1 \neq n_2$. According to the definition of g, we have

$$\begin{aligned} g(n_1) = g(n_2) &\text{ iff } 2n_1 + 3 = 2n_2 + 3 \\ &\text{ iff } 2n_1 = 2n_2 \\ &\text{ iff } n_1 = n_2. \end{aligned}$$

   Thus there can be no integers n_1 and n_2 for which $g(n_1) = g(n_2)$ and $n_1 \neq n_2$. In other words, g is one-to-one. However, g is not onto because, for example, there is no integer n for which $g(n) = 0$. To see why, suppose n is an integer and $g(n) = 0$. Then by the definition of g we have $2n + 3 = 0$, so $n = -3/2$. But this contradicts the fact that n is an integer. Note that the domain of g is $\mathbb{Z}$, so for g to be onto it must be the case that for every real number y there is an *integer* n such that $g(n) = y$. Since we have seen that there is no integer n such that $g(n) = 0$, we can conclude that g is not onto.
5. This function is both one-to-one and onto. The verification that h is one-to-one is very similar to the verification in solution 4 that g is one-to-one, and it is left to the reader. To see that h is onto, we must show that $\forall y \in \mathbb{R} \exists x \in \mathbb{R}(h(x) = y)$. Here is a brief proof of this statement. Let y be an arbitrary real number. Let $x = (y-3)/2$. Then $g(x) = 2x + 3 = 2 \cdot ((y-3)/2) + 3 = y - 3 + 3 = y$. Thus, $\forall y \in \mathbb{R} \exists x \in \mathbb{R}(h(x) = y)$, so h is onto.

Although the definition of one-to-one is easiest to understand when it is stated as a negative statement, as in Definition 5.2.1, we know from Chapter 3 that the definition will be easier to use in proofs if we reexpress it as an equivalent positive statement. The following theorem shows how to do this. It also gives a useful equivalence for the definition of onto.

Theorem 5.2.3. *Suppose* $f : A \to B$.

1. f *is one-to-one iff* $\forall a_1 \in A \forall a_2 \in A(f(a_1) = f(a_2) \to a_1 = a_2)$.
2. f *is onto iff* $\text{Ran}(f) = B$.

Proof.

1. We use the rules from Chapters 1 and 2 for reexpressing negative statements as positive ones.

$$\begin{aligned} f \text{ is one-to-one iff } & \neg\exists a_1 \in A\, \exists a_2 \in A(f(a_1) = f(a_2) \wedge a_1 \neq a_2) \\ \text{iff } & \forall a_1 \in A\, \forall a_2 \in A\, \neg(f(a_1) = f(a_2) \wedge a_1 \neq a_2) \\ \text{iff } & \forall a_1 \in A\, \forall a_2 \in A(f(a_1) \neq f(a_2) \vee a_1 = a_2) \\ \text{iff } & \forall a_1 \in A\, \forall a_2 \in A(f(a_1) = f(a_2) \rightarrow a_1 = a_2). \end{aligned}$$

2. First we relate the definition of onto to the definition of range.

$$\begin{aligned} f \text{ is onto iff } & \forall b \in B\, \exists a \in A(f(a) = b) \\ \text{iff } & \forall b \in B\, \exists a \in A((a, b) \in f) \\ \text{iff } & \forall b \in B(b \in \text{Ran}(f)) \\ \text{iff } & B \subseteq \text{Ran}(f). \end{aligned}$$

   Now we are ready to prove part 2 of the theorem.

   ($\rightarrow$) Suppose f is onto. By the equivalence just derived we have $B \subseteq \text{Ran}(f)$, and by the definition of range we have $\text{Ran}(f) \subseteq B$. Thus, it follows that $\text{Ran}(f) = B$.

   ($\leftarrow$) Suppose $\text{Ran}(f) = B$. Then certainly $B \subseteq \text{Ran}(f)$, so by the equivalence, f is onto. □

Commentary. It is often most efficient to write the proof of an iff statement as a string of equivalences, if this can be done. In the case of statement 1 this is easy, using rules of logic. For statement 2 this strategy doesn't quite work, but it does give us an equivalence that turns out to be useful in the proof.

Example 5.2.4. Let $A = \mathbb{R} \setminus \{-1\}$, and define $f : A \rightarrow \mathbb{R}$ by the formula

$$f(a) = \frac{2a}{a+1}.$$

Prove that f is one-to-one but not onto.

Scratch work

By part 1 of Theorem 5.2.3, we can prove that f is one-to-one by proving the equivalent statement $\forall a_1 \in A\, \forall a_2 \in A(f(a_1) = f(a_2) \rightarrow a_1 = a_2)$. Thus, we let a_1 and a_2 be arbitrary elements of A, assume $f(a_1) = f(a_2)$, and then prove $a_1 = a_2$. This is the strategy that is almost always used when proving that a function is one-to-one. The remaining details of the proof involve only simple algebra and are given later.

To show that f is not onto we must prove $\neg\forall x \in \mathbb{R}\exists a \in A(f(a) = x)$. Reexpressing this as a positive statement, we see that we must prove $\exists x \in \mathbb{R}\forall a \in A(f(a) \neq x)$, so we should try to find a particular real number x such that $\forall a \in A(f(a) \neq x)$. Unfortunately, it is not at all clear what value we should use for x. We'll use a somewhat unusual procedure to overcome this difficulty. Instead of trying to prove that f is not onto, let's try to prove that it *is* onto! Of course, we're expecting that this proof won't work, but maybe seeing *why* it won't work will help us figure out what value of x to use in the proof that f is *not* onto.

To prove that f is onto we would have to prove $\forall x \in \mathbb{R}\exists a \in A(f(a) = x)$, so we should let x be an arbitrary real number and try to find some $a \in A$ such that $f(a) = x$. Filling in the definition of f, we see that we must find $a \in A$ such that

$$\frac{2a}{a+1} = x.$$

To find this value of a, we simply solve the equation for a:

$$\frac{2a}{a+1} = x \quad\Rightarrow\quad 2a = ax + x \quad\Rightarrow\quad a(2 - x) = x \quad\Rightarrow\quad a = \frac{x}{2 - x}.$$

Aha! The last step in this derivation wouldn't work if $x = 2$, because then we would be dividing by 0. This is the only value of x that seems to cause trouble when we try to find a value of a for which $f(a) = x$. Perhaps $x = 2$ is the value to use in the proof that f is *not* onto.

Let's return now to the proof that f is not onto. If we let $x = 2$, then to complete the proof we must show that $\forall a \in A(f(a) \neq 2)$. We'll do this by letting a be an arbitrary element of A, assuming $f(a) = 2$, and then trying to derive a contradiction. The remaining details of the proof are not hard.

Solution

Proof. To see that f is one-to-one, let a_1 and a_2 be arbitrary elements of A and assume that $f(a_1) = f(a_2)$. Applying the definition of f, it follows that $2a_1/(a_1 + 1) = 2a_2/(a_2 + 1)$. Thus, $2a_1(a_2 + 1) = 2a_2(a_1 + 1)$. Multiplying out both sides gives us $2a_1a_2 + 2a_1 = 2a_1a_2 + 2a_2$, so $2a_1 = 2a_2$ and therefore $a_1 = a_2$.

To show that f is not onto we will prove that $\forall a \in A(f(a) \neq 2)$. Suppose $a \in A$ and $f(a) = 2$. Applying the definition of f, we get $2a/(a + 1) = 2$. Thus, $2a = 2a + 2$, which is clearly impossible. Thus, $2 \notin \mathrm{Ran}(f)$, so $\mathrm{Ran}(f) \neq \mathbb{R}$ and therefore f is not onto. □

As we saw in the preceding example, when proving that a function f is one-to-one it is usually easiest to prove the equivalent statement $\forall a_1 \in A\, \forall a_2 \in A(f(a_1) = f(a_2) \to a_1 = a_2)$ given in part 1 of Theorem 5.2.3. Of course, this is just an example of the fact that it is generally easier to prove a positive statement than a negative one. This equivalence is also often used in proofs in which we are *given* that a function is one-to-one, as you will see in the proof of part 1 of the following theorem.

Theorem 5.2.5. *Suppose $f : A \to B$ and $g : B \to C$. As we saw in Theorem 5.1.5, it follows that $g \circ f : A \to C$.*

1. *If f and g are both one-to-one, then so is $g \circ f$.*
2. *If f and g are both onto, then so is $g \circ f$.*

Proof.

1. Suppose f and g are both one-to-one. Let a_1 and a_2 be arbitrary elements of A and suppose that $(g \circ f)(a_1) = (g \circ f)(a_2)$. By Theorem 5.1.5 this means that $g(f(a_1)) = g(f(a_2))$. Since g is one-to-one it follows that $f(a_1) = f(a_2)$, and similarly since f is one-to-one we can then conclude that $a_1 = a_2$. Thus, $g \circ f$ is one-to-one.
2. Suppose f and g are both onto, and let c be an arbitrary element of C. Since g is onto, we can find some $b \in B$ such that $g(b) = c$. Similarly, since f is onto, there is some $a \in A$ such that $f(a) = b$. Then $(g \circ f)(a) = g(f(a)) = g(b) = c$. Thus, $g \circ f$ is onto. □

Commentary.

1. As in Example 5.2.4, we prove that $g \circ f$ is one-to-one by proving that $\forall a_1 \in A\, \forall a_2 \in A((g \circ f)(a_1) = (g \circ f)(a_2) \to a_1 = a_2)$. Thus, we let a_1 and a_2 be arbitrary elements of A, assume that $(g \circ f)(a_1) = (g \circ f)(a_2)$, which means $g(f(a_1)) = g(f(a_2))$, and then prove that $a_1 = a_2$. The next sentence of the proof says that the assumption that g is one-to-one is being used, but it might not be clear *how* it is being used. To understand this step, let's write out what it means to say that g is one-to-one. As we observed before, rather than using the original definition, which is a negative statement, we are probably better off using the equivalent positive statement $\forall b_1 \in B\, \forall b_2 \in B(g(b_1) = g(b_2) \to b_1 = b_2)$. The natural way to use a given of this form is to plug something in for b_1 and b_2. Plugging in $f(a_1)$ and $f(a_2)$, we get $g(f(a_1)) = g(f(a_2)) \to f(a_1) = f(a_2)$, and since we know $g(f(a_1)) = g(f(a_2))$, it follows by modus ponens that $f(a_1) = f(a_2)$. None of this was explained in the proof; readers of the

proof are expected to work it out for themselves. Make sure you understand how, using similar reasoning, you can get from $f(a_1) = f(a_2)$ to $a_1 = a_2$ by applying the fact that f is one-to-one.

2. After the assumption that f and g are both onto, the form of the rest of the proof is entirely guided by the logical form of the goal of proving that $g \circ f$ is onto. Because this means $\forall c \in C\, \exists a \in A((g \circ f)(a) = c)$, we let c be an arbitrary element of C and then find some $a \in A$ for which we can prove $(g \circ f)(a) = c$.

Functions that are both one-to-one and onto are particularly important in mathematics. Such functions are sometimes called *one-to-one correspondences* or *bijections*. Figure 5.5(b) shows an example of a one-to-one correspondence. Notice that in this figure both A and B have four elements. In fact, you should be able to convince yourself that if there is a one-to-one correspondence between two finite sets, then the sets must have the same number of elements. This is one of the reasons why one-to-one correspondences are so important. We will discuss one-to-one correspondences between infinite sets in Chapter 8.

Here's another example of a one-to-one correspondence. Suppose A is the set of all members of the audience at a sold-out concert and S is the set of all seats in the concert hall. Let $f : A \to S$ be the function defined by the rule

$$f(a) = \text{the seat in which } a \text{ is sitting.}$$

Because different people would not be sitting in the same seat, f is one-to-one. Because the concert is sold out, every seat is taken, so f is onto. Thus, f is a one-to-one correspondence. Even without counting people or seats, we can tell that the number of people in the audience must be the same as the number of seats in the concert hall.

Exercises

1. Which of the functions in exercise 1 of Section 5.1 are one-to-one? Which are onto?

*2. Which of the functions in exercise 2 of Section 5.1 are one-to-one? Which are onto?

3. Which of the functions in exercise 3 of Section 5.1 are one-to-one? Which are onto?

4. Which of the functions in exercise 4 of Section 5.1 are one-to-one? Which are onto?

*5. Let $A = \mathbb{R} \setminus \{1\}$, and let $f : A \to A$ be defined as follows:

$$f(x) = \frac{x+1}{x-1}.$$

(a) Show that f is one-to-one and onto.

(b) Show that $f \circ f = i_A$.

6. Suppose a and b are real numbers and $a \neq 0$. Define $f : \mathbb{R} \to \mathbb{R}$ by the formula $f(x) = ax + b$. Show that f is one-to-one and onto.

7. Define $f : \mathbb{R}^+ \to \mathbb{R}$ by the formula $f(x) = 1/x - x$.

(a) Show that f is one-to-one. (Hint: You may find it useful to prove first that if $0 < a < b$ then $f(a) > f(b)$.)

(b) Show that f is onto.

(c) Define $g : \mathbb{R}^+ \to \mathbb{R}$ by the formula $g(x) = 1/x + x$. Is g one-to-one? Is it onto?

8. Let $A = \mathscr{P}(\mathbb{R})$. Define $f : \mathbb{R} \to A$ by the formula $f(x) = \{y \in \mathbb{R} \mid y^2 < x\}$.

(a) Find $f(2)$.

(b) Is f one-to-one? Is it onto?

*9. Let $A = \mathscr{P}(\mathbb{R})$ and $B = \mathscr{P}(A)$. Define $f : B \to A$ by the formula $f(\mathcal{F}) = \bigcup \mathcal{F}$.

(a) Find $f(\{\{1, 2\}, \{3, 4\}\})$.

(b) Is f one-to-one? Is it onto?

10. Suppose $f : A \to B$ and $g : B \to C$.

(a) Prove that if $g \circ f$ is onto then g is onto.

(b) Prove that if $g \circ f$ is one-to-one then f is one-to-one.

11. Suppose $f : A \to B$ and $g : B \to C$.

(a) Prove that if f is onto and g is not one-to-one, then $g \circ f$ is not one-to-one.

(b) Prove that if f is not onto and g is one-to-one, then $g \circ f$ is not onto.

12. Suppose $f : A \to B$. Define a function $g : B \to \mathscr{P}(A)$ by the formula $g(b) = \{a \in A \mid f(a) = b\}$. Prove that if f is onto then g is one-to-one. What if f is not onto?

*13. Suppose $f : A \to B$ and $C \subseteq A$. In exercise 7 of Section 5.1 we defined $f \restriction C$ (the restriction of f to C), and you showed that $f \restriction C : C \to B$.

(a) Prove that if f is one-to-one, then so is $f \restriction C$.

(b) Prove that if $f \restriction C$ is onto, then so is f.

(c) Give examples to show that the converses of parts (a) and (b) are not always true.

14. Suppose $f : A \to B$, and there is some $b \in B$ such that $\forall x \in A(f(x) = b)$. (Thus, f is a *constant* function.)

(a) Prove that if A has more than one element then f is not one-to-one.

(b) Prove that if B has more than one element then f is not onto.

15. Suppose $f : A \to C$, $g : B \to C$, and A and B are disjoint. In exercise 12(a) of Section 5.1 you proved that $f \cup g : A \cup B \to C$. Now suppose that f and g are both one-to-one. Prove that $f \cup g$ is one-to-one iff $\text{Ran}(f)$ and $\text{Ran}(g)$ are disjoint.

16. Suppose R is a relation from A to B, S is a relation from B to C, $\text{Ran}(R) = \text{Dom}(S) = B$, and $S \circ R : A \to C$. In exercise 13(a) of Section 5.1 you proved that $S : B \to C$. Now prove that if S is one-to-one then $R : A \to B$.

*17. Suppose $f : A \to B$ and R is a relation on A. As in exercise 15 of Section 5.1, define a relation S on B as follows:

$$S = \{(x, y) \in B \times B \mid \exists u \in A \, \exists v \in A(f(u) = x \wedge f(v) = y \wedge (u, v) \in R)\}.$$

(a) Prove that if R is reflexive and f is onto then S is reflexive.

(b) Prove that if R is transitive and f is one-to-one then S is transitive.

18. Suppose R is an equivalence relation on A, and let $g : A \to A/R$ be defined by the formula $g(x) = [x]_R$, as in exercise 20(b) in Section 5.1.

(a) Show that g is onto.

(b) Show that g is one-to-one iff $R = i_A$.

19. Suppose $f : A \to B$, R is an equivalence relation on A, and f is compatible with R. (See exercise 21 of Section 5.1 for the definition of *compatible*.) In exercise 21(a) of Section 5.1 you proved that there is a unique function $h : A/R \to B$ such that for all $x \in A$, $h([x]_R) = f(x)$. Now prove that h is one-to-one iff $\forall x \in A \, \forall y \in A(f(x) = f(y) \to xRy)$.

*20. Suppose A, B, and C are sets and $f : A \to B$.

(a) Prove that if f is onto, $g : B \to C$, $h : B \to C$, and $g \circ f = h \circ f$, then $g = h$.

(b) Suppose that C has at least two elements, and for all functions g and h from B to C, if $g \circ f = h \circ f$ then $g = h$. Prove that f is onto.

21. Suppose A, B, and C are sets and $f : B \to C$.

(a) Prove that if f is one-to-one, $g : A \to B$, $h : A \to B$, and $f \circ g = f \circ h$, then $g = h$.

(b) Suppose that $A \neq \varnothing$, and for all functions g and h from A to B, if $f \circ g = f \circ h$ then $g = h$. Prove that f is one-to-one.

22. Let $\mathcal{F} = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$, and define a relation R on $\mathcal{F}$ as follows:

$$R = \{(f, g) \in \mathcal{F} \times \mathcal{F} \mid \exists h \in \mathcal{F}(f = h \circ g)\}.$$

(a) Let f, g, and h be the functions from $\mathbb{R}$ to $\mathbb{R}$ defined by the formulas $f(x) = x^2 + 1$, $g(x) = x^3 + 1$, and $h(x) = x^4 + 1$. Prove that hRf, but it is not the case that gRf.

(b) Prove that R is a preorder. (See exercise 25 of Section 4.5 for the definition of *preorder*.)

(c) Prove that for all $f \in \mathcal{F}$, $f R i_{\mathbb{R}}$.

(d) Prove that for all $f \in \mathcal{F}$, $i_{\mathbb{R}} R f$ iff f is one-to-one. (Hint for right-to-left direction: Suppose f is one-to-one. Let $A = \mathrm{Ran}(f)$, and let $h = f^{-1} \cup ((\mathbb{R} \setminus A) \times \{0\})$. Now prove that $h : \mathbb{R} \to \mathbb{R}$ and $i_{\mathbb{R}} = h \circ f$.)

(e) Suppose that $g \in \mathcal{F}$ is a constant function; in other words, there is some real number c such that $\forall x \in \mathbb{R}(g(x) = c)$. Prove that for all $f \in \mathcal{F}$, gRf. (Hint: See exercise 17 of Section 5.1.)

(f) Suppose that $g \in \mathcal{F}$ is a constant function. Prove that for all $f \in \mathcal{F}$, $f R g$ iff f is a constant function.

(g) As in exercise 25 of Section 4.5, if we let $S = R \cap R^{-1}$, then S is an equivalence relation on $\mathcal{F}$. Also, there is a unique relation T on $\mathcal{F}/S$ such that for all f and g in $\mathcal{F}$, $[f]_S T [g]_S$ iff $f R g$, and T is a partial order on $\mathcal{F}/S$. Prove that the set of all one-to-one functions from $\mathbb{R}$ to $\mathbb{R}$ is the largest element of $\mathcal{F}/S$ in the partial order T, and the set of all constant functions from $\mathbb{R}$ to $\mathbb{R}$ is the smallest element.

23. Let $f : \mathbb{N} \to \mathbb{N}$ be defined by the formula $f(n) = n$. Note that we could also say that $f : \mathbb{N} \to \mathbb{Z}$. This exercise will illustrate why, in Definition 5.2.1, we defined the phrase "f maps onto B," rather than simply "f is onto."

(a) Does f map onto $\mathbb{N}$?

(b) Does f map onto $\mathbb{Z}$?

5.3. Inverses of Functions

We are now ready to return to the question of whether the inverse of a function from A to B is always a function from B to A. Consider again the function F from part 1 of Example 5.1.2. Recall that in that example we had $A = \{1, 2, 3\}$, $B = \{4, 5, 6\}$, and $F = \{(1, 5), (2, 4), (3, 5)\}$. As we saw in Example 5.1.2, F is a function from A to B. According to the definition of the inverse of a relation, $F^{-1} = \{(5, 1), (4, 2), (5, 3)\}$, which is clearly a relation from B to A. But F^{-1} fails to be a function from B to A for two reasons. First of all, $6 \in B$, but 6 isn't paired with any element of A in the relation F^{-1}. Second, 5 is paired with two different elements of A, 1 and 3. Thus, this example shows that the inverse of a function from A to B is not always a function from B to A.

You may have noticed that the reasons why F^{-1} isn't a function from B to A are related to the reasons why F is neither one-to-one nor onto, which were discussed in part 1 of Example 5.2.2. This suggests the following theorem.

Theorem 5.3.1. *Suppose $f : A \to B$. If f is one-to-one and onto, then $f^{-1} : B \to A$.*

Proof. Suppose f is one-to-one and onto, and let b be an arbitrary element of B. To show that f^{-1} is a function from B to A, we must prove that $\exists! a \in A((b, a) \in f^{-1})$, so we prove existence and uniqueness separately.

Existence: Since f is onto, there is some $a \in A$ such that $f(a) = b$. Thus, $(a, b) \in f$, so $(b, a) \in f^{-1}$.

Uniqueness: Suppose $(b, a_1) \in f^{-1}$ and $(b, a_2) \in f^{-1}$ for some $a_1, a_2 \in A$. Then $(a_1, b) \in f$ and $(a_2, b) \in f$, so $f(a_1) = b = f(a_2)$. Since f is one-to-one, it follows that $a_1 = a_2$. □

Commentary. The form of the proof is guided by the logical form of the statement that $f^{-1} : B \to A$. Because this means $\forall b \in B \exists! a \in A((b, a) \in f^{-1})$, we let b be an arbitrary element of B and then prove existence and uniqueness for the required $a \in A$ separately. Note that the assumption that f is onto is the key to the existence half of the proof, and the assumption that f is one-to-one is the key to the uniqueness half.

Suppose f is any function from a set A to a set B. Theorem 5.3.1 says that a sufficient condition for f^{-1} to be a function from B to A is that f be one-to-one and onto. Is it also a necessary condition? In other words, is the converse of Theorem 5.3.1 true? (If you don't remember what the words *sufficient*, *necessary*, and *converse* mean, you should review Section 1.5!) We will show in Theorem 5.3.4 that the answer to this question is yes. In other words, if f^{-1} is a function from B to A, then f must be one-to-one and onto.

If $f^{-1} : B \to A$ then, by the definition of function, for every $b \in B$ there is exactly one $a \in A$ such that $(b, a) \in f^{-1}$, and

$$\begin{aligned} f^{-1}(b) &= \text{the unique } a \in A \text{ such that } (b, a) \in f^{-1} \\ &= \text{the unique } a \in A \text{ such that } (a, b) \in f \\ &= \text{the unique } a \in A \text{ such that } f(a) = b. \end{aligned}$$

This gives another useful way to think about f^{-1}. If f^{-1} is a function from B to A, then it is the function that assigns, to each $b \in B$, the unique $a \in A$

such that $f(a) = b$. The assumption in Theorem 5.3.1 that f is one-to-one and onto guarantees that there is exactly one such a.

As an example, consider again the function f that assigns, to each person in the audience at a sold-out concert, the seat in which that person is sitting. As we saw at the end of the last section, f is a one-to-one, onto function from the set A of all members of the audience to the set S of all seats in the concert hall. Thus, f^{-1} must be a function from S to A, and for each $s \in S$,

$$\begin{aligned} f^{-1}(s) &= \text{the unique } a \in A \text{ such that } f(a) = s \\ &= \text{the unique person } a \text{ such that the seat in which } a \text{ is sitting is } s \\ &= \text{the person who is sitting in the seat } s. \end{aligned}$$

In other words, the function f assigns to each person the seat in which that person is sitting, and the function f^{-1} assigns to each seat the person sitting in that seat.

Because $f : A \to S$ and $f^{-1} : S \to A$, it follows by Theorem 5.1.5 that $f^{-1} \circ f : A \to A$ and $f \circ f^{-1} : S \to S$. What are these functions? To figure out what the first function is, let's let a be an arbitrary element of A and compute $(f^{-1} \circ f)(a)$.

$$\begin{aligned} (f^{-1} \circ f)(a) &= f^{-1}(f(a)) \\ &= f^{-1}(\text{the seat in which } a \text{ is sitting}) \\ &= \text{the person sitting in the seat in which } a \text{ is sitting} \\ &= a. \end{aligned}$$

But recall that for every $a \in A$, $i_A(a) = a$. Thus, we have shown that $\forall a \in A((f^{-1} \circ f)(a) = i_A(a))$, so by Theorem 5.1.4, $f^{-1} \circ f = i_A$. Similarly, you should be able to check that $f \circ f^{-1} = i_S$.

When mathematicians find an unusual phenomenon like this in an example, they always wonder whether it's just a coincidence or if it's part of a more general pattern. In other words, can we prove a theorem that says that what happened in this example will happen in other examples too? In this case, it turns out that we can.

Theorem 5.3.2. *Suppose f is a function from A to B, and suppose that f^{-1} is a function from B to A. Then $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$.*

Proof. Let a be an arbitrary element of A. Let $b = f(a) \in B$. Then $(a, b) \in f$, so $(b, a) \in f^{-1}$ and therefore $f^{-1}(b) = a$. Thus,

$$(f^{-1} \circ f)(a) = f^{-1}(f(a)) = f^{-1}(b) = a = i_A(a).$$

Since a was arbitrary, we have shown that $\forall a \in A((f^{-1} \circ f)(a) = i_A(a))$, so $f^{-1} \circ f = i_A$. The proof of the second half of the theorem is similar and is left as an exercise (see exercise 8). □

Commentary. To prove that two functions are equal, we usually apply Theorem 5.1.4. Thus, since $f^{-1} \circ f$ and i_A are both functions from A to A, to prove that they are equal we prove that $\forall a \in A((f^{-1} \circ f)(a) = i_A(a))$.

Theorem 5.3.2 says that if $f : A \to B$ and $f^{-1} : B \to A$, then each function undoes the effect of the other. For any $a \in A$, applying the function f gives us $f(a) \in B$. According to Theorem 5.3.2, $f^{-1}(f(a)) = (f^{-1} \circ f)(a) = i_A(a) = a$. Thus, applying f^{-1} to $f(a)$ undoes the effect of applying f, giving us back the original element a. Similarly, for any $b \in B$, applying f^{-1} we get $f^{-1}(b) \in A$, and we can undo the effect of applying f^{-1} by applying f, since $f(f^{-1}(b)) = b$.

For example, let $f : \mathbb{R} \to \mathbb{R}$ be defined by the formula $f(x) = 2x$. You should be able to check that f is one-to-one and onto, so $f^{-1} : \mathbb{R} \to \mathbb{R}$, and for any $x \in \mathbb{R}$,

$$f^{-1}(x) = \text{the unique } y \text{ such that } f(y) = x.$$

Because $f^{-1}(x)$ is the unique solution for y in the equation $f(y) = x$, we can find a formula for $f^{-1}(x)$ by solving this equation for y. Filling in the definition of f in the equation gives us $2y = x$, so $y = x/2$. Thus, for every $x \in \mathbb{R}$, $f^{-1}(x) = x/2$. Notice that applying f to any number doubles the number and applying f^{-1} halves the number, and each of these operations undoes the effect of the other. In other words, if you double a number and then halve the result, you get back the number you started with. Similarly, halving any number and then doubling the result gives you back the original number.

Are there other circumstances in which the composition of two functions is equal to the identity function? Investigation of this question leads to the following theorem.

Theorem 5.3.3. *Suppose $f : A \to B$.*

1. *If there is a function $g : B \to A$ such that $g \circ f = i_A$ then f is one-to-one.*
2. *If there is a function $g : B \to A$ such that $f \circ g = i_B$ then f is onto.*

Proof.

1. Suppose $g : B \to A$ and $g \circ f = i_A$. Let a_1 and a_2 be arbitrary elements of A, and suppose that $f(a_1) = f(a_2)$. Applying g to both sides of this

equation we get $g(f(a_1)) = g(f(a_2))$. But $g(f(a_1)) = (g \circ f)(a_1) = i_A(a_1) = a_1$, and similarly, $g(f(a_2)) = a_2$. Thus, we can conclude that $a_1 = a_2$, and therefore f is one-to-one.
2. See exercise 9. □

Commentary. The assumption that there is a $g : B \to A$ such that $g \circ f = i_A$ is an existential statement, so we immediately imagine that a particular function g has been chosen. The proof that f is one-to-one follows the usual pattern for such proofs, based on Theorem 5.2.3.

We have come full circle. In Theorem 5.3.1 we found that if f is a one-to-one, onto function from A to B, then f^{-1} is a function from B to A. From this conclusion it follows, as we showed in Theorem 5.3.2, that the composition of f with its inverse must be the identity function. And in Theorem 5.3.3 we found that when the composition of two functions is the identity function, we are led back to the properties one-to-one and onto! Thus, combining Theorems 5.3.1–5.3.3, we get the following theorem.

Theorem 5.3.4. *Suppose $f : A \to B$. Then the following statements are equivalent.*

1. *f is one-to-one and onto.*
2. *$f^{-1} : B \to A$.*
3. *There is a function $g : B \to A$ such that $g \circ f = i_A$ and $f \circ g = i_B$.*

Proof. 1 → 2. This is precisely what Theorem 5.3.1 says.
2 → 3. Suppose $f^{-1} : B \to A$. Let $g = f^{-1}$ and apply Theorem 5.3.2.
3 → 1. Apply Theorem 5.3.3. □

Commentary. As we saw in Section 3.6, the easiest way to prove that several statements are equivalent is to prove a circle of implications. In this case we have proven the circle $1 \to 2 \to 3 \to 1$. Note that the proofs of these implications are quite sketchy. You should make sure you know how to fill in all the details.

For example, let f and g be functions from $\mathbb{R}$ to $\mathbb{R}$ defined by the following formulas:

$$f(x) = \frac{x+7}{5}, \qquad g(x) = 5x - 7.$$

Then for any real number x,

$$(g \circ f)(x) = g(f(x)) = g\left(\frac{x+7}{5}\right) = 5 \cdot \frac{x+7}{5} - 7 = x + 7 - 7 = x.$$

Thus, $g \circ f = i_{\mathbb{R}}$. A similar computation shows that $f \circ g = i_{\mathbb{R}}$. Thus, it follows from Theorem 5.3.4 that f must be one-to-one and onto, and f^{-1} must also be a function from $\mathbb{R}$ to $\mathbb{R}$. What is f^{-1}? Of course, a logical guess would be that $f^{-1} = g$, but this doesn't actually follow from the theorems we've proven. You could check it directly by solving for $f^{-1}(x)$, using the fact that $f^{-1}(x)$ must be the unique solution for y in the equation $f(y) = x$. However, there is no need to check. The next theorem shows that f^{-1} must be equal to g.

Theorem 5.3.5. *Suppose $f : A \to B$, $g : B \to A$, $g \circ f = i_A$, and $f \circ g = i_B$. Then $g = f^{-1}$.*

Proof. By Theorem 5.3.4, $f^{-1} : B \to A$. Therefore, by Theorem 5.3.2, $f^{-1} \circ f = i_A$. Thus,

$$\begin{aligned} g &= i_A \circ g && \text{(exercise 9 of Section 4.3)} \\ &= (f^{-1} \circ f) \circ g \\ &= f^{-1} \circ (f \circ g) && \text{(Theorem 4.2.5)} \\ &= f^{-1} \circ i_B \\ &= f^{-1} && \text{(exercise 9 of Section 4.3).} \end{aligned}$$

□

Commentary. This proof gets the desired conclusion quickly by clever use of previous theorems and exercises. For a more direct but somewhat longer proof, see exercise 10.

Example 5.3.6. In each part, determine whether or not f is one-to-one and onto. If it is, find f^{-1}.

1. Let $A = \mathbb{R} \setminus \{0\}$ and $B = \mathbb{R} \setminus \{2\}$, and define $f : A \to B$ by the formula
$$f(x) = \frac{1}{x} + 2.$$
(Note that for all $x \in A$, $1/x$ is defined and nonzero, so $f(x) \neq 2$ and therefore $f(x) \in B$.)
2. Let $A = \mathbb{R}$ and $B = \{x \in \mathbb{R} \mid x \geq 0\}$, and define $f : A \to B$ by the formula
$$f(x) = x^2.$$

Solutions

1. You can check directly that f is one-to-one and onto, but we won't bother to check. Instead, we'll simply try to find a function $g : B \to A$ such that

$g \circ f = i_A$ and $f \circ g = i_B$. We know by Theorems 5.3.4 and 5.3.5 that if we find such a g, then we can conclude that f is one-to-one and onto and $g = f^{-1}$.

Because we're hoping to have $g = f^{-1}$, we know that for any $x \in B = \mathbb{R} \setminus \{2\}$, $g(x)$ must be the unique $y \in A$ such that $f(y) = x$. Thus, to find a formula for $g(x)$, we solve for y in the equation $f(y) = x$. Filling in the definition of f, we see that the equation we must solve is

$$\frac{1}{y} + 2 = x.$$

Solving this equation we get

$$\frac{1}{y} + 2 = x \quad \Rightarrow \quad \frac{1}{y} = x - 2 \quad \Rightarrow \quad y = \frac{1}{x-2}.$$

Thus, we define $g : B \to A$ by the formula

$$g(x) = \frac{1}{x-2}.$$

(Note that for all $x \in B$, $x \neq 2$, so $1/(x-2)$ is defined and nonzero, and therefore $g(x) \in A$.) Let's check that g has the required properties. For any $x \in A$, we have

$$g(f(x)) = g\left(\frac{1}{x} + 2\right) = \frac{1}{1/x + 2 - 2} = \frac{1}{1/x} = x.$$

Thus, $g \circ f = i_A$. Similarly, for any $x \in B$,

$$f(g(x)) = f\left(\frac{1}{x-2}\right) = \frac{1}{1/(x-2)} + 2 = x - 2 + 2 = x,$$

so $f \circ g = i_B$. Therefore, as we observed earlier, f must be one-to-one and onto, and $g = f^{-1}$.

2. Imitating the solution to part 1, let's try to find a function $g : B \to A$ such that $g \circ f = i_A$ and $f \circ g = i_B$. Because applying f to a number squares the number and we want g to undo the effect of f, a reasonable guess would be to let $g(x) = \sqrt{x}$. Let's see if this works.

   For any $x \in B$ we have

$$f(g(x)) = f(\sqrt{x}) = (\sqrt{x})^2 = x,$$

so $f \circ g = i_B$. But for $x \in A$ we have

$$g(f(x)) = g(x^2) = \sqrt{x^2},$$

and this is *not* always equal to x. For example, $g(f(-3)) = \sqrt{(-3)^2} = \sqrt{9} = 3 \neq -3$. Thus, $g \circ f \neq i_A$. This example illustrates that you must check *both* $f \circ g = i_B$ *and* $g \circ f = i_A$. It is possible for one to work but not the other.

What went wrong? We know that if f^{-1} is a function from B to A, then for any $x \in B$, $f^{-1}(x)$ must be the unique solution for y in the equation $f(y) = x$. Applying the definition of f gives us $y^2 = x$, so $y = \pm\sqrt{x}$. Thus, there is not a *unique* solution for y in the equation $f(y) = x$; there are two solutions. For example, when $x = 9$ we get $y = \pm 3$. In other words, $f(3) = f(-3) = 9$. But this means that f is not one-to-one! Thus, f^{-1} is not a function from B to A.

Functions that undo each other come up often in mathematics. For example, if you are familiar with logarithms, then you will recognize the formulas $10^{\log x} = x$ and $\log 10^x = x$. (We are using base-10 logarithms here.) We can rephrase these formulas in the language of this section by defining functions $f : \mathbb{R} \to \mathbb{R}^+$ and $g : \mathbb{R}^+ \to \mathbb{R}$ as follows:

$$f(x) = 10^x, \qquad g(x) = \log x.$$

Then for any $x \in \mathbb{R}$ we have $g(f(x)) = \log 10^x = x$, and for any $x \in \mathbb{R}^+$, $f(g(x)) = 10^{\log x} = x$. Thus, $g \circ f = i_{\mathbb{R}}$ and $f \circ g = i_{\mathbb{R}^+}$, so $g = f^{-1}$. In other words, the logarithm function is the inverse of the "raise 10 to the power" function.

We saw another example of functions that undo each other in Section 4.5. Suppose A is any set, let $\mathcal{E}$ be the set of all equivalence relations on A, and let $\mathcal{P}$ be the set of all partitions of A. Define a function $f : \mathcal{E} \to \mathcal{P}$ by the formula $f(R) = A/R$, and define another function $g : \mathcal{P} \to \mathcal{E}$ by the formula

$$\begin{aligned} g(\mathcal{F}) &= \text{the equivalence relation determined by } \mathcal{F} \\ &= \bigcup_{X \in \mathcal{F}} (X \times X). \end{aligned}$$

You should verify that the proof of Theorem 4.5.6 shows that $f \circ g = i_{\mathcal{P}}$, and exercise 10 in Section 4.5 shows that $g \circ f = i_{\mathcal{E}}$. Thus, f is one-to-one and onto, and $g = f^{-1}$. One interesting consequence of this is that if A has a finite number of elements, then we can say that the number of equivalence relations on A is exactly the same as the number of partitions of A, even though we don't know what this number is.

Exercises

*1. Let R be the function defined in exercise 2(c) of Section 5.1. In exercise 2 of Section 5.2, you showed that R is one-to-one and onto, so $R^{-1} : P \to P$. If $p \in P$, what is $R^{-1}(p)$?

2. Let F be the function defined in exercise 4(b) of Section 5.1. In exercise 4 of Section 5.2, you showed that F is one-to-one and onto, so $F^{-1} : B \to B$. If $X \in B$, what is $F^{-1}(X)$?

*3. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by the formula

$$f(x) = \frac{2x + 5}{3}.$$

Show that f is one-to-one and onto, and find a formula for $f^{-1}(x)$. (You may want to imitate the method used in the example after Theorem 5.3.2, or in Example 5.3.6.)

4. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by the formula $f(x) = 2x^3 - 3$. Show that f is one-to-one and onto, and find a formula for $f^{-1}(x)$.

*5. Let $f : \mathbb{R} \to \mathbb{R}^+$ be defined by the formula $f(x) = 10^{2-x}$. Show that f is one-to-one and onto, and find a formula for $f^{-1}(x)$.

6. Let $A = \mathbb{R} \setminus \{2\}$, and let f be the function with domain A defined by the formula

$$f(x) = \frac{3x}{x - 2}.$$

(a) Show that f is a one-to-one, onto function from A to B for some set $B \subseteq \mathbb{R}$. What is the set B?

(b) Find a formula for $f^{-1}(x)$.

7. In the example after Theorem 5.3.4, we had $f(x) = (x + 7)/5$ and found that $f^{-1}(x) = 5x - 7$. Let f_1 and f_2 be functions from $\mathbb{R}$ to $\mathbb{R}$ defined by the formulas

$$f_1(x) = x + 7, \qquad f_2(x) = \frac{x}{5}.$$

(a) Show that $f = f_2 \circ f_1$.

(b) According to part 5 of Theorem 4.2.5, $f^{-1} = (f_2 \circ f_1)^{-1} = f_1^{-1} \circ f_2^{-1}$. Verify that this is true by computing $f_1^{-1} \circ f_2^{-1}$ directly.

8. (a) Prove the second half of Theorem 5.3.2 by imitating the proof of the first half.

(b) Give an alternative proof of the second half of Theorem 5.3.2 by applying the first half to f^{-1}.

*9. Prove part 2 of Theorem 5.3.3.

10. Use the following strategy to give an alternative proof of Theorem 5.3.5: Let (b, a) be an arbitrary element of $B \times A$. Assume $(b, a) \in g$ and prove $(b, a) \in f^{-1}$. Then assume $(b, a) \in f^{-1}$ and prove $(b, a) \in g$.

*11. Suppose $f : A \rightarrow B$ and $g : B \rightarrow A$.
   (a) Prove that if f is one-to-one and $f \circ g = i_B$, then $g = f^{-1}$.
   (b) Prove that if f is onto and $g \circ f = i_A$, then $g = f^{-1}$.
   (c) Prove that if $f \circ g = i_B$ but $g \circ f \neq i_A$, then f is onto but not one-to-one, and g is one-to-one but not onto.

12. Suppose $f : A \rightarrow B$ and f is one-to-one. Prove that there is some set $B^{\pm} \subseteq B$ such that $f^{-1} : B^{\pm} \rightarrow A$.

13. Suppose $f : A \rightarrow B$ and f is onto. Let $R = \{(x, y) \in A \times A \mid f(x) = f(y)\}$. By exercise 20(a) of Section 5.1, R is an equivalence relation on A.
   (a) Prove that there is a function $h : A/R \rightarrow B$ such that for all $x \in A$, $h([x]_R) = f(x)$. (Hint: See exercise 21 of Section 5.1.)
   (b) Prove that h is one-to-one and onto. (Hint: See exercise 19 of Section 5.2.)
   (c) It follows from part (b) that $h^{-1} : B \rightarrow A/R$. Prove that for all $b \in B$, $h^{-1}(b) = \{x \in A \mid f(x) = b\}$.
   (d) Suppose $g : B \rightarrow A$. Prove that $f \circ g = i_B$ iff $\forall b \in B(g(b) \in h^{-1}(b))$.

*14. Suppose $f : A \rightarrow B$, $g : B \rightarrow A$, and $f \circ g = i_B$. Let $A^{\pm} = \text{Ran}(g) \subseteq A$.
   (a) Prove that for all $x \in A^{\pm}$, $(g \circ f)(x) = x$.
   (b) Prove that $f \restriction A^{\pm}$ is a one-to-one, onto function from $A^{\pm}$ to B and $g = (f \restriction A^{\pm})^{-1}$. (See exercise 7 of Section 5.1 for the meaning of the notation used here.)

15. Let $B = \{x \in \mathbb{R} \mid x \geq 0\}$. Let $f : \mathbb{R} \rightarrow B$ and $g : B \rightarrow \mathbb{R}$ be defined by the formulas $f(x) = x^2$ and $g(x) = \sqrt{x}$. As we saw in part 2 of Example 5.3.6, $g \neq f^{-1}$. Show that $g = (f \restriction B)^{-1}$. (Hint: See exercise 14.)

*16. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be defined by the formula $f(x) = 4x - x^2$. Let $B = \text{Ran}(f)$.
   (a) Find B.
   (b) Find a set $A \subseteq \mathbb{R}$ such that $f \restriction A$ is a one-to-one, onto function from A to B, and find a formula for $(f \restriction A)^{-1}(x)$. (Hint: See exercise 14.)

17. Suppose A is a set, and let $\mathcal{F} = \{f \mid f : A \rightarrow A\}$ and $\mathcal{P} = \{f \in \mathcal{F} \mid f$ is one-to-one and onto$\}$. Define a relation R on $\mathcal{F}$ as follows:

$$R = \{(f, g) \in \mathcal{F} \times \mathcal{F} \mid \exists h \in \mathcal{P}(f = h^{-1} \circ g \circ h)\}.$$

(a) Prove that R is an equivalence relation.
(b) Prove that if $f\,R\,g$ then $(f \circ f)R(g \circ g)$.
(c) For any $f \in \mathcal{F}$ and $a \in A$, if $f(a) = a$ then we say that a is a *fixed point* of f. Prove that if f has a fixed point and $f\,R\,g$, then g also has a fixed point.

*18. Suppose $f : A \to C$, $g : B \to C$, and g is one-to-one and onto. Prove that there is a function $h : A \to B$ such that $g \circ h = f$.

5.4. Closures

Often in mathematics we work with a function from a set to itself. In that situation, the following concept can be useful.

Definition 5.4.1. Suppose $f : A \to A$ and $C \subseteq A$. We will say that C is *closed under* f if $\forall x \in C(f(x) \in C)$.

Example 5.4.2.

1. Let $A = \{a, b, c, d\}$ and $f = \{(a, c), (b, b), (c, d), (d, c)\}$. Then $f : A \to A$. Let $C_1 = \{a, c, d\}$ and $C_2 = \{a, b\}$. Is C_1 closed under f? Is C_2?
2. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be defined by the formulas $f(x) = x + 1$ and $g(x) = x - 1$. Is $\mathbb{N}$ closed under f? Is it closed under g?
3. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by the formula $f(x) = x^2$. Let $C_1 = \{x \in \mathbb{R} \mid 0 < x < 1\}$ and $C_2 = \{x \in \mathbb{R} \mid 0 < x < 2\}$. Is C_1 closed under f? Is C_2?

Solutions

1. The set C_1 is closed under f, because $f(a) = f(d) = c \in C_1$ and $f(c) = d \in C_1$. However, C_2 is not closed under f, because $a \in C_2$ but $f(a) = c \notin C_2$.
2. For every natural number n, $n + 1$ is also a natural number, so $\mathbb{N}$ is closed under f. However, $\mathbb{N}$ is not closed under g, because $0 \in \mathbb{N}$ but $g(0) = -1 \notin \mathbb{N}$.
3. For every real number x, if $0 < x < 1$ then $0 < x^2 < 1$ (see Example 3.1.2), so C_1 is closed under f. But $1.5 \in C_2$ and $f(1.5) = 1.5^2 = 2.25 \notin C_2$, so C_2 is not closed under f.

We saw in part 2 of Example 5.4.2 that $\mathbb{N}$ is not closed under the function $g : \mathbb{R} \to \mathbb{R}$ defined by the formula $g(x) = x - 1$. Suppose we wanted to add elements to $\mathbb{N}$ to get a set that is closed under g. Since $0 \in \mathbb{N}$, we'd need to add $g(0) = -1$. But if -1 were added to the set, then it would also have to contain $g(-1) = -2$, and if we threw in -2 then we'd also have to add $g(-2) = -3$. Continuing in this way, it should be clear that we'd have to add all of the negative integers to $\mathbb{N}$, giving us the set of all integers, $\mathbb{Z}$. But notice that $\mathbb{Z}$ is closed under g, because for every integer n, $n - 1$ is also an integer. So we have succeeded in our task of enlarging $\mathbb{N}$ to get a set closed under g.

When we enlarged $\mathbb{N}$ to $\mathbb{Z}$, the numbers we added – the negative integers – were numbers that *had* to be added if we wanted the resulting set to be closed under g. It follows that $\mathbb{Z}$ is the smallest set containing $\mathbb{N}$ that is closed under g. We are using the word *smallest* here in exactly the way we defined it in Section 4.4. If we let $\mathcal{F} = \{C \subseteq \mathbb{R} \mid \mathbb{N} \subseteq C$ and C is closed under $g\}$, then $\mathbb{Z}$ is the smallest element of $\mathcal{F}$, where as usual it is understood that we mean smallest in the sense of the subset partial order. In other words, $\mathbb{Z}$ is an element of $\mathcal{F}$, and it's a subset of every element of $\mathcal{F}$. We will say that $\mathbb{Z}$ is the *closure* of $\mathbb{N}$ under g.

Definition 5.4.3. Suppose $f : A \to A$ and $B \subseteq A$. Then the *closure* of B under f is the smallest set $C \subseteq A$ such that $B \subseteq C$ and C is closed under f, if there is such a smallest set. In other words, a set $C \subseteq A$ is the closure of B under f if it has the following properties:

1. $B \subseteq C$.
2. C is closed under f.
3. For every set $D \subseteq A$, if $B \subseteq D$ and D is closed under f then $C \subseteq D$.

According to Theorem 4.4.6, if a set has a smallest element, then it can have only one smallest element. Thus, if a set B has a closure under a function f, then this closure must be unique, so it makes sense to call it *the* closure rather than *a* closure. However, as we saw in Example 4.4.7, some families of sets don't have smallest elements, so it is not immediately clear if sets always have closures under functions. In fact they do, as we will show in our proof of Theorem 5.4.5 below. But first let's look at a few more examples of closures.

Example 5.4.4.

1. In part 1 of Example 5.4.2, the set $C_2 = \{a, b\}$ was not closed under f. What is the closure of C_2 under f?
2. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by the formula $f(x) = x + 1$, and let $B = \{0\}$. What is the closure of B under f?

Solutions

1. Since $a \in C_2$, to get a set closed under f we will need to add in $f(a) = c$. But then we'll also have to add $f(c) = d$, giving us the entire set $A = \{a, b, c, d\}$. Clearly A is closed under f, so the closure of C_2 under f is A.
2. Since $0 \in B$, the closure of B under f will have to contain $f(0) = 1$. But then it must also contain $f(1) = 2$, $f(2) = 3$, $f(3) = 4$, and in fact all positive integers. Adding all the positive integers to B gives us the set $\mathbb{N}$, which we already know from part 2 of Example 5.4.2 is closed under f. Thus the closure of $\{0\}$ under f is $\mathbb{N}$.

Here's an example that illustrates the usefulness of the concepts we have been discussing. Let P be a set of people, and suppose that each person in the set P has a best friend who is also in P. Then we can define a function $f : P \to P$ by let $f(p) = p$'s best friend. Suppose that whenever someone in the set P hears a piece of gossip, he or she tells it to his or her best friend (but no one else). Now consider any set $C \subseteq P$, and suppose that C is closed under f. Then for any person $p \in C$, p's best friend is also in C. Thus, if any person in C hears a piece of gossip, the only person he or she will tell the gossip to is also in C. No one in C will ever transmit gossip to a person who is not in C. Thus, if we tell some people in C a bit of gossip, it may spread to other people in C, but it will never leave C. If you want to track the spread of gossip in this population, you should be interested in recognizing which subsets of P are closed under f.

Suppose we tell a piece of gossip to all of the people in some set $B \subseteq P$. How will the gossip spread? The people in B will tell their best friends, and then they will tell their best friends, who will tell their best friends, and so on. Based on our previous examples, you might guess that the set H of people who eventually hear the gossip will be the closure of B under f. Let's see if we can give a careful proof that H has the three properties listed in Definition 5.4.3.

Clearly $B \subseteq H$, since the people in B hear the gossip right at the start of the process. This confirms property 1 of Definition 5.4.3. If p is any element of H, then p eventually hears the gossip. But as soon as p hears the gossip, he or she will tell $f(p)$, so $f(p) \in H$ as well. Thus H is closed under f, as required by property 2 of the definition. Finally, suppose $B \subseteq C \subseteq P$ and C is closed under f. Then as we observed earlier, any gossip that is told to the people in B may spread to others in C, but it will never leave C. Thus, everyone who ever hears the gossip must belong to C, which means that $H \subseteq C$. This confirms property 3, so H is indeed the closure of B under f.

We turn now to the proof that closures always exist. Suppose $f : A \to A$ and $B \subseteq A$. One way to try to prove the existence of the closure of B under f is to add to B those elements that must be added to make it closed under f, as we did in earlier examples, and then prove that the result is closed under f. Although this can be done, a careful treatment of the details of this proof would require the method of mathematical induction, which we have not yet discussed. We will present this proof in Section 6.5, after we've discussed mathematical induction. But there is another approach to the proof that uses only ideas that we have already studied. We know that the closure of B under f, if it exists, must be the smallest element of the family $\mathcal{F} = \{C \subseteq A \mid B \subseteq C$ and C is closed under $f\}$. According to exercise 20 of Section 4.4, the smallest element of a set is also always the greatest lower bound of the set, and by Theorem 4.4.11, the g.l.b. of any nonempty family of sets $\mathcal{F}$ is $\bigcap \mathcal{F}$. This is the motivation for our next proof.

Theorem 5.4.5. *Suppose that $f : A \to A$ and $B \subseteq A$. Then B has a closure under f.*

Proof. Let $\mathcal{F} = \{C \subseteq A \mid B \subseteq C$ and C is closed under $f\}$. You should be able to check that $A \in \mathcal{F}$, and therefore $\mathcal{F} \neq \varnothing$. Thus, we can let $C = \bigcap \mathcal{F}$, and by exercise 9 of Section 3.3, $C \subseteq A$. We will show that C is the closure of B under f by proving the three properties in Definition 5.4.3.

To prove the first property, suppose $x \in B$. Let D be an arbitrary element of $\mathcal{F}$. Then by the definition of $\mathcal{F}$, $B \subseteq D$, so $x \in D$. Since D was arbitrary, this shows that $\forall D \in \mathcal{F}(x \in D)$, so $x \in \bigcap \mathcal{F} = C$. Thus, $B \subseteq C$.

Next, suppose $x \in C$ and again let D be an arbitrary element of $\mathcal{F}$. Then since $x \in C = \bigcap \mathcal{F}$, $x \in D$. But since $D \in \mathcal{F}$, D is closed under f, so $f(x) \in D$. Since D was arbitrary, we can conclude that $\forall D \in \mathcal{F}(f(x) \in D)$, so $f(x) \in \bigcap \mathcal{F} = C$. Thus, we have shown that C is closed under f, which is the second property in Definition 5.4.3.

Finally, to prove the third property, suppose $B \subseteq D \subseteq A$ and D is closed under f. Then $D \in \mathcal{F}$, and applying exercise 9 of Section 3.3 again we can conclude that $C = \bigcap \mathcal{F} \subseteq D$. □

Commentary. Our goal is $\exists C(C$ is the closure of B under $f)$, so we should begin by defining C. However, the definition $C = \bigcap \mathcal{F}$ doesn't make sense unless we know $\mathcal{F} \neq \varnothing$, so we must prove this first. Because $\mathcal{F} \neq \varnothing$ means $\exists D(D \in \mathcal{F})$, we prove it by giving an example of an element of $\mathcal{F}$. The example is A, so we must prove $A \in \mathcal{F}$. The statement in the proof that "you should be able to check" that $A \in \mathcal{F}$ really does mean that you should do the checking. According to the definition of $\mathcal{F}$, to say that $A \in \mathcal{F}$ means that $A \subseteq A$, $B \subseteq A$, and A is closed under f. You should make sure you see why all three of these statements are true.

Having defined C and verified that $C \subseteq A$, we must prove that C has the three properties in the definition of the closure of B under f. To prove the first statement, $B \subseteq C$, we let x be an arbitrary element of B and prove $x \in C$. Since $C = \bigcap \mathcal{F}$, the goal $x \in C$ means $\forall D \in \mathcal{F}(x \in D)$, so to prove it we let D be an arbitrary element of $\mathcal{F}$ and prove $x \in D$. To prove that C is closed under f, we assume that $x \in C$ and prove $f(x) \in C$. Once again, by the definition of C this goal means $\forall D \in \mathcal{F}(f(x) \in D)$, so we let D be an arbitrary element of $\mathcal{F}$ and prove $f(x) \in D$. Finally, to prove the third goal we assume that $D \subseteq A$, $B \subseteq D$, and D is closed under f and prove $C \subseteq D$. Fortunately, an exercise from an earlier section takes care of this proof.

Closed sets and closures also come up in the study of functions of more than one variable. If $f : A \times A \to A$, then f is called a *function of two variables*. An element of the domain of f would be an ordered pair (x, y), where $x, y \in A$. The result of applying f to this pair should be written $f((x, y))$, but it is customary to leave out one pair of parentheses and just write $f(x, y)$.

Definition 5.4.6. Suppose $f : A \times A \to A$ and $C \subseteq A$. We will say that C is *closed under* f if $\forall x \in C\, \forall y \in C(f(x, y) \in C)$.

Example 5.4.7.

1. Let $f : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ and $g : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ be defined by the formulas $f(x, y) = x/y$ and $g(x, y) = x^y$. Is $\mathbb{Q}^+$ closed under f? Is it closed under g?
2. Let $f : \mathscr{P}(\mathbb{N}) \times \mathscr{P}(\mathbb{N}) \to \mathscr{P}(\mathbb{N})$ and $g : \mathscr{P}(\mathbb{N}) \times \mathscr{P}(\mathbb{N}) \to \mathscr{P}(\mathbb{N})$ be defined by the formulas $f(X, Y) = X \cup Y$ and $g(X, Y) = X \cap Y$. Let $\mathcal{I} = \{X \in \mathscr{P}(\mathbb{N}) \mid X \text{ is infinite}\}$. Is $\mathcal{I}$ closed under f? Is it closed under g?

Solutions

1. If $x, y \in \mathbb{Q}^+$, then there are positive integers p, q, r, and s such that $x = p/q$ and $y = r/s$. Therefore
$$f(x, y) = \frac{x}{y} = \frac{p/q}{r/s} = \frac{p}{q} \cdot \frac{s}{r} = \frac{ps}{qr} \in \mathbb{Q}^+.$$
This shows that $\mathbb{Q}^+$ is closed under f. However, 2 and $1/2$ are elements of $\mathbb{Q}^+$ and $g(2, 1/2) = 2^{1/2} = \sqrt{2} \notin \mathbb{Q}^+$ (see Theorem 6.4.5), so $\mathbb{Q}^+$ is not closed under g.

2. If X and Y are infinite sets of natural numbers, then $f(X, Y) = X \cup Y$ is also infinite, so $\mathcal{I}$ is closed under f. On the other hand, let E be the set of even natural numbers and let P be the set of prime numbers. Then E and P are both infinite, but $g(E, P) = E \cap P = \{2\}$, which is finite. Therefore $\mathcal{I}$ is not closed under g.

As before, we can define the closure of a set under a function of two variables to be the smallest closed set containing it, and we can prove that such closures always exist.

Definition 5.4.8. Suppose $f : A \times A \to A$ and $B \subseteq A$. Then the *closure of B under f* is the smallest set $C \subseteq A$ such that $B \subseteq C$ and C is closed under f, if there is such a smallest set. In other words, a set $C \subseteq A$ is the closure of B under f if it has the following properties:

1. $B \subseteq C$.
2. C is closed under f.
3. For every set $D \subseteq A$, if $B \subseteq D$ and D is closed under f then $C \subseteq D$.

Theorem 5.4.9. *Suppose that $f : A \times A \to A$ and $B \subseteq A$. Then B has a closure under f.*

Proof. See exercise 11. □

A function from $A \times A$ to A could be thought of as an operation that can be applied to a pair of objects $(x, y) \in A \times A$ to produce another element of A. Often in mathematics an operation to be performed on a pair of mathematical objects (x, y) is represented by a symbol that we write between x and y. For example, if x and y are real numbers then $x + y$ denotes another number, and if x and y are sets then $x \cup y$ is another set. Imitating this notation, when mathematicians define a function from $A \times A$ to A they sometimes represent it with a symbol rather than a letter, and they write the result of applying the function to a pair (x, y) by putting the symbol between x and y, rather than by putting a letter before (x, y). When a function from $A \times A$ to A is written in this way, it is usually called a *binary operation on A*.

For example, in part 2 of Example 5.4.7 we defined $g : \mathscr{P}(\mathbb{N}) \times \mathscr{P}(\mathbb{N}) \to \mathscr{P}(\mathbb{N})$ by the formula $g(X, Y) = X \cap Y$. Instead of introducing the name g for this function, we could have talked about $\cap$ as a binary operation on $\mathscr{P}(\mathbb{N})$. We showed in the example that the set $\mathcal{I}$ of all infinite subsets of $\mathbb{N}$ is not closed under g. Another way to say this is that $\mathcal{I}$ is not closed under the binary operation $\cap$. What is the closure of $\mathcal{I}$ under $\cap$? For the answer, see exercise 16.

Here's another example. We could define a binary operation $*$ on $\mathbb{Z}$ by saying that for any integers x and y, $x * y = x^2 - y^2$. Is the set $\{0, 1\}$ closed under the binary operation $*$? The answer is no, because $0 * 1 = 0^2 - 1^2 = -1 \notin \{0, 1\}$. Thus, the closure of $\{0, 1\}$ under $*$ must include -1. But as you can easily check, $\{-1, 0, 1\}$ is closed under $*$. Therefore the closure of $\{0, 1\}$ under $*$ is $\{-1, 0, 1\}$.

Exercises

*1. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by the formula $f(x) = (x + 1)/2$. Are the following sets closed under f?
   (a) $\mathbb{Z}$.
   (b) $\mathbb{Q}$.
   (c) $\{x \in \mathbb{R} \mid 0 \leq x < 4\}$.
   (d) $\{x \in \mathbb{R} \mid 2 \leq x < 4\}$.

2. Let $f : \mathscr{P}(\mathbb{N}) \to \mathscr{P}(\mathbb{N})$ be defined by the formula $f(X) = X \cup \{17\}$. Are the following sets closed under f?
   (a) $\{X \subseteq \mathbb{N} \mid X$ is infinite$\}$.
   (b) $\{X \subseteq \mathbb{N} \mid X$ is finite$\}$.
   (c) $\{X \subseteq \mathbb{N} \mid X$ has at most 100 elements$\}$.
   (d) $\{X \subseteq \mathbb{N} \mid 16 \in X\}$.

*3. Let $f : \mathbb{Z} \to \mathbb{Z}$ be defined by the formula $f(n) = n^2 - n$. Find the closure of $\{-1, 1\}$ under f.

4. For any set A, the set of all relations on A is $\mathscr{P}(A \times A)$. Let $f : \mathscr{P}(A \times A) \to \mathscr{P}(A \times A)$ be defined by the formula $f(R) = R^{-1}$. Is the set of reflexive relations on A closed under f? What about the set of symmetric relations and the set of transitive relations? (Hint: See exercise 12 of Section 4.3.)

5. Suppose $f : A \to A$. Is $\varnothing$ closed under f?

6. Suppose $f : A \to A$.
   (a) Prove that if $\mathrm{Ran}(f) \subseteq C \subseteq A$ then C is closed under f.
   (b) Prove that for every set $B \subseteq A$, the closure of B under f is a subset of $B \cup \mathrm{Ran}(f)$.

*7. Suppose $f : A \to A$ and f is one-to-one and onto. Then by Theorem 5.3.1, $f^{-1} : A \to A$. Prove that if $C \subseteq A$ and C is closed under f, then $A \setminus C$ is closed under f^{-1}.

8. Suppose $f : A \to A$ and $C \subseteq A$. Prove that C is closed under f iff the closure of C under f is C.

*9. Suppose $f : A \to A$ and C_1 and C_2 are subsets of A that are closed under f.
(a) Prove that $C_1 \cup C_2$ is closed under f.
(b) Must $C_1 \cap C_2$ be closed under f? Justify your answer.
(c) Must $C_1 \setminus C_2$ be closed under f? Justify your answer.

10. Suppose $f : A \to A$, $B_1 \subseteq A$, and $B_2 \subseteq A$. Let C_1 be the closure of B_1 under f, and let C_2 be the closure of B_2.
(a) Prove that if $B_1 \subseteq B_2$ then $C_1 \subseteq C_2$.
(b) Prove that the closure of $B_1 \cup B_2$ under f is $C_1 \cup C_2$.
(c) Must the closure of $B_1 \cap B_2$ be $C_1 \cap C_2$? Justify your answer.
(d) Must the closure of $B_1 \setminus B_2$ be $C_1 \setminus C_2$? Justify your answer.

11. Prove Theorem 5.4.9.

*12. If $\mathcal{F}$ is a set of functions from A to A and $C \subseteq A$, then we will say that C is *closed under* $\mathcal{F}$ if $\forall f \in \mathcal{F} \forall x \in C(f(x) \in C)$. In other words, C is closed under $\mathcal{F}$ iff for all $f \in \mathcal{F}$, C is closed under f. If $B \subseteq A$, then the *closure* of B under $\mathcal{F}$ is the smallest set $C \subseteq A$ such that $B \subseteq C$ and C is closed under $\mathcal{F}$. (You are asked to prove in the next exercise that the closure always exists.)
(a) Let f and g be the functions from $\mathbb{R}$ to $\mathbb{R}$ defined by the formulas $f(x) = x + 1$ and $g(x) = x - 1$. Find the closure of $\{0\}$ under $\{f, g\}$.
(b) For each natural number n, let $f_n : \mathscr{P}(\mathbb{N}) \to \mathscr{P}(\mathbb{N})$ be defined by the formula $f_n(X) = X \cup \{n\}$, and let $\mathcal{F} = \{f_n \mid n \in \mathbb{N}\}$. Find the closure of $\{\varnothing\}$ under $\mathcal{F}$.

13. Suppose $\mathcal{F}$ is a set of functions from A to A and $B \subseteq A$. See the previous exercise for the definition of the closure of B under $\mathcal{F}$.
(a) Prove that B has a closure under $\mathcal{F}$.
For each $f \in \mathcal{F}$, let C_f be the closure of B under f, and let C be the closure of B under $\mathcal{F}$.
(b) Prove that $\bigcup_{f \in \mathcal{F}} C_f \subseteq C$.
(c) Must $\bigcup_{f \in \mathcal{F}} C_f$ be closed under $\mathcal{F}$? Justify your answer with either a proof or a counterexample.
(d) Must $\bigcup_{f \in \mathcal{F}} C_f = C$? Justify your answer with either a proof or a counterexample.

*14. Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by the formula $f(x, y) = x - y$. What is the closure of $\mathbb{N}$ under f?

15. Let $f : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ be defined by the formula $f(x, y) = x/y$. What is the closure of $\mathbb{Z}^+$ under f?

16. As in part 2 of Example 5.4.7, let $\mathcal{I} = \{X \in \mathscr{P}(\mathbb{N}) \mid X \text{ is infinite}\}$.
(a) Prove that for every set $X \subseteq \mathbb{N}$ there are sets $Y, Z \in \mathcal{I}$ such that $Y \cap Z = X$.

(b) What is the closure of $\mathcal{I}$ under the binary operation $\cap$?

*17. Let $\mathcal{F} = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$. Then for any $f, g \in \mathcal{F}$, $f \circ g \in \mathcal{F}$, so $\circ$ is a binary operation on $\mathcal{F}$. Are the following sets closed under $\circ$?

(a) $\{f \in \mathcal{F} \mid f$ is one-to-one$\}$. (Hint: See Theorem 5.2.5.)

(b) $\{f \in \mathcal{F} \mid f$ is onto$\}$.

(c) $\{f \in \mathcal{F} \mid f$ is strictly increasing$\}$. (A function $f : \mathbb{R} \to \mathbb{R}$ is *strictly increasing* if $\forall x \in \mathbb{R}\, \forall y \in \mathbb{R}(x < y \to f(x) < f(y))$.)

(d) $\{f \in \mathcal{F} \mid f$ is strictly decreasing$\}$. (A function $f : \mathbb{R} \to \mathbb{R}$ is *strictly decreasing* if $\forall x \in \mathbb{R}\, \forall y \in \mathbb{R}(x < y \to f(x) > f(y))$.)

18. Let $\mathcal{F} = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$. If $f, g \in \mathcal{F}$, then we define the function $f + g : \mathbb{R} \to \mathbb{R}$ by the formula $(f + g)(x) = f(x) + g(x)$. Note that $+$ is a binary operation on $\mathcal{F}$. Are the following sets closed under $+$?

(a) $\{f \in \mathcal{F} \mid f$ is one-to-one$\}$.

(b) $\{f \in \mathcal{F} \mid f$ is onto$\}$.

(c) $\{f \in \mathcal{F} \mid f$ is strictly increasing$\}$. (See the previous exercise for the definition of strictly increasing.)

(d) $\{f \in \mathcal{F} \mid f$ is strictly decreasing$\}$. (See the previous exercise for the definition of strictly decreasing.)

19. For any set A, the set of all relations on A is $\mathscr{P}(A \times A)$, and $\circ$ is a binary operation on $\mathscr{P}(A \times A)$. Is the set of reflexive relations on A closed under $\circ$? What about the set of symmetric relations and the set of transitive relations?

*20. Division is not a binary operation on $\mathbb{R}$, because you can't divide by 0. But we can fix this problem. We begin by adding a new element to $\mathbb{R}$. We will call the new element "NaN" (for "Not a Number"). Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{\text{NaN}\}$, and define $f : \overline{\mathbb{R}} \times \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ as follows:

$$f(x, y) = \begin{cases} x/y, & \text{if } x, y \in \mathbb{R} \text{ and } y \neq 0, \\ \text{NaN}, & \text{otherwise.} \end{cases}$$

This notation means that if $x, y \in \mathbb{R}$ and $y \neq 0$ then $f(x, y) = x/y$, and otherwise $f(x, y) = \text{NaN}$. Thus, for example, $f(3, 7) = 3/7$, $f(3, 0) = \text{NaN}$, and $f(\text{NaN}, 7) = \text{NaN}$. Which of the following sets are closed under f?

(a) $\mathbb{R}$.

(b) $\mathbb{R}^+$.

(c) $\mathbb{R}^-$.

(d) $\mathbb{Q}$.

(e) $\mathbb{Q} \cup \{\text{NaN}\}$.

21. If $\mathcal{F}$ is a set of functions from $A \times A$ to A and $C \subseteq A$, then we will say that C is *closed under* $\mathcal{F}$ if $\forall f \in \mathcal{F} \forall x \in C \forall y \in C(f(x, y) \in C)$. In other words, C is closed under $\mathcal{F}$ iff for all $f \in \mathcal{F}$, C is closed under f. If $B \subseteq A$, then the *closure* of B under $\mathcal{F}$ is the smallest set $C \subseteq A$ such that $B \subseteq C$ and C is closed under $\mathcal{F}$, if there is such a smallest set. (Compare these definitions to the definitions in exercise 12.)
    (a) Prove that the closure of B under $\mathcal{F}$ exists.
    (b) Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by the formulas $f(x, y) = x + y$ and $g(x, y) = xy$. Prove that the closure of $\mathbb{Q} \cup \{\sqrt{2}\}$ under $\{f, g\}$ is the set $\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. (This set is called $\mathbb{Q}$ *with* $\sqrt{2}$ *adjoined*, and is denoted $\mathbb{Q}(\sqrt{2})$.)
    (c) With f and g defined as in part (b), what is the closure of $\mathbb{Q} \cup \{\sqrt[3]{2}\}$ under $\{f, g\}$?

5.5. Images and Inverse Images: A Research Project

Suppose $f : A \to B$. We have already seen that we can think of f as matching each element of A with exactly one element of B. In this section we will see that f can also be thought of as matching *subsets* of A with subsets of B and vice-versa.

Definition 5.5.1. Suppose $f : A \to B$ and $X \subseteq A$. Then the *image* of X under f is the set $f(X)$ defined as follows:

$$\begin{aligned} f(X) &= \{f(x) \mid x \in X\} \\ &= \{b \in B \mid \exists x \in X(f(x) = b)\}. \end{aligned}$$

(Note that the image of the whole domain A under f is $\{f(a) \mid a \in A\}$, and as we saw in Section 5.1 this is the same as the range of f.)

If $Y \subseteq B$, then the *inverse image* of Y under f is the set $f^{-1}(Y)$ defined as follows:

$$f^{-1}(Y) = \{a \in A \mid f(a) \in Y\}.$$

Note that the function f in Definition 5.5.1 may fail to be one-to-one or onto, and as a result f^{-1} may not be a function from B to A, and for $y \in B$, the notation "$f^{-1}(y)$" may be meaningless. However, even in this case Definition 5.5.1 still assigns a meaning to the notation "$f^{-1}(Y)$" for $Y \subseteq B$. If you find this surprising, look again at the definition of $f^{-1}(Y)$, and notice that it does not treat f^{-1} as a function. The definition refers only to the results of applying f to elements of A, not the results of applying f^{-1} to elements of B.

For example, let L be the function defined in part 3 of Example 5.1.2, which assigns to each city the country in which that city is located. As in Example 5.1.2, let C be the set of all cities and N the set of all countries. If B is the set of all cities with population at least one million, then B is a subset of C, and the image of B under L would be the set

$$\begin{aligned} L(B) &= \{L(b) \mid b \in B\} \\ &= \{n \in N \mid \exists b \in B(L(b) = n)\} \\ &= \{n \in N \mid \text{there is some city with population at least} \\ &\qquad \text{one million that is located in the country } n\}. \end{aligned}$$

Thus, $L(B)$ is the set of all countries that contain a city with population at least one million. Now let A be the subset of N consisting of all countries in Africa. Then the inverse image of A under L is the set

$$\begin{aligned} L^{-1}(A) &= \{c \in C \mid L(c) \in A\} \\ &= \{c \in C \mid \text{the country in which } c \text{ is located is in Africa}\}. \end{aligned}$$

Thus, $L^{-1}(A)$ is the set of all cities in African countries.

Let's do one more example. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by the formula $f(x) = x^2$, and let $X = \{x \in \mathbb{R} \mid 0 \leq x < 2\}$. Then

$$f(X) = \{f(x) \mid x \in X\} = \{x^2 \mid 0 \leq x < 2\}.$$

Thus, $f(X)$ is the set of all squares of real numbers between 0 and 2 (including 0 but not 2). A moment's reflection should convince you that this set is $\{x \in \mathbb{R} \mid 0 \leq x < 4\}$. Now let's let $Y = \{x \in \mathbb{R} \mid 0 \leq x < 4\}$ and compute $f^{-1}(Y)$. According to the definition of inverse image,

$$\begin{aligned} f^{-1}(Y) &= \{x \in \mathbb{R} \mid f(x) \in Y\} \\ &= \{x \in \mathbb{R} \mid 0 \leq f(x) < 4\} \\ &= \{x \in \mathbb{R} \mid 0 \leq x^2 < 4\} \\ &= \{x \in \mathbb{R} \mid -2 < x < 2\}. \end{aligned}$$

By now you have had enough experience writing proofs that you should be ready to put your proof-writing skills to work in answering mathematical questions. Thus, most of this section will be devoted to a research project in which you will discover for yourself the answers to basic mathematical questions about images and inverse images. To get you started, we'll work out the answer to the first question.

Suppose $f : A \to B$, and W and X are subsets of A. A natural question you might ask is whether or not $f(W \cap X)$ must be the same as $f(W) \cap f(X)$. It seems plausible that the answer is yes, so let's see if we can prove it. Thus, our goal will be to prove that $f(W \cap X) = f(W) \cap f(X)$. Because this is an equation between two sets, we proceed by taking an arbitrary element of each set and trying to prove that it is an element of the other.

Suppose first that y is an arbitrary element of $f(W \cap X)$. By the definition of $f(W \cap X)$, this means that $y = f(x)$ for some $x \in W \cap X$. Since $x \in W \cap X$, it follows that $x \in W$ and $x \in X$. But now we have $y = f(x)$ and $x \in W$, so we can conclude that $y \in f(W)$. Similarly, since $y = f(x)$ and $x \in X$, it follows that $y \in f(X)$. Thus, $y \in f(W) \cap f(X)$. This completes the first half of the proof.

Now suppose that $y \in f(W) \cap f(X)$. Then $y \in f(W)$, so there is some $w \in W$ such that $f(w) = y$, and also $y \in f(X)$, so there is some $x \in X$ such that $y = f(x)$. If only we knew that w and x were equal, we could conclude that $w = x \in W \cap X$, so $y = f(x) \in f(W \cap X)$. But the best we can do is to say that $f(w) = y = f(x)$. This should remind you of the definition of one-to-one. If we knew that f was one-to-one, we could conclude from the fact that $f(w) = f(x)$ that $w = x$, and the proof would be done. But without this information we seem to be stuck.

Let's summarize what we've discovered. First of all, the first half of the proof worked fine, so we can certainly say that in general $f(W \cap X) \subseteq f(W) \cap f(X)$. The second half worked *if* we knew that f was one-to-one, so we can also say that if f is one-to-one, then $f(W \cap X) = f(W) \cap f(X)$. But what if f isn't one-to-one? There might be some way of fixing up the proof to show that the equation $f(W \cap X) = f(W) \cap f(X)$ is still true even if f isn't one-to-one. But by now you have probably come to suspect that perhaps $f(W \cap X)$ and $f(W) \cap f(X)$ are not always equal, so maybe we should devote some time to trying to show that the proposed theorem is incorrect. In other words, let's see if we can find a counterexample – an example of a function f and sets W and X for which $f(W \cap X) \neq f(W) \cap f(X)$.

Fortunately, we can do better than just trying examples at random. Of course, we know we'd better use a function that isn't one-to-one, but by examining our attempt at a proof, we can tell more than that. The attempted proof that $f(W \cap X) = f(W) \cap f(X)$ ran into trouble only when W and X contained elements w and x such that $w \neq x$ but $f(w) = f(x)$, so we should choose an example in which this happens. In other words, not only should we make sure f isn't one-to-one, we should also make sure W and X contain elements that *show* that f isn't one-to-one.

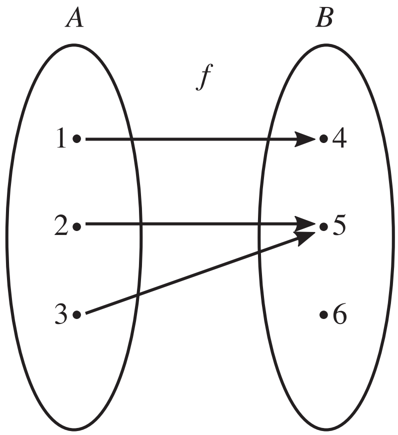


Figure 5.6.

The graph in Figure 5.6 shows a simple function that isn't one-to-one. Writing it as a set of ordered pairs, we could say $f = \{(1,4),(2,5),(3,5)\}$ and $f : A \to B$, where $A = \{1,2,3\}$ and $B = \{4,5,6\}$. The two elements of A that show that f is not one-to-one are 2 and 3, so these should be elements of W and X, respectively. Why not just try letting $W = \{2\}$ and $X = \{3\}$? With these choices we get $f(W) = \{f(2)\} = \{5\}$ and $f(X) = \{f(3)\} = \{5\}$, so $f(W) \cap f(X) = \{5\} \cap \{5\} = \{5\}$. But $f(W \cap X) = f(\varnothing) = \varnothing$, so $f(W \cap X) \neq f(W) \cap f(X)$. (If you're not sure why $f(\varnothing) = \varnothing$, work it out using Definition 5.5.1!) If you want to see an example in which $W \cap X \neq \varnothing$, try $W = \{1,2\}$ and $X = \{1,3\}$.

This example shows that it would be incorrect to state a theorem saying that $f(W \cap X)$ and $f(W) \cap f(X)$ are always equal. But our proof shows that the following theorem is correct:

Theorem 5.5.2. *Suppose $f : A \to B$, and W and X are subsets of A. Then $f(W \cap X) \subseteq f(W) \cap f(X)$. Furthermore, if f is one-to-one, then $f(W \cap X) = f(W) \cap f(X)$.*

Now, here are some questions for you to try to answer. In each case, try to figure out as much as you can. Justify your answers with proofs and counterexamples.

1. Suppose $f : A \to B$ and W and X are subsets of A.
   (a) Will it always be true that $f(W \cup X) = f(W) \cup f(X)$?
   (b) Will it always be true that $f(W \setminus X) = f(W) \setminus f(X)$?
   (c) Will it always be true that $W \subseteq X \leftrightarrow f(W) \subseteq f(X)$?

2. Suppose $f : A \to B$ and Y and Z are subsets of B.
   (a) Will it always be true that $f^{-1}(Y \cap Z) = f^{-1}(Y) \cap f^{-1}(Z)$?
   (b) Will it always be true that $f^{-1}(Y \cup Z) = f^{-1}(Y) \cup f^{-1}(Z)$?
   (c) Will it always be true that $f^{-1}(Y \setminus Z) = f^{-1}(Y) \setminus f^{-1}(Z)$?
   (d) Will it always be true that $Y \subseteq Z \leftrightarrow f^{-1}(Y) \subseteq f^{-1}(Z)$?
3. Suppose $f : A \to B$ and $X \subseteq A$. Will it always be true that $f^{-1}(f(X)) = X$?
4. Suppose $f : A \to B$ and $Y \subseteq B$. Will it always be true that $f(f^{-1}(Y)) = Y$?
5. Suppose $f : A \to A$ and $C \subseteq A$. Prove that the following statements are equivalent:
   (a) C is closed under f.
   (b) $f(C) \subseteq C$.
   (c) $C \subseteq f^{-1}(C)$.
6. Suppose $f : A \to B$ and $g : B \to C$. Can you prove any interesting theorems about images and inverse images of sets under $g \circ f$?

Note: An observant reader may have noticed an ambiguity in our notation for images and inverse images. If $f : A \to B$ and $Y \subseteq B$, then we have used the notation $f^{-1}(Y)$ to stand for the inverse image of Y under f. But if f is one-to-one and onto, then, as we saw in Section 5.3, f^{-1} is a function from B to A. Thus, $f^{-1}(Y)$ could also be interpreted as the image of Y under the function f^{-1}. Fortunately, this ambiguity is harmless, as the next problem shows.

7. Suppose $f : A \to B$, f is one-to-one and onto, and $Y \subseteq B$. Show that the inverse image of Y under f and the image of Y under f^{-1} are equal. (Hint: First write out the definitions of the two sets carefully!)