Very 'final' version.

# The Development and Use of Variables in Mathematics and Computer Science

*Meurig Beynon*  *Steve Russ*

*Department of Computer Science, University of Warwick*

*Abstract*

There are a wide variety of uses of variables in mathematics which we cope with in practice through conventions and tacit assumptions. Experience with computers has made us articulate, criticise and develop these assumptions much more carefully. Historically the term 'variable quantity' was introduced in the context of describing and calculating *changing quantities* which corresponded to phenomena in the observable world (eg the velocity, or fluxion, of a body moving under the inverse square law). The evolution of the concept has divorced it from these 'roots of reference' and required us to establish the formal apparatus of interpretation and valuation. While the changes considered are highly structured this may be satisfactory, but computing power invites us to cope with change in vastly more complex, unstructured situations such as in simulation of 'real world' processes. We relate this challenge to the distinctive differences in the use of variables in mathematics and practical computing, and we develop a general framework in which all uses of variables can be described in a unified way.

## §1. A general framework for variables

*State and the notion of primitive variable*

From about 3000 BC until about 10 years ago people occupied with engineering, navigation, commerce, science or mathematics depended to a great extent on the construction and use of large tables of data. These tables formed a vital part of their capacity to organise, explain and predict changes in the world. They ranged from Babylonian tablets recording astronomical data alongside data about the weather, the frequency of earthquakes and prices of wheat, through later tables for such matters as high-tides, mortality and compound interest, up to the 4-figure logarithm tables used by generations of school-children until the late 1970's. Such systematic association of different values with the same family of symbols illustrates the use of a primitive concept of variable that will be significant in the general framework to be proposed. Such *primitive variables* are typically used for describing *state*.

There are two fundamental elements involved in the notion of a primitive variable: a symbol and the capacity for that symbol to be associated with a value. A variable (of any description) is not the name of a value, like $\pi$, since it is characteristic that different values may be associated with the same variable (for example, choosing different points on a parabola or computing successive values of a counter in a loop). It is often convenient and powerful not to associate any value with a variable but only to manipulate it in ways appropriate to its type. This capacity to be associated with different values or no value is reflected in the picture of a variable (common in computing) in which we associate with a symbol a unique 'location' which may, or may not, 'contain' a value. A primitive variable serves this function of a 'marker' for a pigeon-hole. This expresses the idea of 'capacity to be associated with a value', while also matching the way our minds seem to deal with different values being associated with concepts. For example, if x metres is the height of a projectile, when we make a program declaration, "x : real;" there is a physical memory location associated with 'x'; and when we say "as x increases, at some time it will be greater than ... ", we mean by 'it', not the symbol 'x' but the value associated with the height of the projectile and so with 'x'.

In the past, our methods for recording complex state information were very laborious, and of limited scope. A major contribution of mathematics was to provide means of simplifying the representation of state. By identifying relationships between variables, it became possible to reduce the amount of information to be recorded in order to characterise the state of a system. In this way, modern mathematics has a central interest in systems whose behaviour can be understood in terms of states in which the possible values of variables form a class: $\{v_1,v_2,v_3,.... \mid P( v_1,v_2,v_3,.... )\}$, where P is a system of relationships that might include explicit definitions of one variable in terms of others, or an equational constraint governing several variables.

In mathematics an isolated primitive variable has only a minor role. The applications for which there are no relationships between variables to be recorded, either because we are unable to perceive them or because they don't exist, have traditionally been divorced from mathematics. The nearest equivalent to the mathematical equation or functional relationship may be the methodologies by which we seek to systematise cataloguing, and the theories by which we organise species. The advent of the computer has revolutionised our power to represent state. By physically constructing arrays of millions of boolean variables that can be freely assigned, we can in principle conveniently record many more states than by any other means. Through our dependence upon computer records, the whole economy of advanced nations has come to rely upon such power. A potent side-effect of practical computing has been to liberate the potentially anarchic primitive variable upon which procedural programming at every level of abstraction depends, and to challenge mathematics to discipline its use.

*The uses of variables and the notion of context*

In modern practice a mathematical variable is used in a wide variety of ways. It is used as an indeterminate, as an arbitrary constant or as an unknown value to be found. It is used to express relationships and functions, as a parameter to a class of problems or objects, as a dummy (bound) variable in the context of operator symbols like summation and integration, and it is used in the context of limit operations. Comparing the notion of a primitive variable with the apparently more complex mathematical variable raises the issue of whether we should be speaking of different kinds of variables as well as different uses of variables. The question will become even more pertinent when we include variables as used in spreadsheets and variables over compound data structures in both computing and mathematics.

By a *valuation* we shall mean a relation between a set of primitive variables and a set of values of appropriate types. We define a *context* as a set of primitive variables together with a means for providing their valuations. An example of a valuation is the state of a computer system or experimental environment in which the set of values have been (or at least could be) 'observed' simultaneously. In a different kind of context some or all of the values in any valuation may be related to one another (eg the co-ordinates of points on a parabola). Contexts and their valuations can be viewed in many ways and form a convenient unifying framework for discussing different uses of