Equations and Functions, Variables and Expressions

Equations and functions are ubiquitous components of mathematical language. Success in mathematics beyond basic arithmetic depends on having a solid working knowledge of both ideas.

Equations are statements that two quantities are "the same". When we say "the same", evidently, we don't mean exactly the same. Two different things are not exactly the same; if they were, they would be one thing, not two. We mean that they are the same for present purposes. Most equations that one encounters in school mathematics assert that two different mathematical expressions represent the same number: they are the same in the sense that, when evaluated according to the rules governing the formation of expressions, they produce the same result.

Functions are recipes or rules for assigning things to other things. Usually in school mathematics, both things are numbers, but there are many other kinds of functions, some having nothing to do with numbers. (Because of the tendency in school mathematics to talk of functions in a numerical context, some things that are studied in K-12 math, and which are functions, are never so identified.) It is because the notion of function is so broad that it has been such a useful tool in mathematics.

We will elaborate on each of these notions below, show how they are connected, and illustrate how they can be used to express many ideas. We will start out discussing

Functions:

The idea of function originated in the intuitive idea that one quantity may depend on another quantity, or several other quantities, as in "The position of the sun in the sky is a function of time of day", or "The area of a rectangle is a function of its height and width." Many things vary with time, and one might want to describe them as functions of time. For example, if a particle moves in space, its position at a given time can be thought of as a function of time. If one has some investments, their value can be thought of as a function of time. The amount of carbon dioxide in the atmosphere can be thought of as a function of time. However, the modern conception of function is extremely general and flexible, and applies in situations far beyond the contexts where the idea was born.

Roughly speaking, a function is a way to assign things to (other things). More formally, to define a function, we must first specify a collection or set of things to receive assignments. This set is called the *domain* of the function. Then we must specify another set that contains the things that will get assigned. This second set is called the *range*.

Having identified the domain D and the range R, a *function* from D to R is a recipe or rule that assigns to every member of D exactly one corresponding member of R. This is the key property of a function: every member of D gets one and only one member of R attached to it. Members of R may or may not be assigned to a member of D, and some of them may be assigned to more than one member of D. But every member of D must be assigned exactly one member of R.

The standard notation for functions uses a symbol, such as f to stand for the function. Then for each member d of D, the symbol $f(d)$ signifies the element of R assigned to D by the function. We say this as "eff of dee", and refer to it as the *value of f at d*. The symbol $f(d)$ is supposed to evoke a process by which d is converted into $f(d)$. We also use the notation $F : D \to R$ to suggest that f sends (the elements of) D to R.

Examples: Here is a collection of functions.

0) The familiar numerical functions.

These are functions of a real number x, and assign another real number $f(x)$ to x. That is, the domain and the range are the real numbers. It was working with these functions that gave birth to the general idea of function. In rough order of appearance in K-12, they are as follows.

a) First order, aka linear, functions: These are functions that take a number, multiply it by a fixed number, and add another fixed number to the result, such as $x \to 2x + 3$. Linear functions can be used to describe many simple phenomena, such as the price or things like a plumber's bill, based on an initial fee, plus a charge per hour for work done, or the price of rentals, that cost a given price for the first day (or hour), then a smaller price for succeeding days, or the price of taxis, that cost a certain amount for getting in, then a certain amount per mile driven.

b) Quadratic functions: These functions involve the square of a variable, as well as a multiple of the variable and a constant. They can be used to describe a variety of things, particularly projectile motion, area of various figures, the sum of consecutive numbers, or more generally, of arithmetic progressions, and so forth. A typical example is $x \to \frac{x(x+1)}{2}$, which gives the sum of the integers from 1 to n, or the number of ways to choose a pair of objects from $n+1$ objects.

c) Exponential functions. These are used to described quantities which grow proportionally to their size. The value of a savings account, or the population of some organisms under conditions of ample resources, or the quantity of a radioactive substance over time are examples. A specific example is $x \to 3 \cdot 2^x$.

d) Trigonometric functions. This refers to a family of functions that are *periodic* - their values repeat after a cetain interval; that is they satisfy a condition that $f(x+P) = f(x)$, for some number P, called the period. The most basic examples are $x \to \sin x$ and $x \to \cos x$.

Because these are typically the main examples of functions that students encounter, and because learning their properties and how to use them can involve considerable labor, and because these functions are all defined on the fixed domain of the real numbers, which is also the range, and because the specification of a domain and a range is somewhat cumbersome, the study of functions centered on learning the important applications of these examples often leads to the suppression of the notion of domain, and also to the idea that a function is essentially a formula. In the context of real-valued functions of a real variable (sometimes referred to simply as *real functions*, this suppression saves time and effort that can be applied to becoming familiar with the distinctive properties of each of these interesting families of functions. However, these functions, even all these several types and others that combine them, do not convey the full flexibility and richness of the function concept. Also, even in a discussion of purely numerical functions, suppression of issues of domain can lead to confusion. Domain issues become of important particularly when discussing inverse functions, starting with square root, and inverse functions tend to be one of the poorest understood topics in the domain of elementary numerical functions.

1) The successor function. Here $D = \mathbf{N}$, the *natural numbers*, aka the positive whole numbers: 1, 2, 3, 4, ... The range is also $\mathbf{N}$. The successor function s attaches to each whole number the next largest whole number. Standard algebraic notation would represent this by the recipe

$$s(n) = n + 1,$$

where n stands for any natural number.

2) The sum function. Here the domain $D = \mathbf{N} \times \mathbf{N}$, the product of two copies of $\mathbf{N}$. By this we mean ordered pairs (n_1, n_2), where both n_1 and n_2 are natural numbers, i.e., are in $\mathbf{N}$. The range is again $\mathbf{N}$. The *sum function* A assigns to the pair (n_1, n_2) their sum $n_1 + n_2$. Again this can be expressed using standard algebraic notation:

$$A(n_1, n_2) = n_1 + n_2.$$

3) The product function. Here again, $D = \mathbf{N} \times \mathbf{N}$ and $R = \mathbf{N}$. The *product function* M assigns to a pair (n_1, n_2) of whole numbers their product $n_1 \cdot n_2$. Once more we can take advantage of conventional algebraic symbolism to express M:

$$M(n_1, n_2) = n_1 \cdot n_2$$

for any natural numbers n_1 and n_2.

These first three examples suggest that functions are connected to numbers and arithmetic, and indeed, arithmetic is a prominent sources of functions. However, functions need not have anything, or very little, to do with numbers.

4) The mother function. For this example, the domain is the set of all living people, and the range is the set of all people who have lived. The function M assigns to each person his or her mother, meaning the person who gave birth to him or her.

5) The postage function. For this example, the domain is the set of all US postal items – letters and packages mailed in the US – and the range is the non-negative integers. The function assigns to each letter and package the value, in cents, of the postage - stamps, matered postage, or both togehter - paid to mail the item.

6) The ticket price function. The domain here is the set of tickets to, say, Disney World. The range is the rational numbers. The ticket function attaches to each ticket to Disney World it price, in dollars.

7) The area function. For this example, the domain is the set of all polygonal figures in the (Euclidean) plane, and the range is the set of non-negative real numbers. The function is the *area function*: it assigns to each (polygonal plane) figure its area, or more precisely, the ratio of its area to a standard unit area. The corresponding function for three-dimensional fiigures is the *volume function*.

Functions as lists: Given a function $F : D \to R$, we could imagine a list of pairs $(d, F(d))$, in which the first member is any element of the domain D, and the second element is a member of R, specifically, the value of F at d. This list L_F would have the property that, for each member of D, there was exactly one pair in the list with first element d. A function defines such a list: if we know the function F, we could check whether a given pair (d, r) was on the list, by simply checking whether $r = F(d)$. But also, such a list L defines a function F_L; for if we know there is exactly one pair on the list L with first element d, then we can define $F_L(d)$ to be the second element of the pair whose first element is d.

So, from a formal point of view, we can think of functions as lists, of the type just described. However, we usually do not describe functions that way. Part of the reason is practical: we often want to define functions in infinite sets, and it is not so practical to make an infinite list. We need some more direct way of checking whether a given element of R is $F(d)$ than looking it up in a huge list. So functions are usually described by some procedure that creates $F(d)$ from knowledge of d. In the case of "real functions", we usually give a formula that tells us how to produce $F(d)$ from d by doing a certain computation. More generally, there is often some rule we can use to determine $F(d)$ without reference to a list, in particular, without having to know any other values of F. For example, as described below, many numerical functions are described by means of expressions.

New functions from old:

A. Restriction and Extension

If f is a function from a domain D to the range R, and if D_1 is a subset of D, then f assigns an unique element of R to each element of D_1, and this gives us a function $f_1 : D_1 \to R$. The function f_1 is called the *restriction* of f to D_1. The function f is callled the *extension* of f_1 to D.

For example, consider the area function (example 7) above). We have defined its domain as the set of all polygonal regions in the plane. On this rather large set, there is no convenient formula for the area of a typical region. However, if we consider the subset of all rectangles, then we know a nice formula for area on this subset. Specifically, a rectangle has two sets of parallel sides, and the lengths of these sides give us a pair of numbers (ℓ, w). We can assume that $\ell \geq w$, so that the order of the numbers is well-defined. This gives us a function (the "sidelength function") from the set of rectangles to the set $\mathbf{R} \times \mathbf{R} = \mathbf{R}^2$ of pairs of real numbers (ℓ, w). The by our convention, the image of the sidelength function is not all of $\mathbf{R}^2$, but only those pairs (x, y) with $x \geq y$). If R is a rectangle, we can write $(\ell(R), w(R))$ for the sidelengths of R. Then the area function A restricted to the set of rectangles can be described by the equation $A(R) = \ell(R) \cdot w(R)$.

The situation described in the first paragraph may make the idea of restriction seem rather dull, but it could happen that we have found the function f_1 defined on D_1 in some context, and then in another context, found the function f on D, and then we might have noticed that every time we evaluate f on an element of D_1, we get the same value as we get from f_1. It might be surprising to find that there was a function defined on a larger domain that takes the same values, i.e., that f_1 could be extended to D.

In fact, this is the situation in when function is described by a table, and we are asked to find "the pattern". "The pattern" amounts to an extension of the function described in the table to a larger domain.

For example, consider the table

x	$f(x)$
1	1
2	3
3	6
4	10

Given this table, one might say to oneself: each time x goes up by one, $f(x)$ goes up by one more. This might lead to creation of more entries in the table:

5	15
6	21
7	28

Or we might think of the situation geometrically, in terms of a shape made of squares, arranged in rows of length one, two, three, etc, and imagine rotating the shape around its middle so that the two shapes would fit together, and make a rectangle of base equal to x, and height equal to $x+1$. This would suggest to us the formula $f(x) = \frac{x(x+1)}{2}$. Then we might further note that this formula actually makes sense, not only for positive integers, but for all real numbers.

This progression amounts to extending the function given in the original table. The domain of the original function consisted of the four numbers $\{1, 2, 3, 4\}$. When we computed more values by "extending the pattern", we created a larger table, describing a function whose domain was $\{1, 2, 3, 4, 5, 6, 7\}$. Then when we reasoned about the general term, we effectively extended the function to the domain of all positive integers. Finally, when we looked at the formula and realized is made sense for any x, we obtained an extension to a function with domain all real numbers. Not bad for starting with a domain of four numbers!

B. Composition

One of the sources of power of the function idea comes from the fact that we can combine functions by the process known as *composition*. It works like this. Suppose we have a function $f : D_1 \to D_2$ from a first set D_1, and another function $g : D_2 \to D_3$. Then we can apply f and g in succession to get a function from D_1 to D_3. This is called the *composite function*, or the *composition* of f and g, and is denoted $g \circ f$. The definition is

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

This makes sense, because the value $f(x)$ belongs to D_2, which is the domain of g.

Composition is a flexible way to create a wide variety of functions from very simple beginnings. Here are a very few examples of composite functions.

1) The area function on the set of rectangles is the composition of the sidelength function from rectangles to $\mathbf{R}^2$, and the product function $M : \mathbf{R}^2 \to \mathbf{R}$, of example 3) above.

2) The ticket cost function. If a group of people visits Disney World, the each person in the group has to figure out what kind of ticket will provide most pleasurable visit. This gives us a function from the set of visitors to the set of Disney World ticket types. Then to figure out how much each person must pay, one applies the Ticket Price function. This gives us a function from the set of visitors to the real numbers, which tells each visitor the price of his or her ticket.

3) If we have two lines in the Euclidean plane, we can reflect across one line, and then across the second line. This composition gives us another transformation of the plane. If we assume that the two lines are typical, and intersect, it turns out that this composite transformation is a rotation around the point of intersection of the two lines.

4) We can embed the positive integers $\mathbf{N}$ in the product $\mathbf{N} \times \mathbf{N}$ of two copies, by letting one component be a constant. That is, given a positive integer a, we can define a function $E_a : \mathbf{N} \to \mathbf{N} \times \mathbf{N}$ by the formula

$$E_a(x) = (a, x).$$

Suppose that we then compose E_2 with the sum function $A : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$. We obtain

$$(A \circ E_2)(x) = A(2, x) = 2 + x.$$

Or suppose that we compose E_2 with the product function M. We then obtain

$$(M \circ E_2)(x) = M(2, n) = 2n.$$

By combining these operations, we can produce any linear expression in x;

$$A(E_3(M(E_2(x))) = A(E_3(2x)) = A(3, 2x) = 3 + 2x.$$

We can also embed $\mathbf{N}$ in $\mathbf{N} \times \mathbf{N}$ by the *diagonal mapping*: $\Delta(x) = (x, x)$. We can compute that $(M \circ \Delta)(x) = M(x, x) = x \cdot x = x^2$.

Thus, we can obtain simple expressions by using composition of the mappings defining the basic arithmetic operations with other simple mappings. In fact, by combining compositions in appropriate ways, we can build arbitrary polynomial expressions. It is not necessary to think of expressions as being constructed out of compositions of mappings, but this construction does illustrate the power and flexibility of composition as an operation on functions.

Expressions. Many functions are described via *expressions*. "Expression" is another widely used term in mathematics, but it has not been formalized at the same level as "function". There are different types of expressions, and some have been defined formally, but there is no accepted overall definition of the term "expression" that covers all cases.

A *polynomial expression* in one variable x is an expression that can be obtained by multiplication and addition of x by itself and with numbers. It is the smallest set of expressions that has the property that

i) The identity function $x \to x$ is a polynomial.
ii) The constant function $x \to a$, for any fixed number a, is a polynomial.
iii) If P and Q are polynomials, then $P + Q$ is a polynomial.
iv) If P and Q are polynomials, then PQ is a polynomial.

One can also consider polynomial expressions in several variables.

A polynomial expression defines a function, but many different expressions can define the same function. For example, because of the Commutative Rule for addition, the expressions $x + 2$ and $2 + x$ define the same function of x, In other words,

$$x + 2 = 2 + x,$$

for all positive integers x. The statement that two expressions yield the same function is called an *identity*.

A way of thinking about identities is that an expression is a specific recipe for producing a function: $2 + x$ says "Take the number 2, and add to it the variable number x." On the other hand, $x + 2$ says "Take the variable number x, and add to it the number 2."" These two recipes yield the same number for any value of x.

Because we are so used to the Commutative Rule for addition, the difference between $2 + x$ and $x + 2$ may seem to involve splitting hairs. An identity that is more interesting, and also clearly involves different processes is

$$x^2 - y^2 = (x - y)(x + y).$$

The left hand side says: "Take any number x and multiply it by itself; then take any (other) number y and multiply it by itself; then subtract y^2 from x^2. The right hand side says: "Take any number x and any (other) number y, and subtract y from x; then add y to x; then multiply the result of the second procedure by the result of the first. The identity states that the two rather different processes involved in forming the two expressions from any given numbers x and y will always give the same number as a result. To demonstrate this involves using several rules of arithmetic, including several applications of the Distributive Rule.