Teaching Math More Effectively, Through the Design of Calculational Proofs

David Gries*
Fred B. Schneider**

TR 94-1415
March 1994

Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

* Supported by Darpa under ONR grant N00014-91-J-4123.
** Supported by ONR under contract N00014-91-J-1219, AFOSR under proposal 93NM312, NSF under grant CCR-8701103, and DARPA/NSF grant CCR-9014363.

Teaching Math More Effectively, Through the Design of Calculational Proofs

David Gries[1] and Fred B. Schneider[2]
Computer Science, Cornell University

February 1994

Lower-level college math courses usually avoid using formalism, in both definitions and proofs. Later, when students have mastered definitions and proofs written largely in English, they may be shown how informal reasoning could be formalized, but the impression is left that such formalization would not be worth the effort. The design of proofs is also not taught. Students see proofs and may be asked to develop a few themselves, but there is little or no discussion of principles or strategies for designing proofs.

Few are happy with the results of these courses. Generally, students' reasoning abilities are poor, even after several math courses. Many students still fear math and notation, and the development of proofs remains a mystery to most. In short, students are not being equipped with the tools needed to employ mathematics in solving new problems.

We believe that this state of affairs can be improved. This article describes our approach.

The inadequacy of informal proofs

A proof of a theorem should provide evidence for belief in the validity of the theorem, where the evidence consists of facts (e.g. previously proved theorems) and an explanation of how they interact to convince. A good presentation of a proof should clearly explain the facts and how they are combined. It will also make the proof appear so obvious that readers can see how it was developed, can explain it to others, and perhaps can prove other theorems in a similar fashion.

Now look at the proof in Table 1, which was taken from a math text and is typical of informal proofs. First, note that this proof does not state the facts on which it rests. (For example, it says, "If $y \notin A$, then, since $y \in A \cup B$ we must have $y \in B$", but there is no reference to the theorem that justifies this inference.) Second, it is difficult to see precisely how the facts interact —the sequence and subsequences of inferences and all the case analyses in the proof cannot be easily digested. The structure of the proof is hidden by all the verbiage. One case analysis is presented in two paragraphs and others by sequential sentences within a paragraph; however, sequential sentences are also used to define steps common to all cases. Finally, this proof yields little insight into its development —how did it arise?

[1] Supported by Darpa under ONR grant N00014-91-J-4123.

[2] Supported by ONR under contract N00014-91-J-1219, AFOSR under proposal 93NM312, NSF under grant CCR-8701103, and DARPA/NSF grant CCR-9014363.

And yet, in spite of its inadequacies, this proof (and others like it) is held up as a model for students to emulate.

Table 1: Conventional Proof of $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

We first show that $A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C)$. If $x \in A \cup (B \cap C)$, then either $x \in A$ or $x \in B \cap C$. If $x \in A$, then certainly $x \in A \cup B$ and $x \in A \cup C$, so $x \in (A \cup B) \cap (A \cup C)$. On the other hand, if $x \in B \cap C$, then $x \in B$ and $x \in C$, so $x \in A \cup B$ and $x \in A \cup C$, so $x \in (A \cup B) \cap (A \cup C)$. Hence, $A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C)$.

Conversely, if $y \in (A \cup B) \cap (A \cup C)$, then $y \in A \cup B$ and $y \in A \cup C$. We consider two cases: $y \in A$ and $y \notin A$. If $y \in A$, then $y \in A \cup (B \cap C)$, and this part is done. If $y \notin A$, then, since $y \in A \cup B$ we must have $y \in B$. Similarly, since $y \in A \cup C$ and $y \notin A$, we have $y \in C$. Thus, $y \in B \cap C$, and this implies $y \in A \cup (B \cap C)$. Hence $(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$. The theorem follows.

Calculational proofs in an equational logic

Our thesis is that mathematics and rigorous thinking can be taught more effectively by first teaching the design of rigorous proofs using a formal logic. However, the choice of logic and the accompanying method of proof is critical to success. In our experience, an equational logic, which is based on equality and Leibniz's "substitution of equals for equals", is most suitable because it has the following characteristics.

- Equational logic is easy to teach, since the style is already familiar to those who have had high-school algebra.
- Equational logic provides an alternative to reasoning in English. Rarely do proofs in equational logic parrot informal English arguments. Instead, proofs are *calculational*, in that they are developed by calculating using the rules of the logic, much as one calculates to solve a problem in high-school algebra. Further, principles and strategies can be used to help discover theorems and proofs.
- The use of equational logic need not lead to overwhelming complexity (as is the case with some logics). On the contrary, its use in a rigorous fashion is often a simplifying force. Typically, calculational proofs are shorter, simpler, and easier to remember than informal English proofs.
- Equational logic is versatile —it can be extended to a wide variety of mathematical domains.

Table 2 contains a calculational proof of theorem $p \vee q \equiv p \vee \neg q \equiv p$. Note that equivalence $\equiv$ is treated associatively, so that this theorem can be viewed either as $(p \vee q \equiv p \vee \neg q) \equiv p$ or as $p \vee q \equiv (p \vee \neg q \equiv p)$. Also, symbol $=$ is used for equality over any type, including type boolean. Symbol $=$ is used conjunctionally: $b = c = d$ is equivalent to $b = c \wedge c = d$. Use

Table 2: Equational proof of $p \vee q \;\equiv\; p \vee \neg q \;\equiv\; p$

$$\begin{array}{ll}
 & p \vee q \;\equiv\; p \vee \neg q \\
= & \quad \langle \text{Distr. of } \vee \text{ over } \equiv,\; p \vee (q \equiv r) \;\equiv\; p \vee q \;\equiv\; p \vee r \rangle \\
 & p \vee (q \;\equiv\; \neg q) \\
= & \quad \langle\, \neg q \equiv q \equiv \mathit{false} \,\rangle \\
 & p \vee \mathit{false} \\
= & \quad \langle \text{Identity of } \vee,\; p \vee \mathit{false} \;\equiv\; p \rangle \\
 & p
\end{array}$$

of associativity of equivalence helps avoid formal detail without sacrificing rigor —our notation is designed with an eye to preventing complexity from overwhelming.

Each step of the proof in Table 2 has the following form.

$$\begin{array}{ll}
 & E[v := P] \\
= & \quad \langle\, P = Q \,\rangle \\
 & E[v := Q]
\end{array}$$

Such a step shows equality of two formulas using the rule of "substitution of equals for equals". The hint between the two formulas shows the equality being used in the substitution ($E[v := P]$ denotes expression E with every free occurrence of variable v replaced by expression P). Transitivity of equality allows us to conclude that the first and last formula of the proof of Table 2 are equal.

Notice that the proof format makes it easy to find the facts on which the proof depends —they are given in the hints that appear after each $=$ sign. Here, we have written out the full text of each fact, but we might use the name or number of an already proved theorem.

A theorem of the logic is either an axiom or a theorem that is proved equal to an already-existing theorem. Also, we have a metatheorem: To prove $P \;\equiv\; Q$ it suffices to translate P into Q (or Q into P) as was done in Table 2.

Explicit principles and strategies drove the calculation of the proof of Table 2. For example, one strategy for proving $P \;\equiv\; Q$ is to transform the more complicated of P and Q into the simpler one. In the proof, we viewed the formula to be proved as $(p \vee q \;\equiv\; p \vee \neg q) \;\equiv\; p$ and started with the more complicated, left-hand term. Second, the proof in Table 2 is "opportunity driven" or "forced", in that at each step, the shape of the formula almost dictates in a unique way what substitution to make. Here, the shape of the first line of the proof cries out for simplification using distribution of $\vee$ over $\equiv$. The second step is an equally obvious simplification, based on the shape of the formula.

Table 3 gives another calculational proof: our proof of distributivity of set union over set intersection. In contrast to the proof of Table 1, this proof exhibits all the good qualities mentioned earlier. It refers to all the facts it uses (e.g. the definition of $\cup$). Its structure is simple, with each step being clearly delineated. And, it is based on a strategy —one that is used over and over in mathematics: To prove something about operators (here, $\cup$ and $\cap$), eliminate them using their definitions, perform some manipulation, and reintroduce the operators.

Table 3: Calculational Proof of $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Below, we prove that $v \in A \cup (B \cap C) \equiv v \in (A \cup B) \cap (A \cup C)$. By Extensionality (the definition of equality of sets), we then conclude $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

$$\begin{array}{ll}
 & v \in A \cup (B \cap C) \\
= & \quad \langle \text{Definition of } \cup \rangle \\
 & v \in A \ \vee \ v \in B \cap C \\
= & \quad \langle \text{Definition of } \cap \rangle \\
 & v \in A \ \vee \ (v \in B \wedge v \in C) \\
= & \quad \langle \text{Distr. of } \vee \text{ over } \wedge \rangle \\
 & (v \in A \vee v \in B) \ \wedge \ (v \in A \vee v \in C) \\
= & \quad \langle \text{Definition of } \cup \text{, twice} \rangle \\
 & (v \in A \cup B) \ \wedge \ (v \in A \cup C) \\
= & \quad \langle \text{Definition of } \cap \rangle \\
 & v \in (A \cup B) \cap (A \cup C)
\end{array}$$

Anyone experienced in such calculational proofs will find the proofs of Tables 2 and 3 obvious and straightforward and will have no difficulty reproducing them. And, although these proofs are rigorous (and could be checked by a mechanical proof checker), there is no overwhelming complexity.

Equational logic and the calculational approach can be extended to all domains typically taught in a first discrete math course —e.g. set theory, mathematical induction, a theory of integers, functions and relations, combinatorics, and recurrence relations. This is done by first defining the pure predicate calculus and then extending it by adding new types, presenting axioms that define the manipulative properties of the operations on those types, and building up a library of theorems.

A key to making rigor and formalism palatable is to keep notation consistent and uniform. Mathematics employs a number of different notations for quantification —see, for example, the left column of Table 4. We replace these different forms by a single notation for all quantifications. For any operator $\star$ that is associative, is symmetric, and has an identity, the notation[3]

$$(\star i \mid R.i : P.i)$$

denotes the "accumulation" using operator $\star$ of the values of expression $P.i$ over all values of i that satisfy range-predicate $R.i$. For example, Table 4 gives the conventional notation and a more uniform notation for four different quantifications. Other operators that can be used for $\star$ are multiplication of integers, reals, and complex numbers, $b \cdot c$; union of sets, $S \cup T$; intersection of

[3] Bound variable i can be annotated with a type to indicate the range of values it may assume. A discussion of types is outside the scope of this article. Also, we write $R.i$ to denote application of function R to argument i.

Table 4: A Uniform Notation for Quantification

Conventional notation	Uniform notation
$\Sigma_{i=1}^{3} i^2$	$(+i \mid 1 \leq i \leq 3 : i^2)$
$(\forall x).1 \leq x \leq 3 \Rightarrow b[x] = 0$	$(\wedge x \mid 1 \leq x \leq 3 : b[x] = 0)$
$(\exists x).1 \leq x \leq 3 \wedge b[x] = 0$	$(\vee x \mid 1 \leq x \leq 3 : b[x] = 0)$
$\bigcup_{i=1}^{3} S_i$	$(\cup i \mid 1 \leq x \leq 3 : S_i)$

sets $S \cap T$; minimum of two values, $b \downarrow c$ (if $\downarrow$ does not have an identity, axioms and theorems that deal with a *false* range $R.i$ are not applicable); maximum of two values, $b \uparrow c$; and greatest common divisor, b **gcd** c.

With a single notation, scope, free occurrence of a variable, and bound occurrence of a variable can be defined for all quantifications just once. More importantly, general axioms and theorems for manipulating all quantifications can be introduced. The issue of quantification is thus simplified.

Note that $\wedge$ and $\vee$ are associative, are symmetric, and have identities, so $(\wedge i \mid R.i : P.i)$ and $(\vee i \mid R.i : P.i)$ makes sense. The first is universal quantification, more conventionally written as $(\forall i \mid R.i : P.i)$; the second is existential quantification, $(\exists i \mid R.i : P.i)$.

Teaching the calculational approach

Equational propositional logic, along with preliminaries (e.g. the definition of textual substitution) can be taught to college freshmen in four weeks. During that time, students will see many proofs and will develop many themselves, in the calculational style. They will also learn strategies and principles for designing proofs. As students develop a skill in proving theorems, they learn that attention to rigor may be a simplifying force —and not an onerous burden.

Four weeks may seem like a long time to spend on propositional logic, but learning the calculational approach and gaining confidence in formal manipulation requires it and is worth it. Initially, most students are troubled by the prospect of uninterpreted manipulation. They want to think about the meanings of mathematical statements. Having meanings for objects is a "safety net", which, students feel, prevents them from performing nonsensical manipulations. Unfortunately, the use of the "meaning" safety net does not scale well to complicated problems. Skill in performing uninterpreted syntactic manipulation does.

Students also have to be convinced that using formalism can be helpful. They must see first hand that a rigorous approach can help them solve problems they could not easily solve without it. This is possible with our approach. After just three days of learning equational logic, one can begin to attack the kinds of word problems that are found in Smullyan's books, for example.

Once logic and proof have been thoroughly presented, other topics can be discussed —e.g. set theory, a theory of integers, and mathematical induction. Each topic is presented using the same calculational approach. In this manner, the notions of proof and proof style become the unifying force, the glue that binds together arguments in all domains.

A discussion of informal versus formal presentations of proofs can impart deeper understanding of both, enabling students to deal more easily with math that they will see in later courses. For example, proof by contradiction in any domain is easily seen to be based on the theorem $p \equiv \neg p \Rightarrow \mathit{false}$ of propositional logic.

As another example, suppose we prove the metatheorem that a formula P is a theorem iff the formula $(\forall x \mid : P)$ is a theorem. Then, the different ways in which theorems are expressed in texts can be discussed, and the following three statements can be seen to be equivalent. In the first, it is assumed informally that a and b are integers —perhaps this is mentioned in the accompanying prose; in the second, the type is given informally; in the third, the type is made formally explicit.

$$a + b = b + a$$
$$a + b = b + a \qquad \text{(for } a, b \text{ integers)}$$
$$(\forall a, b{:}\mathbb{Z} \mid : a + b = b + a)$$

To make rigor and formalism palatable, every new notation must be explained and the rules must be given for manipulating it. Fear of formalism comes from having to use a formalism without knowing rules for its use, and attention to basic detail overcomes this fear. For example, traditionally, students are not shown rules for manipulating summations like $\Sigma_{i=1}^{3} i^2$; consequently, they have trouble with mathematical induction, where problems require manipulation of such summations.

When formal notations are presented properly, as a repository of the facts and a means of clarification, students begin to like formalism and to rely on it. It is the formalism that provides rules for judging between sound and unsound inference and that helps expose ambiguity and eliminate it.

Here is an example in which attention to rigor and formal detail provides a measure of clarity that is impossible to obtain otherwise. Consider proving $b^{m+n} = b^m \cdot b^n$, for n, m natural numbers, by mathematical induction. Without formalizing quantification and having rules for manipulating it, no amount of informal explanation will clarify for students the different roles of m and n in the proof. However, $b^{m+n} = b^m \cdot b^n$ is equivalent to $(\forall m, n \mid 0 \le n \wedge 0 \le m : b^{m+n} = b^m \cdot b^n)$, which can be rewritten (using an axiom of quantification and the ability to name a formula) as

$$(\forall n \mid 0 \le n : P.n) \quad \text{where} \quad P.n : \quad (\forall m \mid 0 \le m : b^{m+n} = b^m \cdot b^n) \quad .$$

Now it is clear that n is the "induction variable" and that induction hypothesis $P.n$ is a universal quantification over m .

Further, once students understand quantification, they can prove the following —using a calculational proof. Let U be a set and $\prec$ a binary relation over U . Then $(U, \prec)$ admits induction iff $(U, \prec)$ is well founded. This theorem, which is rarely mentioned in informal presentations, gives deeper insight into induction.

Discussion

The rigorous approach to teaching math has not, as yet, been accepted. Two criticisms are heard frequently: (1) students can't handle rigor and formalism, and (2) teaching syntactic manipulation impedes understanding that a more semantic and informal approach provides.

Our own experience belies the first criticism; in fact, the criticism should go the other way. Teaching mathematics through informalism is like driving in a fog. One sees dim figures in the distance, and every once in a while some of them suddenly appear clearly, but usually everything is veiled and mysterious. It's dangerous to drive in the fog, especially in a strange territory, and one must drive slowly. Even so, one may not always be sure where one is. Teaching rigor and precision, provided it is done without the veil of complexity interfering, burns away the fog, leaving everything crisp and clear and making it possible to drive faster and to enter uncharted lands.

We can rebut the criticism concerning semantics versus syntactics as well. An informal proof, like that in Table 1, can be translated into a proof in a natural-deduction or Hilbert-style logic. The resulting proof every bit as syntactic as ours. The English proof is simply an informal version of a syntactic proof —and, as we have seen, a poor one at that. Therefore, the informal proof has no more meaning or semantics than a formal calculational proof.

Perhaps this criticism concerning semantics comes about because formal statements are sometimes difficult to understand. However, presenting a formal definition or theorem does not preclude giving alternative views as well. For example, a presentation of the axiomatic definition of set union can be supplemented with a Venn diagram, an English description, and an informal notion of evaluation. Nevertheless, it should be realized that for purposes of reasoning —constructing proofs— it is the axiomatic definition that is important. In fact, the axiomatic definition should be viewed as encoding all the meaning of the object being defined.

We also hear complaints that our approach suppresses intuition, that everything begins to appear mechanical. By "intuition" one usually means direct perception of truth or fact, independent of any reasoning process; keen and quick direct insight; or pure, untaught, noninferential knowledge (*Webster's Encyclopedic Unabridged Dictionary*, 1989). There is simply no hope of teaching this —how can one teach something that is untaught, noninferential, and independent of any reasoning process? Of course, one can hope that students will develop an ability to intuit by watching instructors in math courses over the years. But this hit-or-miss prospect cannot be called *teaching* intuition.

On the other hand, a good part of mathematics is concerned with the opposite of intuition: with new and different reasoning processes that complement our ability to reason in English. This part of mathematics can be taught, and our approach to logic is an excellent vehicle for that task. Further, using the calculational approach to proofs, we are able to teach aids to discovery. In particular, with our disciplined, syntactic, proof style, we can teach principles and strategies whose application can indeed lead to the discovery of some (but not all) theorems and proofs.. We have yet to see comparable principles and strategies for conventional English proofs.

New ideas in teaching are slow to catch on. People don't like changing their habits —especially if it requires them to change their own way of thinking. However, current teaching methods are not exciting students or even educating them well, and alternatives should be seriously considered. Our approach bears looking into by all who want to teach mathematics effectively.[4]

[4] The authors' 500-page text *A Logical Approach to Discrete Math* (Springer Verlag, NY, 1993) uses the approach described in this article in teaching the usual topics in discrete math —logic, set theory, a theory of integers, induction, functions and relations, combinatorics, solving recurrence relations, and graph theory. The 300-page Instructor's Manual contains other essays that concern the approach, as well as answers to the exercises. Together, the text and Instructor's Manual contain over 700 calculational proofs, most of which are short and simple. Contact Gries at gries@cs.cornell.edu to obtain the Instructor's Manual.