VARIABLES: SYNTAX, SEMANTICS AND SITUATIONS

JOHN T. BALDWIN

Developing the concept of variable seems to be a major obstacle for students in learning algebra [KN04, MC01, LM99, Usi88]. In this article we describe in the context of high school algebra a framework for understanding the notion of variable. Two components of this framework, syntax and semantics, are standard mathematical notions. Since their precise formulation might not be familiar to the reader, we summarize the ideas in the high school algebra context and place them in a historical framework. But this analysis suffices only for 'naked math'. We then discuss the third component: situation. If we view these three notions as the vertices of a triangle, our claim is that all three vertices of the triangle are essential elements of learning algebra. The goal of this paper is to lay out this mathematical and epistemological framework; in future work we intend to employ this analysis as a tool to explore specific lessons and student work.

This paper expands one section of our joint work with Hyung Sook Lee and Alexander Radosavljevic [BLR00]. We acknowledge helpful comments by Sergei Abramovich, Daniel Chazan, William Howard, Cathy Kessel, Lena Khisty, Mara Martinez, Aria Razfar, Kelly Rivette, Judah Schwartz and Kristin Umland.

1. Theoretical Background

This paper draws from a number of perspectives including logic, mathematics education, history of mathematics, and philosophy. Our first goal is expository. We just describe the syntax and semantics of first order logic as it been accepted by philosophers, mathematicians, and logicians for the last seventy-five years. We argue that this notion of variable is simpler and more appropriate for high school algebra than the notion of variable quantity that remains as vague now as it was when introduced in the 17th century. But we further argue that an analysis of quantity is necessary to connect pure algebra with concrete problems. And we lay out such an analysis. It is only at this stage that questions of the psychology of learning enter our discussion.

Arcavi and Schoenfeld [AS88] describe many different uses of the word variable by professionals in various fields. Other mathematics educators [Usi88, Cha00] also stress the many uses of variables; many of these accounts suggest that these are really different meanings of the word variable. We attempt to provide a single explanation by insisting that no use of the word variable can be fully understood without specifying a context: syntax, semantics, and, in applications, situation. Rather than positing many meanings of 'variable' we fix a particular notion with several components; in fixing a particular context each of these components must be specified. In discussing mathematical syntax and semantics we are just rehearsing the standard 'Tarski semantics'[1] for the

Date: January 8, 2009.

The research for this article was partially supported by NSF grant DMS-0500841. Further partial support is from a grant from the National Science Foundation to the Center for the Mathematics Education of Latinos (Cemela) (No. ESI-0424983). Dr. Lena Licn Khisty, is Principal Investigator at University of Illinois Chicago (UIC) for Cemela. The findings and opinions expressed here are those of the author(s) and do not necessarily reflect the views of the funding agency.

[1]Tarski gave a mathematical definition of truth in an arbitrary mathematical structure in [Tar35].

special case of high school algebra. However, in order to describe applications of mathematics, we add a third component: situation[2]. For this analysis we draw heavily on the Judah Schwartz study [Sch] of the semantic aspects of quantity.

Among the potpourri of definitions of variables advanced in [AS88], they find two major types: polyvalent names and variable objects. We identify these complexes of ideas as the 'substitutional approach' and the 'function approach'. We discuss two aspects of the function approach. First we distinguish the modern notion of function from the earlier notion[3] of 'variable quantities'. Dominguez [Dom04] gives a nice and short summary of how these notions were understood in the 18th and early 19th century. The 'variable quantity' viewpoint has largely died out in advanced mathematics with the acceptance of the modern view of function. But the tradition persists in some mathematics education literature e.g. [Cha00, AS88] and in some basic texts [Loo82]. We then show how the function approach is subsumed by the substitutional approach and argue that 'variable quantity' mystifies an idea that is now well understood.

Usiskin [Usi88] separates the kinds of use on another axis. He lists four conceptions, algebra as generalized arithmetic, procedures for equation solving, study of relationship among quantities, and study of structures. The first three of these conceptions correspond to our categories discussed in Section 2: laws of algebra, kind 2: unknowns, and kind 1,3,4: function arguments, curves, function families. His fourth category of structures corresponds to the role in our analysis of choosing different models (or structures) to interpret an inscription.

Our analysis is rooted in the investigations of the notion of variable by such philosophers and mathematicians as Peirce, Frege, Hilbert, Löwenheim, Skolem, Gödel, and Tarski. We will first describe in Section 2 the application to high school algebra of the general procedure for defining truth in a mathematical structure that resulted from these investigations anc can be found in any undergraduate text in symbolic logic (e.g. [Hed04, Mat72]. There is no thought that a fully formal explanation of the meaning of variables as begun in Section 2 is part of the K-12 curriculum; rather it is a way to describe one aspect of that curriculum. Note however that the description we give below of the interpretation of expressions and equations is implicit in many high school algebra books, e.g. ([Edu09, Bea00, Mea94]).

We will comment several times on Chazan's work in [Cha00]; despite our critique of some his analysis, we find his classroom suggestions extremely helpful. We adapt Schwartz [Sch] description of quantity as key to our analysis of situation. But we realize that this account does not represent the broader scope of his inquiry.

Our approach draws on Emily Grossholz's [Gro07] concept of 'productive ambiguity', which we will explain further below. Roughly speaking, productive ambiguity refers to a phrase or inscription being capable of distinct informative interpretations. We argue that the productive ambiguity of a symbolic expression can be disambiguated by specifying the mathematical interpretation and the physical situation where it is being used. Nevertheless, the ability to transfer not only individual inscriptions but their relations from one context to another is at the heart of mathematical applicability.

In general our departure from the mathematics education literature discussed above that distinguished many kinds of variable is that we seize on one description of the relation between

[2] The term situation was first chosen for alliterative effect. We then realized that the term had been subliminally picked up the natural language semantics of [BP99]; that is a far more sophisticated study than we attempt here. Another source is the productive use of the term *problem situation* in [DW90].

[3] [AS88] refer to a 1710 definition of variable quantity.

syntactic objects including variables and mathematical structures. In Section 2.1 we show how this relationship can be interpreted in various ways to encompass the meanings these other writers have elaborated. Our claim is that this unifying procedure makes the mathematics at various levels easier to understand.

Logicians divide the elements of mathematical language into 'terms' which denote objects of a model (a set of numbers) and 'formulas' which denote[4] truth or falsehood. The language of high school algebra generally uses 'expressions' for terms and 'sentences or statements' for formulas. We adopt this second convention and write *inscription* when we mean either an expression or a statement.

2. Syntax and Semantics

In this section we describe in the context of high school algebra one of the insights of modern logic: the notion of variable can cogently be explained by describing both a system of formal inscriptions (a mathematical language) involving variables and the interpretation of these inscriptions in number systems. Connecting this mathematical notion with concrete problems requires detailing the *situation* connecting the inscriptions and the numbers with specific entities. We explore this connection in Section 3.

As we discuss in the historical section (Section 2.6 much of the difficulty in understanding variable comes from difficulty in disentangling the related notions of *function* and *variable*. Before beginning our more detailed analysis of variables we quickly sketch the two notions.

Functions: It is no accident that many introductory mathematics books or sets of mathematical standards include sections on functions *and* variables. But, such a linkage is not inevitable. In the simplest sense a function is a rule that assigns to each member of its domain a unique value.[5] Thus the domain might be the words (strings of letters) in English and the function f could assign to each word the number of distinct letters occurring in it. We have just described functions without using variables. Frequently, $f(x)$ is written rather than f although the x adds no information. Karl Menger [Men53] argued powerfully but futilely against writing the x more than 50 years ago. We discuss below the reasons for introducing variables into the description of function. The development of intuition for the notion of function is an important subject for study. The use of tables, graphs, in-out machines and other activities may all help to instill intuition for functions. But developing this intuition is not our concern here. We are arguing for a specific way to understand the symbolic representation of algebraic functions and children's transition to writing expressions for functions. We might write A^3 for the 'add three' function. This kind of idea has been explored extensively for developing function intuition in children (e.g. David Page [PC95] and Robert Moses [MC01]). But when the function is defined by a more complicated combination of a given set of operations on a specified domain (e.g. polynomials), it is useful to introduce a symbol such as x to represent the argument of the function. We illustrate the versatility of this notation in the examples below.

Variables The term 'variable' is used in many ways. For example, the words independent and dependent variable are introduced to describe the argument and range of a function. This use of 'variable' developed since the 17th century in an attempt to explicate calculus. Alternatively a variable can be viewed as symbol that can be replaced in the formal statement by a name for a number. This last usage, which we refer to as the *substitutional approach* appears much earlier, stemming

[4]A formula denotes a truth value only if each free variable has been replaced by a name; see Section2.1

[5]More precisely a function f from A to B is subset of $A \times B$ such that if $\langle a, b\rangle$ and $\langle a, b'\rangle$ are both in f then $b = b'$.

from at least the 16th century (Viete). Section 2.7 explains the sense in which it encompasses the first and enables the description of polynomial functions.

The crux of our analysis in that 'variable' cannot be understood in isolation. We describe the use of variable in terms of one syntactical and two semantical contexts. Syntax refers to a formal language of mathematics involving sequences of symbols (inscriptions). Mathematical semantics connects the formal language with mathematical structures. Some of these symbols represent numbers; different sorts of formal expressions involving the symbols describe numbers, functions, and truth values. 'Real world' semantics attaches these symbols and numbers to objects in the world (e.g. the number or cost of pizza). This section of the paper expounds syntax and mathematical semantics; the interface with real world semantics occupies Section 3.

2.1. **Four kinds of use of variables.** We describe below, specifically for the algebra of the real numbers, how to interpret various uses of variable in terms of the substitutional approach. In this note, variable refers (as in most mathematics) to a symbol such as x, y, ... There are three components to the use of such a symbol: an expression or equation containing the symbol, a set of numbers for the variable to represent (domain of the variable)[6], and the assignment of a concrete quantity that the numbers measure (Mary's age in years). More precisely, there is a fixed vocabulary of mathematical operation symbols; expressions and equations are formed using that vocabulary. They are interpreted in a structure for that vocabulary: a set equipped with an operation (with an appropriate number of arguments) for each operation symbol in the vocabulary. For convenience we use the same symbol, e.g., $+$ for the symbol in the formal language and for its interpretation as the usual addition on the reals. Our discussion of syntax is very close to the explication of Frege in 9.3 of [DT98].

'Algebra' generally refers to contexts where a set of numbers (e.g. $\Re$, the reals), is equipped with a set of operations called the vocabulary (e.g. $+, \cdot, 0, 1$), mapping $\Re^2$ to $\Re$. When we contrast algebra with arithmetic,[7] we are contrasting the calculation of specific operations on numbers with more general formulations. Thus expressions in arithmetic are formal strings of symbols; each symbol is either a name for numbers or a name for a fundamental arithmetical operation such as addition or multiplication. We explain below how to assign meaning to such expressions. In arithmetic we write expressions such as $1 + 1$ and equations[8] such as

$$3 + 4 = 5 + 2. \tag{1}$$

We have a set of numbers, say the real numbers, in mind and the symbols $1, 2, 3, \ldots, 1/2, 1/3, \pi \ldots$ naturally denote particular real numbers. And an equation is either true (Equation 1) or false:

$$3 + 4 = 5 + 3.$$

In studying algebra, we introduce a new group of symbols, called variables; they usually are letters such as $x, y, z, \ldots$.

[6]The term domain of the variable is well-entrenched. But the domain of the the variable is the range of the interpretation functions and the usage 'as x ranges over the reals' is also common.

[7]There is a second meaning of arithmetic as referring to operations on the natural numbers. We adopt a common mathematical phrase where one speaks of the 'arithmetic of the real numbers' to mean the calculation with $+$ and $\cdot$ on the reals.

[8]Technically, the 3 in Equation 1 is a numeral, a name for a number. Trying to make this distinction in the lower grades was one of the notorious follies of the 'new math'. But it is essential in algebra to distinguish between expressions or equations which are formal statements that either represent numbers, functions, or have a truth value. We have used the same symbol for the numeral and the number it denotes.

This allows us to write new expressions [9] such as $x + 3$ or $3x^2 + 5x + 2$ and new equations such as

(1) $y = x + 3$
(2) $3x^2 + 5x + 2 = 0$
(3) $x^2 + y^2 = 1$
(4) $b = \frac{3}{4}d$

These equations appear similar but are used in different ways. We will discuss the four *kinds* of use in turn. In each of these equations, the variables are *free* (not quantified; see Section 2.2). Equations with free variables determine relations on the real numbers (solution sets). We stress again that one cannot describe variables without specifying the domain of interpretation; in these examples that domain is the reals.

(1) **Function arguments** Life is now more complicated than when we considered arithmetic. The expression $x+3$ does not denote a number; for each particular value that is substituted for x, we get another number (the first plus 3). An expression like $x + 3$ determines (or represents) a function. In fact, we take advantage of this and write the equation $y = x + 3$. This equation is neither true nor false. Rather, it defines a subset of $\Re \times \Re$: the collection of pairs $\langle a, b\rangle$ such that $b = a + 3$. And so we compute the 'add 3' function by substituting a value for x and evaluating the expression.

(2) **Unknowns** The *solution set* of an equation in one variable is a subset of the real numbers. That is, $3x^2+5x+2 = 0$ defines the subset of those numbers a such that $3a^2+5a+2 = 0$[10]. Now since the real numbers satisfy the distributive law: $3a^2 + 5a + 2 = (3a + 2)(a + 1)$. And since the real numbers satisfy the zero product property[11] $(3a+2)(a+1) = 0$ implies that $3a + 2 = 0$ or $a + 1 = 0$. So the only two numbers that satisfy the given equation are $-2/3$ and -1. So $3x^2 + 5x + 2 = 0$ is a fancy way to describe the set $\{-2/3, -1\}$.

In this context, the word *unknown* is often used instead of variable. We are trying to find what values can be substituted for x to make the equation true. In the contexts where high school students first meet unknowns there are usually only a finite number of values for the unknown that satisfy the given equation(s). But as the next examples shows, this is misleading.

(3) **Curves** The equation $x^2 + y^2 = 1$, is a less trivial example. It defines the unit circle; all pairs of numbers (a, b) such that $a^2 + b^2 = 1$. In example (1) we have defined the graph of a function. Here we define the graph of a relation that is not a function.

(4) **Function families** How does the word 'vary' enter the picture? In the first context we *vary* the argument by choosing which number to substitute for x and then we compute the value of 'add 3' at that argument. Consider the bouncing ball experiment [Gol00]. A ball is dropped from a various heights and each time we measure the height to which it bounces back. We collect data and to analyze it we fix the following vocabulary. The 'manipulated variable', d, is drop height - the distance above the ground from which we drop the ball. The 'responding variable', b, is bounce height - the distance above the ground the ball rises

[9] In fact, if we introduce x^n as an abbreviation for the product of n x's, we have defined the class of polynomials as done in high school algebra.

[10] One may write $\{a : 3a^2 + 5a + 2 = 0\}$ to denote this set.

[11] Mathematicians would say 'have no non-trivial zero divisors'

to. Suppose that the data shows the bounce height is 3/4 of the drop height. How do we represent that information as an equation? We write

$$b = \frac{3}{4}d$$

and interpret this equation exactly as in kind 1). But this example illustrates the flexibility of our notation. $b = \frac{3}{4}d$, is the result of substituting $3/4$ for the variable k in the equation in three variables $b = kd$. For any particular ball, we find that the 'bounciness' (more formally 'coefficient of resiliency') k is constant. Thus we have a family of equations with the *parameter* k; we say the bounce height is *proportional* to the drop height.

So our analysis of the bouncing ball represents a more general phenomena. We have an equation in several variables (for simplicity: k, d, b); thus it defines a subset of $\Re^3$. For any particular choice (substitution) of a value for k, we get an equation with a 'manipulated' (or independent) variable d and a 'responding' (or dependent) variable b. To describe the graphs of these equations we consider substitutions of real numbers into the equation $b = kd$; these give us a subset of $\Re^2$.

2.2. **Quantifiers. Free and bound variables** An occurrence of a variable x in a statement is bound if it lies in the scope of a quantifier ($\forall$ or $\exists$)[12]. Otherwise the occurrence is *free*. Thus for example in the statement

$$(\exists y)y < z,$$

y is bound and z is free. A statement with n distinct free variables defines a subset of $\Re^n$. Logicians call a statement with no free variables a sentence; like a declarative sentence in English it is either true or false (in the particular mathematical structure under consideration). We now describe the most common use of quantifiers in algebra.

Laws of Algebra The equation $x(y + z) = xy + xz$ is a problematic notation. If we interpret it in the same way as the examples above we see that it defines $\Re \times \Re \times \Re$. So we really meant to write:

$$(\forall x)(\forall y)(\forall z)x(y + z) = xy + xz.$$

This sentence is true because it is true no matter what triple of real numbers is substituted for the variables. The universal quantifiers $\forall$ have *bound* the variables. Recall that the solution of a finite set (system) of equations is those tuples of numbers that satisfy each of the equations. Generally speaking, in high school algebra only universal quantifications of systems of equations or systems of equations with no quantifiers appear. And often the universal quantifiers are omitted for convenience despite the ambiguity.

Limits One does not need more complicated inscriptions involving quantifiers to study algebra. But the notion of limit does use several quantifiers. To say that the limit of a function f at the point a is b, we write:

$$(\forall \epsilon)\epsilon > 0 \rightarrow (\exists \delta)\delta > 0 \wedge [(\forall x)|x - a| < \delta \rightarrow |f(x) - b| < \epsilon.$$

[12] $\forall$ is read 'for all' and $\forall x\phi(x)$ is true in the reals if $\phi(r)$ is true for *each* real number r. Similarly, $\exists$ is read 'there exists' and $\exists x\phi(x)$ is true in the reals if $\phi(r)$ is true for *some* real number r. This description is the inductive step in the definition of truth in first order logic. The further development does not appear in 'algebra' and so we pursue it here. See any introductory text such as [Hed04, Mat72].

It is exactly this greater complexity that led to the confusion around the notion of variable in the 17th and 18th centuries.

2.3. **Polynomial rings and Polynomial Functions.** As we noted above, for the purpose of algebra, expression is another name for polynomial. And by definition, the sum or product of two expressions is another expression. Naturally, one wants to simplify notation by identifying two expressions which induce the same function on the reals. The laws of rings then justify the usual procedures in algebra for 'simplifying expressions'. The resulting structure can be described in other ways. The expressions discussed above correspond to the elements of the ring $\Re[x]$ studied in abstract algebra. The operation of each element of $\Re[x]$ on $\Re$ by substitution is then called a *polynomial function*. As an alternative to the concrete definition of operations on $\Re[x]$ we have just given, the notion of variable can be avoided by appealing to more sophisticated algebraic constructions (e.g. [Lan64] pages 180-182). A different construction more allied to the ideas here is common in universal algebra [Grä68].

2.4. **The choice of interpreting structure.** In Section 2.1 we gave four examples of how to interpret equations in the vocabulary $(+, \cdot, 0, 1)$ in the real numbers. The choice of structure is as essential aspect in understanding a particular use of variable and very much dependent on the situation. And while the usual choice in high school algebra is the real numbers, the correct solution of some problems in high school algebra depend essentially on a choosing a different structure for the value of the variables. Consider for example the question of how many buses, each carrying 45 students, are needed to transport 128 students. The problem situation calls for an integer answer. A natural first step is to construct the function $f(b) = 45b$ which tells us the number of students that are carried by b buses. We want to know for what value of b, $f(b)$ is 187. But the equation $45b = 128$ has no integer solution; 45 does not have a multiplicative inverse in the integers. But we can find the answer by operations on the integers: quotient and remainder. We must look more closely at the specific situation to determine how the quotient determines the actual solution. In the bus problem we increase the quotient by 1 since we must accommodate the remaining students. If the problem asked how many figurines requiring 45 ounces of powder can be made from a 128 ounce cannister of powder the answer would be the quotient. This sort of problem is very sensitive to practical meaning of the quotient and the remainder.

The distinction between discrete and continuous quantity is reflected in the mathematical difference between the reals and the integers (viewed as structures for $(+, \cdot, 0, 1)$. The reals satisfy[13]

$$(\forall x)x = 0 \vee (\exists y)yx = 1;$$

the integers do not. But we can (indeed must) consider the meaning of quotient and remainder in the division algorithm: $m = dq + r$ for the integers. This example also illustrates the advantage of considering two domains for the same equation. We can graph in the reals and see an approximation to the integer solution.

2.5. **Productive Ambiguity.** Grosholz introduces the notion of productive ambiguity. On page 24 of [Gro07] she writes 'When distinct representations are juxtaposed and superimposed, the result is often a productive ambiguity that expresses and generates new knowledge.'

A basic example that causes much confusion in Algebra I is the minus sign. '$-$' denotes in various contexts the unary operator denoting a specific negative number (-3), the unary operation

[13]In the following we use $\vee$ for 'or'; dually we would write $\wedge$ for 'and'.

of additive inverse $((-)(-)3 = 3$, and the binary subtraction operator $(5 - 3)$. And an example fundamental to the discussion here is functional notation: $f(1)$ means the value of f at 1; $f(a)$ means the value of f at a for some unspecified a; $f(x)$ means f. But students learning the notation may well think that each notation denotes multiplication by f.

The equals sign: The confusion between the 'evaluation' meaning of equality and the relational value of equality (which we used implicitly in Section 2.1) is well-known (e.g. [HK80, AS88]). For those who know this is a productive ambiguity. As studies such as [HK80] show, excessive emphasis on the evaluative interpretation of equality in the early grades can have a disastrous effect on student understanding. A third meaning of equality is given by Example 2.1.2. We construed this earlier as finding the set of x on which the expressions the left and right hand side of an equation have the same value. In view of the discussion in Example 2.1.2 (see also Section 2.7.1), this can also be viewed as asking when the functions defined on each side have the same value (i.e. when their graphs intersect). Arcavi and Schoenfeld [AS88] earlier provided the following example concerning this notion:

$$(1) \qquad (\forall x)[\frac{1}{x-1} - \frac{1}{x+1} = \frac{2}{x^2-1}]$$

$$(2) \qquad (\forall x)[\frac{2}{x^2-1} = \frac{1}{x-1} - \frac{1}{x+1}]$$

They point out that in one context Statement 1 represents the subtraction of two rational expressions while in another context Statement 2 represents the partial fraction decomposition of the rational function $\frac{2}{x^2-1}$. In order to make sense of the second statement we must modify our meaning of equality to mean: either the expression on the left and right are both defined and equal or neither is defined. When studying integration in calculus, we regard the rational expressions in Statement 2 as defining functions (just as in Example 1). Thus we could (though not in high school) regard this as a statement about function spaces.[14] Again, we see that specifying the domain in which inscriptions are interpreted is necessary to clarify the productive use of the same symbolic expression in different situations.

2.6. **A brief history of variables.** According to Mates [Mat72] the first use of letters as variables occurs in Aristotle: in the Prior Analytics A stands for an arbitrary term of syllogism; in the Physics it is a measure of force. Letters also stand for arbitrary points in Greek geometry and pairs of letters stand for line segments. But only with Diophantus do letters represent pure numbers (i.e. positive integers giving rise to the term Diophantine equation). Mahoney ([Mah94], pages 35-36) notes that Viéte in the 17th century denotes both unknowns and parameters by letters and the significant consequence that he can develop formulas in the coefficients for the solutions to equations. The semantics is of the substitutional type but Viéte allows only positive numbers as values. But he retains the Greek concept of dimension and rather than having a operator for powers, he speaks of square and cubic by special symbols. Descartes ([Mah94], page 43-45) makes several crucial extensions: exponentiation is represented by superscripts; the notion that the powers of variable having different dimensions is abandoned: the expressions can take values anywhere in the real field. But, new notions of variable are introduced with the beginning of calculus and the study of limits. Newton spoke of the ultimate ratio of quantities, of nascent and evanescent quantities. Several 18th century variants on the notion of variable quantities are

[14] The elements of a function space are functions; these are standard tools in advanced analysis

described in [Dom04]. There seems to be confusion at that time between what developed into our modern notions of variable and function. Thus many 18th century writers [Dom04] speak of the limit of a variable where we would speak of the limit of a function. Leibnitz introduced the word function and gradually through the work of e.g. Euler, D'Alembert, Cauchy and Dirichlet the function notion triumphs.[15] But the metaphysical notion of variable quantity remained. It is only in Frege's 1879 *Begriffsschrift*[16] that Church ([Chu56], page 23) finds, 'the elimination of the dubious notion of variable quantity in favor of the variable as a kind of symbol'. And finally the extension to truth in arbitrary mathematical structures, which clarifies the distinction between discrete and continuous quantity is taken during the early 20th century. This culminates in Tarski's 1933 definition of truth. Considering different structures is what makes the discussion here a Tarskian as opposed to a Fregean treatment.

2.7. **Two Claims.** We make two claims,[17] one mathematical and one pedagogical.

2.7.1. *Claim 1: In algebra, the substitutional approach subsumes the functional approach.* We take as understood the notion of a unary function[18] on the reals as a rule that determines a value for each argument. Moreover, in algebra (by definition) each rule is given by a (piece-wise) polynomial function $p(x)$.[19] In high school algebra, we frequently define functions by saying $y = x + 3$ or $f(x) = x + 3$ or more generally $f(x) = p(x)$ where f is a function symbols introduced at this point and $p(x)$ is a specific polynomial such as $x^2 + 7x + 3$. And as noted in Example 1, the graph of the function f is the set of pairs $\langle a, p(a)\rangle = \langle a, a^2 + 7a + 3\rangle$ for $a \in \Re$. In Example 1, we wrote y for the value of the function; we could as well write $f(x) = x + 3$. [20]

2.7.2. *Claim 2: Don't use the phrase 'variable quantity'.* We have just noted the standard fact that the substitutional approach provides a mathematical explanation of the connections between functions and variables in algebra. Despite the several hundred year history of 'variable quantities', they are a mischievous notion in this context. By focusing on input and output rather than rule, 'Variable quantities' emphasizes the wrong notion. As we described in Section 2.6, the notion of variable quantity arose as a mixture of the modern notions of variable and function. The key notion to explain 'regular variation'is function. The rule for a function provides a definite value for each input. Rather than quantities changing, different arguments and values are considered. We have a definite correspondence rather some indefinite object changing. The modern formulation of the limit concept (either in terms of $\epsilon - \delta$ or nonstandard analysis) is based on this understanding of

[15]At first, a function is given by a formula although possibly one involving an infinite series. Dirichlet in 1837 allows more generalized notions of law defining a function. And the current set theoretic notion comes only after Cantor.

[16]This booklet of Frege is often seen as the foundation of modern symbolic logic and quantification theory. 'Begriffsschrift' means 'formula language' or as we now say formal language.

[17]Neither of these claims in original. In fact, I can provide no source; I have 'known' them for as long as I can remember. I do recall working out how to prove the first when teaching a course for middle school teachers.

[18]A unary function depends on a single input; our remarks extend easily to functions of several variables

[19]A function p is *piece-wise* polynomial if the reals are partitioned into intervals and the restriction of p to each interval is given by a polynomial. It is an easy extension to consider rational expressions (ratios of polynomials).

[20]Formally, we add a unary function symbol f to the language. But the specification of the formal language is not appropriate in high school.

variable. In fact, it was only this understanding of variable that allowed Abraham Robinson to put nonstandard analysis[21] on a solid footing.

To emphasize this point we contrast it with the view of Chazan who writes:

> In common parlance, there is a view of quantity as amount, a static view, whereas mathematicians think of quantity as something whose value can change. ([Cha00] p. 85)

This contrast is less sharp than the historical distinction between the static and dynamic view of variable/quantity. For example in the late 18th century encyclopedia of Diderot and D'alembert (translation in [Dom04]), we find:

> In geometry we call *variable quantities* those quantities that vary following any law. Such are the abscissae and the ordinates of curves, their osculating radii, etc. They are called that way as opposite to constant quantities which are those that never change, like the diameter of a circle, the parameter of a parabola, etc.

We see a central distinction: the static view assigns to some object (variable[22] or quantity) different values; the dynamic view sees the quantity or variable assuming values (usually with respect to a rule). Thus the dynamic view tends to conflate variable and function.

Chazan goes on to quote Kaput :

> Quantities are conceptual entities that exist in people's conception of situations. A person is thinking of a quantity when he or she conceives a quality or event in such a way that this quality is measurable or countable. A quantity is composed of the object, an appropriate unit or dimension and a process by which to assign a numerical value to the quality. ([Cha00] p. 85, [Kap89], p. 45)

We agree with Chazan that the common view of quantity is static. And indeed careful reading of the quote indicates Kaput shares this static meaning.[23] The point is made even more sharply by the mathematician Susanna Epp[24]:

> It is especially in connection with functions that people describe a variable as 'a quantity that can change' or say that variables x and y 'truly vary'. But this terminology can cause students to think of variables as a bizzare new kind of being.

It seems that both 'common parlance' and mathematicians share the static view of quantity. In common language one speaks both of a quantity of wheat and more specifically of four bushels. So it hard to understand the contrast Chazan makes. If he means, one assigns values to a quantity, the two clauses describe the same concept. If he means variable quantity in the dynamic sense described above then the notion is not the modern mathematicians notion of quantity.

We should apply Occam's razor: one should not increase, beyond what is necessary, the number of entities required to explain anything. The student arrives with a static view of quantity as amount. This view is reflected in the modern understanding of how to interpret variables and functions. So rather than reintroduce the confusions of the 17th century, we should simply use the substitutional approach to variable.

---

[21]Abraham Robinson famously established in the 1960's a framework for studying infinitesimals that is as rigorous as the ϵ-δ foundation for calculus.

[22]And on the Fregean view espoused here a variable is just a letter.

[23]A bit further on he writes, 'While we can in principle assign a numerical value to a quantity, we needn't do so in order to engage in quantitative reasoning.' ([Kap89], p. 45)

[24]They were made equally forcefully 75 years ago by Tarski [Tar65] on page 4 and again on page 100.

Our argument is not with such manners of speaking as, 'as time varies'. Our objection is to the notion that a 'varying quantity' provides some new and different idea of variable than that studied in solving equations.

Kaput raises a further issue about quantities: measurement versus counting. Note that purely mathematically this issue is addressed by the two components, syntax and semantics, that we have used to describe variables. If we write $3x + 2 = 7$, we can interpret and attempt to solve that equation either in the real numbers as we described for concreteness above or in a discrete structure (the natural numbers or the integers). Which is chosen depends on the third vertex of our triangle: situation. Measurement calls for interpretation in the reals or rationals; counting in the integers or natural numbers.

3. Situation and Quantity

In Section 2, we examined two vertices of our fundamental triangle, syntax and semantics, and the edge, interpretation, that connects them. This analysis did not require the word quantity; we contrasted our approach with others that use that concept. But Section 2 concerned the relations between a language of mathematics and numbers. In this section, we consider how to connect these domains with concrete problem situations. And for this the notion of quantity is central.

Schwartz [Sch] distinguishes between 'nominal' and 'adjectival' quantity. By nominal (or noun-like) quantity or 'pure quantity', he means numbers; we'll just say numbers. But his notion of adjectival quantity is much more subtle. Unlike the view of Chazan that we critiqued above, Schwartz's notion is of *definite* quantity. He asserts (page 9) that 'adjectival quantity may be thought of as having the following structure: {measure, attribute}.' He further distinguishes that the measure may be by either 'count nouns' (cardinal numbers) or 'mass nouns'. In order to specify, e.g. the attribute width, which is measured by a mass noun, one must assign a unit and measure. So we make the representation (cardinality of set, definition of set):

There are four apples. {4, apples}

or

((magnitude, unit), attribute):

The door is three feet wide. {(3,ft), door's width}

We reflect the distinction between mass nouns and count nouns by the choice of the mathematical structure in which the inscription is interpreted. Thus, count nouns are interpreted in the natural numbers. Schwartz emphasizes that the *measurement* of a mass noun yields a non-negative rational number. He then studies the difficulties of calculation with adjectival quantities. An immediate obstacle is that the results of measurement are not closed under arithmetical operations. The natural numbers are not closed under subtraction or division. And the positive rationals are victims of the Pythagorean paradox: if we suppose to have measured the two legs of a right triangle, we cannot precisely measure the third side.

Schwartz makes an insightful analysis of many operations on adjectival quantities. But there are many cases to consider (at least four operations, at least three aspects of each quantity that is an argument for the operation). We argue that the notion of adjectival quantity is extremely useful for mathematizing a problem. And Schwartz's analysis is valuable in the teaching students to understand the use and meaning of arithmetic operation. But once the mathematical equations expressing a specific situation have been set up, the solution of the problem should proceed entirely in the mathematical domain. Let's examine this in a particular case.

3.1. **Pizza problem.** This problem adapted to a local situation an example in [LM99] and has been given both on exams and for class discussion to several cohorts of students in Methods of Teaching Secondary Mathematics. These were either graduate students or advanced undergraduates. These future teachers are asked to analyze a student's reasoning; the exact problem the student was asked to solve is deliberately left vague.

I went to the Pompeii restaurant and bought the same number of salads and small pizzas. Salads cost two dollars each and pizzas cost six dollars each. I spent \$40 all together. Assume that the equation $2S + 6P = 40$ is correct. Then,

$$2S + 6P = 40.$$

Since $S = P$, I can write

$$2P + 6P = 40.$$

So

$$8P = 40.$$

The last equation says 8 pizzas is equal to \$40 so each pizza costs \$5.

What is wrong with the above reasoning? Be as detailed as possible. How would you try to help a student who made this mistake?

In each case only about one-half of the students identified the source of the difficulty: P is the *number of pizzas* I bought; not the cost of the pizza, and not 'pizzas'. This problem proved an effective way of drawing the students attention to the need to identify the variable verbally (e.g. number of salads) and determine what set of numbers it ranges over.

In Schwartz's notation, the variable P should identified as the currently unknown count of the number of in the set: pizzas.[25] If we attempt to follow the analysis and identify the appropriate variable at each stage in the solution, we run into trouble. After applying the distributive law P, we have the equation

$$8P = 40.$$

This no longer directly corresponds to our original data. We have at least two ways to describe the quantities in this situation. We can think of P as the number of 'meals' (where a meal is a pizza and a salad) bought. In that case the coefficient 8 is the price of a meal. Or we can retain P as the number of pizzas and indeed retain that P is the 'amount spent when buying each pizza'. But now the amount is not cost of a pizza but the cost of a pizza-salad combination.

Note that we in interpreting the situation we must carefully describe not only the quantity correlated to each variable but to also the quantity correlated to each coefficient. It is only when both of correspondences have been established that the correctness of the equation can be assessed by unit analysis. Agreement of units on the two sides of the equation is a necessary but by no means sufficient test for a correct mathematization.

Schwartz's notion ([Sch], page 21) of *intensive quantity* provides further insight into the assignment of quantities to coefficients in such situations. These are descriptors like price per pound, miles per gallon, miles per hour. A first test for the adequacy of a formula purporting to describe a physical situation is that each coefficient and variable can be labeled by a meaningful quantity. This is a good test of purported descriptions of physical relationships obtained by curve fitting.

[25] In the light of Schwartz's analysis, we offer a different analysis of this situation than in our earlier paper [BLR00]. We retain the original moral; algebra is safer and easier than too-fine an analysis.

One advantage of algebra is that we do *not* necessarily have an assigned meaning for each variable that remains the same throughout the computation. We had said in [BLR00] that the example shows that this is often impossible, even in very simple situations. But a more careful examination disputes this claim. If we attempt to provide a semantics of quantity in the Schwartz sense for each equation in the calculation, we must determine the ((magnitude, unit), attribute) of each variable *and* of each coefficient. And, at least in this case, we can maintain the quantity assigned to the variable P by carefully construing the coefficient.

To elaborate a bit, in the given equation $2S + 6P = 40$, the coefficients 2 and 6 indicate unit prices of salads and pizza. 2 and 6 are not 2 dollars and 6 dollars; they represent 2 dollars/salad and 6 dollars/pizza. So, the addition of them doesn't result in 8 dollars; it is 8 dollars/(salad and pizza), or meals as we wrote above. For arithmetical as opposed to algebraic reasoning, it is important to know that the sum of the unit prices of different units requires a new unit for the sum. If we continue this fine analysis, one might think that in the equations $2P + 6P = 40$ and $8P = 40$, each P refers to a different unit: the number of salads, the number of pizzas, and the number of (salads and pizza) in order. But this neglects the fact that in the problem situation these are the same number. Again, one power of algebra is that we do not have to worry about such matters. If our formal algebraic reasoning is correct and we have assigned the proper meaning to P at the beginning of the calculation, the value of P that we find has that meaning, the number of pizzas. Of course, this power is not needed for all problems. Koedinger and Nathan [KN04] argue that for certain problems intuitive argument is more effective for algebra learners than symbolic manipulation.

Recall we have a triangle: equations, mathematical structure, situation. In order to set up the equations, we have to make a semantic analysis of the situation and establish what measure of each attribute we want to assign to each variable. If these attributes are measured by mass nouns we want to interpret the variables in the reals; if they are count nouns we want to interpret them in the integers.

The key point is that the role of the semantics of quantity is to set up mathematical equations. If the variables have been properly assigned an attribute and its measure, the equations correctly reflect the situation, and the algebraic calculation is performed correctly, the final value of the variable is the number that gives the required measurement. There is no need to keep track of the semantics of the attributes during the calculation. As Schwartz has pointed out; nevertheless, such an analysis may sometimes clarify understanding.

4. Summary and consequences for teaching

We have made four points. There is an established semantics for mathematical language. It provides a cogent and complete explanation of the purely mathematical use of variables and functions. However, to connect concrete problems with algebra one must not only interpret variables in numbers but connect the variable with the concrete situation. The notion of quantity (and indeed of adjectival quantity) play an important role in establishing a mathematical description of a concrete situation. This use does not fall back on the vague notion of a variable quantity. Rather, it assigns a measure to some attribute. There is no need, and it is often counter productive, to continue the semantical analysis of quantity during an algebraic calculation.

How does this analysis affect teaching? The point is not a change in content but a greater realization of the unity of the various topics. The crucial point that a variable is a symbol is standard in many algebra texts ([Edu09, Bea00, Mea94]). By using the unifying substitutional approach, we can present students with different aspects of the same notion rather a parade of distinct notions. The generalized arithmetic and functions approach to algebra are seen to be two sides of the same coin. Let us answer Chazan's question ([Cha00], page 89), 'What is an expression?'. An expression is a concrete object, a polynomial, such as $x^2 + 7x + 3$. Two things we can do with a polynomial are the two notions Chazan sees as conflicting: evaluate the polynomial at a particular number by substitution and define a function by writing $f(x) = x^2 + 7x + 3$. The uses are complementary not conflicting. And indeed we should continue as Chazan does in his development to motivate 'simplification' as providing a more useful expression describing the same function (See the Section 2.3). Rather than introducing confusing notions of variable quantity, focus on the role of functions in representing change. The same notion of variable underlies the solution of equations and writing formulas for functions. Not only does the substitutional viewpoint emphasize the reinforcing aspects of function approach and the generalized arithmetic approach to algebra but it provides a clear foundation for analytic geometry. The connections between solutions of equations in one variable and several variables was already explicit in Examples 2 and 3 of Examples 2.1. The development of functions in the Integrated Mathematics Program with the table of values in Overland Trail ([AR03], page 258) is a way of introducing quantity and variable that coheres with our analysis. Indeed, the exercises on 'ox expressions' in [AR03] help instill the notion of substitution as well as providing symbolic representations for functions. In another direction the teaching experiment of Mara Martinez [MB08], which combines mathematical experimentation with the beginnings of proof, makes the substitutions explicit as students develop their own expressions for a numerical relationship.

Daniel Chazan suggested comparing the treatment here with his study of the 'ship problem' on page 77-85 of [Cha00]. The problem deals with two ships that proceed in opposite directions from a given point at speeds of 22 and 28 knots. It asks when they will be 125 nautical miles apart. Chazan contrasts what he calls the 'traditional' (static) and the 'dynamic' method. In the traditional solution, one takes the distance at time x as $50x$ and sets $50x = 125$. In Chazan's dynamic approach one considers the two functions: $f(x)$ is the distance apart of the ships after x hours; the second is the function $g(x)$ which is a constant 125.[26] Then one determines where these two functions agree. Under our analysis these are not two solution methods but two steps in the solution. Chazan's dynamic method explainins in clear and welcome detail why the equation solved by the static method is appropriate. Defining the functions correctly requires attention to quantity. That is to define the function $f(x)$ as the distance between the ships at time x, we must look at the expressions $22x$ and $28x$ and realize that adding them gives this distance. That decision refers again to the specific situation and requires the geometric input that the ships are going in opposite directions. But, having established that f and g are functions whose graphs intersect at the required value of x, to actually find the solution one must have a method for determining the intersection point. And any exact solution depends ultimately on solving the equation in the traditional sense. (Graphical methods can give approximations.) We discuss a similar approach to 'sink problems' in [BLR00].

In [BLR00], we discuss in more detail the role of word problems in introductory algebra. In general, these problems are not studied because someone (indeed anyone) wants to know the answer. Nor are the techniques of algebra necessarily the best method of finding the answer to a specific problem. For certain simpler problems Koedinger and Nathan [KN04] showed a solution in arithmetic is often more natural and easier for 9th graders than an algebraic formulation. But the

[26] Interestingly, in the 18th century meaning, f would be a variable quantity and g a constant quantity.

real problems of science and engineering require that passing from the set-up in terms of quantities to purely mathematical solutions. And these mathematical solutions eventually require much more complicated technical manipulations. In beginning algebra students are provided examples they can solve in multiple ways to develop confidence in the algebraic method.

Our analysis of quantity has several consequences: it provides a framework for setting up equations when solving word problems; it emphasizes that the effort in understanding quantity is essential for mathematizing but not for calculation. Further, if the problem is phrased so that the solution calls for discrete quantities, the third vertex of our analysis comes into play. The variables should be interpreted in the integers rather than the reals.

Variables are a core notion in algebra. They can be described only in a context which includes a syntax, what operations may appear in expressions using them and a choice of the mathematical structure in which the variables are to be interpreted. In order to study concrete problems each variable must be taken to measure (or count) some attribute or object. Separating these three aspects of variable can clarify algebraic applications. Our analysis emphasizes the complementary nature of the traditional algebraic use of variables as unknowns and their more modern use as arguments of functions. Thus, the development of the function concept as recommended in (e.g. [OCP08, ST98]) is compatible with this analysis. The two interpretations in our discussion of the ship problem of an equations such as

$$28x + 22x = 125$$

provides one more instance of the productive ambiguity discussed in Section 2.5. Recognition of these productive ambiguities can be a crucial teaching tool. By realizing that different interpretations will arise and not stressing too much the particular one occurring in a particular class, a teacher can prepare students for different uses of notation that the student may confront later.

References

[AR03] Fraser Alper, Fendel and Resnick. *Interactive Mathematics Program: Integrated High School Mathematics : Year 1*. Key Curriculum Press, 2003.

[AS88] A. Arcavi and A. Shoenfeld. On the meaning of variable. *Mathematics Teacher*, 81:420–470, 1988.

[Bea00] Dolciani Brown and et al. *Algebra: Structure and Method*. McDougald-Littel, 2000.

[BLR00] J. Baldwin, H. Lee, and A. Radosavljevic. p is not pizza: Variables from grade 3 to 13. In A. White and D. Berlin, editors, *Proceedings of International Consortium for Research in Science and Mathematics Education, (ICRSME XII) Quito, Ecuador 2008*. Columbian Mathematical Society, 200? to appear.

[BP99] J. Barwise and J. Perry, editors. *Situations and Attitudes*. David Hume series. CLSI Press, 1999.

[Cha00] D. Chazan. *Beyond formulas in mathematics teaching*. Teachers College Press, 2000.

[Chu56] A. Church. *Introduction to Mathematical Logic: volume 1*. Annals of Mathematics Studies. Princeton University Press, 1956.

[Dom04] J. Domingues. Variables, limits, and infinitesimals in portugal in the late 18th century. *Historia Mathematica*, 31:1533, 2004.

[DT98] J.P. Drouhard and A.R. Teppo. Symbols and language. In Stacey, Chick, and Kendal, editors, *The Future of the Teaching and Learning of Algebra: the 12th ICMI Study*, pages 225–264. Kluwer Academic, 1998.

[DW90] F. Demana and B. Waits. *Precalculus, Functions and Graphs*. Addison-Wesley, 1990.

[Edu09] Educational Development Corporation. *CME Algebra I*. Pearson, 2009. EDC: Educational Development Corporation.

[Gol00] Howard Goldberg. The bouncing ball. In *TIMS Laboratory Investigations*. Kendall-Hunt, 2000. `http://www.math.uic.edu/~imse/IMSE/LAB/labcdrom.html`.

[Grä68] George Grätzer. *Universal Algebra*. Van Nostrand, 1968.

[Gro07] E. Grossholz. *Representation and Productive Ambiguity in Mathematics and the Sciences*. Oxford University Press, 2007.

[Hed04] Shawn Hedman. *A first course in logic*. Oxford Unversity Press, 2004.

[HK80] N. Herscovics and C. Kiernan. Constructing meaning for the concept of equation. *Mathematics Teacher*, 1980.

[Kap89] J. Kaput. Long term algeba reform: Democratizing access to big ideas. In C. Lacampagne, W. Blair, and J.Kaput, editors, *The Algebra Initiative Colloquium*, pages 33–53. U.S. Department of Education, 1989. available online.

[KN04] K. Koedinger and M. Nathan. The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*, 13:129–164, 2004.

[Lan64] Serge Lang. *Algebraic Geometry*. Interscience, 1964.

[LM99] Jack Lochhead and José P. Mestre. From words to algebra: Mending misconceptions. In *Algebraic Thinking, Grades K-12*, pages 321–328. National Council of Teachers of Mathematics, 1999.

[Loo82] L. Loomis. *Calculus, 3rd Edition*. Addison-Wesley, 1982.

[Mah94] Michael Sean Mahoney. *The Mathematical Career of Pierre de Fermat, 1601-1665*. Princeton University Press, 1994.

[Mat72] Benson Mates. *Elementary Logic*. Oxford University Press, 1972.

[MB08] M. Martinez and B. Brizuela. Integrating algebra with proof in high school, students work with variables on the calendar sequence. submitted, 2008.

[MC01] Robert Moses and Charles Cobb. *Radical Equations, Math Literacy and Civil Rights*. Beacon Press, 2001.

[Mea94] McConnel and et al. *UCSMP Algebra*. Scott-Foresman, 1994.

[Men53] Karl Menger. The ideas of variable and function. *Proceedings of the National Academy of Science*, 1953. `http://www.jstor.org.proxy.cc.uic.edu/stable/pdfplus/88557.pdf`.

[OCP08] M. Oehrtman, M. Carlson, and Thompson P. Foundational reasoning abilities that promote coherence in students understanding of function. In M. Carlson and C. Rasmussen, editors, *Making the Connection: Research and practice in undergraduate mathematics*, pages 27–42. Mathematical Association of America, Washington, D.C., 2008.

[PC95] David Page and Kathryn Chval. *Maneuvers with Number Patterns*. Dale-Seymour Publications, Palo Alto, CA, 1995.

[Sch] J. Schwartz. Semantics of quantity. `http://gseweb.harvard.edu/~faculty/schwartz/papers.htm`.

[ST98] L. Saldanha and P. Thompson. Re-thinking covariation from a quantitative perspective: Simultaneous continuous variation. In S.B. Berensah and W.N. Coulombe, editors, *Proceedings of the annual meeting of Psychology of Mathematics Education-North America*. North Carolina State University, Raleigh, N.C., 1998.

[Tar35] Alfred Tarski. Der Wahrheitsbegriff in den formaliserten Sprachen. *Studia Philosophica*, 1:261–405, 1935. translated as: The concept of truth in formalized languages in Logic, [Tar56].

[Tar56] Alfred Tarski. *Logic, semantics and metamathematics: Papers from 1923-38*. Oxford: Clarendon Press, 1956. translated by J.H. Woodger.

[Tar65] Alfred Tarski. *Introduction to Logic and to the Methodology of the Deductive Sciences*. Oxford: Galaxy Book, 1965. First edition: Oxford 1941.

[Usi88] Zalman Usiskin. Conceptions of school algebra and uses of variables. In *The Ideas of Algebra, K-12. 1988 Yearbook*, pages 8–19. NCTM, 1988.