Some things I have always wanted you to know about expressions, equations, and functions, but couldn't be bothered to tell you

William McCallum

3 October 2008

Expressions

An *expression* gets its name from the fact that it expresses something. A mathematical expression expresses the result of a calculation with numbers. Some of the numbers might be explicitly given, like 2 or π or 1.25. Other numbers in the expression might be represented by letters, such as x, y, P, n, or ξ. The numbers these letters stand for might or might not be specified by the mathematical narrative surrounding the expression. The expression itself has nothing to say on this issue.

Sometimes we are in a situation where we want to allow the possibility that the letters might stand for a range of possible numbers, and for this reason we sometimes call the letters *variables*. In other situations, they might be called *constants* or *parameters*. The different vocabulary terms do not refer to any differences that are intrinsic to the letters used or to the expressions they are used in. Rather they refer to differences in the mathematical context for the expression. It is useful to think of expressions as sentence fragments, which can be used in many different sorts of sentences. Different types of sentences cause us to give different names to the letters. At any rate, no matter what the letters in an expression are called, it is understood that at any given moment when we are considering an expression, we are supposed to regard all the occurrences of the same letter in that expression as having the same value.

What sorts of operations are allowed in an equation? We will come back to this question later, but for now let us limit the operations to the basic operations of arithmetic: addition, subtraction, multiplication, and division. Expressions have various syntax rules, which are captured in conventions about the order of operations and the use of parentheses.

The calculation an expression represents might use only a single operation, such as in $4 + 3$, or $3x$, or it might use a series of nested or parallel operations, such as in $3(a^2 + 9) - 9/b$. The expression might not have any operations at all. Both 3 and A are expressions. If the expression contains letters, then we do not know its value until we know the values of the letters. Choosing specific values for the letters and calculating the resulting value of an expression is called *evaluating the expression*.

Equations

An *equation* is a statement that two expressions have the same value. The statement need not be true: both $3 = 4$ and $2 + 2 = 4$ are equations. If the expressions on either side of an equation contain variables, then the truth of the statement cannot be determined until we assign values to those letters. Thus, in general, an equation has no instrinsic meaning or truth-value. As with expressions, the meaning is determined by the mathematical context. The typical sorts of sentences in which an equation might occur contain, either explicitly or in coded form, logical quantifiers for the variables (Does there exist an x such that $x^2 = -1$? Is $a^2 - b^2 = (a - b)(a + b)$ for all a and b?)

A set of values for the variables in an equation that makes the equation into a true statement is called a *solution* of the equation. *Solving* the equation is understood to mean finding all the solutions. For this to be a well-defined request, we need to specify the domain of numbers from which the solutions can be taken. This can be indicated in various ways. We might say "find all real solutions to $3x^2 + x = 1$" or "consider the equation $z^2 = -1$ in a complex variable z". Often the domain is not stated explicitly.

An equation whose solution set consists of all numbers in a certain domain is called an *identity*.

The solutions of an equation in real variables can be graphed in a Cartesian space of the same dimension as the number of variables. For example, the solutions to an equation in one real variable can be plotted on the number line, and the solutions to the equation $y = 3x + 2$ form a line in the Cartesian plane. We say two equations are *equivalent* if they have the same solution set.

Functions

An expression can naturally be regarded as an input-output machine: you input values of the variables and the expression outputs the value of the calculation it represents with those numbers. [1] This leads to the notion of a function: something that accepts inputs and gives outputs. The difference between a function and the expression that defines it is that, in the formal notion of a function, the actual calculation used to produce the outputs is irrelevant. Any calculation that always gives the same result gives the same function. Indeed, the formal definition of a function is that it is just a set of input-output pairs, a subset of the Cartesian product of a certain set of possible inputs and another set of possible outputs. The metaphor of the input-output machine, with its implicit suggestion that we must always get the same output every time we give the same input, is captured in the formal definition by a condition imposed on the set of input-output pairs, that no two pairs can have the same input and

[1] Expressions that are used for this purpose, such as πr^2, are often called formulas. But the term formula is also sometimes used to refer to the equation $A = \pi r^2$.

different outputs. This simple insight is often disguised as then evil vertical line test.

The set of possible inputs is called the *domain* of the function. If the set of inputs is a subset of the real numbers, we say we have a *function of a real variable.*

Since a function is formally just a set of ordered pairs, we can plot the set in the Cartesian plane (if the input and output sets are subsets of the real numbers) to get the *graph* of the function.

The conceptual leap from looking at expressions to looking at the functions they define is often under-estimated. This is exacerbated when we start using letters stand for functions. If f is a function, then $f(x)$ is the output of f when given input x. For example, if f stands for the function of a real variable whose output is the square of its input, we write $f(x) = x^2$ to indicate this arrangement. Notice that it is also true that $f(x) = (-x)^2$. The expressions x^2 and $(-x)^2$ are different expressions for the same function. Notice that x^2 is not itself a function. It is important to maintain the distinction between expressions and the functions they define. If two expressions define the same function, we say that they are *equivalent expressions.*

Functions and Equations

Once we introduce functions and function notation, we open ourselves up to fair amount of confusion. First, when we define a function f using an expression, we write something that looks very like an equation: consider the function f defined by $f(x) = x^2$. Is this last object an equation or not? Second, functions and equations both have graphs. The function in one variable defined by $f(x) = x^2$ has the same graph as the equation in two variables $y = x^2$. This often leads us to talk about "the function $y = x^2$", which is, strictly speaking, incorrect. For one thing, although there is a strong convention that x stands for the input variable and y stands for the output variable, an equation in two variables does not, in and of itself, specify which of its variables is which. The tangles in the curriculum caused by confusing functions and equations reach their climax when we start dealing with inverse functions, and the mysterious incantation to "swap the variables" appears.

Functions and Expressions

Once we have introduced functions and function notation, we have the possibility of expanding the repertoire of expressions. A function can be thought of as an operation on numbers, and we can add functions to the allowable operations in forming expressions. Thus, for example, we have exponential expressions like $P(1 + r/n)^{nT}$, and trigonometric expressions like $a^2 + b^2 - 2ab\cos\theta$. From this point of view, $f(x)$ itself an be regarded as an expression. This answers the question we asked in the previous section: $f(x) = x^2$ is an equation, in fact it is the defining equation for the function f.