

from Boolean Algebra to Unified Algebra

[Eric C. R. Hehner](#)

University of Toronto

Abstract

Boolean algebra is simpler than number algebra, with applications in programming, circuit design, law, specifications, mathematical proof, and reasoning in any domain. So why is number algebra taught in primary school and used routinely by scientists, engineers, economists, and the general public, while boolean algebra is not taught until university, and not routinely used by anyone? A large part of the answer may be in the terminology and symbols used, and in the explanations of boolean algebra found in textbooks. This paper points out some of the problems delaying the acceptance and use of boolean algebra, and suggests some solutions.

Introduction

This paper is about the symbols and notations of boolean algebra, and about the way the subject is explained. It is about education, and about putting boolean algebra into general use and practice. To make the scope clear, by “boolean algebra” I mean the algebra whose expressions are of type boolean. I mean to include the expressions of propositional calculus and predicate calculus. The expressions of this algebra are called “boolean expressions”. Analogously, the expressions of number algebra are called “number expressions”.

Boolean algebra is the basic algebra for much of computer science. Other applications include digital circuit design, law, reasoning about any subject, and any kind of specifications, as well as providing a foundation for all of mathematics. Boolean algebra is inherently simpler than number algebra. There are only two boolean values and a few boolean operators, and they can be explained by a small table. There are infinitely many number values and number operators, and even the simplest, counting, is inductively defined. So why is number algebra taught in primary school, and boolean algebra in university? Why isn't boolean algebra better known, better accepted, and better used?

One reason may be that, although boolean algebra is just as useful as number algebra, it isn't as necessary. Informal methods of reckoning quantity became intolerable several thousand years ago, but we still get along with informal methods of specification, design, and reasoning. Another reason may be just an accident of educational history, and still another may be our continuing mistreatment of boolean algebra.

Historical Perspective

To start to answer these questions, I'm going to look briefly at the history of number algebra. Long after the invention of numbers and arithmetic, quantitative reasoning was still a matter of trial and error, and still conducted in natural language. If a man died leaving his 3 goats and 20 chickens to be divided equally between his 2 sons, and it was agreed that a goat is worth 8 chickens, the solution was determined by iterative approximations, probably using the goats and chickens themselves in the calculation. The arithmetic needed for verification was well

understood long before the algebra needed to find a solution.

The advent of algebra provided a more effective way of finding solutions to such problems, but it was a difficult step up in abstraction. The step from constants to variables is as large as the step from chickens to numbers. In English 500 years ago, constants were called “numbers denominate” [concrete numbers], and variables were called “numbers abstracte”. One of the simplest, most general laws, sometimes called “substitution of equals for equals”,

$$x=y \Rightarrow fx = fy$$

seems to have been discovered a little at a time. Here is one special case [20]:

“In the firste there appeareth 2 numbers, that is $14x + 15y$ equalle to one number, whiche is $71y$. But if you marke them well, you maie see one denomination, on bothe sides of the equation, which never ought to stand. Wherfore abating [subtracting] the lesser, that is $15y$ out of bothe the numbers, there will remain $14x = 56y$ that is, by reduction, $1x = 4y$.

Scholar. I see, you abate $15y$ from them bothe. And then are thei equalle still, seyng thei wer equalle before. According to the thirde common sentence, in the patthewaie: If you abate even [equal] portions, from thynges that bee equalle, the partes that remain shall be equall also.

Master. You doe well remember the firste grounds of this arte.”

And then, a paragraph later, another special case:

“If you adde equalle portions, to thynges that bee equalle, what so amounteth of them shall be equalle.”

Each step in an abstract calculation was accompanied by a concrete justification. For example, we have the Commutative Law [0]:

When the chekyns of two gentle menne are counted, we may count first the chekyns of the gentylman having fewer chekyns, and after the chekyns of the gentylman having the greater portion. If the number of the greater portion be counted first, and then that of the lesser portion, the denomination so determined shall be the same.

This version of the Commutative Law includes an unnecessary case analysis, and it has missed a case: when the two gentlemen have the same number of chickens, it does not say whether the order matters. The Associative Law [0]:

When thynges to be counted are divided in two partes, and lately are found moare thynges to be counted in the same generall quantitie, it matters not whether the thynges lately added be counted together with the lesser parte or with the greater parte, or that there are severalle partes and the thynges lately added be counted together with any one of them.

As you can imagine, the distance from $2x + 3 = 3x + 2$ to $x=1$ was likely to be several pages. The reason for all the discussion in between formulas was that algebra was not yet fully trusted. Algebra replaces meaning with symbol manipulation; the loss of meaning is not easy to accept. The author constantly had to reassure those readers who had not yet freed themselves from thinking about the objects represented by numbers and variables. Those who were skilled in the art of informal reasoning about quantity were convinced that thinking about the objects helps to calculate correctly, because that is how they did it. As with any technological advance, those who are most skilled in the old way are the most reluctant to see it replaced by the new.

Today, of course, we expect a quantitative calculation to be conducted entirely in algebra, without reference to thynges. Although we justify each step in a calculation by reference to an algebraic law, we do not have to justify the laws continually. We can go farther, faster, more succinctly, and with much greater certainty. In a typical modern proof (see the first Appendix) we see lines like

$$\begin{aligned} \lambda^r a^r &= (bab^{-1})^r = ba^r b^{-1} = a^r \\ b^r &= \lambda^r b^r = (\lambda b)^r = (a^{-1} b a)^r = a^{-1} b^r a \\ (a_1^{-1} b_1)^2 &= a_1^{-1} b_1 a_1^{-1} b_1 = a_1^{-1} (b_1 a_1^{-1}) b_1 = a_1^{-1} (\mu a_1^{-1} b_1) b_1 = \mu a_1^{-2} b_1^2 \\ (a_1^{-1} b_1)^r &= \mu^{1+2+\dots+(r-1)} a_1^{-r} b_1^r = \mu^{1+2+\dots+(r-1)} = \mu^{r(r-1)/2} \end{aligned}$$

These lines were taken from a proof of Wedderburn's Theorem (a finite division ring is a commutative field) in [15] (the text used when I studied algebra). Before we start to feel pleased with ourselves at the improvement, let me point out that there is another kind of calculation, a boolean calculation, occurring in the English text between the formulas. In the example proof [15] we find the words “consequently”, “implying”, “there is/are”, “however”, “thus”, “hence”, “since”, “forces”, “if...then”, “in consequence of which”, “from which we get”, “whence”, “would imply”, “contrary to”, “so that”, “contradicting”; all these words suggest boolean operators. We also find bookkeeping sentences like “We first remark ...”, “We must now rule out the case ...”; these suggest the structure of a boolean expression. It will be quite a large expression, perhaps taking an entire page. If written in the usual unformatted fashion of proofs in current algebra texts, it will be quite unreadable. The same problem occurs with computer programs, which can be thousands of pages long; to make them readable they must be carefully formatted, with indentation to indicate structure. We will have to do likewise with proofs.

A formal proof is a boolean calculation using boolean algebra; when we learn to use it well, it will enable us to go farther, faster, more succinctly, and with much greater certainty. But there is a great resistance in the mathematical community to formal proof, especially from those who are most expert at informal proof. They complain that formal proof loses meaning, replacing it with symbol manipulation. The current state of boolean algebra, not as an object of study but as a tool for use, is much the same as number algebra was 5 centuries ago.

Boolean Calculation

Given an expression, it is often useful to find an equivalent but simpler expression. For example, in number algebra

$$\begin{aligned} & x \times (z+1) - y \times (z-1) - z \times (x-y) && \text{distribute} \\ = & (x \times z + x \times 1) - (y \times z - y \times 1) - (z \times x - z \times y) && \text{unity and double negation} \\ = & x \times z + x - y \times z + y - z \times x + z \times y && \text{symmetry and associativity} \\ = & x + y + (x \times z - x \times z) + (y \times z - y \times z) && \text{zero and identity} \\ = & x + y \end{aligned}$$

We might sometimes want to find an equivalent expression that isn't simpler; to remove the directionality I'll say “calculation” rather than “simplification”. We can use operators other than = down the left side of the calculation; we can even use a mixture of operators, as long as there is transitivity. For example, the calculation (for real x)

$$\begin{aligned} & x \times (x + 2) && \text{distribute} \\ = & x^2 + 2 \times x && \text{add and subtract 1} \\ = & x^2 + 2 \times x + 1 - 1 && \text{factor} \\ = & (x + 1)^2 - 1 && \text{a square is nonnegative} \\ \geq & -1 \end{aligned}$$

tells us

$$x \times (x + 2) \geq -1$$

Boolean calculation is similar. For example,

$$\begin{aligned}
& (a \Rightarrow b) \vee (b \Rightarrow a) && \text{replace implications} \\
\equiv & \neg a \vee b \vee \neg b \vee a && \vee \text{ is symmetric} \\
\equiv & a \vee \neg a \vee b \vee \neg b && \text{excluded middle, twice} \\
\equiv & \text{true} \vee \text{true} && \vee \text{ is idempotent} \\
\equiv & \text{true}
\end{aligned}$$

And so $(a \Rightarrow b) \vee (b \Rightarrow a)$ has been simplified to *true*, which is to say it has been proven. Here is another example.

$$\begin{aligned}
& \exists n \cdot n + n^2 = n^3 && \text{instance} \\
\Leftarrow & 0 + 0^2 = 0^3 && \text{arithmetic} \\
\equiv & \text{true}
\end{aligned}$$

And so $(\exists n \cdot n + n^2 = n^3) \Leftarrow \text{true}$, and so $\exists n \cdot n + n^2 = n^3$ is proven.

Solving simultaneous equations can also be done as a boolean calculation. For example,

$$\begin{aligned}
& x + xxy + y = 5 \wedge x - xxy + y = 1 && \text{subtract and add } 2 \times xxy \text{ in first equation} \\
\equiv & x - xxy + y + 2 \times xxy = 5 \wedge x - xxy + y = 1 && \text{use second equation to simplify first} \\
\equiv & 1 + 2 \times xxy = 5 \wedge x - xxy + y = 1 \\
\equiv & 2 \times xxy = 4 \wedge x - xxy + y = 1 \\
\equiv & xxy = 2 \wedge x - xxy + y = 1 && \text{use first equation to simplify second} \\
\equiv & xxy = 2 \wedge x - 2 + y = 1 \\
\equiv & xxy = 2 \wedge x + y = 3 \\
\equiv & x=1 \wedge y=2 \vee x=2 \wedge y=1 \\
\Leftarrow & x=1 \wedge y=2
\end{aligned}$$

These examples show that simplifying, proving, and solving are all the same: they are all just calculation.

When an expression is too long to fit on one line, it must be nicely formatted for easy reading, and when a hint is too long to fit on the remainder of a line, it can be written on as many lines as it takes, but I do not consider formatting further here. One point worth mentioning is that subcalculations (if boolean, they are called subproofs or lemmas) can save copying unchanged parts of a calculation through many lines. These subcalculations can be done in another place and referenced, or they can be done in-place, nicely formatted, to provide a structured calculation (structured proof). By far the best way to handle subcalculations is provided by window inference systems [21] [2], which open a new window for each subcalculation, keep track of its direction, and make its context available. For example, in solving the simultaneous equations, I used the second equation to simplify the first, and then the first to simplify the second.

In this brief introduction to boolean calculation, I have not taken the time to present all the rules. For a complete presentation, the reader is referred to [14]. A research monograph that uses calculational proof is [7]. A textbook on discrete math that uses calculational proof is [10]. For further discussion of calculational proofs see [9] [17].

Traditional Terminology

Formal logic has developed a complicated terminology that its students are forced to learn. There are terms, which are said to have values. There are formulas, also known as propositions or sentences, which are said not to have values, but instead to be true or false. Operators (+, -) join terms, while connectives (\wedge , \vee) join formulas. Some terms are boolean, and they have the value *true* or *false*, but that's different from being true or false. It is difficult to find a definition of predicate, but it seems that a boolean term like $x=y$ stops being a boolean term and

mysteriously starts being a predicate when we admit the possibility of using quantifiers (\exists , \forall). Does $x+y$ stop being a number term if we admit the possibility of using summation and product (Σ , Π)? There are at least three different equal signs: $=$ for terms, and \Leftrightarrow and \equiv for formulas and predicates, with one of them carrying an implicit universal quantification. We can even find a peculiar mixture in some textbooks, such as the following:

$$a+b = a \vee a+b = b$$

Here, a and b are boolean variables, $+$ is a boolean operator (disjunction), $a+b$ is a boolean term (having value *true* or *false*), $a+b = a$ and $a+b = b$ are formulas (so they are true or false), and finally \vee is a logical connective.

Fortunately, in the past few decades there has been a noticeable shift toward erasing the distinction between being true or false and having the value *true* or *false*. It is a shift toward the calculational style of proof. But we have a long way to go yet, as I find whenever I ask my beginning students to prove something of the form $a \oplus b$ where \oplus is pronounced “exclusive or”. They cannot even start because they expect something that looks grammatically like a sentence. If I change it to either of the equivalent forms $(a \oplus b) \equiv \text{true}$ or $a \neq b$ they are happy because they can read it as a sentence with a verb. But $(a \neq b) \equiv \text{true}$ confuses them again because it seems to have too many verbs. If I ask them to prove something of the form $a \vee b$, they take an unwittingly constructivist interpretation, and suppose I want them to prove a or prove b because that is what “do a or b ” means in English. The same lack of understanding can be found in many introductory programming texts where boolean expressions are not taught in their generality but as comparisons because comparisons have verbs. We find

while *flag=true* **do** *something* **od**

but not the equivalent, simpler, more efficient

while *flag* **do** *something* **od**

because *flag* isn't the right part of speech to follow **while**. Our dependence on natural language for the understanding of boolean expressions is a serious impediment.

Traditional Notations

Arithmetic notations are reasonably standard throughout the world. The expression

$$738 + 45 = 783$$

is recognized and understood by schoolchildren almost everywhere. But there are no standard boolean notations. Even the two boolean constants have no standard symbols. Symbols in use include

<i>true</i>	t	T	1	0
<i>false</i>	f	F	0	1

Quite often the boolean constants are written as 1 and 0, with $+$ for disjunction, adjacency for conjunction, and perhaps $-$ for negation. With this notation, here are some laws.

$$x(y+z) = xy + xz$$

$$x + yz = (x+y)(x+z)$$

$$x + -x = 1$$

$$x(-x) = 0$$

The first law above coincides with number algebra, but the next three clash with number algebra. The overwhelming reaction of algebraists to notational criticisms is: it doesn't matter which symbols are used; just introduce them, and get on with it. But to apply an algebra, one must recognize the patterns, matching laws to the expression at hand. The laws have to be familiar. It takes an extra moment to think which algebra I am using as I apply a law. The logician R.L. Goodstein [8] chose to use 0 and 1 the other way around, which slows me down a little more.

A big change, like using x as a variable and $+$ as an operator, would slow me down a lot. I think it matters even to algebraists because they too have to recognize patterns. To a larger public, the reuse of arithmetic symbols with different meanings is an insurmountable obstacle. And when we mix arithmetic and boolean operators in one expression, as we often do, it is impossible to disambiguate.

The most common notations for the two boolean constants found in programming languages and in programming textbooks seem to be *true* and *false*. I have two objections to these symbols. The first is that they are English-based and clumsy. Number algebra could never have advanced to its present state if we had to write out words for numbers.

seven three eight + four five = seven eight three

is just too clumsy, and so is

true \wedge false \vee true \equiv true

Clumsiness may seem minor, but it can be the difference between success and failure in a mathematical formalism.

My second, and more serious, objection is that the words *true* and *false* confuse the algebra with an application. One of the primary applications of boolean algebra is to formalize reasoning, to determine the truth or falsity of some statements from the truth or falsity of others. In that application, we use one of the boolean constants to represent truth, and the other to represent falsity. So for that application, it seems reasonable to call them *true* and *false*. The algebra arose from that application, and it is so much identified with it that many people cannot separate them; they think the boolean values really are *true* and *false*. But of course boolean expressions are useful for describing anything that comes in two kinds. We apply boolean algebra to circuits in which there are two voltages. We sometimes say that there are 0s and 1s in a computer's memory, or that there are *true*s and *false*s. Of course that's nonsense; there are neither 0s and 1s nor *true*s and *false*s in there; there are low and high voltages. We need symbols that can represent truth values and voltages equally well.

Boolean expressions have other applications, and the notations we choose should be equally appropriate for all of them. Computer programs are written to make computers work in some desired way. Before writing a program, a programmer should know which ways are desirable and which are not. That divides computer behavior into two kinds, and we can use boolean expressions to represent them. A boolean expression used this way is called a specification. We can specify anything, not just computer behavior, using boolean expressions. For example, if you would like to buy a table, then tables are of two kinds: those you find desirable and are willing to buy, and those you find undesirable and are not willing to buy. So you can use a boolean expression as a table specification. Acceptable and unacceptable human behavior is specified by laws, and boolean expressions have been proposed as a better way than legal language for writing laws [1]. They can be used to calculate the attractions and repulsions among a set of magnets.

For symbols that are independent of the application, I propose the lattice symbols \top and \perp , pronounced “top” and “bottom”. Since boolean algebra is the mother of all lattices, I think it is appropriate, not a misuse of those symbols. They can equally well be used for true and false statements, for high and low voltages (power and ground), for satisfactory and unsatisfactory tables, for innocent and guilty behavior, or any other opposites.

For disjunction, the symbol \vee is fairly standard, coming from the Latin word “vel” for “or”.

For conjunction, the symbol is less standard, the two most common choices being $\&$ and \wedge . We are even less settled on a symbol for implication. Symbols in use include

$\rightarrow \Rightarrow \therefore \supset$

The usual explanation says it means “if then”, followed by a discussion about the meaning of “if then”. Apparently, people find it difficult to understand an implication whose antecedent is *false*; for example, “If my mother had been a man, I’d be the king of France.” [19]. Such an implication is called “counter-factual”. Some people are uneasy with the idea that *false* implies anything, so some researchers in Artificial Intelligence have proposed a new definition of implication. The following truth table shows both the old and new definitions.

a	b	old $a \Rightarrow b$	new $a \Rightarrow b$
<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>
<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>
<i>false</i>	<i>true</i>	<i>true</i>	<i>unknown</i>
<i>false</i>	<i>false</i>	<i>true</i>	<i>unknown</i>

where *unknown* is a third boolean value. When the antecedent is *false*, the result of the new kind of implication is *unknown*. This is argued to be more intuitive. I believe this proposal betrays a serious misunderstanding of logic. When someone makes a statement, they are saying that the statement is true. Even if the statement is “if a then b ” and a is known to be false, nonetheless we are being told that “if a then b ” is true. It is the consequent b that is unknown. And that is represented perfectly by the old implication: there are two rows in which a is *false* and $a \Rightarrow b$ is *true*; on one of these rows, b is *true*, and on the other b is *false*.

Debate about implication has been going on for a long time; 22 centuries ago, Callimachus, the librarian at Alexandria, said “Even the crows on the roof croak about what implications are sound.”[3] [18]. In case you think that confusion is past, or just for beginners, consider the explanation of implication in *Contemporary Logic Design*, 1994 [16]:

“As an example, let’s look at the following logic statement:

IF the garage door is open

AND the car is running

THEN the car can be backed out of the garage

It states that the conditions — the garage is open and the car is running — must be true before the car can be backed out. If either or both are false, then the car cannot be backed out.”

Even a Berkeley computer science and electrical engineering professor can get implication wrong.

Implication is best presented as an ordering. If we are calling the boolean values “top” and “bottom”, we can say “lower than or equal to” for implication. It is easy, even for primary school students, to accept that \perp is lower than or equal to \top , and that \perp is lower than or equal to \perp . With this new pronunciation and explanation, three other neglected boolean operators become familiar and usable; they are “higher than or equal to”, “lower than”, and “higher than”. For lack of a name and symbol, the last two operators have been treated like shameful secrets, and shunned. If we are still calling the boolean values “true” and “false”, then we shall have to call implication “falsier than or equal to”. As we get into boolean expressions that use other types, ordering remains a good explanation: $x < 4$ is falsier than or equal to $x < 6$, as a sampling of evaluations illustrates (try $x = 3, 5, 7$). I have tried using the standard words “stronger” and “weaker”, saying $x < 4$ is stronger than $x < 6$; but I find that some of my students

have an ethical objection to saying that falsity is stronger than truth.

That implication is the boolean ordering, with \top and \perp at the extremes, is not appreciated by all who use boolean algebra. In the specification language Z [24], boolean expressions are used as specifications. Specification A refines specification B if all behavior satisfying A also satisfies B . Although increasing satisfaction is exactly the implication ordering, the designers of Z defined a different ordering for refinement where \top is not satisfied by all computations, only by terminating computations, and \perp is satisfied by some computations, namely nonterminating computations. They chose to embed a new lattice within boolean algebra, rather than to use the lattice that it provides.

Implication has often been defined as a “secondary” operator in terms of the “primary” operators negation and disjunction:

$$(a \Rightarrow b) \equiv \neg a \vee b$$

Proofs about implications proceed by getting rid of them in favor of the more familiar negation and disjunction, as I did earlier in an example. This avoids the informal explanation, but it makes an unsupportable distinction between “primary” and “secondary” operators, and hides the fact that it is an ordering. When we learn that implication is an ordering, proofs about implications become shorter and easier.

If we present implication as an ordering, as I prefer, then we face the problem of how to use this ordering in the formalization of natural language reasoning. To what extent does the algebraic operator “lower than or equal to” correspond to the English word “implication”? Philosophers and linguists are welcome to consider this question. But we shouldn't let the complexities of this application of boolean algebra complicate the algebra, any more than we let the complexities of the banking industry complicate the definition of arithmetic.

Symmetry and Duality

In choosing infix symbols, there is a simple principle that really helps our ability to calculate: we should choose symmetric symbols for symmetric operators, and asymmetric symbols for asymmetric operators, and choose the reverse of an asymmetric symbol for the reverse operator. The benefit is that a lot of laws become visual: we can write an expression backwards and get an equivalent expression. For example, $x + y < z$ is equivalent to $z > y + x$. By this principle, the arithmetic symbols $+ \times < > =$ are well chosen but $-$ and \neq are not. The boolean symbols $\wedge \vee \Rightarrow \Leftarrow \equiv \oplus$ are well chosen, but \neq is not.

Duality can be put to use, just like symmetry, if we use vertically symmetric symbols for self-dual operators, and vertically asymmetric symbols for non-self-dual operators with the vertical reverse for their duals. The laws that become visual are: to negate an expression, turn it upside down. For example, $(\top \wedge \neg \perp) \vee \perp$ is the negation of $(\perp \vee \neg \top) \wedge \top$ if you allow me to use the vertically symmetric symbol \neg for negation, which is self-dual. There are two points that require attention when using this rule. One is that parentheses may need to be added to maintain the precedence; but if we give dual operators the same precedence, there's no problem. The other point is that variables cannot be flipped, so we negate them instead (since flipping is equivalent to negation). The well-known example is deMorgan's law: to negate $a \vee b$, turn it upside down and negate the variables to get $\neg a \wedge \neg b$. By this principle, the symbols $\top \perp \neg \wedge \vee$ are well chosen, but $\Rightarrow \Leftarrow \equiv \neq \oplus$ are not. By choosing better symbols we can let the symbols do some of the work of calculation, moving it to the level of visual processing.

from Booleans to Numbers

Some boolean expressions are laws: they have value \top no matter what values are assigned to the variables. Some boolean expressions are unsatisfiable: they have value \perp no matter what values are assigned to the variables. The remaining boolean expressions are in between, and “solving” means finding an assignment of values for the variables for which the boolean expression has value \top . (Solving is not just for equations but for any kind of boolean expression.) A lot of mathematics is concerned with solving. And in particular, number algebra has developed by the desire to solve. To caricature the development, we choose an unsatisfiable boolean expression and say “What a pity that it has no solutions. Let's give it one.”. This has resulted in an increasing sequence of domains, from naturals to integers to rationals to reals to complex numbers. The boolean expression $x+1 = 0$ is unsatisfiable in the natural numbers, but we give it a solution and thereby invent the integers. Similarly we choose to give solutions to $x \times 2 = 1$, $x^2 = 2$, $x^2 = -1$, and thereby progress to larger domains. This progression is both historical and pedagogical. At the same time as we gain solutions, we lose laws, since the laws and unsatisfiable expressions are each other's negations. For example, when we gain a solution to $x^2 = 2$, we lose the law $x^2 \neq 2$.

As the domain of an operator or function grows, we do not change its symbol; addition is still denoted $+$ as we go from naturals to complex numbers. I will not argue whether the naturals are a subset of the complex numbers or just isomorphic to a subset; for me the question has no meaning. But I do argue that it is important to use the same notation for natural 1 and complex 1 because they behave the same way, and for natural $+$ and complex $+$ because they behave the same way on their common domain. To be more precise, all boolean expressions over the naturals retain the same solutions over the complex numbers, and all laws of complex arithmetic that can be interpreted over the naturals are laws of natural arithmetic. The reason we must use the same symbols is so that we do not have to relearn all the solutions and laws as we enlarge or shrink the domain. And indeed, it is standard mathematical practice to use the same symbols.

For exactly the same good reasons that we have a unified treatment of number algebras, we must now unify boolean and number algebras. The question whether boolean is a different type from number is no more relevant than the question whether natural and integer are different types. What's important is that solutions and laws are learned once, in a unified system, not twice in conflicting systems. And that matters both to primary school students who must struggle to learn what will be useful to them, and to professional mathematicians who must solve and apply laws. Historically, number algebra did not grow from boolean algebra; but pedagogically it can do so. As already argued, the use of 0 1 $+$ \times for \perp \top \vee \wedge doesn't work. To find an association between booleans and numbers that works for unification, we must use a number system extended with an infinite number. Such a system is useful for many purposes; for example, it is used in [13] to prove things about the execution time of programs (some execution times are infinite). For a list of axioms of this arithmetic, please see [13] [14]. The association that works is as follows.

	<u>boolean</u>	<u>number</u>	
top	\top	∞	infinity
bottom	\perp	$-\infty$	minus infinity
negation	\neg	$-$	negation
conjunction	\wedge	\downarrow	minimum
disjunction	\vee	\uparrow	maximum
implication	\Rightarrow	\leq	order
equivalence	\equiv	$=$	equality
exclusive or	\oplus	\neq	inequality

With this association, all number laws employing only these operators correspond to boolean laws. For example,

<u>boolean law</u>	<u>number law</u>
$\top \oplus \perp$	$\infty \neq -\infty$
$a \equiv \neg \neg a$	$x = - -x$
$a \vee \top \equiv \top$	$x \uparrow \infty = \infty$
$a \wedge \perp \equiv \perp$	$x \downarrow -\infty = -\infty$
$a \vee \perp \equiv a$	$x \uparrow -\infty = x$
$a \wedge \top \equiv a$	$x \downarrow \infty = x$
$a \Rightarrow \top$	$x \leq \infty$
$\perp \Rightarrow a$	$-\infty \leq x$
$a \vee (b \wedge c) \equiv (a \vee b) \wedge (a \vee c)$	$x \uparrow (y \downarrow z) = (x \uparrow y) \downarrow (x \uparrow z)$
$a \wedge (b \vee c) \equiv (a \wedge b) \vee (a \wedge c)$	$x \downarrow (y \uparrow z) = (x \downarrow y) \uparrow (x \downarrow z)$
$a \vee b \equiv \neg(\neg a \wedge \neg b)$	$x \uparrow y = -(-x \downarrow -y)$
$a \wedge b \equiv \neg(\neg a \vee \neg b)$	$x \downarrow y = -(-x \uparrow -y)$

There are boolean laws that do not correspond to number laws, just as there are integer laws that are not real laws. That's another way of saying that there are unsatisfiable boolean expressions that correspond to satisfiable number expressions. We will use this for our unified development.

Unified Algebra

Here is my proposal for the symbols of a unified algebra.

	<u>unified</u>	
top	\top	infinity
bottom	\perp	minus infinity
negation	$-$	negation
conjunction	\wedge	minimum
disjunction	\vee	maximum
“nand”	\triangle	negation of minimum
“nor”	∇	negation of maximum
implication	\leq	order
reverse implication	\geq	reverse order
strict implication	$<$	strict order
strict reverse implication	$>$	strict reverse order
equivalence	$=$	equality
exclusive or	\neq	inequality

The symbols $- \leq \geq < > =$ are world-wide standards, used by school children in all countries,

so I dare not suggest any change to them. The symbol \neq for inequality is the next best known, but I have dared to stand up the slash so that all symmetric operators have symmetric symbols and all asymmetric operators have asymmetric symbols. (Although it was not a consideration, \neq also looks more like \oplus .) Duality has been sacrificed to standards; the pair $\leq <$ are duals, so they ought to be vertical reflections of each other; similarly the pair $\geq >$, and also $= \neq$; addition and subtraction are self-dual, and happily $+$ and $-$ are vertically symmetric; multiplication is not self-dual, but \times is unfortunately vertically symmetric. The symbols \vee for maximum and \wedge for minimum seem to point the wrong way, but consider that \vee has its arms pointing up and it holds water, and \wedge has its arms pointing down and it doesn't hold water.

Having unified the symbols, I suppose we should also unify the terminology. I vote for the number terminology in the right column, except that I prefer to call \top and \perp "top" and "bottom".

The association between booleans and numbers suggested here allows the greatest number of boolean laws to be generalized to all numbers. For example, if a , b , and c are boolean, then it is usual to define **if a then b else c fi** by the law (in standard boolean notation):

$$\mathbf{if\ } a \mathbf{\ then\ } b \mathbf{\ else\ } c \mathbf{\ fi} \equiv (a \wedge b) \vee (\neg a \wedge c)$$

If a remains boolean but b and c are numbers, the **if**-expression on the left is still sensible (the Algol **if**), and this is still a law. In the unified notation it is

$$\mathbf{if\ } a \mathbf{\ then\ } b \mathbf{\ else\ } c \mathbf{\ fi} = (a \wedge b) \vee (\neg a \wedge c)$$

The next examples, written in boolean notation, are the laws

$$(a \wedge b \Rightarrow c) \equiv (a \Rightarrow c) \vee (b \Rightarrow c)$$

$$(a \vee b \Rightarrow c) \equiv (a \Rightarrow c) \wedge (b \Rightarrow c)$$

A common error is to use conjunction twice, or disjunction twice. The boolean reading " a and b implies c if and only if a implies c or b implies c " sounds no more reasonable than " a and b implies c if and only if a implies c and b implies c ". In unified notation,

$$(a \wedge b \leq c) = (a \leq c) \vee (b \leq c)$$

$$(a \vee b \leq c) = (a \leq c) \wedge (b \leq c)$$

it is obvious that the minimum of a and b is less than or equal to c when at least one of a or b is less than or equal to c , and the maximum of a and b is less than or equal to c when both a and b are less than or equal to c . They are laws for all numbers, not just the booleans.

The arithmetic expression $x-y$ varies directly with x and inversely with y . Thus if we increase x , we increase $x-y$, and if we decrease y we increase $x-y$. We calculate:

$$\begin{array}{ll} x-y & \text{increase } x \text{ to } x+1 \text{ and so increase the whole expression} \\ \leq & (x+1)-y \quad \text{decrease } y \text{ to } y-1 \text{ and so increase the whole expression} \\ \leq & (x+1)-(y-1) \end{array}$$

Similarly the boolean expression $x \geq y$ varies directly with x and inversely with y (no matter whether x and y are numbers and \geq is number comparison, or x and y are boolean and \geq is reverse implication, or x and y are a mixture of number and boolean). We calculate as follows:

$$\begin{array}{ll} x \geq y & \text{increase } x \text{ to } x+1 \text{ and so increase the whole expression} \\ \leq & (x+1) \geq y \quad \text{decrease } y \text{ to } y-1 \text{ and so increase the whole expression} \\ \leq & (x+1) \geq (y-1) \end{array}$$

It is exactly the same calculation. By unifying number algebra with boolean algebra we carry our ability to calculate over from numbers to booleans.

Unified Development

Suppose we start with boolean algebra in the unified notation, with the terminology “top”, “bottom”, “minimum”, “maximum”, “less than”, and so on. Now we say: what a pity that $x = \neg x$ has no solution; let's give it one. The new solution is denoted 0 . While gaining a solution to some boolean expressions, we have lost some laws such as the law of the excluded middle $x \vee \neg x$.

Now we have an algebra of three values: \top , 0 , \perp . In one application they can be used to represent “yes”, “maybe”, and “no”; in another they can be used to represent “large”, “medium”, and “small”. This algebra has 27 one-operand operators, one of which is \neg , defined as

x	\top	0	\perp
$\neg x$	\perp	0	\top

It has 19683 two-operand operators, four of which are:

xy	$\top\top$	$\top 0$	$\top\perp$	$0\top$	00	$0\perp$	$\perp\top$	$\perp 0$	$\perp\perp$
$x=y$	\top	\perp	\perp	\perp	\top	\perp	\perp	\perp	\top
$x \leq y$	\top	\perp	\perp	\top	\top	\perp	\top	\top	\top
$x \ll y$	\top	0	\perp	\top	0	0	0	0	0
$x \oplus y$	\perp	\top	0	\top	0	\perp	0	\perp	\top

Whether \leq or \ll or another operator represents implication in the presence of uncertainty can be debated, but the algebra is not affected by the debate. The operator \oplus is modular (or circular) addition, and the other operators of modular arithmetic can be given similarly.

We might continue our development with a four-valued algebra and five-valued algebra, but at this point I recommend filling in the space between \top and 0 , and between 0 and \perp , with all the integers. And then we progress to the rationals, the reals, and the complex numbers as usual.

The argument in favor of this unification of boolean algebra and number algebra is just as strong as the argument in favor of using the same notations for the different number algebras. But the latter is familiar, and so it seems right, while the former is unfamiliar, and for that reason alone it may seem wrong. Ultimately, the benefits will outweigh the unfamiliarity. For example, the data structure known as AND-OR trees and the algorithm that uses them become the same as the data structure and algorithm known as minimax methods; they should not have to be learned twice.

A different unification of boolean algebra and number algebra that aims at the same goal (using the same calculations for booleans and numbers), but emphasizes traditional modular arithmetic along the way, can be found in [5], a provocative work of grand scope.

from Informal to Formal

Many mathematical notations began their life as abbreviations for some words. For example, $=$ was introduced in [20] to mean “is equal to” (see the second Appendix):

“And to avoide the tedious repetition of these woordes “is equalle to” I will sette as I doe often in woorke use, a paire of paralleles or Gemowe [twin] lines of one lengthe, thus: $=$, because noe 2 thynge, can be moare equalle.”

Later, $=$ became associated with some algebraic properties, namely reflexivity, symmetry, transitivity, and substitutivity. Today, it is defined by those properties, not as an abbreviation for some words. Someone might say that Alice and Bob are equal tennis players because they have played each other 10 times, and each has won 5 matches. They might similarly say that Bob and Carol are equal tennis players because they too have played each other 10 times, and each has won 5 matches. But this kind of equality is not transitive. As it happens, Alice and Carol are unequal tennis players: they have played each other 10 times, and Alice has won 8 matches. Because of the lack of transitivity, no mathematician today would use $=$ for tennis equality.

In the notation commonly used for small sets, such as $\{1, 3, 7\}$, the comma was introduced as just punctuation, not as a mathematical operator. As soon as the notation is introduced, we must say that the order in which elements are written is irrelevant so that $\{1, 2\} = \{2, 1\}$; the way to say that formally is $A,B = B,A$ (comma is commutative). We must also say that repetitions of elements are irrelevant so that $\{3, 3\} = \{3\}$; the way to say that formally is $A,A = A$ (comma is idempotent). And we should say that comma is associative $A,(B,C) = (A,B),C$ so that parentheses are unnecessary. Evidently the comma can be seen as a mathematical operator with algebraic properties that aggregates elements into a structure that is simpler, more primitive, than sets; let us call them bunches. Even the curly braces can be seen as an operator that applies to a bunch and makes a set; its inverse \sim applies to a set and makes a bunch: $\sim\{1,2\} = 1,2$.

When a child first learns about sets, there is often an initial hurdle: that a set with one element is not the same as the element. It would be easier to present a set as packaging: a package with an apple in it is obviously not the same as the apple. Just as $\{1\}$ and 1 differ, so $\{1,2\}$ and 1,2 differ. Bunch theory tells us about aggregation; set theory tells us about packaging. The two are independent.

Apart from being cute, are bunches useful? The subject of functional programming has suffered from an inability to express nondeterminism conveniently. To say something about a value, but not pin it down completely, one can express the set of possible values. Unfortunately, sets do not reduce properly to the deterministic case; in this context it is again a problem that a set containing one element is not equal to the element. What is wanted is bunches. One can always regard a bunch as a “nondeterministic value”. Bunches can also be used as a “type theory” with the advantage that it is unnecessary to duplicate the operators of the value space at the type level because the two are unified. And finally, the easiest way to present sets is via bunches. For details see [13] [14]. Formalization of the lowly comma leads to a beautiful and useful algebra.

We have just seen two examples of formalization, one from the past and one from the future. Now here's an example of a formalization gone astray: functions defined as sets of ordered pairs. This way of defining functions is part of the interesting demonstration that all of mathematics can be based on sets. The demonstration requires us to implement functions, numbers, and everything else in terms of sets (logicians say “model”). For example, the natural numbers can be equated to sets, with no inconsistency, as follows:

$$0 = \emptyset \quad (\text{the empty set})$$

$$n+1 = n \cup \{n\}$$

So, for example, $3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$. Few people would say that 3 really is the set $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$; the set-implementation of natural numbers was constructed by John von Neumann just to serve this one demonstration. Numbers are best formalized, not by a set-implementation, but by an algebra showing how they participate in arithmetic operations. Similarly, functions are best formalized by showing the laws of application and function composition (in general, set union and intersection are not useful ways of combining functions). But the set-implementation of functions has somehow taken root in the current mathematical culture; many people (and textbooks) say that a function really is a set of ordered pairs. A useful formalization is not one that answers the question “what is it?”, but one that answers the question “how do we use it?”.

I write a function, or local scope, according to the following example:

$$\langle n: \text{nat} \cdot n+1 \rangle$$

This is essentially a “lambda-expression”[6], although Church did not use angle brackets. He borrowed a “hat” notation from Whitehead and Russell, but for typesetting convenience moved the hat down in front; the most similar available character in the typesetter's tray was λ ; thus the lambda calculus was born [22]. Following van de Snepscheut [23], I use angle brackets to delimit the scope of the variable. Next, I want to get rid of the idea that all possible variables (infinitely many of them) already “exist”, and that the function notation “binds” a variable, and any variable that is not bound remains “free”. I prefer the programmer's terminology of “local” and “nonlocal” variables. Variables do not automatically “exist”; they are introduced (rather than bound) with a limited scope by the function notation.

Two notations that have not yet made the transition from informal beginning to formal, calculational tool are the quantifiers \forall and \exists . For most mathematicians today they remain abbreviations for the words “for all” and “there exists”, and their meaning is just whatever can be understood from those words. The word “all” sounds clear and unambiguous, but there is debate as to whether so-called “nonstandard” elements are included. Existence is even more contentious, as can be seen from the debate between classical and constructive mathematicians. Only a formal definition, equivalent to an automated theorem prover, is clear and unambiguous. Only a formal definition gives us calculation.

Quantifiers

There are several commonly used notations that introduce a local (bound, dummy) variable. For examples,

$$\sum_{x=0}^{\infty} f(x) \quad \int_a^b f(x) dx \quad \forall x: D \cdot P(x) \quad \{f(x) \mid x \in D\}$$

The introduction of the local variable and its domain are exactly the job of the function notation, so all expressions requiring a local variable can be uniformly expressed as an operator applied to a function. If the body of a function is a number expression, then we can apply $+$ to obtain the sum of the function results. For example,

$$+\langle n: \text{nat} \cdot 1/2^n \rangle$$

There is no syntactic ambiguity caused by this use of $+$, so no need to employ another symbol Σ for addition. We can apply any associative symmetric operator, such as

$$\begin{aligned} &\times \langle n: \text{nat} \cdot 1/2^n \rangle \\ &\wedge \langle n: \text{nat} \cdot n > 5 \rangle \\ &\vee \langle n: \text{nat} \cdot n > 5 \rangle \end{aligned}$$

The minimum operator \wedge replaces “for all”, and the maximum operator \vee replaces “there exists”. By applying $=$ and \neq to functions we obtain the two independent parity operators. Set comprehension and integrals can be treated this same way.

If function f has domain D , then $f = \langle x: D \cdot f x \rangle$, so quantifications traditionally written

$$\sum_{x: D} f(x) \qquad \forall x: D \cdot P(x)$$

which we can now write as

$$+\langle x: D \cdot f x \rangle \qquad \wedge \langle x: D \cdot P x \rangle$$

can be written even more succinctly as

$$+f \qquad \wedge P$$

Using adjacency for composition, deMorgan's laws

$$\neg(\forall x: D \cdot P x) \equiv (\exists x: D \cdot \neg P x) \qquad \neg(\exists x: D \cdot P x) \equiv (\forall x: D \cdot \neg P x)$$

become

$$\neg \wedge P = \vee \neg P \qquad \neg \vee P = \wedge \neg P$$

The Specialization and Generalization laws say that if y is an element of D ,

$$(\forall x: D \cdot P x) \Rightarrow P y \qquad P y \Rightarrow (\exists x: D \cdot P x)$$

They now become

$$\wedge P \leq P y \qquad P y \leq \vee P$$

which say that the minimum item is less than or equal to any item, and any item is less than or equal to the maximum item. These laws hold for all numbers, not just for the booleans.

Given a function f , we can subtract y from each function value, and then find the minimum, or we can save a lot of subtractions by finding the minimum function value and then subtracting y from that.

$$\begin{aligned} &\wedge \langle x \cdot f x - y \rangle && \text{factor out } -y \\ = &\wedge \langle x \cdot f x \rangle - y \\ = &\wedge f - y \end{aligned}$$

The same calculation can be performed if $-$ is replaced by \geq .

$$\begin{aligned} &\wedge \langle x \cdot f x \geq y \rangle && \text{factor out } \geq y \\ = &\wedge \langle x \cdot f x \rangle \geq y \\ = &\wedge f \geq y \end{aligned}$$

The traditional reading of the top line is: for all x , f of x is at least y . But the middle line is read: the minimum function value is at least y . It is the same \wedge operator on both lines, but traditionally it is pronounced “for all” on one line and “minimum” on the other. Or maybe $f x$ and y are both boolean, and \geq is reverse implication, in which case the top line would traditionally be read: for all x , (f of x is implied by y); and the middle line would traditionally be read: (for all x , f of x) is implied by y . No matter whether f is a numeric or boolean function, no matter whether \geq is a numeric or boolean comparison, the law is the same.

If we factor from the other side of the $-$ sign, we have to change minimum to maximum:

$$\begin{aligned}
 & \wedge \langle x \cdot y - f x \rangle && \text{factor out } y - \\
 = & y - \vee \langle x \cdot f x \rangle \\
 = & y - \vee f
 \end{aligned}$$

Once again replacing $-$ with \geq , we have the same calculation.

$$\begin{aligned}
 & \wedge \langle x \cdot y \geq f x \rangle && \text{factor out } y \geq \\
 = & y \geq \vee \langle x \cdot f x \rangle \\
 = & y \geq \vee f
 \end{aligned}$$

This has several traditional readings depending on what's boolean and what's numeric. But it is just one law. Unified algebra gives us many other factoring/distribution laws just like these (see [14]).

The goal is to create an algebra that is easy to learn and easy to use. That goal is not always consistent with traditional mathematical terminology and symbology. The reader is cautioned against matching the algebra directly against their own familiar terms and symbols. Although I have been using the words “minimum” and “maximum” for \wedge and \vee , the words “greatest lower bound” and “least upper bound”, or “infimum” and “supremum”, may be more traditional in some contexts. For example,

$$\wedge \langle n: \text{nat} \cdot 1/n \rangle = 0$$

Even more caution must be used with the words “all” and “exists”. Intuition about existence in mathematics (like intuition about anything else) depends on what you have learned. We tend to believe that what we have learned is true. But mathematical truth is constructed, and we must be open to the possibility of constructing it differently. Unlearning can be more difficult than learning.

Quantifier Examples

Is $(\exists x \cdot P x) \Rightarrow (\forall y \cdot Q y)$ equivalent to $\forall x \cdot \forall y \cdot (P x \Rightarrow Q y)$? Even experienced logicians don't find it obvious. To see if they are equivalent, those who reason informally say things like “suppose some x has property P ”, and “suppose all y have property Q ”. They are led into case analyses by treating \forall and \exists as abbreviations for “for all” and “there exists” (as they originally were). Of the few who reason formally, most don't know many laws; perhaps they start by getting rid of the implications in favor of negation and disjunction, then use deMorgan's laws. Let me rewrite the question in the new notations.

$$(\vee P \leq \wedge Q) = \wedge \langle x \cdot \wedge \langle y \cdot P x \leq Q y \rangle \rangle$$

We might read the left side as saying that the maximum P is less than or equal to the minimum Q , and we might read the right side as saying that all P are less than or equal to all Q . Informal readings can be misleading, and we should never attach our understanding to an informal reading, but sometimes we can get inspiration from it. In this case, the reading sounds reasonable enough to suggest we might prove it, and not just for booleans, but for all numbers. Leaving the non-null domains implicit, here's the two-step proof:

$$\begin{aligned}
 & \wedge \langle x \cdot \wedge \langle y \cdot P x \leq Q y \rangle \rangle && \text{factor out } P x \leq \\
 = & \wedge \langle x \cdot P x \leq \wedge Q \rangle && \text{factor out } \leq \wedge Q \\
 = & \vee P \leq \wedge Q
 \end{aligned}$$

Let L be a nonempty list (a function whose domain is an initial segment of the naturals). $+L$ is its sum, and $\vee L$ is its maximum; let $\#L$ be its length. We can say that the average item in the list is less than or equal to the maximum item as follows (leaving the domain implicit).

$$\begin{aligned}
& +L/\#L \leq \vee L && \text{now apply } >1 \text{ to both sides of the inequality} \\
\leq & (+L/\#L > 1) \leq (\vee L > 1) && \text{multiply by } \#L ; \text{ distribute } >1 \\
= & (+L > \#L) \leq \vee \langle i \cdot L \ i > 1 \rangle
\end{aligned}$$

The bottom line is the “pigeon-hole principle”; it says that if the total number of things is greater than the number of places to put them, then some place has more than one thing in it. Notice what has happened: we read \vee as “maximum” on the top line, and as “some” on the bottom line; we read \leq as “less than or equal to” on the top line, and as “if then” on the bottom line.

Here is a further illustration of the benefits of unified algebra. Let f be a function from the naturals to the reals. If f is nondecreasing, then $f0$ is its minimum. Traditionally, this might be written (leaving the domain implicit) as

$$(\forall n \cdot f n \leq f(n+1)) \Rightarrow (f0 = \text{MIN} \{f n \mid 0 \leq n < \infty\})$$

Rewriting this in the new notation, and weakening it to say that $f0$ is less than or equal to the minimum, we get

$$\wedge \langle n \cdot f n \leq f(n+1) \rangle \leq (f0 \leq \wedge f)$$

Now we apply the portation law, which says that for boolean a and any b and c ,

$$(a \leq (b \leq c)) = (a \wedge b \leq c)$$

to obtain

$$f0 \wedge \wedge \langle n \cdot f n \leq f(n+1) \rangle \leq \wedge f$$

If f happens to have a boolean range, this is induction, more traditionally written

$$f0 \wedge (\forall n \cdot f n \Rightarrow f(n+1)) \Rightarrow (\forall n \cdot f n)$$

Thus we see induction as a special case of a more general law saying that the first item in a nondecreasing sequence is its minimum.

Probability

The seminal work [4] by Boole on boolean algebra refers to both logic and probability. The standard theory of probability assigns 0 to an event that cannot happen, 1/2 to an event that is equally likely to happen or not happen, and 1 to an event that is certain to happen. In a set of events in which exactly one event must happen, the probabilities sum to 1. The integral of a probability distribution (density) must be 1.

Perhaps there is another way to develop probability theory based on unified algebra. Perhaps an event that cannot happen has probability \perp , an event that is equally likely to happen or not happen has probability 0, and an event that is certain to happen has probability \top . In a set of events in which exactly one event must happen, the average probability is 0. The integral of a probability distribution must be 0. Perhaps the new probability space is related to the logarithm of the old space; essentially, probabilities are replaced by information content. My hope is that the complicated formulas for distributions in the standard theory can be simplified by transforming the space of probabilities.

Metalogic

In the study of logic, at or near the beginning, logicians present the symbol \vdash to represent theoremhood. I ask you to put yourself in the place of a beginning student. This symbol is applied to a boolean expression just like the boolean operators; but we know all the boolean operators and this isn't one of them. It sometimes has a left operand as well as a right operand, and then the explanation makes it seem just like implication. To say that it is a “meta-operator” just labels it, and doesn't explain it. Saying that it applies to the form, rather than the meaning, is

confusing too, since the entire point of the algebra is to enable us to work with the form and ignore the meaning. The distinction between metanotations and the object notations is not easily seen.

To make things worse, there are different levels of meta-operators. Proof rules are sometimes presented using a horizontal line, which is yet another level of implication. Consider, for example, the Modus Ponens proof rule, which uses all three kinds of implication:

$$\frac{A \vdash x, B \vdash x \Rightarrow y}{A, B \vdash y}$$

Rewriting comma as conjunction, and turnstile and line as implication, we get a tautology:

$$(A \leq x) \wedge (B \leq (x \leq y)) \leq (A \wedge B \leq y)$$

Rewriting any proof rule this way gives a tautology (if \vdash has nothing to its left, use \top). Rewriting any tautology whose main connective is implication gives a valid proof rule. It is hard to see the difference between the meta-operators and the object-level operators because there is no formal difference! The proof rules are used to explain how to use the boolean expressions; natural language is used to explain how to use the proof rules. For beginners (and others) it would be better to skip the meta-notations altogether and just use natural language to explain how to use the boolean expressions.

At a more advanced level, when we want a formalism to study formalisms, we will need an operator that applies to the form of an expression. For that purpose, we do not need any new kind or level of operator. Rather, we need to do exactly what Gödel did when he encoded expressions, but we can use a better encoding. We need to do exactly what programmers do: distinguish program from data. One person's program may be a compiler writer's data, but when it is data, it is a character string. The character string “ $a \vee \neg a$ ” can be used as a code for the expression $a \vee \neg a$. We apply \vdash to character strings so that $\vdash s$ is a theorem when the boolean expression represented by string s is a theorem.

We have a name, “theorem”, for a boolean expression that can be simplified to \top , and an operator, \vdash , whose purpose is to identify theorems. Strangely, logicians have not introduced a name, say “antitheorem”, for a boolean expression that can be simplified to \perp , and no operator such as \neg , whose purpose is to identify antitheorems. Perhaps that's because “antitheorem” just means “negation of a theorem” in those logics having negation and an appropriate proof rule. But we bother to name both booleans, even though one is just the negation of the other.

I propose that logicians can improve metalogic by taking another lesson from programming. Instead of \vdash and \neg , we need only one operator to serve both purposes. It is called an interpreter. I want $\mathbb{I} s$ to be a theorem if and only if s represents a theorem, and an antitheorem if and only if s represents an antitheorem. It is related to \vdash and \neg by the two implications

$$\vdash s \leq \mathbb{I} s \leq \neg \neg s$$

In fact, if we have defined \vdash and \neg , those implications define \mathbb{I} . But I want \mathbb{I} to replace \vdash and \neg so I shall instead define it by showing how it applies to every form of boolean expression. Here is the beginning of its definition.

$$\begin{aligned} \mathbb{I} \text{“}\top\text{”} &= \top \\ \mathbb{I} \text{“}\perp\text{”} &= \perp \\ \mathbb{I} \text{“}\neg s\text{”} &= \neg \mathbb{I} s \\ \mathbb{I} (s \text{“}\wedge\text{”} t) &= \mathbb{I} s \wedge \mathbb{I} t \\ \mathbb{I} (s \text{“}\vee\text{”} t) &= \mathbb{I} s \vee \mathbb{I} t \end{aligned}$$

And so on. For example,

$$\mathbb{I} \text{“} \neg \top \wedge \perp \text{”} = \neg \top \wedge \perp = \perp$$

In a vague sense \mathbb{I} acts as the inverse of quotation marks; it “unquotes” its operand. That is what an interpreter does: it turns passive data into active program. It is a familiar fact to programmers that we can write an interpreter for a language in that same language, and that is just what I am doing here. Interpreting (unquoting) is exactly what logicians call Tarskian semantics. In summary, an interpreter is a better version of \vdash , and strings make metalevel operators unnecessary.

Using \mathbb{I} , the famous Gödel incompleteness proof is just 3 lines. Suppose that every boolean expression is either a theorem or an antitheorem (a complete logic), and define Q by

$$Q = \text{“} \neg \mathbb{I} Q \text{”}$$

Then

$$\begin{aligned} & \mathbb{I} Q && \text{replace } Q \text{ with its equal} \\ = & \mathbb{I} \text{“} \neg \mathbb{I} Q \text{”} && \mathbb{I} \text{ unquotes} \\ = & \neg \mathbb{I} Q \end{aligned}$$

which proves a boolean expression equal to its negation, showing the logic to be inconsistent. A logic in which we can define an interpreter, and in which we can replace an expression with its equal, must be inconsistent or incomplete. We choose consistency, and we choose to allow the replacement of an expression with its equal, so we are forced to give up the ability to define a complete interpreter; in particular, \mathbb{I} cannot unquote “ $\neg \mathbb{I} Q$ ”. For further details of this version of Gödel's incompleteness theorem, see [12].

You cannot learn a programming language by reading an interpreter for it written in that same language. And you cannot learn logic, or a logic, by reading an interpreter for it written in logic. Not only is it inscrutable to a novice, but also it may be subject to more than one interpretation. Logic is better presented as algebra [11]. We don't present number algebra with the aid of a metaoperator that applies to number expressions and results in their values, and we should not present boolean algebra that way. I think boolean algebra should be presented with a little natural language and a lot of laws, because laws don't use any metanotations.

Terms of Honor

My final comment concerns mathematical terminology intended to honor mathematicians. In some parts of mathematics it is standard: Lie algebra, Stone algebra, Cartesian product, Jordan decomposition, Cayley transform, Hilbert space, Banach space, Hausdorff space, Borel measure, Lebesgue integration, Fredholm index, Wedderburn's Theorem, and so on. It is well known that the person so honored is sometimes the wrong person; often it is only one of many who equally deserve to have their names attached to the idea. I suspect that sometimes the intention is not so much to honor a person as to use the person's prestige to lend respectability to an idea. Whatever the reason, the effect is to obscure and make the mathematics forbidding and inaccessible. It may be argued that this is good, keeping the uninitiated from thinking they understand when they don't, but I reject that argument as elitist. I know what nand and nor are, but I forget which is the Scheffer stroke and which the Peirce arrow. To say that an operator is symmetric or commutative is much more descriptive and understandable than calling it Abelian. DeMorgan's laws would be better named duality laws. We who are used to the terms forget what a barrier they pose to beginners.

The term “boolean algebra” honors George Boole. (It used to be thought that the word “algebra” honors someone, but it comes from an Arabic word meaning “the reintegration and reunion of broken parts”. In any case, the word is now standard, known by people everywhere.) The best way to honor George Boole is to make the algebra that he created [4] a well known and well used tool, and to do that we might have to remove his name from it, and give it a more descriptive and accessible name, like “binary algebra”.

Conclusions

Logic has been well studied and is now well understood, but it is not well used. Programmers learn that logic is a foundation of programming, but they don't often use it to program. Mathematicians study about logic, but they don't often use it in their proofs. Logic is a tool, like a knife. People have looked at it from every angle; they've described how it works at great length; now it's time to pick it up and use it. To use logic well, one must learn it early, and practice a lot. Fancy versions of logic, such as modal logic and metalogic, can be left to university study, but there is a simple basic algebra that can be taught early and used widely.

Number algebra is used by scientists and engineers everywhere. It is used by economists and architects. It is taught first to 6-year olds, without a metanotation, concretely as addition and subtraction of numbers. Then variables and equations are introduced, and always the applications are emphasized. As a result of that early and long education, scientists and engineers and mathematicians are comfortable with it. Boolean algebra can be equally useful if it is taught the same way. At present, it is not in a good state for presentation to a wide audience. We need to simplify the terminology, get rid of the metanotations, adopt the view that proof is calculation, choose some good symbols, detach it from its dominant application in which the boolean values represent true and false statements, and explain it as algebra.

There is a small advantage to choosing uniquely boolean symbols: we can give them a precedence after the arithmetic operators, which reduces the need for parentheses. On the other hand, there is a large advantage to uniting boolean and number symbols in the way I have suggested: the laws and solutions are familiar and can be interpreted either as booleans or numbers. In addition, by placing booleans in the same context as numbers, we move quickly away from debates about the meanings of operators. The fact that the booleans can be embedded in the extended integers just as smoothly as the integers are embedded in the rationals seems a compelling reason to do so.

Quantifiers can be simplified, made uniform, and generalized by treating them as operators on functions. We should stop speaking about “existence”, and speak instead about the maximum of a function. Similarly, we should stop speaking about “all”, and speak instead about the minimum of a function. We should stop trying to say what functions and other mathematical ideas are, and say instead how to write them and use them.

An interpreter serves the same purpose as the metalevel theoremhood operator with the added advantage that it gives antitheoremhood as well as theoremhood. And by applying it to strings, we don't need to introduce a separate metalevel of operators. Metalogic is an advanced topic, not a good introduction to boolean algebra for those who are new to the subject.

This paper has not presented a detailed proposal for a change to our primary and secondary mathematics curriculum, but it has presented the case for making a change, and several

suggestions. The main suggestion is to unify boolean algebra with number algebra so that we can begin with the simplest algebra and move smoothly to the more complicated algebras, all using the same notations and in the same calculational framework.

Acknowledgment

Theo Norvell provided references [16] and [19]. Benet Devereux provided references [3] and [18]. Rutger M. Dijkstra, Wim Hesselink, and Jim Grundy corrected some errors and caused me to improve some explanations.

References

- [0] Unfortunately, 500 year old algebra texts are hard to find. This is not a quotation, but my own creation. I think it is representative of the work of the time, such as [20]; see the second Appendix.
- [1] L.E.Allen: “Symbolic Logic: a Razor-Edged Tool for Drafting and Interpreting Legal Documents”, *the Yale Law Journal* v.66 p.833-879, 1957
- [2] R.Back, J.Grun dy, J.vonWright: “Structured Calculational Proof”, *Formal Aspects of Computing* v.9 n.5 p.469-483, 1997
- [3] J.M.Bochinsky: *a History of Formal Logic* second edition, translated and edited by Ivo Thomas, Chelsea Publishing, New York, 1970
- [4] G.Boole: *an Investigation of the Laws of Thought, on which are founded the Mathematical Theories of Logic and Probabilities*, MacMillan, 1854, reprinted by Dover, 1973
- [5] R.Boute: “Binary Algebra and Functional Predicate Calculus: a Practical Approach”, University of Ghent, Belgium, 1999
- [6] A.Church: “the Calculi of Lambda-Conversion”, *Annals of Mathematical Studies* v.6, Princeton University Press 1941
- [7] E.W.Dijkstra, C.S.Scholten: *Predicate Calculus and Program Semantics*, Springer-Verlag 1990
- [8] R.L.Goodstein: *Development of Mathematical Logic*, Springer-Verlag 1971
- [9] D.Gries: “Improving the Curriculum through the teaching of Calculation and Discrimination”, *Communications of the ACM* v.34 n.3 p.45-55, 1991 March
- [10] D.Gries, F.B.Schneider: *a Logical Approach to Discrete Math*, Springer-Verlag, 1993
- [11] P.Halmos, S.Givant: *Logic as Algebra*, the Mathematical Association of America, 1998
- [12] E.C.R.Hehner: “Beautifying Gödel”, chapter 18 in *Beauty is our Business, a birthday tribute to Edsger Dijkstra*, Springer-Verlag, 1990; hehner.ca/God.pdf
- [13] E.C.R.Hehner: *a Practical Theory of Programming*, Springer-Verlag 1993; second edition 2004, hehner.ca/aPToP
- [14] E.C.R.Hehner: “Unified Algebra”, hehner.ca/UA.pdf
- [15] I.N.Herstein: *Topics in Algebra* p.323, Blaisdell 1964.
- [16] R.H.Katz: *Contemporary Logic Design*, Benjamin Cummings 1994 p.10
- [17] L.Lamport: “How to Write a Proof”, *American Mathematical Monthly* v.102 n.7 p.600-608 1995 Aug.
- [18] J.Lukasiewicz: “on the History of the Logic of Propositions”, *Selected Works*, citing Sextus Empiricus in *Adversus Mathematicos*, circa 200
- [19] C.Navarre: quoted in Barbara W. Tuchman: *a Distant Mirror: the Calamitous Fourteenth Century*, Knopf 1978
- [20] R.Recorde: *the Whetstone of Witte*, London 1557, reprinted by Da Capo Press Amsterdam 1969.

- [21] P.J.Robinson, J.Staples: “Formalizing a Hierarchical Structure of Practical Mathematical Reasoning”, *Journal of Logic and Computation* 3(1):47-61, February 1993
- [22] J.B.Rosser: “Highlights of the History of the Lambda-Calculus”, *Annals of the History of Computing* v.6 n.4 p.337-349, 1984 October
- [23] J.L.A.van de Snepscheut: *What Computing is All About*, Springer-Verlag, 1993
- [24] J.M.Spivey: *the Z Notation: a Reference Manual*, Prentice-Hall, 1989

Appendix

Wedderburn's Theorem: a finite division ring is a commutative field (verbatim from [15])

Proof: Let D be a finite division ring and let Z be its center. By induction we may assume that any division ring having fewer elements than D is a commutative field.

We first remark that if $a, b \in D$ are such that $b'a = ab'$ but $ba \neq ab$ then $b' \in Z$. For, consider $N(b') = \{x \in D \mid b'x = xb'\}$. $N(b')$ is a subdivision ring of D ; if it were not D , by our induction hypothesis, it would be commutative. However, both a and b are in $N(b')$ and these do not commute; consequently, $N(b')$ is not commutative so must be all of D . Thus $b' \in Z$.

Every nonzero element in D has finite order, so some positive power of it falls in Z . Given $w \in D$ let the order of w relative to Z be the smallest positive integer $m(w)$ such that $w^{m(w)} \in Z$. Pick an element a in D but not in Z having minimal possible order relative to Z , and let this order be r . We claim that r is a prime number for if $r = pq$ with $1 < p < r$ then a^p is not in Z . Yet $(a^p)^q = a^r \in Z$, implying that a^p has an order relative to Z smaller than that of a .

By the corollary to Lemma 7.9 there is an $x \in D$ such that $xax^{-1} = a^i \neq a$; thus $x^2ax^{-2} = x(xax^{-1})x^{-1} = xax^ix^{-1} = (xax^{-1})^i = (a^i)^i = a^{i^2}$. Similarly, we get $x^{r-1}ax^{-(r-1)} = a^{i^{r-1}}$. However, r is a prime number thus by the little Fermat theorem (corollary to Theorem 2.a), $i^{r-1} = 1 + ur$, hence $a^{i^{r-1}} = a^{1+ur} = aa^{ur} = \lambda a$ where $\lambda = a^{ur} \in Z$. Thus $x^{r-1}a = \lambda azx^{r-1}$. Since $x \notin Z$, by the minimal nature of r , x^{r-1} cannot be in Z . By the remark of the earlier paragraph since $xa \neq ax$, $x^{r-1}a \neq ax^{r-1}$ and so $\lambda \neq 1$. Let $b = x^{r-1}$; thus $bab^{-1} = \lambda a$; consequently, $\lambda^r a^r = (bab^{-1})^r = ba^r b^{-1} = a^r$ since $a^r \in Z$. This relation forces $\lambda^r = 1$.

We claim that if $y \in D$ then whenever $y^r = 1$, then $y = \lambda^i$ for some i , for in the field $Z(y)$ there are at most r roots of the polynomial $u^r - 1$; the elements $1, \lambda, \lambda^2, \dots, \lambda^{r-1}$ in Z are all distinct since λ is of the prime order r and they already account for r roots of $u^r - 1$ in $Z(y)$, in consequence of which $y = \lambda^i$.

Since $\lambda^r = 1$, $b^r = \lambda^r b^r = (\lambda b)^r = (a^{-1}ba)^r = a^{-1}b^r a$ from which we get $ab^r = b^r a$. Since a commutes with b^r but does not commute with b , by the remark made earlier, b^r must be in Z . By Theorem 7.b the multiplicative group of nonzero elements of Z is cyclic; let $\gamma \in Z$ be a generator. Thus $a^r = \gamma^j$, $b^r = \gamma^k$; if $j = sr$ then $a^r = \gamma^{sr}$; whence $(a/\gamma^s)^r = 1$; this would imply that $a/\gamma^s = \lambda^i$, leading to $a \in Z$, contrary to $a \notin Z$. Hence, r does not divide j ; similarly r does not divide k . Let $a_1 = a^k$ and $b_1 = b^j$; a direct computation from $ba = \lambda ab$ leads to $a_1 b_1 = \mu b_1 a_1$ where $\mu = \lambda^{-jk} \in Z$. Since the prime number r which is the order of λ does not divide j or k , $\lambda^{jk} \neq 1$ whence $\mu \neq 1$. Note that $\mu^r = 1$.

Let us see where we are. We have produced two elements a_1, b_1 such that:

- (1) $a_1^r = b_1^r = \alpha \in Z$.
- (2) $a_1 b_1 = \mu b_1 a_1$ with $\mu \neq 1$ in Z .
- (3) $\mu^r = 1$.

We compute $(a_1^{-1}b_1)^r$; $(a_1^{-1}b_1)^2 = a_1^{-1}b_1 a_1^{-1}b_1 = a_1^{-1}(b_1 a_1^{-1})b_1 = a_1^{-1}(\mu a_1^{-1}b_1)b_1 = \mu a_1^{-2}b_1^2$. If we compute $(a_1^{-1}b_1)^3$ we find it equal to $\mu^{1+2}a_1^{-3}b_1^3$. Continuing we obtain $(a_1^{-1}b_1)^r = \mu^{1+2+\dots+(r-1)}a_1^{-r}b_1^r = \mu^{1+2+\dots+(r-1)} = \mu^{r(r-1)/2}$. If r is an odd prime, since $\mu^r = 1$, we get $\mu^{r(r-1)/2} = 1$, whence $(a_1^{-1}b_1)^r = 1$. Being a solution of $y^r = 1$, $a_1^{-1}b_1 = \lambda^i$ so that $b_1 = \lambda^i a_1$; but then $\mu b_1 a_1 = a_1 b_1 = b_1 a_1$, contradicting $\mu \neq 1$. Thus if r is an odd prime number, the theorem is proved.

We must now rule out the case $r=2$. In that special situation we have two elements $a_1, b_1 \in D$ such that $a_1^2 = b_1^2 = \alpha \in Z$, $a_1 b_1 = \mu b_1 a_1$ where $\mu^2 = 1$ and $\mu \neq 1$. Thus $\mu = -1$ and $a_1 b_1 = -b_1 a_1 \neq b_1 a_1$; in consequence, the characteristic of D is not 2. By Lemma 7.7 we can find elements $\zeta, \eta \in Z$ such that $1 + \zeta^2 - \alpha \eta^2 = 0$. Consider $(a_1 + \zeta b_1 + \eta a_1 b_1)^2$; on computing this out we find that $(a_1 + \zeta b_1 + \eta a_1 b_1)^2 = \alpha(1 + \zeta^2 - \alpha \eta^2) = 0$. Being in a division ring this yields that $a_1 + \zeta b_1 + \eta a_1 b_1 = 0$; thus $0 \neq 2a_1^2 = a_1(a_1 + \zeta b_1 + \eta a_1 b_1) + (a_1 + \zeta b_1 + \eta a_1 b_1)a_1 = 0$. This contradiction finishes the proof and Wedderburn's theorem is established.

Appendix

Four pages from R. Recorde: *the Whetstone of Witte*, London 1557, reprinted by Da Capo Press Amsterdam 1969.

The Arte The rule of equation, commonly called Algebers Rule.

The rule of
equation.



Hereto haue I taughte you, the common formes of worke, in numbers Denominate. Whiche rules are vsed also in numbers Abstrakte, & likewise in Sarme numbers. Although the formes of these worke be scueralle, in eche kinde of number. But now will I teache you that rule, that is the principall in Cosike worke: and for whiche all the other dooe serue.

This Rule is called the Rule of *Algebra*, after the name of the inuentoure, as some men thinke: or by a name of singular excellencie, as other iudge. But of his vse it is rightly called, the rule of *equation*: because that by *equation* of numbers, it doeth dissolue doubtfull questions: And vnfolde intricate riddles. And this is the order of it.

The somme of the rule of equation:



When any question is propounded, apperteyning to this rule, you shall imagine a name for the number, that is to bee soughte, as you remember, that you learned in the rule of false position. And with that number shall you procede, accordyng to the question, vntil you finde a Cosike number, equalle to that number, that the question expresseth, whiche you shal reduce

of Cosike numbers.

reduce euer more to the leaſte numbers. And then diuide the number of the lesser denomination, by the number of the greateste denomination, and the quotient doeth answer to the question. Except the greater denomination, doe beare the signe of some rooted number. For then must you extract the roote of that quotiente, accordyng to that signe of denomination.

Scholar. It seemeth that this rule is all one, with the rule of false position: and therefore mighte so be called: sepyng it taketh a false number, to worke withal.

Master. This rule doeth farre excell that other. And dooeth not take a false number, but a true number for his position, as it shall bee declared anon. Wherby it maie bee thoughte, to bee a rule of wonderfull inuention, that teacheth a manne at the firste worde, to name a true number, befoze he knoweth resolutely, what he hath named.

But because that name is common to many numbers (although not in one question) and therefore the name is obscure, till the worke doe deteat it, I thinke this rule might well bee called, the rule of darke position, or of strange position: but not of false position.

And for the moze easie and apte worke in this arte wec dooe commonly name that darke position. i. *Ze*. And with it doe we worke, as the question intendeth, till we come to the equation.

This rule of equation, is diuided by some men, into diuerse partes. As namely *Schenbelius* dooeth make. 3. generalle cannons. Some other men make a greater number of distinctiōs in this rule. But I intende (as I thinke beste for this treatice, whiche maie serue as farre
ff. i. as

The partes
of the rule.

The Arte

as their woordes doe extende) to distinate it onely into twoo partes. Wherof the firste is, when our number is equalle vnto one other. And the seconde is, when one number is compared as equalle vnto 2. other numbers.

Alwaies willing you to remember, that you reduce your numbers, to their leaste denominations, and smalleste formes, befoze you procede any farther.

And again, if your equation be soche, that the greatest denomination *Cosike*, be ioined to any parte of a compoude number, you shall tourne it so, that the number of the greatest signe alone, maie stande as equalle to the reste.

And this is all that needeth to be taughte, concerning this woorde.

Wheretof, so; easie alteration of equations. I will propounde a fewe examples, bicause the extraction of their rootes, maie the more aptly bee wroughte. And to avoide the tedious repetition of these woordes: is equalle to: I will sette as I doe often in woordes use, a paire of paraleles, or twoo lines of one lengthe, thus: ———, bicause noe. 2. thynges, can be moare equalle. And now marke these numbers.

1. $14ze + 15g = 71g$
 2. $20ze = 18g = 102g$
 3. $26z + 10ze = 9z + 10ze + 213g$
 4. $19ze + 192g = 10z + 108g + 19ze$
 5. $18ze + 24g = 8z + 2ze$
 6. $34z = 12ze = 40ze + 480g + 9z$
- In the firste there appeareth. 2. numbers, that is
14ze.

of Cosike numbers.

$14ze + 15g$. equalle to one number, whiche is $71g$. But if you marke therein well, you maie see one denomination, on bothe sides of the equation, which neuer ought to stand. Wherfoze abating the lesser, that is. $15g$. out of bothe the numbers, there will remaine. $14ze = 56g$. that is, by reduction, $1ze = 4g$.

Scholar. I see, you abate. $15g$. fram therein bothe. And then are they equalle still, seing they wer equalle befoze. Acco:dyng to the thirde common sentence, in the pathewaic:

If you abate euen portions, from thynges that bee equalle, the partes that remain shall be equall also.

Master. You doe well remember, the firste grounds of this arte. For all syngeth of those principles Geometricalls. Wherfoze call to your minde like waies the secunde common sentence, in the same booke, and then haue you another reason, whiche will helpe you not onely, in the other formes of woordes here, but also beery often in the practise of this arte.

Scholar. What is this.

If you adde equalle portions, to thynges that bee equalle, what so amounteth of them shall be equalle.

Master. These twoo sentences doe instruct you that when you see on bothe the sides of the equation, any one denomination *Cosike*, you shall marke the signe that is annexed to the lesser of them bothe: and if it be the signe of addition. — + — then shall you abate that lesser number, from bothe the partes of the equation.

As I did in this firste example. But if the signe be of abatemente — — then shall you adde that lesser number, to bothe partes. And so shall you doe, till there be noe one denomination on bothe partes, but diuerse and distinct.

So the seconde number will be. $20ze = 120g$ and in the laste termes. $1ze = 6g$.

Scholar. I see that you adde. $18g$. to bothe partes.