Chapter 6

Functions

> *Home computers are being called upon to perform many new functions, including the consumption of homework formerly eaten by the dog.*
>
> \- Doug Larson.

We regularly want to associate to each value of one set A some particular value taken from another set B (which may be the same set A). Such a mapping of values in A to values in B is referred to as a *function*.

Functions arise everywhere in people's lives. For example, shoppers are ever calculating (or at least estimating) for themselves the cost of their basket of goods from the number and unit costs (plus relevant sales taxes). Functions are especially relevant to the computer scientist's world. Computer programs are written to turn input values into output values, and the design and implementation of Boolean circuits will inevitably start from a definition of the function of the circuit which describes its behaviour on each possible input. For this reason, it is necessary to take a careful look at what a function is and understand its definition and the various properties that it may enjoy.

6.1 Basic Definitions

A ***function*** f from a set A to a set B is an assignment of exactly one element of B to each element of A. We write $f : A \to B$ to denote that f is a function from A to B, and we write $f(a)$ to refer to the unique element of B assigned to the element a of A by the function f. Thus f maps each element a of A to an element $b = f(a)$ of B, which we will also denote by $f : a \mapsto b$. Figure 6.1 gives a pictorial representation of such a function.

Example 6.1

Each person in a class of twelve students is assigned a particular grade, in the form of an integer percentage between 0 and 100, which appears as

F. Moller, G. Struth, *Modelling Computing Systems*,
Undergraduate Topics in Computer Science,
DOI 10.1007/978-1-84800-322-4_7, © Springer-Verlag London 2013

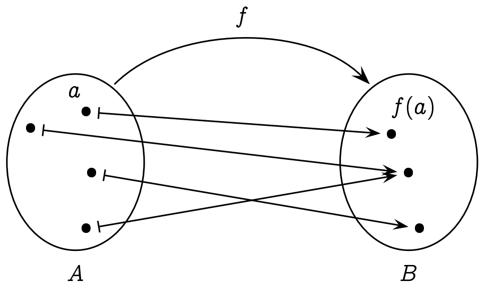


Figure 6.1: A function $f : A \to B$ from A to B.

follows on a list posted on a bulletin board:

Andrews	75	Evans	78	Parker	64
Archer	92	Fletcher	46	Smith	59
Collins	64	Greene	68	Taylor	100
Davies	88	Lewis	54	Williams	78

Here, each person in the set

$$\begin{aligned}\mathsf{Class} = \{\,&\text{Andrews, Archer, Collins, Davies, Evans, Fletcher,}\\ &\text{Greene, Lewis, Parker Smith, Taylor, Williams}\,\}\end{aligned}$$

is assigned a value from the set

$$\mathsf{Marks} = \{\,0, 1, 2, 3, 4, \ldots, 100\,\}.$$

This describes a function

$$\mathsf{score} : \mathsf{Class} \to \mathsf{Marks}$$

in which, for example, score(Greene) = 68; the function score maps the value Greene to the value 68, that is, score : Greene $\mapsto$ 68.

It is possible for a function $f : A \to B$ to assign the same value from B to two different values of A. In the above example,

$$\text{score(Collins)} = \text{score(Parker)} = 64.$$

However, only one value of B may be assigned to any value of A. In this sense, a function $f : A \to B$ may be viewed as a machine into which you input a value $x \in A$ and – depending only on that value – some value $f(x) \in B$ will be output in response:

$$\text{input}\quad x \longrightarrow \boxed{\;f\;} \longrightarrow f(x)\quad\text{output}$$

If the same value for x is input on two separate occasions on the left, then the same value will be output on the right for $f(x)$ on both occasions.

If $f : A \to B$ is a function from A to B, we refer to A as the ***domain*** of f and B as its ***codomain***. If $f(a) = b$ we refer to a as an ***argument*** of the function f, and to b as the ***value*** of the function f on argument a.

If the domain of the function f is the Cartesian product $A_1 \times A_2 \times \cdots \times A_n$, then we say that f has ***arity*** n or that f takes n arguments. A function which takes two arguments is called a ***binary function***. Common binary functions are often written in ***infix*** form $x\,f\,y$ rather than $f(x, y)$. For example, we would naturally write $2 + 2 = 4$ rather than $+(2, 2) = 4$.

The ***range*** of the function $f : A \to B$, denoted $\mathsf{range}(f)$, is the subset of the codomain B consisting of all values that the function f can produce:

$$\mathsf{range}(f) \;=\; \{\, f(a) \;:\; a \in A \,\}.$$

Given a subset $S \subseteq A$ of the domain of f, the ***image*** of S under f, denoted by $f(S)$, is the subset of the codomain B consisting of all values that the function f can produce: from arguments in S:

$$f(S) \;=\; \{\, f(a) \;:\; a \in S \,\}.$$

Thus, for example, $\mathsf{range}(f) = f(A)$.

Given a subset $T \subseteq B$ of the codomain of f, the ***preimage*** of T under f, denoted by $f^{-1}(T)$, is the subset of the domain A consisting of all arguments of f which produce values in T:

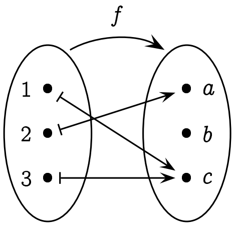
$$f^{-1}(T) \;=\; \{\, a \in A \;:\; f(a) \in T \,\}.$$

Notice in particular that $f^{-1}(B) = A$, since every argument in A produces some value in B. We can also note that images and preimages allow us to view f and f^{-1} as functions between the powersets $\mathcal{P}(A)$ and $\mathcal{P}(B)$:

$$f \;:\; \mathcal{P}(A) \to \mathcal{P}(B) \qquad \text{and} \qquad f^{-1} \;:\; \mathcal{P}(B) \to \mathcal{P}(A).$$

Example 6.2

Consider the function $f : \{\,1,\,2,\,3\} \to \{\,a,\,b,\,c\,\}$ defined, as depicted below, by $f(1) = c$, $f(2) = a$ and $f(3) = c$.



The domain of f is $\{\,1,\,2,\,3\}$.

The codomain of f is $\{\,a,\,b,\,c\,\}$.

The range of f is $\{\,a,\,c\,\}$.

$f(\{1,2\}) = \{\,a, c\,\}$ and $f(\{1,3\}) = \{\,c\,\}$.

$f^{-1}(\{b,c\}) = \{\,1,3\,\}$ and $f^{-1}(\{c\}) = \{\,1,3\,\}$.

Exercise 6.2 (Solution on page 431)

1. What is the range of the function score from Example 6.1?
2. If a score of 70 or higher is considered to be a first-class mark, express the set of students who have scored a first-class mark as a preimage of an appropriate set.

Example 6.3

Here are three example functions defined with respect to an arbitrary set A.

1. The *identity function* $\mathsf{id}_A : A \to A$ is the function which maps each element a of A to itself: $\mathsf{id}_A(x) = x$ for all $x \in A$.
2. the *cardinality function* $|\cdot| : \mathcal{P}_{\text{fin}}(A) \to \mathbb{N}$ maps each finite subset of A to the number of elements in that subset: $|X|$ = the number of elements of X. (The cardinality of a set is simply the number of elements in the set.)

   Note that this function is only well-defined on finite sets. For example, there is no natural number n which denotes the number of elements in the set $\mathbb{N}$.
3. Given a subset $S \subseteq A$ of A, its *characteristic function* $\chi_S : A \to \mathbb{B}$ indicates whether or not an object is an element of S:

$$\chi_S(x) \;=\; \begin{cases} 1, & \textit{if } x \in S; \\ 0, & \textit{if } x \notin S. \end{cases}$$

Exercise 6.3 (Solution on page 432)

Indicate which of the following are functions from the set Humans of all humans to itself. For each that is not a function, indicate why it fails to be a function.

1. $\mathit{Mother}(x)$ represents the mother of x.
2. $\mathit{Parent}(x)$ represents the parent of x.
3. $\mathit{Child}(x)$ represents the child of x.
4. $\mathit{FirstBornChild}(x)$ represents the first-born child of x.

Example 6.4

Functions are common in mathematics, where they are typically given by a formula. For example, the function $f : \mathbb{R} \to \mathbb{R}$ defined by

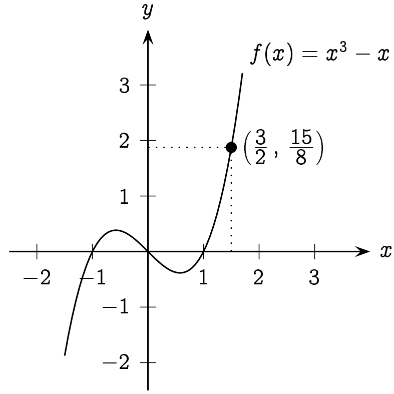


Figure 6.2: The graph of the function $f(x) = x^3 - x$.

$$f(x) \quad = \quad x^3 - x$$

takes a real value $x \in \mathbb{R}$ and returns another real value $f(x) \in \mathbb{R}$ which is computed from x by the formula $x^3 - x$. We can use this formula to calculate the value of $f(x)$ when $x = \frac{3}{2}$:

$$f\left(\frac{3}{2}\right) \quad = \quad \left(\frac{3}{2}\right)^3 - \frac{3}{2} \quad = \quad \frac{27}{8} - \frac{3}{2} \quad = \quad \frac{15}{8}.$$

Such functions are typically plotted as a *graph* on the xy-plane as in Figure 6.2, where we have indicated the point $\left(\frac{3}{2}, \frac{15}{8}\right)$ on the graph.

Motivated by the above example, we can represent a function $f : A \to B$ from A to B as a set of pairs over the Cartesian product $A \times B$. The ***graph*** of f, denoted $\mathsf{graph}(f)$, is the set of all pairs $(a, b) \in A \times B$ such that $b = f(a)$. Thus, for every $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in \mathsf{graph}(f)$, namely $b = f(a)$. As an example, for the score function from Example 6.1,

$$\begin{aligned} \mathsf{graph}(\mathsf{score}) \quad = \quad \{ & (\text{Andrews}, 75), (\text{Archer}, 92), (\text{Collins}, 64), \\ & (\text{Davies}, 88), (\text{Evans}, 78), (\text{Fletcher}, 46), \\ & (\text{Greene}, 68), (\text{Lewis}, 54), (\text{Parker}, 64), \\ & (\text{Smith}, 59), (\text{Taylor}, 100), (\text{Williams}, 78) \}; \end{aligned}$$

and for $f(x) = x^3 - x$,

$$\mathsf{graph}(f) \quad = \quad \{ (x, x^3 - x) \ : \ x \in \mathbb{R} \}.$$

The graph of a function provides a complete description of the function, in that two functions defined over the same domain and codomain are equal if, and only if, their graphs are equal. This is easily proven in the following.

Theorem 6.4

Let $f, g : A \to B$ be two functions defined on the same domain and codomain. Then $f(a) = g(a)$ for all $a \in A$ if, and only if, $\mathsf{graph}(f) = \mathsf{graph}(g)$.

Proof: Suppose that $f(a) = g(a)$ for all $a \in A$, and let $(a, b) \in A \times B$ be arbitrary. We need to show that $(a, b) \in \mathsf{graph}(f) \Leftrightarrow (a, b) \in \mathsf{graph}(g)$. But

$$\begin{aligned}(a,b) \in \mathsf{graph}(f) &\Leftrightarrow b = f(a)\\ &\Leftrightarrow b = g(a) \Leftrightarrow (a,b) \in \mathsf{graph}(g).\end{aligned}$$

Suppose now that $\mathsf{graph}(f) = \mathsf{graph}(g)$, and let $a \in A$ be arbitrary. We need to show that $f(a) = g(a)$. But $(a, f(a)) \in \mathsf{graph}(f)$, and since $\mathsf{graph}(f) = \mathsf{graph}(g)$, we have $(a, f(a)) \in \mathsf{graph}(g)$, and hence $f(a) = g(a)$. □

Exercise 6.4 (Solution on page 432)

What is the graph of the function f from Example 6.2?

6.2 One-To-One and Onto Functions

A function $f : A \to B$ associates a single value $b \in B$ to each value $a \in A$, but the same value $b \in B$ may be associated to more than one value in A; that is, we may have two different values $a, a' \in A$ such that $f(a) = f(a')$. For example, given the function $f(x) = x^3 - x$ there are *three* values of x for which $f(x) = 0$, namely $x = -1$, $x = 0$ and $x = 1$.

If a function does *not* assign the same value to two different inputs, it is said to be *one-to-one* (*1-1*), or *injective*.

Definition 6.4

*A function $f : A \to B$ is **one-to-one** (**1-1**), or **injective**, if, and only if, $f(a) = f(a')$ implies that $a = a'$ for all $a, a' \in A$. More formally:*

$$\forall\, a, a' \in A\, \Big(f(a) = f(a') \;\to\; a = a'\Big).$$

In other words, there do not exist two different values in A which f maps to the same value in B:

$$\neg\exists\, a \in A\, \exists\, a' \in A \left(f(a) = f(a') \ \wedge\ a \neq a'\right).$$

Exercise 6.5 (Solution on page 432)

Indicate which of the following functions are one-to-one. For those that are not one-to-one, indicate the reason that they fail to be one-to-one.

1. The function score : Class → Marks from Example 6.1.
2. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$.
3. The function $f : \mathbb{N} \to \mathbb{N}$ defined by $f(x) = x^2$.

Definition 6.5

*A function $f : A \to B$ is **onto**, or **surjective**, if, and only if, its range is equal to its codomain, range$(f) = B$; that is, every value $b \in B$ is the image of some value $a \in A$:*

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

Exercise 6.6 (Solution on page 432)

Indicate which of the following functions are onto. For those that are not onto, indicate the reason that they fail to be onto.

1. The function score : Class → Marks from Example 6.1.
2. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$.
3. The function $f : \mathbb{N} \to \mathbb{N}$ defined by $f(x) = x^2$.

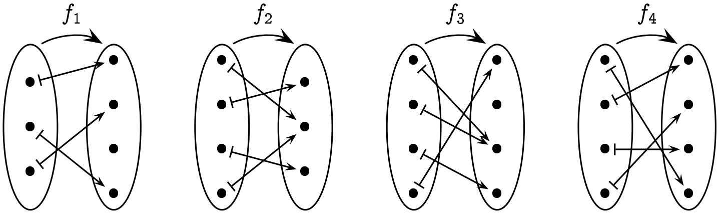
A function $f : A \to B$ which is both one-to-one and onto is particularly special: it defines a perfect correspondence between the sets A and B, in that the function f pairs up the elements of A and B, with each element of one set paired to exactly one element of the other set. Such a function is referred to as a *bijection*.

Definition 6.6

*A function f is a **bijection** if it is both one-to-one and onto.*

Exercise 6.7 (Solution on page 432)

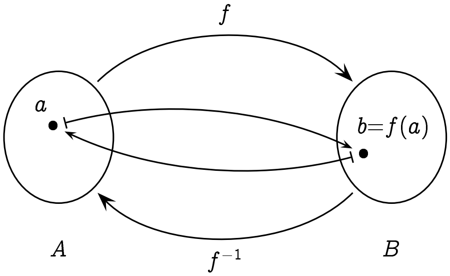
Indicate which of the functions f_1, f_2, f_3 or f_4 depicted by the following diagrams are one-to-one, which of them is onto, and which of them is both (i.e., a bijection).



Let $f : A \to B$ be a bijection. Since f is onto, every element $b \in B$ is the image of some element $a \in A$; and since f is one-to-one, every element $b \in B$ is the image of a *unique* element $a \in A$. This suggest that we can turn the mapping around, to *invert* it, and associate a unique element of A with each element of B.

Definition 6.7

If f is a bijection, then the ***inverse*** *function $f^{-1} : B \to A$ is the function that assigns to each element $b \in B$ the unique element $a \in A$ such that $f(a) = b$. That is, $f^{-1}(b) = a$ if, and only if, $f(a) = b$. This can be pictured as follows:*



The function $f^{-1} : B \to A$ is also a bijection, and $(f^{-1})^{-1} = f$.

Example 6.7

The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x + 3$ is a bijection, with $f^{-1}(x) = (y - 3)/2$. For example,

$$f(5) \;=\; 2 \cdot 5 + 3 \;=\; 13$$

and

$$f^{-1}(13) \;=\; (13 - 3)/2 \;=\; 5.$$

More generally, if $f : A \to B$ is one-to-one, then f provides a bijection from A to range(f), and we can define the inverse function f^{-1} : range(f) $\to A$.

Example 6.8

Let $A = \{a, b, c, \ldots, z\}$ be the set consisting of the usual 26 characters of the alphabet. We can use a bijection $f : A \to A$ as the basis of a simple encryption scheme. For example, suppose we take the bijection f defined as follows:

$$\begin{tikzcd}
a \arrow[d, mapsto] & b \arrow[d, mapsto] & c \arrow[d, mapsto] & d \arrow[d, mapsto] & e \arrow[d, mapsto] & f \arrow[d, mapsto] & g \arrow[d, mapsto] & h \arrow[d, mapsto] & i \arrow[d, mapsto] & j \arrow[d, mapsto] & k \arrow[d, mapsto] & l \arrow[d, mapsto] & m \arrow[d, mapsto] \\
y & k & t & e & c & s & w & u & b & m & z & v & l
\end{tikzcd}$$

$$\begin{tikzcd}
n \arrow[d, mapsto] & o \arrow[d, mapsto] & p \arrow[d, mapsto] & q \arrow[d, mapsto] & r \arrow[d, mapsto] & s \arrow[d, mapsto] & t \arrow[d, mapsto] & u \arrow[d, mapsto] & v \arrow[d, mapsto] & w \arrow[d, mapsto] & x \arrow[d, mapsto] & y \arrow[d, mapsto] & z \arrow[d, mapsto] \\
q & g & d & p & o & j & f & r & h & n & a & x & i
\end{tikzcd}$$

To encode a message we apply the function f to each letter of the message. For example, the message

WE ATTACK AT DAWN

would be encoded as

NC YFFYTZ YF EYNQ

It is important that the function f is a bijection. No two letters can be mapped to the same letter, as otherwise it would be impossible to decode since different messages would give rise to the same encrypted text.

In order to decode messages that we receive which are encoded as above, we simply apply the inverse function f^{-1} to each of the letters of the encrypted text.

This encryption method is insecure; it is very easy to decode encrypted messages even if you don't know the function f with which they are encrypted. However, the idea of using a bijection f to encode messages, thus allowing such messages to be decoded with the inverse function f^{-1}, is fundamental.

Exercise 6.8 (Solution on page 433)

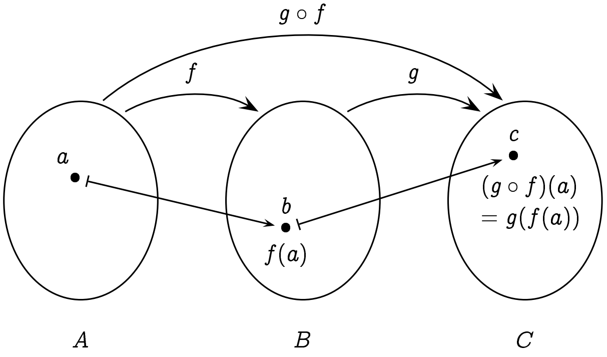
What is the inverse of the function f of Example 6.8?

6.3 Composing Functions

If we have a function $f : A \to B$ from A to B and another function $g : B \to C$ from B to C, we can:

- first apply the function f to some argument $a \in A$ to arrive at a value $b = f(a) \in B$;
- then use the value $b = f(a) \in B$ as an argument to the function g to arrive at a value $c = g(f(a)) \in C$.

Composing two function applications, one after the other, is a very common thing to do; it is commonly denoted by $g \circ f$, and can be pictured as follows:



Definition 6.8

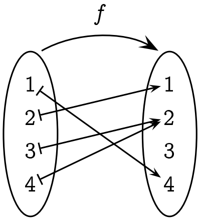
Given a function $f : A \to B$ from A to B and a function $g : B \to C$ from B to C, the ***composition*** *of g and f is the function $g \circ f : A \to C$ from A to C defined by*

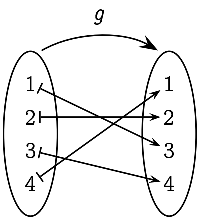
$$(g \circ f)(x) \;=\; g(f(x)).$$

Note that the co-domain of the function f must be the same as the domain of the function g in order to form the composition. Also note carefully the order of the functions: the composition $g \circ f$ of the functions g and f first applies the function f to its input before applying the function g to the result. The reason for writing $g \circ f$ rather than $f \circ g$ is to coincide with the order in which the individual function applications appear: $(g \circ f)(x) = g(f(x))$.

Exercise 6.9 (Solution on page 433)

Consider the following two functions f and g from $\{1, 2, 3, 4\}$ to itself:





Find $f \circ g$ and $g \circ f$.

If $f : A \to A$ then we can compose f with itself. In this case we typically write f^2 for $f \circ f$, and more generally $f^{n+1} = f \circ f^n$. In other words,

$$f^n = \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times}}$$

As special cases we have $f^0 = \mathrm{id}_A$ and $f^1 = f$, noting that $f \circ \mathrm{id}_A = f$ (see exercise 8, page 177).

If we compose two one-to-one functions, we will arrive at yet another one-to-one function. The same is true of onto functions. These facts are demonstrated in the following two theorems.

Theorem 6.9

If $f : A \to B$ and $g : B \to C$ are both one-to-one, then so is $g \circ f : A \to C$.

Proof: Suppose $(g \circ f)(x) = (g \circ f)(y)$; that is, $g(f(x)) = g(f(y))$.
What we need to demonstrate is that $x = y$.

Since g is one-to-one, $f(x) = f(y)$.

Hence, since f is one-to-one, $x = y$. □

Theorem 6.10

If $f : A \to B$ and $g : B \to C$ are both onto, then so is $g \circ f : A \to C$.

Proof: Suppose $c \in C$.
What we need to demonstrate is that $c = (g \circ f)(a)$ for some $a \in A$.

Since g is onto, $c = g(b)$ for some $b \in B$.

Since f is onto, $b = f(a)$ for some $a \in A$.

Hence $c = g(f(a)) = (g \circ f)(a)$. □

Exercise 6.10 (Solution on page 433)

Prove that if $f : A \to B$ and $g : B \to C$ are both bijections, then so is $g \circ f : A \to C$.

Exercise 6.11 (Solution on page 433)

Prove that if $f : A \to B$ is a bijection, then $f^{-1} \circ f = \mathrm{id}_A$ and $f \circ f^{-1} = \mathrm{id}_B$.

Exercise 6.12 (Solution on page 433)

Prove that function composition is associative: if $f : A \to B$, $g : B \to C$ and $h : C \to D$ then $h \circ (g \circ f) = (h \circ g) \circ f$.

⋆ 6.4 Comparing the Sizes of Sets

We can easily compare the sizes (the cardinalities) of two finite sets simply by counting their elements; the size of one is greater than the size of the other if it contains more elements, and the two sets are the same size if they contain the same number of elements.

Counting the number of elements in a finite set involves listing them in some arbitrary order, denoting one of them to be the first element, another to be the second element, and so on to the last element. For example, we would conclude that the set $\{$ Joel, Felix, Oskar, Amanda $\}$ has four elements by virtue of the fact that we could find a one-to-one and onto function (a bijection)

$$f \ : \ \{1, 2, 3, 4\} \to \{\text{ Joel, Felix, Oskar, Amanda }\}$$

which effectively lists the elements of the set. For example, the function f may list this set (alphabetically) as follows:

$$\begin{aligned} f \ : \quad & 1 \mapsto \text{Amanda} \\ & 2 \mapsto \text{Felix} \\ & 3 \mapsto \text{Joel} \\ & 4 \mapsto \text{Oskar} \end{aligned}$$

This bijection demonstrates that the two sets $\{$ Joel, Felix, Oskar, Amanda $\}$ and $\{1, 2, 3, 4\}$ are the same size (i.e., have the same cardinality).

We can compare the sizes of any two sets by trying to find a bijection between them which would demonstrate that the two sets are the same size. If such a bijection doesn't exist, then one set must be bigger than the other. For example, if we try to find a bijection

$$f \ : \ \{\text{ Joel, Felix, Oskar, Amanda }\} \to \{\text{ cola, fanta, sprite }\}$$

we would quickly realise that this would be impossible, as no such function could be one-to-one: some element of the second set would have to be the image of more than one element of the first set since there are not enough elements in the second set to go around. If this function was aimed at providing each child with a drink, then it is clear that some drink would have to be shared.

For the same reason, no function

$$f : \{\text{cola, fanta, sprite}\} \to \{\text{Joel, Felix, Oskar, Amanda}\}$$

could be onto. If this function was aimed at distributing drinks to children, then it is clear that at least one child would not get a drink.

Given two arbitrary sets A and B, we would naturally consider B to be at least as large as A if we could find a *one-to-one* function $f : A \to B$, since f would associate each element of A with its own element of B, so intuitively there would have to be at least as many elements of B as there are of A. On the other hand, we would naturally consider A to be at least as large as B if we could find an *onto* function $f : A \to B$, since f would associate each element of B with at least one element of A which is not associated with any other element of B, so intuitively there would have to be at least as many elements of A as there are of B. Finally, we would naturally consider the two sets to be of the same size (cardinality) if we could find a bijection $f : A \to B$ giving a direct correspondence associating each element of one of the sets with its own element of the other set. We will denote that a set A is no bigger than, no smaller than, and the same size as B by $A \preceq B$, $A \succeq B$, and $A \cong B$, respectively, and summarise this discussion as follows.

Definition 6.12

- $A \preceq B$ *if, and only if, there exists a one-to-one function* $f : A \to B$.
- $A \succeq B$ *if, and only if, there exists an onto function* $f : A \to B$.
- $A \cong B$ *if, and only if, there exists a bijection* $f : A \to B$.

The following results show that these definitions make sense in terms of comparing sizes of sets. The first result says that one set is no bigger than a second if, and only if, the second is no smaller than the first. The second result says that two sets are the same size if, and only if, each is no larger than the other.

Theorem 6.12

$A \preceq B$ if, and only if, $B \succeq A$. That is, there exists a one-to-one function $f : A \to B$ if, and only if, there exists an onto function $g : B \to A$.

Proof: Suppose that $f : A \to B$ is one-to-one, and fix some element $a_0 \in A$.[†] We can define the function $g : B \to A$ as follows:

- if $b \in \text{range}(f)$ then $b = f(a)$ for a unique value $a \in A$, and we define $g(b)$ to be this unique value a;
- if $b \notin \text{range}(f)$ then we define $g(b)$ to be a_0.

[†] If no such a_0 exists, that is if $A = \emptyset$, then $\emptyset$ trivially represents the graph of a one-to-one function from A to B, as well as the graph of an onto function from B to A.

This function g is onto, as $a = g(f(a))$ for each element $a \in A$.

Suppose now that $g : B \to A$ is onto. For each value $a \in A$, fix some value b_a such that $g(b_a) = a$. Then the function $f : A \to B$ defined as $f(a) = b_a$ for each $a \in A$ is clearly one-to-one. □

Theorem 6.13 Schröder-Bernstein Theorem

$A \cong B$ if, and only if, $A \preceq B$ and $B \preceq A$.

Proof: Suppose we have functions $f : A \to B$ and $g : B \to A$ which are both one-to-one; we wish to construct a bijection $h : A \to B$.

For any $a \in A$, consider the sequence generated from a by alternately applying g^{-1} and f^{-1} whenever possible:

$$a \;\mapsto\; g^{-1}(a) \;\mapsto\; f^{-1}(g^{-1}(a)) \;\mapsto\; g^{-1}\big(f^{-1}(g^{-1}(a))\big) \;\mapsto\; \cdots$$

This is possible since f and g are one-to-one, and hence $f^{-1} : \mathsf{range}(f) \to A$ and $g^{-1} : \mathsf{range}(g) \to B$ are well-define functions. However, this sequence may stop at some point, either at an element of A not in the range of g (and hence for which g^{-1} is not defined) or at an element of B not in the range of f (and hence for which f^{-1} is not defined).

We can then define our bijection $h : A \to B$ as follows:

$$h(a) \;=\; \begin{cases} g^{-1}(a), & \text{if the sequence generated by } a \\ & \text{ends at an element of } B; \\ f(a), & \text{otherwise.} \end{cases}$$

This is a well-defined function, since $g^{-1}(a)$ will be defined if the sequence generated by a ends at an element of B (in particular, not at a). It remains to demonstrate that this function is one-to-one and onto.

To demonstrate that h is one-to-one, let us assume that $h(x) = h(y)$, and show that we must have $x = y$.

- If the sequences generated by x and y both end at elements in B, then $h(x) = g^{-1}(x)$ and $h(y) = g^{-1}(y)$, so $g^{-1}(x) = g^{-1}(y)$, and hence $x = y$.
- If neither sequence generated by x and y ends at an element of B, then $h(x) = f(x)$ and $h(y) = f(y)$, so $f(x) = f(y)$, and hence $x = y$.
- If the sequence generated by x ends at an element of B, but not so for the sequence generated by y, then $h(x) = g^{-1}(x)$ and $h(y) = f(y)$, so $g^{-1}(x) = f(y)$. But then $y = f^{-1}(g^{-1}(x))$ would appear (as the third element) in the sequence generated by x, contradicting the assumption that its sequence ends differently to that generated by x.

- If the sequence generated by y ends at an element of B, but not so for the sequence generated by x, then $h(y) = g^{-1}(y)$ and $h(x) = f(x)$, so $g^{-1}(y) = f(x)$. But then $x = f^{-1}(g^{-1}(y))$ would appear (as the third element) in the sequence generated by y, contradicting the assumption that its sequence ends differently to that generated by y.

To demonstrate that h is onto, let us assume that $b \in B$, and show that we must have $b = h(a)$ for some $a \in A$.

- If the sequence generated by $g(b)$ ends at an element of B, then $h(g(b)) = g^{-1}(g(b)) = b$.
- If the sequence generated by $g(b)$ does not end at an element of B, then $f^{-1}(b)$ must be defined and appear (as the third element) in the sequence generated by $g(b)$, and hence $h(f^{-1}(b)) = f(f^{-1}(b)) = b$. □

These definitions are unremarkable for finite sets, but reveal surprising relationships between infinite sets, as the following example demonstrates.

Example 6.13

The set $\mathbb{N}$ of nonnegative integers in some sense contains almost twice as many elements as the set $\mathbb{E} = \{\,0, 2, 4, \ldots\,\}$ of nonnegative even integers. However, the function $f : \mathbb{N} \to \mathbb{E}$ defined by $f(n) = 2n$ provides a bijection from $\mathbb{N}$ to $\mathbb{E}$, demonstrating that there are in fact the same "number" of even integers as there are integers. This bijection can be pictured as follows:

$$\begin{array}{rcccccccccccc}
f: & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \cdots \\
 & 0 & 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 &
\end{array}$$

The confusion arising from the above example is with the idea of the "number" of elements of an infinite set. There are in fact an infinite number of objects in each of the sets, and as such there is no problem with considering them to have the same cardinality.

Realising that the set of even integers is no smaller than the set of all integers, it may seem that one infinite set is as big as any other. In fact, some infinite sets are larger than others. To explore this idea, we start with the following definitions.

Definition 6.13

A set A is said to be ***finite*** *if, and only if, there is a bijection*

$$f : \{1, 2, 3, \ldots, n\} \to A$$

for some $n \in \mathbb{N}$. This function effectively lists all of the elements of A, and the value n is the ***cardinality*** *of A: $|A| = n$.*

A set A is said to be ***countably infinite*** *if, and only if, there is a bijection*

$$f : \mathbb{N} \to A.$$

This function lists the elements of A in an infinite list.

Finally, a set is said to be ***countable*** *if, and only if, it is finite or countably infinite; and it is said to be* ***uncountable*** *if, and only if, it is not countable.*

Example 6.14

The set of integers $\mathbb{Z}$ is countable. A bijection $f : \mathbb{N} \to \mathbb{Z}$ witnessing this fact can be defined as

$$f(n) \quad = \quad \begin{cases} \frac{n+1}{2} & \textit{if } n \textit{ is odd;} \\[1ex] -\frac{n}{2} & \textit{if } n \textit{ is even.} \end{cases}$$

This function would list the integers as follows:

$$\begin{tikzcd}
f : 0 \arrow[d, mapsto] & 1 \arrow[d, mapsto] & 2 \arrow[d, mapsto] & 3 \arrow[d, mapsto] & 4 \arrow[d, mapsto] & 5 \arrow[d, mapsto] & 6 \arrow[d, mapsto] & 7 \arrow[d, mapsto] & 8 \arrow[d, mapsto] & 9 \arrow[d, mapsto] & 10 \arrow[d, mapsto] & \\
0 & 1 & -1 & 2 & -2 & 3 & -3 & 4 & -4 & 5 & -5 & \cdots
\end{tikzcd}$$

Clearly this function is one-to-one and onto, as every integer will appear exactly once in this list.

Exercise 6.14 (Solution on page 433)

What is the inverse $f^{-1} : \mathbb{Z} \to \mathbb{N}$ of the bijection $f : \mathbb{N} \to \mathbb{Z}$ given in Example 6.14?

Exercise 6.15 (Solution on page 434)

Prove that $A \cong B$ for any two countable sets A and B.

That is, given bijections $f : \mathbb{N} \to A$ and $g : \mathbb{N} \to B$, show how to construct a bijection $h : A \to B$.

As an example of the difference between countable and uncountable sets, we shall see that there are far more numbers on the real number line than just the integers; that is, the set of real numbers $\mathbb{R}$ is uncountable. This may seem perfectly sensible, as these numbers fill the number line: between any two different real numbers, no matter how close they are to each other, you can find a third. The integers, on the other hand, are relatively few and far between.

While such an intuitive argument gives rise to a valid result in this case, the same intuition would lead you to believe that there are uncountably-many rational numbers, as between any two different rational numbers, no matter how close they are to each other, you can find a third. However, before we demonstrate that there are uncountably-many reals, we first demonstrate that this intuition about the rationals is faulty; the rationals are countable, and hence no more numerous that the integers.

Example 6.15

The set $\mathbb{Q}^+$ of positive rational numbers is countable. To see this, we need to find a bijection

$$f \;:\; \mathbb{N} \to \mathbb{Q}^+$$

which completely lists them. To this end, we first note that a positive rational is a number of the form $\frac{p}{q}$, where p and q are positive integers, and we can arrange these in an infinite number of infinite rows by:

- listing all the rationals with numerator $p = 1$ in the first row,
- listing all the rationals with numerator $p = 2$ in the second row,
- listing all the rationals with numerator $p = 3$ in the third row,

and so on as depicted in Figure 6.3. We can then zigzag diagonally through this arrangement as depicted in Figure 6.3, listing the rationals in the order in which they are encountered. However, we only list rationals that appear in lowest form, and ignore those (depicted crossed out in grey circles in Figure 6.3) that are not in lowest form; for example, we do not include $\frac{4}{6}$ in our listing as it will have already appeared earlier in our list as $\frac{2}{3}$.

The resulting listing provides the required bijection $f : \mathbb{N} \to \mathbb{Q}^+$:

$$f \;:\; \begin{array}{cccccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \cdots \\ \frac{1}{1} & \frac{2}{1} & \frac{1}{2} & \frac{1}{3} & \frac{3}{1} & \frac{4}{1} & \frac{3}{2} & \frac{2}{3} & \frac{1}{4} & \frac{1}{5} & \frac{5}{1} & \end{array}$$

This function is one-to-one and onto as only rationals $\frac{p}{q}$ in lowest form appear in the list and each of these is encountered once and only once while zigzagging through the arrangement.

Extending this result to show that the set $\mathbb{Q}$ of all the rational numbers is countable is straightforward.

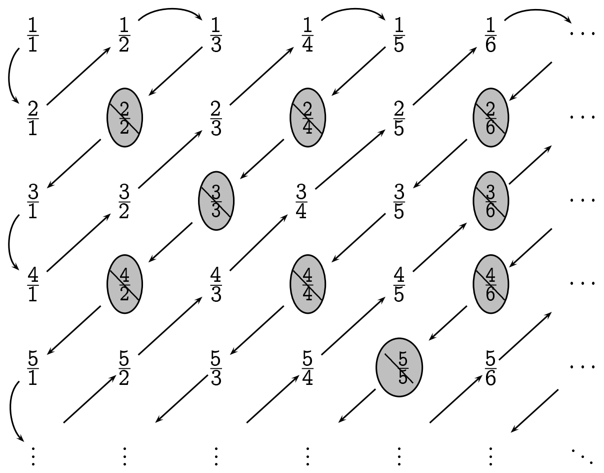


Figure 6.3: Enumerating the positive rationals.

Exercise 6.16 (Solution on page 434)

Prove that the set $\mathbb{Q}$ of all rationals is countable.

Example 6.16

The set $[0,1]$ of nonnegative real numbers no greater than 1 is uncountable. To see this we must show that no bijection $f : \mathbb{N} \to [0,1]$ exists. To this end, assume that f is such a function, and consider the listing of the real numbers that it gives:

$$\begin{array}{cccccccccccc}
f: & 0 & \mapsto & 0\,. & \boxed{d_{00}} & d_{01} & d_{02} & d_{03} & d_{04} & d_{05} & \cdots \\
 & 1 & \mapsto & 0\,. & d_{10} & \boxed{d_{11}} & d_{12} & d_{13} & d_{14} & d_{15} & \cdots \\
 & 2 & \mapsto & 0\,. & d_{20} & d_{21} & \boxed{d_{22}} & d_{23} & d_{24} & d_{25} & \cdots \\
 & 3 & \mapsto & 0\,. & d_{30} & d_{31} & d_{32} & \boxed{d_{33}} & d_{34} & d_{35} & \cdots \\
 & 4 & \mapsto & 0\,. & d_{40} & d_{41} & d_{42} & d_{43} & \boxed{d_{44}} & d_{45} & \cdots \\
 & 5 & \mapsto & 0\,. & d_{50} & d_{51} & d_{52} & d_{53} & d_{54} & \boxed{d_{55}} & \cdots \\
 & \vdots & & \vdots\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

Each number in this list, being a nonnegative real number no greater than 1, is given by an infinite decimal expansion with a leading 0. In particular, the value 0 appears as $0.00000\cdots$ and the value 1 appears as $0.99999\cdots$.

Consider now the real number

$$r \;=\; 0\,.\,r_1\,r_2\,r_3\,r_4\,r_5\,\cdots$$

in which the ith decimal digit r_i is given by

$$r_i = (d_{ii} + 5) \bmod 10.$$

That is, the ith decimal digit of r is defined to differ by 5 from the ith decimal digit of $f(i)$.

Assuming that the function f above is indeed a bijection, and in particular onto, the value r must appear somewhere in the list; that is, we must have $r = f(n)$ for some $n \in \mathbb{N}$. However, for each n, r differs (by 5) from $f(n)$ in the nth decimal place, meaning that we cannot have $r = f(n)$.

An infinite set may thus be either countably infinite or uncountably infinite. In Exercise 6.15 we saw that any two countably-infinite sets are the same size, but the same is not true of two uncountable sets. The following exercise demonstrates that given any set, no matter how big, you can always construct an even bigger set by merely taking its powerset.

Exercise 6.17 (Solution on page 434)

Show that the powerset $\mathcal{P}(A)$ of any set A is strictly larger than A, by showing that no function $f : A \to \mathcal{P}(A)$ can be onto.

(Hint: Show that the set $B = \{\, x \in A \;:\; x \notin f(x) \,\}$ is different from $f(a)$ for all $a \in A$.)

⋆ 6.5 The Knaster-Tarski Theorem

In this section, as an example in working with sets, we prove an important result on the existence of (greatest and least) fixed points of monotonic functions defined on the powerset of a given set. We also describe a procedure for calculating these fixed points.

Definition 6.17

Let S be a set, and let $f : \mathcal{P}(S) \to \mathcal{P}(S)$ be a function which maps subsets of S to subsets of S.

- *f is **monotonic** if, and only if, $f(A) \subseteq f(B)$ whenever $A \subseteq B$.*
- *$A \subseteq S$ is a **fixed point** of f if, and only if, $f(A) = A$.*

- *$A \subseteq S$ is the* ***greatest fixed point*** *of f – denoted gfp(f) – if, and only if, A is a fixed point (ie, $f(A) = A$) and A is larger than all other fixed points: if $f(B) = B$ then $B \subseteq A$.*
- *$A \subseteq S$ is the* ***least fixed point*** *of f – denoted lfp(f) – if, and only if, A is a fixed point (ie, $f(A) = A$) and A is smaller than all other fixed points: if $f(B) = B$ then $A \subseteq B$.*

Note that fixed points need not exist; and even if they do exist, then there is no guarantee that greatest and/or least fixed points exist.

Example 6.17

Let $S = \{0\}$ and define $f : \mathcal{P}(S) \to \mathcal{P}(S)$ by $f(\emptyset) = S$ and $f(S) = \emptyset$.

Clearly f does not have a fixed point.

Exercise 6.18 (Solution on page 434)

Define a function $f : \mathcal{P}(S) \to \mathcal{P}(S)$ over the set $S = \{1, 2\}$ which has two fixed points which are neither greatest nor least fixed points.

The following result, however, shows that both greatest and least fixed points exist for monotonic functions.

Theorem 6.18 Knaster-Tarski Theorem

If $f : \mathcal{P}(S) \to \mathcal{P}(S)$ is monotonic, then f has both greatest and least fixed points. Furthermore, these can be defined as follows:

- *gfp(f) $=$ $\bigcup\{A \subseteq S : A \subseteq f(A)\}$; and*
- *lfp(f) $=$ $\bigcap\{A \subseteq S : f(A) \subseteq A\}$.*

Proof: We will prove the result about the greatest fixed point *gfp*(f) and leave the result about the least fixed point *lfp*(f) as an exercise (Exercise 12, page 178).

To this end, let $\mathsf{G} = \bigcup\{A \subseteq S : A \subseteq f(A)\}$ as in the Theorem. We first demonstrate that $\mathsf{G} \subseteq f(\mathsf{G})$ by showing that given any $a \in \mathsf{G}$ we must have that $a \in f(\mathsf{G})$.

> Suppose $a \in \mathsf{G}$. By the definition of G, this means that $a \in A$ for some $A \subseteq S$ such that $A \subseteq f(A)$. Hence $a \in f(A)$. Moreover, $A \subseteq \mathsf{G}$ (as G is the union of all such sets), so by the monotonicity of f we have that $f(A) \subseteq f(\mathsf{G})$. Hence $a \in f(\mathsf{G})$ as required.

Next, we demonstrate the reverse inclusion, that $f(\mathsf{G}) \subseteq \mathsf{G}$.

> Since we've shown that $\mathrm{G} \subseteq f(\mathrm{G})$, by the monotonicity of f we have that $f(\mathrm{G}) \subseteq f(f(\mathrm{G}))$. This means that $f(\mathrm{G})$ is one of the sets in the family of sets whose union is G, and hence $f(\mathrm{G}) \subseteq \mathrm{G}$.

We've thus shown that G is a fixed point of f. It remains to show that it is the greatest fixed point. To this end, suppose that X is any fixed point of f. Since $X \subseteq f(X)$, X is one of the sets in the family of sets whose union is G, and hence $X \subseteq \mathrm{G}$. □

Beyond knowing that greatest and least fixed points of f exist, we would like to know how to calculate them without having to calculate $f(A)$ for all subsets $A \subseteq S$. To do this, we can exploit the following observations.

Theorem 6.19

For all $n \in \mathbb{N}$,

1. $f^n(\emptyset) \subseteq f^{n+1}(\emptyset)$ *and* $f^n(\emptyset) \subseteq \mathit{lfp}(f)$;
2. $f^n(S) \supseteq f^{n+1}(S)$ *and* $f^n(S) \supseteq \mathit{gfp}(f)$.

From this we can deduce the following:

(a) $\bigcup_{n\in\mathbb{N}} f^n(\emptyset) \subseteq \mathit{lfp}(f)$ *and* $\bigcup_{n\in\mathbb{N}} f^n(S) \supseteq \mathit{gfp}(f)$;

(b) *If* $f^n(\emptyset) = f^{n+1}(\emptyset)$ *then* $\mathit{lfp}(f) = f^n(\emptyset)$;

(c) *If* $f^n(S) = f^{n+1}(S)$ *then* $\mathit{gfp}(f) = f^n(S)$;

(d) *If* $|S| = n$ *then* $\mathit{lfp}(f) = f^n(\emptyset)$ *and* $\mathit{gfp}(f) = f^n(S)$.

Proof: We prove only *1.*, by straightforward induction, and leave *2.* and the corollaries *(a)-(d)* as exercises (Exercise 13, page 178).

For the base case, $f^0(\emptyset) = \emptyset$, so clearly $f^0(\emptyset) \subseteq f^1(\emptyset)$ and $f^0(\emptyset) \subseteq \mathit{lfp}(f)$.

For the induction case, assuming that $f^{n-1}(\emptyset) \subseteq f^n(\emptyset)$ and that $f^{n-1}(\emptyset) \subseteq \mathit{lfp}(f)$,

- $f^n(\emptyset) = f(f^{n-1}(\emptyset)) \subseteq f(f^n(\emptyset)) = f^{n+1}(\emptyset)$; and
- $f^n(\emptyset) = f(f^{n-1}(\emptyset)) \subseteq f(\mathit{lfp}(f)) = \mathit{lfp}(f)$. □

Thus, in order to calculate the least fixed point $\mathit{lfp}(f)$ of f, we can repeatedly apply f starting from the empty set $\emptyset$ until we arrive at a fixed point, which by above will be $\mathit{lfp}(f)$:

$$\emptyset = f^0(\emptyset) \subset f^1(\emptyset) \subset f^2(\emptyset) \subset \cdots \subset f^n(\emptyset) = f^{n+1}(\emptyset) = \mathit{lfp}(f).$$

A similar procedure, starting from S, will give us the greatest fixed point. This is guaranteed to work if the set S is finite; however, if S is infinite, we may generate infinite sequences of sets which approach yet never reach the fixed points.

Exercise 6.20 (Solution on page 434)

Let $f : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{N})$ be defined by $f(S) = \{0\} \cup \{n+2 \ : \ n \in S\}$.

1. Prove that f is monotonic.
2. Show that $f^n(\emptyset) \subset f^{n+1}(\emptyset)$ and $f^n(\mathbb{N}) \supset f^{n+1}(\mathbb{N})$ for each $n \in \mathbb{N}$.
3. Determine the least and greatest fixed points $lfp(f)$ and $gfp(f)$.

6.6 Additional Exercises

1. Identify the domain, codomain, and range of the following functions.
   (a) the function that assigns to each nonnegative integer the least prime number greater than it.
   (b) the function that assigns to each pair of positive integers the maximum of these two values.

2. Let $A = \{1, 2, 3, 4\}$ and $B = \{a, b, c, d\}$, and let f_1, f_2 and f_3 be functions from A to B with the following graphs:

$$\begin{aligned} \mathsf{graph}(f_1) &= \{(1,d), (2,a), (3,c), (4,c)\} \\ \mathsf{graph}(f_2) &= \{(1,d), (2,c), (3,a), (4,b)\} \\ \mathsf{graph}(f_3) &= \{(1,b), (2,c), (3,a), (4,d)\} \end{aligned}$$

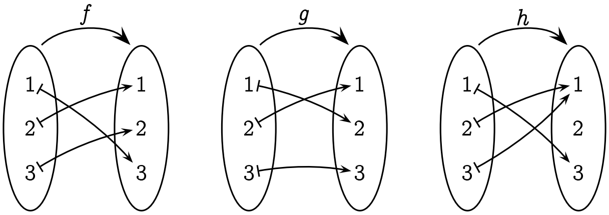
   Indicate which of these functions are one-to-one, which are onto, and which are bijections.

3. Give an example of a function from $\mathbb{N}$ to $\mathbb{N}$ that is
   (a) one-to-one but not onto.
   (b) onto but not one-to-one.
   (c) one-to-one and onto, but which does not map any value to itself.
   (d) neither one-to-one nor onto.

4. Find all functions from $X = \{a, b\}$ to $Y = \{1, 2, 3\}$. In each case, indicate whether or not the function is one-to-one, and whether or not it is onto.

5. Define the function $f : [0,1] \to (0,1)$ by: $f(0) = 1/2$; $f(1/n) = 1/(n+2)$ for all positive integers n; and $f(x) = x$ otherwise. Prove that f is a bijection.

6. Consider the following three functions f, g and h from $\{1, 2, 3\}$ to itself:



   Find $f \circ g$, $g \circ f$, $f \circ h$, $h \circ f$, $g \circ h$, $h \circ g$.

7. Find $g \circ f$ and $f \circ g$, where $f(x) = x^2+1$ and $g(x) = x-2$ are functions from $\mathbb{R}$ to $\mathbb{R}$.

8. Prove that for any function $f : A \to B$, $f = f \circ \mathsf{id}_A$ and $f = \mathsf{id}_B \circ f$, where $\mathsf{id}_X : X \to X$ is the identity function on X, that is, $f(x) = x$ for all $x \in X$.

9. Assuming that $f : A \to B$ and $g : B \to C$, prove or disprove the following.

   (a) If f and $g \circ f$ are both one-to-one, then g must also be one-to-one.
   (b) If g and $g \circ f$ are both one-to-one, then g must also be one-to-one.
   (c) If f and $g \circ f$ are both onto, then g must also be onto.
   (d) If g and $g \circ f$ are both onto, then g must also be onto.

10. Prove that if $A \subseteq B$ and B is countable, then A is countable.

11. In Example 6.15 we saw how to construct a function $f : \mathbb{N} \to \mathbb{Q}^+$ which listed all of the positive rational numbers by zigzagging through an infinite array of rational numbers. However, we had to disregard the rational numbers that we came across which were not in lowest terms. In this exercise we explore an alternative approach which avoids this complication.

    Consider the tree-like diagram in Figure 6.4 which is constructed by starting with $1/1$ at the top, and from each branching point labelled i/j drawing a left branch labelled $i/(i+j)$ and a right branch labelled $(i+j)/j$.

    Argue that every positive rational number appears exactly once in this tree, by arguing that each of the following is true.

    (a) Every node is labelled by a rational number in lowest form.
    (b) Every rational number appears somewhere in the tree.
    (c) No rational number appears twice in the tree.

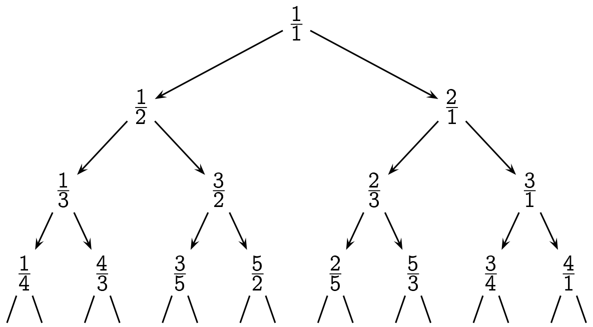


Figure 6.4: Enumerating the rational numbers

Thus, to list the rational numbers without repetition we need merely list the successive rows of the tree.

12. Prove the second part of Theorem 6.18 from page 174, that L as defined there is the least fixed point of f.

13. (a) Prove the second part of Theorem 6.19 (page 175).
    (b) Prove the four corollaries *(a)-(d)* to Theorem 6.19 (page 175).