How to Calculate Proofs: Bridging the Cultural Divide

Raymond T. Boute

This article argues that one of the most neglected opportunities in many branches of mathematics is symbolic reasoning for the "logical" parts rather than just for the algebraic, analytic, etc., parts, where symbolic calculation has been a well-established routine since Viète and Descartes [4], [5].

This is far from suggesting that symbolic reasoning should be promoted from neglect to supremacy or that it is always the best choice: we will discuss its place among other styles. One of several things we *do* want to emphasize is the advantage of being able to calculate with predicates and quantifiers [10], [23], [25] as fluently as is common for derivatives, integrals, and sums (∀ and ∃ being much simpler than Σ). Even if, for a specific problem, this ability is not exploited by choice, at least it will have provided a broader basis for making choices.

The treatment is introductory, addressing anyone interested in augmenting their palette of reasoning strategies, including mathematics educators and, directly or indirectly, students.

As a matter of terminology, one of the styles we will discuss, and also endorse for logic, is the *calculational* style [22], [24] typical in algebra and analysis. The leftmost example in Figure 1 is taken from an engineering textbook [6, p. 91], intentionally leaving symbols unexplained to focus only on the shape of the argument as expressions chained by relational operators (=, ≤). Normally at each step one also explicitly mentions the rule used (definition, theorem, ...).

Raymond T. Boute is professor emeritus at Ghent University. His email address is Raymond.Boute@pandora.be.

DOI: http://dx.doi.org/10.1090/noti945

$$\begin{aligned}
&\frac{1}{n}\sum_{\mathbf{x}} p^n(\mathbf{x}|\theta) l_n(\mathbf{x}) \\
&\le \frac{1}{n}\sum_{\mathbf{x}} p^n(\mathbf{x}|\theta)[1-\log q^n(\mathbf{x})] \\
&= \frac{1}{n} + \frac{1}{n} L(\mathbf{p}^n;\mathbf{q}^n) + H_n(\theta) \\
&= \frac{1}{n} + \frac{1}{n} d(\mathbf{p}^n,\mathcal{G}) + H_n(\theta) \\
&\le \frac{2}{n} + H_n(\theta)
\end{aligned}$$

$$\begin{aligned}
F(s) &= \int_{-\infty}^{+\infty} e^{-|x|} e^{-i2\pi xs}\, dx \\
&= 2\int_0^{+\infty} e^{-x}\cos 2\pi xs\, dx \\
&= 2\ \mathrm{Re}\int_0^{+\infty} e^{-x} e^{i2\pi xs}\, dx \\
&= 2\ \mathrm{Re}\ \frac{-1}{i2\pi s - 1} \\
&= \frac{2}{4\pi^2 s^2 + 1}
\end{aligned}$$

Figure 1. Calculational (left) and equational (right) derivations in common mathematics.

This style of argument is called *calculational*, and if all relational operators are pure equality, as in the rightmost example [12], it is *equational*. Such arguments are commonly appreciated for their sense of "smoothness", especially if there are no interruptions by prose in the chain.

Examples appear in every nonspecialist mathematics journal, a random sample being the CMJ issue at the time of this writing [14, p. 381]. Of course, more sizeable calculations are structured hierarchically, often form part of a larger theory or project, and are interspersed with expository text. Such issues are orthogonal to the one at hand and are discussed later.

Equations form such a powerful tool in problem solving because they delegate substantial parts of reasoning (especially the tricky ones) to symbolic calculation. Practicing fosters a parallel intuition for symbol manipulation, enhancing the intuition in the domain of discourse (or even preceding it when exploring new domains). For instance, in deriving $(a+b)\cdot(a-b) = a^2 - b^2$, the essence is the cancellation in the intermediate form $a^2 + b\cdot a - a\cdot b - b^2$, and the underlying intuition is

of a symbolic nature, not explained as simply in another way.

Setting the Scene: High School Algebra

In a problem-solving view of theorem proving, the demonstrandum is not always known in advance but can be a solution to a problem (to be *discovered*). Educationally, this teaches students to think beyond just proving what is given.

For reference, we show a typical high school algebra calculation in great detail to highlight how, in later examples, logical calculations use exactly the same style, just with other rules.

One of the first topics one learns in high school algebra is "simplifying" expressions by algebraic rules such as *commutativity*, *associativity*, and *distributivity*. An example is *"Simplify* $(a+b)\cdot(a-b)$*"* as a variant of *"Prove* $(a+b)\cdot(a-b) = a^2-b^2$*"*.

Of course, "simplify" is context-dependent; here *"Convert to* sum of products *form"* would be more specific. The reverse problem (starting from a^2-b^2) requires a little "inventive" step. In *"Prove"* variants, a heuristic [23] is starting from the side that minimizes such hindsight.

(a) How head calculation would appear as a step in a larger context

$$(a+b)\cdot(a-b) = \langle\text{Algebra}\rangle\ a^2-b^2$$

(b) How one *might* write this calculation to expose the crucial step

$$\begin{aligned}(a+b)\cdot(a-b) &= \langle\text{Distr. twice}\rangle\ a\cdot a+b\cdot a-a\cdot b-b\cdot b\\ &= \langle\text{Cancellation}\rangle\ a\cdot a-b\cdot b\\ &= \langle\text{Def. square}\rangle\ a^2-b^2\end{aligned}$$

(c) How one can write it as the top hierarchical level in a detailed calculation

$$\begin{aligned}(a+b)\cdot(a-b) &= \langle\text{Definit. infix }-\rangle\ (a+b)\cdot(a+{}^{-}b)\\ &= \langle\text{Rearrange}\rangle\quad a\cdot a+((b\cdot a+a\cdot{}^{-}b)+b\cdot{}^{-}b)\\ &= \langle\text{Cancellation}\rangle\quad a\cdot a+{}^{-}(b\cdot b)\\ &= \langle\text{Definit. infix }-\rangle\ a\cdot a-b\cdot b\\ &= \langle\text{Definit. square}\rangle\ a^2-b^2\end{aligned}$$

(d) Calculation details of (c)

$$\begin{aligned}(a+b)\cdot(a-b) &= \langle\text{Definit. infix }-\rangle\ (a+b)\cdot(a+{}^{-}b)\\ &= \langle\text{Rearrange}\rangle\\ &\quad\left|\begin{aligned}&= \langle\text{Left distrib. }\cdot/+\rangle\ (a+b)\cdot a+(a+b)\cdot{}^{-}b\\ &= \langle\text{Right distr. }\cdot/+\rangle\ (a\cdot a+b\cdot a)+(a\cdot{}^{-}b+b\cdot{}^{-}b)\\ &= \langle\text{Associativity }+\rangle\ a\cdot a+(b\cdot a+(a\cdot{}^{-}b+b\cdot{}^{-}b))\\ &= \langle\text{Associativity }+\rangle\ a\cdot a+((b\cdot a+a\cdot{}^{-}b)+b\cdot{}^{-}b)\end{aligned}\right.\\ &= \langle\text{Cancellation}\rangle\\ &\quad\left|\begin{aligned}&= \langle a\cdot{}^{-}b={}^{-}(a\cdot b)\rangle\ a\cdot a+((b\cdot a+{}^{-}(a\cdot b))+{}^{-}(b\cdot b))\\ &= \langle\text{Commutativ. }\cdot\rangle\ a\cdot a+((a\cdot b+{}^{-}(a\cdot b))+{}^{-}(b\cdot b))\\ &= \langle\text{Inverse }+\rangle\quad a\cdot a+(0+{}^{-}(b\cdot b))\\ &= \langle\text{Unit }+\rangle\quad a\cdot a+{}^{-}(b\cdot b)\end{aligned}\right.\\ &= \langle\text{Definit. infix }-\rangle\ a\cdot a-b\cdot b\\ &= \langle\text{Definit. square}\rangle\ a^2-b^2\end{aligned}$$

Figure 2. The calculational nature of common high school algebra: an example.

Figure 2 shows the calculation to various degrees of detail, exposing the hierarchy [2], [33], [35]. We briefly comment. Even in high school one soon learns to do symbolic head calculation, and in larger contexts one would write the result as Figure 2(a). For exposing the crucial step, namely the cancellation, Figure 2(b) is a possibility. However, Figure 2(b) involves tacit conventions such as omitting parentheses and using "$-$" for (additive) inversion as well as subtraction (overloading).

Figure 2(c) is more explicit. We reserve ⟨Rearrange⟩ for any mix of associativity, commutativity, and distributivity. Figure 2(d) shows the most detail we give here. In a hypertext environment [2], [33], [35], (d) can be seen as obtained by "clicking" on the ⟨*justification*⟩s in (c). The property $a\cdot{}^{-}b = {}^{-}(a\cdot b)$ is proven for rings (abstract) in [28, p. 89], but in high school it is proven in literally the same way for numbers (concrete). Here are some crucial observations.

- The calculations are essentially formal: using rules without interpreting the expressions.
- Notation and rules are sufficiently robust to make formal use safe (mature formalism).
- The theorem $(a+b)\cdot(a-b) = a^2-b^2$ is *discovered*, not set in advance as a goal.

Another advantage is clarity: in the detailed calculation, the steps use the rules so explicitly that one can infer these rules by "reverse engineering", comparing consecutive lines.

All of this is common practice in high school. Who would derive $(a+b)\cdot(a-b) = a^2-b^2$ using prose? The same question can be asked for logical reasoning.

Rationale: The Roles of Diversity, Logic as Algebra, and Formality

Sangwin [46] emphasizes the role of equations in problem solving from secondary school onwards. He quotes Pólya as saying that [...] the most important task of mathematical instruction in the secondary school is to teach the setting up of equations to solve word problems.

Yet, there is a caveat. Students may become so exclusively focused on equations as "the" easy shortcut in problem solving that some teachers advise them to explore different methods as well, since diversity of methods and perspectives enhances insight and understanding. My favorite example is deriving simple geometric solutions to typical calculus problems—also because calculus deserves appreciation for its intrinsic beauty rather than just for its applications.

Diversity notwithstanding, often a problem is stated in symbolic form either as given or by translation from prose to facilitate the solution. In that case, it would be a setback not to exploit the formal properties of the symbolism insofar as it is mature for practical use.

Indeed, Boyer [11, p. 180] distinguishes three stages in the development of algebra. He presents it as an oversimplified first approximation, but it is relevant to any branch of mathematics and can be made more accurate by the qualifications added between [ ].

1. *rhetorical*: writing everything out in words [and reasoning correspondingly in words];

2. *syncopation*: adopting abbreviations [but still without systematic formal rules];
3. *symbolic*: final stage [with systematic syntax plus systematic formal calculation rules].

Algebra for polynomials reached the syncopation stage with Diophantus, while the symbolic stage was achieved by Viète and Descartes [4], [5]. It was Leibniz's dream for logical reasoning to become symbolic in a similar way (*"calculemus"*). Progress in this direction started with Boole, and the developments in formal logic over the subsequent one hundred fifty years are well documented.

Yet most logic textbooks present symbolic logic at first but regress to rhetoric logic in later chapters, illustrating what we meant earlier by "mature for practical use". Not surprisingly, symbolic logic is even less used in nonspecialist mathematics. For instance, Taylor [49] deplores that many mathematicians still use $\forall$ and $\exists$ as syncopation, often obscuring the logical structure of arguments. With the typical δ-ϵ arguments from analysis as an example, he concludes that *examiners fully deserve the garbage that they get in return*.

The explanation may well be that mathematicians miss the smoothness of algebra in the formal logic of most logic textbooks and might better appreciate Halmos and Givant's *Logic as Algebra* [25]. (Halmos also stated years earlier that $\forall$ and $\exists$ are better used not at all than as mere shorthand.) How to bring the idea of logic as algebra into practice is shown in mathematics journals by Gries and others [24], and additional introductory material can be found in the computer science literature [2], [10], [22], [23], some of it even usable in high school [3].

Many will remember the disasters caused in mathematics education around 1960 by the injudicious and coarse manner in which formality and abstraction were introduced [1]. Mathematics has recovered, but, unfortunately for mathematics, formality itself was blamed and has not recovered, as generations of mathematicians still seem to be smarting from the trauma.

Indeed, the backlash was severe, including a long period of deemphasizing proofs. Steven Krantz [32] noted that classic texts in the style of Rudin [43] are *often no longer suitable, or appear to be inaccessible, to the present crop of students* and clarified this in an email saying, *Many of the students who show up in a real analysis course do not know what a proof is*.

Many texts on proofs in recent years indicate a revived interest, yet some still reflect the "formalism trauma" in urging students to write proofs in prose [51]. This perpetuates the double standard of symbolic algebra/calculus versus rhetoric logic, causing a stylistic rift within mathematics. Lamport [33] notes that *the structure of mathematical proofs has not changed in 300 years. . . . Proofs are still written like essays, in a stilted form of ordinary prose*.

The predilection of students towards equations in problem solving may be turned into an advantage by teaching logic with the familiar calculational flavor of algebra and thus considerably lowering the threshold for formality, in particular symbolic reasoning.

An important technical observation [2] is that the calculational style typical in algebra and calculus (Figures 1 and 2) and illustrated for logic in later examples can be seen as a streamlined form of so-called natural deduction [42], a formalization of the informal reasoning style used in mathematics since antiquity. Whenever it facilitates reasoning, forms can be intermixed [23].

Whatever style is used, it is advantageous to embrace rather than eschew formality. Lamport shows in [33] and with different examples in [35] how proofs in the classical deductive style can be vastly improved by formalizing definitions and theorems and structuring proofs hierarchically.

There can be no doubt that proof assistants and other computer support for logic reasoning [26], [27] are potentially as useful in everyday mathematics as Mathematica, Maple, etc.

However, currently the threshold is rather high. Mathematicians routinely use the formal notation from calculus, linear algebra, and so on, which are quite faithfully supported (including deficiencies!) by the software tools. By contrast, few are used to formalize the logical parts of their arguments, and using a proof assistant requires diligent formalization at every stage, as for using Maple and Mathematica, but there the formality looks "normal".

The threshold can be eliminated by routinely formalizing logic arguments. The motivation is not that "the software requires it" but because it leads to better arguments also "on paper".

A Mini-Gallery Illustrating Logic Reasoning by Calculation

The examples cover a variety of topics and different aspects of the calculational style.

A Detailed Example

Example 3.3.5 from [51, p. 119] is chosen because its author notes that it is the most complex proof thus far, and some readers on the Amazon website mention that it was the first place in [51] where they experienced difficulties.

The operators used are defined in Figure 3. The reference definitions from [51] are transliterated into a more uniform notation in view of later examples from other sources. Of course, among the

wide variety of quantifier notations in the literature, the differences between the form $\forall v \in S(p)$ in [51] and the form $\forall v{:}S\,.\,p$ used here may appear negligible, yet they are a matter of language orthogonality and robustness. The equivalent variants for $\wp$ and $\cup$ are obtained by eliminating $\{\mid\}$, using $y \in \{x \mid P(x)\} \equiv P(y)$. The actual example concerns the following.

> *Task.* Suppose B is a set and $\mathcal{F}$ is a family of sets. Prove that if $\cup\mathcal{F} \subseteq B$, then $\mathcal{F} \subseteq \wp B$.

How this is done in [51] and the resulting rhetorical proof are shown in Figure 4. An equational derivation, simultaneously discovering the theorem and its converse, is shown in Figure 5.

Clarification and Comparison of the Strategies

The scratch work in Figure 4(a) is a condensed version of [51, p. 119] and lists successive *Givens-Goal* tables and explanations. The various strategies used are developed in [51] but are beyond the scope of this article. Some involve introducing arbitrary objects at various points into the proof [51, p. 119]. The discussion by Lamport [35, p. 3] of the questions this may raise for beginners is enlightening.

For calculational derivations, the simple strategy from high school algebra suffices: for each step, look at the shape of the expressions to determine what yields progress: using a definition or a rule from logic. All definitions used in Figure 5 are easily identified in Figure 3. In fact, by the reverse engineering as in Figure 2(d), one can reconstruct the rightmost column of Figure 3 from Figure 5. This even holds for the logic rules used, for instance, by comparing successive lines:

Name	*Symbol*	*Location in* [50]	*Reference definition*	*Equivalent variant*
Subset	$\subseteq$	p. 58, item 5	$A \subseteq B \equiv \forall x\,.\,x \in A \Rightarrow x \in B$	$A \subseteq B \equiv \forall x{:}A\,.\,x \in B$
Powerset	$\wp$	p. 75, Def. 2.3.2	$\wp A = \{x \mid x \subseteq A\}$	$X \in \wp A \equiv X \subseteq A$
Union	$\bigcup$	p. 77, Def. 2.3.5	$\bigcup\mathcal{F} = \{x \mid \exists A{:}\mathcal{F}\,.\,x \in A\}$	$x \in \bigcup\mathcal{F} \equiv \exists A{:}\mathcal{F}\,.\,x \in A$
			Legend: A and B are sets and $\mathcal{F}$ is a family (as a set) of sets	

Figure 3. Definition of subset, powerset, and union (of a family of sets), from [50].

	Explanation	*Givens*	*Goal*
0.	Proving $\bigcup\mathcal{F} \subseteq B \Rightarrow \mathcal{F} \subseteq \wp B$	$\bigcup\mathcal{F} \subseteq B$	$\mathcal{F} \subseteq \wp B$
1.	Goal 0 is $\forall x\,.\,x \in \mathcal{F} \Rightarrow x \in \wp B$	$\bigcup\mathcal{F} \subseteq B$ $x \in \mathcal{F}$	$x \in \wp B$
2.	Goal 1 is $\forall y\,.\,y \in x \Rightarrow y \in B$	$\bigcup\mathcal{F} \subseteq B$ $x \in \mathcal{F}$ $y \in x$	$y \in B$
3.	Try stronger goal $y \in \bigcup\mathcal{F}$	$\bigcup\mathcal{F} \subseteq B$ $x \in \mathcal{F}$ $y \in x$	$y \in \bigcup\mathcal{F}$
4.	Expand goal 3	$\bigcup\mathcal{F} \subseteq B$ $x \in \mathcal{F}$ $y \in x$	$\exists A \in \mathcal{F}\,.\,y \in A$
5.	Observe that x is a witness for A in goal 4, so the proof is done.		

(a) Scratch work

Theorem. *Suppose B is a set and $\mathcal{F}$ is a family of sets. If $\bigcup\mathcal{F} \subseteq B$ then, $\mathcal{F} \subseteq \wp B$.*
Proof. Suppose $\bigcup\mathcal{F} \subseteq B$. Let x be an arbitrary element of $\mathcal{F}$. Let y be an arbitrary element of x. Since $y \in x$ and $x \in \mathcal{F}$, clearly $y \in \bigcup\mathcal{F}$. But then since $\bigcup\mathcal{F} \subseteq B$, $y \in B$. Since y was an arbitrary element of x, we can conclude that $x \subseteq B$, so $x \in \wp B$. But x was an arbitrary element of $\mathcal{F}$, so this shows that $\mathcal{F} \subseteq \wp B$, as required. □

(b) Final writeup

Figure 4. Construction of a rhetorical proof and the final writeup, from [50, pp. 119–120].

(a) How $\bigcup \mathcal{F} \subseteq B$ first led to $\mathcal{F} \subseteq \wp B$ with a few head calculations.

$$
\begin{array}{l}
\bigcup \mathcal{F} \subseteq B \\
\quad \equiv \langle \bigcup\text{-conversion} \rangle \\
\qquad \left|\begin{array}{ll}
\equiv \langle \text{Expand definitions} \rangle & \forall\, x : \bigcup \mathcal{F} \,.\, (\exists\, A : \mathcal{F} \,.\, x \in A) \Rightarrow x \in B \\
\equiv \langle \text{Right distrib. } \Rightarrow/\exists \rangle & \forall\, x : \bigcup \mathcal{F} \,.\, \forall\, A : \mathcal{F} \,.\, x \in A \Rightarrow x \in B \\
\equiv \langle \text{Simplify} \rangle & \forall\, A : \mathcal{F} \,.\, \forall\, x : A \,.\, x \in B \\
\equiv \langle \text{Definition } \subseteq \rangle & \forall\, A : \mathcal{F} \,.\, A \subseteq B
\end{array}\right. \\
\qquad \equiv \langle \text{Definition } \wp \rangle \quad \forall\, A : \mathcal{F} \,.\, A \in \wp B \\
\qquad \equiv \langle \text{Definition } \subseteq \rangle \quad \mathcal{F} \subseteq \wp B
\end{array}
$$

(b) How the head calculations can be exposed for the interested reader.

$$
\begin{array}{l}
\bigcup \mathcal{F} \subseteq B \\
\quad \equiv \langle \bigcup\text{-conversion} \rangle \\
\qquad \left|\begin{array}{ll}
\equiv \langle \text{Expand definitions} \rangle & \text{— exposing 1st head calculation:} \\
\qquad \left|\begin{array}{ll}
\equiv \langle \text{Definition } \subseteq \rangle & \forall\, x : \bigcup \mathcal{F} \,.\, x \in B \\
\equiv \langle \text{Idempotency } \cap \rangle & \forall\, x : \bigcup \mathcal{F} \cap \bigcup \mathcal{F} \,.\, x \in B \\
\equiv \langle \text{Set trading } \forall \rangle & \forall\, x : \bigcup \mathcal{F} \,.\, x \in \bigcup \mathcal{F} \Rightarrow x \in B \\
\equiv \langle \text{Definition } \bigcup \rangle & \forall\, x : \bigcup \mathcal{F} \,.\, (\exists\, A : \mathcal{F} \,.\, x \in A) \Rightarrow x \in B
\end{array}\right. \\
\equiv \langle \text{Right distrib. } \Rightarrow/\exists \rangle & \forall\, x : \bigcup \mathcal{F} \,.\, \forall\, A : \mathcal{F} \,.\, x \in A \Rightarrow x \in B \\
\equiv \langle \text{Simplify} \rangle & \text{— exposing 2nd head calculation:} \\
\qquad \left|\begin{array}{ll}
\equiv \langle \text{Swapping } \forall \rangle & \forall\, A : \mathcal{F} \,.\, \forall\, x : \bigcup \mathcal{F} \,.\, x \in A \Rightarrow x \in B \\
\equiv \langle \text{Set trading } \forall \rangle & \forall\, A : \mathcal{F} \,.\, \forall\, x : \bigcup \mathcal{F} \cap A \,.\, x \in B \\
\equiv \langle \bigcup\text{-subsets} \rangle & \forall\, A : \mathcal{F} \,.\, \forall\, x : A \,.\, x \in B \\
\qquad \lozenge\ A \in \mathcal{F} \Rightarrow A \subseteq \bigcup \mathcal{F} &
\end{array}\right. \\
\equiv \langle \text{Definition } \subseteq \rangle & \forall\, A : \mathcal{F} \,.\, A \subseteq B
\end{array}\right. \\
\qquad \equiv \langle \text{Definition } \wp \rangle \quad \forall\, A : \mathcal{F} \,.\, A \in \wp B \\
\qquad \equiv \langle \text{Definition } \subseteq \rangle \quad \mathcal{F} \subseteq \wp B
\end{array}
$$

Figure 5. Equational derivation of the theorem of Figure 4 and its converse.

20 THEOREM *If there is a one-to-one function on a set A to a subset of a set B and there is also a one-to-one function on B to a subset of A, then A and B are equipollent.*

PROOF Suppose that f is a one-to-one map of A into B and g is one-to-one on B to A. It may be supposed that A and B are disjoint. The proof of the theorem is accomplished by decomposing A and B into classes which are most easily described in terms of parthenogenesis. A point x (of either A or B) is an ancestor of a point y iff y can be obtained from x by successive application of f and g (or g and f). Now decompose A into three sets: let A_E consist of all points of A which have an even number of ancestors, let A_O consist of points which have an odd number of ancestors, and let A_I consist of points which have infinitely many ancestors. Decompose B similarly and observe: f maps A_E onto B_O and A_I onto B_I, and g^{-1} maps A_O onto B_E. Hence the function which agrees on f on $A_E \cup A_I$ and agrees with g^{-1} on A_O is a one-to-one map from A onto B. ∎

21 [...] The intuitively elegant form of the proof of theorem 0.20 is due to G. Birkhoff and S. MacLane. John L. Kelley, *General Topology*, pp. 28–29. Van Nostrand (1955)

Figure 6. The Schröder-Bernstein theorem and the proof in Kelley's text [29].

$$
\begin{array}{ll}
(\exists\, f : A \to B \,.\, \mathsf{inj}\, f) \wedge (\exists\, g : B \to A \,.\, \mathsf{inj}\, g) \Rightarrow \exists\, h : A \to B \,.\, \mathsf{bij}_B\, h & \\
\equiv \langle \text{Enter scope} \rangle & \underline{\forall\, f : (A \to B)_{\mathsf{inj}} \,.\, \forall\, g : (B \to A)_{\mathsf{inj}} \,.}\ \exists\, h : A \to B \,.\, \mathsf{bij}_B\, h \\
\Leftarrow \langle \text{Introd. } H \rangle & \vartriangle\ \exists\, h : A \to B \,.\, h \in \mathcal{R}\, H \wedge \mathsf{bij}_B\, h \ \ \underline{\textbf{where } H\_ := F : \wp A \,.\, f_F \cup g^-_{A \setminus F}} \\
\equiv \langle \text{Change var. } \exists \rangle & \vartriangle\ \exists\, F : \mathcal{D}\, H \,.\, H_F \in A \to B \wedge \mathsf{bij}_B\, H_F\ \vartriangle \\
\equiv \langle \text{Elimin. } H \rangle & \vartriangle\ \underline{\exists\, F : \wp A \,.}\ \widetilde{g} B \setminus F = A \setminus F \wedge \widetilde{g^-}(A \setminus F) = B \setminus \widetilde{f} F \\
\equiv \langle \text{Elimin. } \widetilde{g^-} \rangle & \vartriangle\vartriangle\ F = A \setminus \widetilde{g}(B \setminus \widetilde{f} F) \\
\Leftarrow \langle \text{Calc. fixpt.} \rangle & \vartriangle\ \exists\, F : \wp A \,.\, F = \bigcup \{ C : \wp A \mid C \subseteq A \setminus \widetilde{g}(B \setminus \widetilde{f} C) \} \\
\equiv \langle \text{One-pt. rule } \exists \rangle & \vartriangle\ \bigcup \{ C : \wp A \mid C \subseteq A \setminus \widetilde{g}(B \setminus \widetilde{f} C) \} \in \wp A \\
\equiv \langle \bigcup \{ X : \wp S \mid p \} \in \wp S \rangle & \vartriangle\ 1
\end{array}
$$

Figure 7. Top level of a calculational proof for the Schröder-Bernstein theorem.

– Right distributivity $\Rightarrow/\exists$: $(\exists\, v : S \,.\, p) \Rightarrow q \equiv \forall\, v : S \,.\, p \Rightarrow q$ (v not free in q)

– Swapping adjacent $\forall$: $(\forall\, u : S \,.\, \forall\, v : T \,.\, p) \equiv (\forall\, v : T \,.\, \forall\, u : S \,.\, p)$

This aspect is mentioned only to illustrate how informative this style is to a reader.

Discussion

In the rhetorical proof of Figure 4, "clearly" and "we can conclude that" stand for omitted steps, justifications, and references to the definitions for the reader to figure out. The structure, visible in the scratch work, is absent in the final writeup. Most of the prose amounts to fragmenting the logic and circumventing quantified expressions instead of using their power. The givens/goal approach covers only one direction at a time.

The calculation in Figure 5 is only slightly longer yet much more detailed; nothing is left unexplained. Of course, shortcuts are possible, similar to Figure 2(c).

Finally, guided by the *shape* of the expressions ("letting the symbols do the work"), the end result was *discovered* by calculation, as in Figure 2, and the resulting theorem is stronger because it is a logical equality, not just implication, as in the task description. We also found a general purpose intermediate result for $\cup$*-conversion*: $\cup\, \mathcal{F} \subseteq B \equiv \forall\, A : \mathcal{F} \,.\, A \subseteq B$ (used in a later example). As a general purpose result, it merits factorizing out as a separate theorem.

A More Advanced Example: The (Cantor-) Schröder-Bernstein Theorem

This example is illustrative for many reasons. Books on set theory mention it as the first theorem whose proof is "difficult" [47, p. 94]. Lamport [33, p. 604] reports a proof by Kelley [30], noting that it is terse because it is written for a "more sophisticated audience", but also that the refinement needed to explain it in an introductory class reveals that it is wrong! This proof should be readable via Google Books, but pages are often hidden, so we quote the text in Figure 6.

Some may call such text a proof, but at best it is only a *proof outline*. Discarding details as "evident" can cause errors. An informal proof has the effect that one may tend to read it only as a series of hints, mentally patching errors along the way by letting some intuitive interpretation prevail over what is actually written. Educationally, of course, deliberate sloppiness sets very poor examples, unlike honest mistakes, since *nothing* can guarantee total absence of errors.

Turning an informal outline into a genuine proof requires formalizing and refining the proof obligations, which increases the length dramatically, in this case several pages [33, p. 604]. Using auxiliary concepts such as even/odd numbers or chains [13] causes extra proof obligations. Therefore we try using only concepts from the theorem's domain of discourse: set theory.

The top level of a calculational proof is shown in Figure 7. Little triangles stand for parts (underlined for easy visual retrieval) not copied from an earlier line to avoid repetition. This top-level proof is shown only to convey the flavor and should be seen in the spirit of Figure 1, since giving a list of all formal definitions and further expansion is far beyond the scope of this paper. Readers familiar with the theorem will infer most notation. The central ones are:

- f_F for restriction of function f to set F (definition via abstraction: $f_F = x : \mathcal{D}\, f \cap F \,.\, f\, x$)
- $\widetilde{f} F$ for the image of set F under function f; thus $\widetilde{f} F = \{ f\, x \mid x : \mathcal{D}\, f \cap F \}$
- $A \setminus F$ for set difference: $x \in A \setminus F \equiv x \in A \wedge x \notin F$
- $f \cup g$ to merge f and g (if functions are seen as sets of pairs, $\mathcal{D}\, f \cap \mathcal{D}\, g = \emptyset \Rightarrow f \cup g = f \cup g$)

A twelve-page note with all formal definitions, the derivation of the associated toolkit, the expansion to lower-level detail of Figure 7, the derivation of several variants for the $\langle$Elimin. $\widetilde{g^-}\rangle$-step, and

THEOREM Let A be a set and $s : \wp A \to \wp A$ isotonic $(\forall C, C' : (\wp A)^2 \,.\, C \subseteq C' \Rightarrow sC \subseteq sC')$. If we define $D := \{C : \wp A \mid C \subseteq sC\}$, then $F := \bigcup D$ satisfies $F = sF$.

PROOF With the given assumptions of the theorem, $F \in \wp A$ (lemma as exercise) and

$$
\begin{array}{lll}
F = \bigcup D \Rightarrow & \langle \text{Weakening to } \subseteq \rangle & \bigcup D \subseteq F \\
\equiv & \langle \bigcup\text{-elimination} \rangle & \forall\, C : D \,.\, C \subseteq F \\
\Rightarrow & \langle \text{Isotony } s \rangle & \forall\, C : D \,.\, sC \subseteq sF \\
\equiv & \langle \text{Definition } D \rangle & \forall\, C : D \,.\, C \subseteq sC \wedge sC \subseteq sF \\
\Rightarrow & \langle \text{Transitivity } \subseteq \rangle & \forall\, C : D \,.\, C \subseteq sF \\
\equiv & \langle \bigcup\text{-introduction} \rangle & \bigcup D \subseteq sF \\
\equiv & \langle F = \bigcup D \rangle & \underline{F \subseteq sF} \quad \text{(a)} \\
\triangleright \Rightarrow & \langle \text{Isotony } s \rangle & sF \subseteq s(sF) \quad \text{— Alternative format: } \triangleq \wedge sF \subseteq s(sF) \\
\equiv & \langle \text{Definition } D \rangle & sF \in D \\
\Rightarrow & \langle \bigcup\text{-subset} \rangle & sF \subseteq \bigcup D \\
\equiv & \langle F = \bigcup D \rangle & sF \subseteq F \\
\Rightarrow & \langle \text{(a), antisym. } \subseteq \rangle & F = sF
\end{array}
$$

Figure 8. A simple fixpoint property.

matching fixpoint calculations can be obtained by sending an email to the author.

Still, as promised, we illustrate the use of $\cup$-conversion in proving the fixpoint property for ⟨Calc. fixpt.⟩ in Figure 8. Although $\cup$-conversion is a logical equality, we rename it $\cup$-elimination or $\cup$-introduction, depending on the direction of the calculation, as extra reading aids.

A few simple conventions, not elaborated here, reduce writing, unless one prefers writing expressions in full as is customary in calculus.

Intermezzo: When Is a "Proof by Contradiction" a *Proof by Contradiction*?

The strategies of proving $p \Rightarrow q$ as $(p \wedge \neg q) \Rightarrow q$ or, by the deduction theorem, assuming p and $\neg q$ and proving q are potentially very powerful due to the "free" extra hypothesis $\neg q$. This free bonus explains the near-lyrical acclamations of famous mathematicians like Godfrey H. Hardy. Hence the following is likely to raise some hackles but hopefully will lead to fruitful discussion.

By reflecting the logical structure explicitly at all levels, the calculational style teaches us to consider more critically how far the potential of the "$\neg q$ hypothesis" is effectively exploited, in particular to obtain simple(r) proofs. It turns out that in many theorems typically presented as examples, the true essence does not depend on the $\neg q$ hypothesis and even provides "free" generalizations by simply discarding that hypothesis. Often the extraction of the central concept is made more informative by recasting it in problem form, making the original theorem an evident corollary. Here are some examples, the first being inspired by [50], yet adapted here.

- Euclid's theorem saying that the number of primes is infinite. More informative is:
  For any prime p, each prime factor of $1 + \prod(q : \mathsf{Primes} \mid q \le p)$ is larger than p.
  The proof of this variant is the essence of Euclid's classic proof yet does not depend on the $\neg q$ hypothesis. The original theorem is a direct corollary.
- The irrationality of $\sqrt{2}$. Two variants stated as problems (discussed later) are:
  When is the square root of an integer rational? (Answer: iff it is an integer)
  Given a prime p, for which positive natural k is $\sqrt[k]{p}$ rational? (Answer: iff $k = 1$)
- Thousands of years more recent are typical proofs in abstract algebra about uniqueness of unit elements or inverses. Many texts typically assume, for instance, that $u \neq u'$, then prove $u = u'$ without using $u \neq u'$ and emphatically conclude "Contradiction!".

A hypothesis is *spurious* in a proof if it can be eliminated while maintaining the original simplicity.

Returning to the irrationality of $\sqrt{2}$, comparing the simplicity for different strategies requires that all details are included (within the proof or as lemmas) down to basic principles.

Most proofs proceed via the existence of a reduced fraction of natural numbers whose square is 2 and derive a contradiction with the fraction

Theorem 6.4.5. $\sqrt{2}$ *is irrational.*
Proof. Suppose that $\sqrt{2}$ is rational. This means that $\exists q \in \mathbb{Z}^+ \exists p \in' \mathbb{Z}^+ (p/q = \sqrt{2})$, so the set $S = \{q \in \mathbb{Z}^+ \mid \exists p \in \mathbb{Z}^+ (p/q = \sqrt{2})\}$ is nonempty. By the well-ordering principle, we can let q be the smallest element of S. Since $q \in S$, we can choose some $p \in \mathbb{Z}^+$ such that $p/q = \sqrt{2}$. Therefore $p^2/q^2 = 2$, so $p^2 = 2q^2$ and therefore p^2 is even. We now apply the theorem from Example 3.4.2, which says that for any integer x, x is even iff x^2 is even. Since p^2 is even, p must be even, so we can choose some $\bar{p} \in \mathbb{Z}^+$ such that $p = 2\bar{p}$. Therefore $p^2 = 4\bar{p}^2$, and substituting this into the equation $p^2 = 2q^2$ we get $4\bar{p}^2 = 2q^2$, so $2\bar{p}^2 = q^2$ and therefore q^2 is even. Appealing to Example 3.4.2 again, this means that q must be even, so we can choose some $\bar{q} \in \mathbb{Z}^+$ such that $q = 2\bar{q}$. But then $\sqrt{2} = p/q = (2\bar{p})/(2\bar{q}) = \bar{p}/\bar{q}$, so $\bar{q} \in S$. Clearly $\bar{q} < q$, so this contradicts the fact that q was chosen to be the *smallest* element of S. Therefore $\sqrt{2}$ is irrational. □

Figure 9. A typical traditional proof of the irrationality of $\sqrt{2}$.

being reduced. Such a proof [51, p. 295] is quoted in Figure 9. The proof of Example 3.4.2 (*x is even iff x^2 is even*) is an essential part, and its scratch work takes nearly two pages. It is based on even/odd distinction and hence is not generalizable to the theorem *p divides a^2 implies p divides a*, which Sagher [45] identifies as a crucial result about primes not available in the classical age.

Lamport [33] gives a slightly different proof, with more detail, using properties of the gcd.

The example in Figure 10(b) of free generalization and avoidance of spurious contradictory hypotheses is inspired by the essence of Sagher's remarkably simple argument, still by contradiction [45], shown in Figure 10(a). Sagher tacitly uses a basic theorem for integers [28]: for $d > 0$, the integer quotient $q := D \div d$ and remainder $r := D - dq$ satisfy $0 \le r < d$. For easy reference, we call this *Euclid's Integer Division Theorem* (IDT).[1]

In Figure 10(b) this is made very explicit by extensive detail and by introducing redundant D, q, r rather than writing $d\sqrt{n} - d(d\sqrt{n} \div d)$ in full, all making the ideas clearer to the reader.

For tidiness and WLOG, rationality issues are better handled via natural numbers ($\mathbb{N} = \mathbb{Z}_{\ge 0}$), but here we decided to remain close to the formulations quoted from the literature sources.

To solve the problem *Given a prime p, for which positive natural k is $\sqrt[k]{p}$ rational?*, the simplest approach is to introduce a *multiplicity* function μ such that $\mu_p\, n$ is the multiplicity of a prime factor p in a nonzero natural number n. Existence and uniqueness follow from the *unique factorization theorem* [28, p. 19]. The properties of interest here are $\mu_p\, p = 1$ and the logarithmic-style rules $\mu_p(n \cdot m) = \mu_p\, n + \mu_p\, m$ and $\mu_p\, n^k = k \cdot \mu_p\, n$. This makes the solution a simple exercise (Figure 11).

[1] *Generalizable to reals: for any* real *D and d, $d \neq 0$, there exist unique* integer *q and* real *r satisfying $D = dq + r$ and $0 \le r < |d|$, which justifies jointly defining div ($\div$) by $D \div d = q$ and mod ($\dagger$) by $D \dagger d = r$* [9].

The rationality predicate Ratpos is the restriction of Rat to $\mathbb{R}_{>0}$.

The prime factorization theorem is known as "The Fundamental Theorem of Arithmetic" for good reason: it is how we were taught in third grade (1953) to calculate least-common multiples and greatest-common divisors (not by Euclid's algorithm). Collison [15] notes that (i) although this theorem was not formulated in its current form before Gauss, its principle was known by Euclid; (ii) it was used by Euler and Legendre as if it were "something already known", and (iii) "students are likely to consider it obvious" (of course risking incorrect generalizations).

By contrast, arguments based on infinite descent (or well-foundedness) are too subtle for third-graders. Sagher reports presenting the proof of Figure 10(a) successfully to ninth-graders. The number 0, which in Figure 10(b) obviates infinite descent, was not used in the classical age.

Returning to the general issue of arguments by contradiction: they are not possible if q.e.d.[2] is not known in advance. In education, it is often instructive to present theorems as problems. Incidentally, some students prove anything you ask, even $2 + 2 = 2$ if properly disguised.

If q.e.d. is known, a proof by contradiction is fine if one wants just a "speedup"—and *if* one obtains it! However, it comes at a price when insight is important. Indeed, intermediate results based on the given hypotheses only are direct theorems, but those that depend on the "free" contradictory hypothesis are *contaminated* in the usual scenario. In other scenarios, they may not yield the intended contradiction but new theories. In any case, yielding interesting side results is sometimes called the *fecundity* of a problem or proof and rarely benefits from rushing.

A brief historical note may be interesting. In the classical era, *reductio ad absurdum* was seen

[2] *Here standing for* "quod *est* demonstrandum".

(a) Sagher's proof, by contradiction (*The Amer. Math. Monthly* **95**, 2, (Feb. 1988)) 117.

THEOREM *The square root of a natural number is either an irrational or an integer.*

PROOF Suppose $\sqrt{k} = m/n$, where m and n are integers with $n > 0$. If k is not a square, there exists an integer q so that $q < m/n < q+1$. Now $m^2 = kn^2$ implies $m(m-qn) = n(kn-qm)$ and, hence, $m/n = (kn-qm)/(m-qn)$. From $q < m/n < q+1$ we get $0 < m-qn < n$. Therefore we have $\sqrt{k} = (kn-qm)/(m-qn)$, where the denominator is positive and smaller than the one in the original fraction. Continuing, we get an infinite decreasing sequence of positive integers, an impossibility.

(b) Variant stated as a problem, solved without contradictory hypothesis.

PROBLEM *When is the square root of an integer rational?*

SOLUTION For any integer n, calculate

$$
\begin{array}{lll}
\mathsf{Rat}\,\sqrt{n} & \equiv \langle \text{Definition } \mathsf{Rat} \rangle & \exists\, d : \mathbb{N}_{\neq 0} \,.\, \exists\, D : \mathbb{Z} \,.\, D = d\sqrt{n} \\
& \equiv \langle \text{One-pt. rule } \exists \rangle & \exists\, d : \mathbb{N}_{\neq 0} \,.\, d\sqrt{n} \in \mathbb{Z} \\
& \equiv \langle \mathbb{N} \text{ well-founded} \rangle & \underline{\exists\, d : \mathbb{N}_{\neq 0} \,.\, \mathbf{let}\ D := d\sqrt{n}\ \mathbf{in}\ D \in \mathbb{Z} \wedge} \ \forall\, d' : \mathbb{N}_{\neq 0} \,.\, d'\sqrt{n} \in \mathbb{Z} \Rightarrow d \leq d' \\
& \equiv \langle \text{From } \mathbb{N}_{\neq 0} \text{ to } \mathbb{N} \rangle & \\
& \quad \left| \begin{array}{ll} \equiv \langle \text{Trading sub } \forall \rangle & \vartriangleleft\ \forall\, d' : \mathbb{N} \,.\, d' \neq 0 \Rightarrow d'\sqrt{n} \in \mathbb{Z} \Rightarrow d \leq d' \\ \equiv \langle \text{Prop. calc., def. } < \rangle & \vartriangleleft\ \forall\, d' : \mathbb{N} \,.\, d' < d \wedge d'\sqrt{n} \in \mathbb{Z} \Rightarrow d' = 0 \end{array} \right. & \\
& \Rightarrow \langle \text{Inst. } d' := r,\ 0 \leq r \rangle & \vartriangleleft\ \underline{\mathbf{let}\ q := D \div d\ ;\ r := D - dq\ \mathbf{in}}\ r < d \wedge r\sqrt{n} \in \mathbb{Z} \Rightarrow r = 0 \\
& \equiv \langle \text{Simplify} \rangle & \\
& \quad \left| \begin{array}{ll} \equiv \langle \text{By IDT: } r < d \rangle & \vartriangleleft\vartriangleleft\ r\sqrt{n} \in \mathbb{Z} \Rightarrow r = 0 \\ \equiv \langle r\sqrt{n} = dn - Dq \rangle & \vartriangleleft\vartriangleleft\ dn - Dq \in \mathbb{Z} \Rightarrow r = 0 \\ \equiv \langle dn - Dq \in \mathbb{Z} \rangle & \vartriangleleft\vartriangleleft\ r = 0 \\ \equiv \langle r = 0 \equiv D/d = q \rangle & \vartriangleleft\ \sqrt{n} = D/d = D \div d \end{array} \right. & \\
& \Rightarrow \langle D \div d \in \mathbb{Z},\ \text{weaken} \rangle & \exists\, d : \mathbb{N}_{\neq 0} \,.\, \sqrt{n} \in \mathbb{Z} \\
& \equiv \langle \text{Const. sub } \exists \rangle & \sqrt{n} \in \mathbb{Z}
\end{array}
$$

The converse, $\sqrt{n} \in \mathbb{Z} \Rightarrow \mathsf{Rat}\,\sqrt{n}$, is trivial (technically). So the answer is $\mathsf{Rat}\,\sqrt{n} \equiv \sqrt{n} \in \mathbb{Z}$.

Figure 10. Illustrating the avoidance of spurious contradictory hypotheses.

PROBLEM *Given a prime p, for which positive natural k is $\sqrt[k]{p}$ rational?*

SOLUTION For any prime number p and any nonzero natural number k,

$$
\begin{array}{lll}
\mathsf{Ratpos}\,\sqrt[k]{p} & \equiv \langle \text{Def. } \mathsf{Ratpos} \rangle & \exists\, n, m : \mathbb{P}^2 \,.\, \sqrt[k]{p} = n/m \ \text{—}\ \mathbb{P} := \mathbb{N}_{\neq 0} \\
& \equiv \langle \text{Arithmetic} \rangle & \exists\, n, m : \mathbb{P}^2 \,.\, p \cdot m^k = n^k \\
& \Rightarrow \langle \text{Leibniz} \rangle & \exists\, n, m : \mathbb{P}^2 \,.\, \mu_p\,(p \cdot m^k) = \mu_p\, n^k \\
& \equiv \langle \text{Properties } \mu \rangle & \exists\, n, m : \mathbb{P}^2 \,.\, 1 + k \cdot \mu_p\, m = k \cdot \mu_p\, n \\
& \equiv \langle \text{Arithmetic} \rangle & \exists\, n, m : \mathbb{P}^2 \,.\, 1 = k \cdot (\mu_p\, n - \mu_p\, m) \\
& \Rightarrow \langle \text{Dummy change} \rangle & \exists\, j : \mathbb{Z}_{\neq 0} \,.\, 1 = k \cdot j \\
& \equiv \langle k > 0 \rangle & k = 1
\end{array}
$$

Figure 11. Characterizing the rational roots of a prime number.

as a rather slick and fashionable way in rhetoric (politics, law) to discredit an opponent's viewpoints and was similarly considered a clever ploy in mathematics. Archimedes [44] is reputed to have occasionally "poisoned" his reports with deliberate mistakes to challenge his reader(s) [41]. Perhaps even more subtle is hiding the intuitive thinking leading to a discovery, and what better way to cover one's tracks than a proof by contradiction? In a fascinating account of rediscovered writings by Archimedes, *The Archimedes Codex* [41, p. 143] contains an example of a spurious contradiction, although it is unclear whether that reflects Archimedes' argument.

To conclude, proofs by contradiction should not be rejected but handled with care.

Overlap (s)

$\equiv$ ⟨Basic form⟩ $\underline{\exists\, k,k' : \mathbb{Z}^2_{\neq} \,.\, \exists\, g : \mathbb{R} \,.\, g \in R_k(s) \cap R_{k'}(s)}$

$\equiv$ ⟨Insert defs.⟩ $\underline{\triangleq\, (\exists\, f : S \,.\, g = f - ks) \wedge (\exists\, f' : S \,.\, g = f' - k's)}$

$\equiv$ ⟨Rearrange $\exists$⟩ $\underline{\exists\, f : S \,.\, \exists\, f' : S \,.\, \exists\, k,k' : \mathbb{Z}^2_{\neq} \,.\, \exists\, g : \mathbb{R} \,.\, g = f - ks \wedge g = f' - k's}$

$\equiv$ ⟨Eliminate g⟩ $\underline{\triangleq\, f - ks = f' - k's}$

$\equiv$ ⟨Definition S⟩ $\exists\, f : P \cup N \,.\, \exists\, f' : P \cup N \,.\, \exists\, k,k' : \mathbb{Z}^2_{\neq} \,.\, f - ks = f' - k's$ (A)

Figure 12. Bandpass sampling: (a) concretizing the support overlap formula.

PROBLEM Given $P := [L,U]$, $N := [-U,-L]$, $s : \mathbb{R}_{>0}$. Simplify as far as possible:

$$\exists\, f : P \cup N \,.\, \exists\, f' : P \cup N \,.\, \exists\, k,k' : \mathbb{Z}^2_{\neq} \,.\, f - ks = f' - k's. \tag{A}$$

SOLUTION The inner quantification is simplified by writing $f - ks = f' - k's$ as $f - f' = (k - k')s$. The change of variables $\ell := k - k'$ results in $\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, f - f' = \ell s$.

In expanding the outer quantifications, we exploit symmetry. Also introducing the convenient shorthand $Q_x \equiv \exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, x = \ell s$, domain split for the first $\exists$ yields

$$(\exists\, f : P \,.\, \exists\, f' : P \cup N \,.\, Q_{f-f'}) \quad \vee \quad (\exists\, f : N \,.\, \exists\, f' : P \cup N \,.\, Q_{f-f'}).$$

Letting $f, f' := -f, -f'$, the second term becomes $\exists\, f : P \,.\, \exists\, f' : N \cup P \,.\, Q_{f'-f}$. Now $Q_{-x} \equiv Q_x$ and $N \cup P = P \cup N$, so both terms are equal and collapse since $p \vee p \equiv p$. By another domain split for $\exists$,

$$(\exists\, f : P \,.\, \exists\, f' : P \,.\, Q_{f-f'}) \quad \vee \quad (\exists\, f : P \,.\, \exists\, f' : N \,.\, Q_{f-f'}). \tag{B}$$

We first simplify the second (more interesting) term. Here $f \in P \wedge f' \in N$ implies $f - f' \in [2L, 2U]$, which we add as a conjunct to help remove f and f' as follows:

$\exists\, f : P \,.\, \exists\, f' : N \,.\, \exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, x = \ell s \wedge f - f' \in [2L, 2U]$

$\equiv$ ⟨Leibniz⟩ $\exists\, f : P \,.\, \exists\, f' : N \,.\, \exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, f - f' = \ell s \wedge \ell s \in [2L, 2U]$

$\equiv$ ⟨Swap $\exists$⟩ $\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \exists\, f : P \,.\, \exists\, f' : N \,.\, f - f' = \ell s \wedge \ell s \in [2L, 2U]$

$\equiv$ ⟨Distrib. $\wedge/\exists$⟩ $\underline{\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [2L, 2U]} \;\wedge\; \exists\, f : P \,.\, \exists\, f' : N \,.\, f - f' = \ell s$

$\equiv$ ⟨Arithmetic⟩ $\underline{\triangleq\, \exists\, f : P \,.\, \exists\, f' : N \,.\, f' = f - \ell s}$

$\equiv$ ⟨One-pt. rule $\exists$⟩ $\triangleq\, \exists\, f : P \,.\, f - \ell s \in N$

$\equiv$ ⟨Eliminate f⟩ $\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [2L, 2U]$

$\Diamond$ By $\forall\, v : [2L, 2U] \,.\, \exists\, f : P \,.\, f - v \in N$ (proof: witness $f := v/2$)

Similarly using $f \in P \wedge f' \in P \Rightarrow f - f' \in [-B, B]$, the first term of (B) becomes

$$(\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [-B, B]) \quad \vee \quad (\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [2L, 2U]).$$

By symmetry (as before), the first term equals $\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [0, B]$. From $s \geq 0$ we get $\ell s \in [0, B] \Rightarrow \ell \geq 0$ and $\ell s \in [2L, 2U] \Rightarrow \ell \geq 0$. Hence, by ⟨Trading sub $\exists$⟩,

$$(\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [0, B]) \quad \vee \quad (\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [2L, 2U]).$$

A last simplification requires checking a hunch that the first term is absorbed.

$\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [0, B]$

$\Rightarrow$ ⟨Eliminate ℓ⟩ $s \leq B$ — Converse simple but not needed

$\Diamond$ By $(\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \leq B) \Rightarrow s \leq B$ (proof: $\forall\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \leq B \Rightarrow s \leq B/\ell \leq B$)

$\Rightarrow$ ⟨Weaken: $B \leq 2B$⟩ $s \leq 2B$

$\equiv$ ⟨Euclid IDT⟩ $0 \leq 2U - (2U \div s)s < s \wedge s \leq 2B$

$\equiv$ ⟨Arithm. $\leq$⟩ $2L < (2U \div s)s \leq 2U \wedge s \leq 2B$

$\Rightarrow$ ⟨$2U \div s \in \mathbb{N}_{\neq 0}$⟩ $\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [2L, 2U]$

Seeing no further simplification, we present the final result as

$$\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, 2L/\ell \leq s \leq 2U/\ell \,. \tag{C}$$

Figure 13. Bandpass sampling: (b) simplification to obtain a practical formula.

Solving a Small But Puzzling Problem in Bandpass Sampling

This example shows that "back-of-an-envelope" calculations (quick calculations to resolve an unclear practical issue) are helpful not only with algebra or calculus but also for problems with a logical flavor.

As noted in a recent historical survey [17], sampling has interested scientists since Cauchy. In principle, a bandlimited signal can be faithfully reconstructed from samples taken at a rate that is at least twice the bandwidth. However, guaranteeing the absence of aliasing requires intricate choices. Indeed, in a renowned textbook, Lyons [39] points out many incorrect results in the literature, showing that the problem is puzzling and not evident to get right. He also presents a correct answer but based on exactly the same kind of informal arguments as the faulty ones. We show here how a formal solution resolves the issue in a straightforward way.

We first recall some well-known results. Let a signal x with Fourier transform $\mathcal{F}x$ be sampled with sampling period T (or *rate* $s := 1/T$). The usual model is pointwise multiplication with delta functions distance T apart. The sampled version $\hat{x}_T$ has Fourier transform $\mathcal{F}\hat{x}_T$

$$(\mathcal{F}\hat{x}_T)(\omega) = \int_{-\infty}^{\infty} x(t) \sum_{k=-\infty}^{\infty} \mathrm{e}^{jk\Omega t} \mathrm{e}^{-j\omega t}\, \mathrm{d}\, t$$

$$= \sum_{k=-\infty}^{\infty} (\mathcal{F}x)(\omega - k\Omega) \text{ where } \Omega := 2\pi/T. \tag{1}$$

This shows that $\mathcal{F}\hat{x}_T$ consists of an infinite number of *replicas* of $\mathcal{F}x$, spaced Ω apart. The usual reconstruction with sinc functions is faithful if and only if the replicas do not overlap.

The resulting requirements for bandpass sampling are formalized as follows. For convenience, we switch from $\omega - k\Omega$ to $f - ks$ ($\omega = 2\pi f$, $\Omega = 2\pi s$, $s > 0$). A signal is *bandlimited* if its spectrum is zero outside an interval $P := [L, U]$, called the *support* (for the positive frequencies), and the *bandwidth* is $B := U - L$. Including the negative frequencies $N := [-U, -L]$, the support is $S := P \cup N$, so the support $R_k(s)$ for the k-th replica is given by $R_k(s) = \{f - ks \mid f : S\}$. Sampling rates s causing (undesirable) overlap are given by

$$\text{Overlap}\,(s) \equiv \exists\, k, k' : \mathbb{Z}^2_{\neq} \,.\, \exists\, f : \mathbb{R} \,.\, f \in R_k(s) \cap R_{k'}(s), \tag{2}$$

which is our *basic form*. The problem is deriving a practical engineering formula.

The calculation is done in two phases. The first is shown in Figure 12. Every step reflects head calculation, with details in the appendix, which also shows the second phase in the hierarchical layout illustrated thus far. For a change, Figure 13 shows how one might present it for readers less familiar with the rules for quantifiers than with those for sums or integrals (without prose, as in Figure 1). Although linked by prose, the steps are purely calculational: transforming expressions, guided by their shape. Proper quantifier rules preserve algebraic style, whereas the common reflex on seeing $\exists\, x : S \,.\, p$ is "Well, let's pick such an x." The chain is even equational, and the reader may wish to compare this with handling rhs(2) $\Rightarrow$rhs(3) and rhs(3) $\Rightarrow$ rhs(2) separately. Anyway, we have proven

$$\text{Overlap}(s) \equiv \exists\, \ell : \mathbb{N}_{\neq 0} \,.\, 2L/\ell \le s \le 2U/\ell. \tag{3}$$

Here we give some useful interpretations for special cases. For baseband sampling ($L = 0$ Hz, $U = B$), the formula yields the usual $s > 2B$ requirement. The steps based on Euclid's IDT show that this is also a necessary condition for bandpass sampling (extractable as a theorem). In view of (3), it is sufficient only if L is a multiple of B and provided the spectrum has no significant content at the edges L and U of the interval $[L, U]$.

Two afterthoughts. The original "back-of-an-envelope" calculation grew significantly by adding detail. Also, given the many errors in the literature, how dependable are the calculations shown, not checked by another human or by computer? Yet, adhering to the diversity principle, we devised a 2D f, s diagram where every formal step has a geometric interpretation, and (3) appears as a diagram via a simple coordinate transformation (horizontal shift, vertical scaling) without redrawing, as described in [8].

Note on Strategy and on Representativeness of Examples

Many examples show that maintaining logical equality ($\equiv$) saves the double effort of a *ping-pong argument* (exploring $\Rightarrow$ and $\Leftarrow$ separately). This is a bonus that is not available when the two directions inherently differ, as in the last line of Figure 10(b). Still, in exploration (outcome not known in advance), it is a good strategy to maintain $\equiv$ as long as possible. Once a weakening or strengthening step is taken, the reverse will have to be handled separately, but that is no reason to abandon ship: internal subchains may still be equational and perhaps interesting as separate theorems.

Another issue is representativeness of illustrations. Readers of an earlier version wondered whether the examples might be too elementary. These were chosen to illustrate style only, and examples must not be so large as to require several articles by themselves (e.g., the Odd Order Theorem). Still, new examples are now added, some matching the least elementary theorems in a representative sample of nonspecialist texts on

theorem proving. In selecting from such texts and in other branches as well (real/complex analysis, from δ-ϵ arguments [10] to Hilbert spaces and applications in communications engineering), we found that the calculational style *consistently* made proofs and problem solving noticeably easier for both readers and writers.

A set of examples representative for all branches of mathematics would fill textbooks, but there is no doubt that every reader will find experimenting with logic as algebra rewarding.

Views on Formality, Calculation, Problem Solving and Proofs: The Cultural Divide

While it is true that the practical use of formal logic developed most strongly in computer science by necessity [16], it is arguably more accurate and constructive to recognize that the actual cultural divide lies *within* mathematics. Indeed, recall the main points of this article:

1. Precise notation deserves to be embraced rather than avoided. We learn symbols like + and = in first grade. Words are used only for reading equalities like 2 + 3 = 5 aloud but not for writing them. Why write "and" rather than $\wedge$ or "for all" rather than "$\forall$"?
2. Formal rules permit us to reason with symbols, overcoming the drawbacks of informal arguments. They make the difference between mere syncopation and mature symbolism. Who would prove $(a+b)\cdot(a-b) = a^2 - b^2$ in words?
3. Propositional and quantifier calculus are as useful for logic-type arguments as algebra and integral calculus are for the kind of problems that interested Descartes and Newton.

Formal rules as perfected by Descartes have replaced word arguments in algebra to a large extent without fully eliminating prose, which also has a useful role. Such balance is equally relevant for logic reasoning, but today it is still heavily weighted on the syncopation side.

With the risk of stating the obvious, we emphasize that the class of problems considered are those whose formulation is already given in a symbolic setting or when translation is rewarding. By contrast, there are many informally stated problems whose formalization would entail a huge loss of abstraction and generality (except with a disproportionate amount of effort) yet which can be solved rigorously without symbolism.[3] Every reader will have some favorite examples.

[3] *There are also many problems where a very direct ad hoc formal axiomatization enables applying formal logic without overhead.*

The cultural divide of interest here is between symbolic reasoning in algebra/calculus and rhetoric logic. In this context, the distinction between formality in general and its embodiment in the calculational style hardly matters, and otherwise is sufficiently clear from the context.

Initial Reactions

Colleagues practicing formal reasoning (calculational or not [33], [35]) in nonspecialist mathematics report mutually similar initial reactions from other engineers and mathematicians who have always done the logical parts in prose. This issue appears even more sensitive than others with a formative element, which, as Mason observed for word problems [40], "sometimes provoke strong emotional reactions from otherwise calm adults."

Interestingly, every single argument *contra* would make a no less valid case against the no less formal style exemplified by Figure 2, well-established for hundreds of years. One could easily leave it at that, letting the reader make the balance. However, a concrete list is instructive.

- *Formal reasoning uses symbols, making it less accessible to readers unfamiliar with them.* (i) Of course, every discourse must take into account the background of the intended audience. (ii) There are sound arguments for introducing basic logic symbols much earlier than is customary: truth tables are even simpler than addition tables.
- *Formal reasoning is not as natural as reasoning with words.* (i) The term "natural" is loaded; it usually refers to custom: no mathematician nowadays would consider Figure 2 less natural than a corresponding argument in words. (ii) Logic itself is not "natural": Aristotle's logic came not from nature but from hard thinking [44]. (iii) Everything "natural" can be improved; we may thank Diophantus for improving on "natural algebra".
- *This is not how mathematicians work.* (i) Not all people work in the same way: many *combine* intuition from the domain of discourse with symbolic intuition, one enriching the other. (ii) Mathematics benefits as much as engineering from diversity in methods.
- *It is unclear how calculational arguments "scale up".* Does algebra or logic scale up? Presumably the remark means *How does one present a huge proof as one big calculation?*. It is doubtful whether anyone would want to do this: proofs with logic content, just like algebraic-style ones (Figure 2), form part of a larger whole. As in engineering, complexity and size are mastered by hierarchical design, modularization, and proper interfaces.

- *Formal reasoning requires adding much detail that may be uninteresting or distracting.* This is discussed in detail in [2], [33]: with modularization and hierarchical structuring, the writer can present his work such that the reader may focus on preferred parts.
- *The traditional mathematical style always worked fine for me.* (equally personal): For me too, but doing logic as algebra increases my pleasure in doing mathematics.

These initial reactions are the most representative ones, being also reported by others. As with any topic, much depends on individual temperament, background, and work area.

Values, Beliefs, and Attitudes

Values one would think to be generally shared by all people interested in mathematics often turn out to be a source of serious discord. Some are directly relevant to the issues in this article and hence deserve some discussion.

Morris Kline aptly noted that *More than anything else mathematics is a method.* [31, p. 454]. The purpose is not stated, but the completion *for effective reasoning* seems most appropriate and sufficiently general, and will be our main (not unique) basis for appreciation.

A seemingly obvious value is correctness. In computing science it is the ultimate criterion. In practice one often has to settle for less. Mathematicians may ask: when can one be certain that a solution or proof is correct? Except for very simple cases (who is to judge?), the only fitting answer is "never". Still, checking by peers or by computer can increase confidence.

Utilitarian values only shift the issue. In the end, "the proof of the pudding is in the eating", which does not mean just survival, but enjoyment. Mathematics is not easy (everyone hits walls), so a truly valid reason for doing it is pleasing the mind, which includes sharing. Arnol'd tells in [38] how very young children used to enjoy mathematical problems. According to Russell,

> *It is very desirable in instruction, not merely to persuade the student of the accuracy of important theorems, put to persuade him in a way which itself has, of all possible ways, the most beauty.*
>
> Bertrand Russell,
> *The Study of Mathematics*

Much has been written about beauty and elegance in mathematics, which are ultimately subjective criteria, but common elements seem to be economy of argument (obviously not through omission, but by conceptual clarity), providing insight, and effort-free generalizability.

Depending on the problem, these elements benefit from obviating formalization via a direct insight conveyed in words or precisely reside in the formalization itself. For instance, the *point-free style*, i.e., expressions without domain variables, shown for set theory by Tarski and Givant [48], is appreciated for its elegant algebraic flavor, and it also works for quantifiers [10]. Lamport notes [35]: *A proof should be beautiful mathematics. Its beauty lies in its logical structure, not in its prose.*

Quoting some sources was felt useful since many people find the above criteria irrelevant.

One would expect less discord about diversity of perspectives, which (arguably more than anything else) enhances understanding and insight. Yet a colleague once criticized it as "confusing to students, since they would not know which method to use on an exam." No comment.

Discord also exists about the value of "monster theorems", the holy grail for some mathematicians. Arnol'd considers some of these efforts as *wasting energy on (rock-climbing-type) exercises* [1], but his viewpoint is selective and generally not as harsh as this quote suggests.

What one expects from a proof or a solution to a problem greatly influences how one approaches it or presents it. The idea of "persuading" or "convincing" someone else by a proof (or by an article!) is illusory: more pertinent is gaining (by the writer) or offering (to the reader) further insight and understanding, often helpful to draw conclusions with greater confidence.

The conclusions are central for typical proofs in computer science (e.g., program correctness), which are characterized by the need for attention to detail rather than conceptual depth. Newly written programs that, at first, seem "obviously correct" to their programmer (in this scenario assumed to be unfamiliar with formal methods) nearly always contain fatal flaws that appear equally obvious when pointed out.

Bentley illustrates this with binary search [7], finding an item in a sorted list using a very simple principle people instinctively use in everyday life: look at any place in the list (most efficiently halfway, but that is another issue), compare to the item sought, and, depending on the outcome, repeat the search in the part before or after the initially chosen place. In a test for experienced programmers, given ample time, only about 10 percent got it right. This example concerns a program of half a dozen lines, so what to think about millions of lines?

Fortunately, huge programs are not that densely populated with pitfalls, but the point is that an attitude of "just give me the main ideas; I am smart enough to work out the details" leads to disaster, and some humility is more appropriate. This is not

always appreciated by mathematicians (Kelley's proof of the Schröder-Bernstein theorem comes to mind), yet sometimes it is driven home: all early ideas for proving Fermat's last theorem failed in a "small" detail, and the story goes that a famous German professor replied to such proposals with polite letters that differed only in mentioning the line number identifying the error.

Since proofs for program correctness contain few internal results relevant for extraction, in the presentation the details are often provided "exactly where needed", as emphasized in [2], [33]. Presenting them as a theorem or lemma before the main proof makes the less curious reader wonder about their purpose; presenting them afterwards forces the impatient reader to search.

In mathematics, physics, and engineering, insight and understanding are at least as important as the conclusions, and some internal results may be interesting by themselves. As is customary in papers and textbooks, such results are integrated as theorems in a little ad hoc theory or large general theory relevant to the domain of discourse. If sufficiently general, these theorems are placed before the "main" theorem (which may even have lost this status!); if routine, they can be placed as lemmas after the theorem (even for the impatient).

Students should learn to write in a way that is helpful to readers. Velleman's *How To Prove It: A Structured Approach* [51] describes many ways to write better proofs. In the preface (page xii) he argues that indenting proofs helps *to make the underlying structure of the proof clear*. On page 92 he observes that *Using the notation and rules of logic can be very helpful when you are figuring out the strategy for a proof*. He emphasizes in both cases that mathematicians do *not* present their proofs that way and also that they just write the steps *with no explanation of how they thought of them*.

Perpetuating abuse of notation is a major factor in the cultural divide. A great disappointment to beginning students (and to anyone with a critical mind) is the discrepancy between the reputation of mathematics as being precise and the many abuses of notation that they are forced to accept during their formative years. The intellectual effort is usually minor, but faith in the "honesty" of mathematics as a scientific discipline is damaged by this double standard.

Perhaps more importantly, abuse of notation is an impediment to precise rules and formal reasoning beyond syncopation. This issue will be taken up later.

On Bridging the Gap

Linguistic Symbiosis

A very interesting remark by one of the reviewers is a reference to the work of Ganesalingam [19] and the conclusion that the problem really is the immaturity of computer processing of natural languages. As discussed next, the premises are wrong, but they do not prevent the work from being very valuable and a continuation of current "good practices" of letting formalisms and formalization attain symbiosis with natural language.

Quite misleading is [20, p. 3], contrasting a very unappetizing formula from Russell's *Principia* with an example from a textbook called "modern" because it is more recent but otherwise uses a style that would have looked familiar two hundred years ago. This is just an anecdotal detail.

According to [19, p. 26], *The primary function of symbolic mathematics is to abbreviate material that would be cumbersome to state with text alone*. However, that precisely describes syncopation, the stage of algebra reached with Diophantus. What really counts is formal manipulation for algebra achieved by Viète and Descartes [4], [5]. This is not an expressiveness issue: $a^2 - b^2 = (a + b) \cdot (a - b)$ can be written in words as *The difference between the squares of two numbers equals the product of their sum and their difference (in the same order)*, which does not sound overly verbose. Not expressiveness, but convenience for reasoning led to symbolism.

Ganesalingam recognizes the superior expressiveness of formal mathematics and gives an example [19, p. 22] to show that formal expressions typical in integral calculus have no analogue in natural, computer, or other languages. Indeed, how would one read formula (1) in words? In integral calculus, expressiveness requires symbolism. Incidentally, in calculations, mathematicians are quite patient in repeating integral signs literally in successive lines, even where only the integrand changes.

By contrast, expressiveness is not a primary issue in quantified expressions: given a proper formalism, they can be conveniently read aloud in words (if there are not too many). At the same time, this may explain why they remained stuck in the syncopation stage [49] and in traditional reasoning are disassembled by the "let's pick an x" reflex. When using quantifier calculus instead, successive lines look much like calculations with integrals, except that we have shown a little less patience with repeated quantifications that do not change.

The above considerations explain the following symbiosis, which we found very helpful. In teaching formalization of problems, we emphasize expressing them in precise natural language

without contorted syntax, yet in such a way that (near-) literal formalization is possible. Conversely, our formalism is designed such that reading logical formulas aloud literally yields grammatically correct sentences. Interestingly, this significantly helped notational engineering.

A Small Example: Set Membership and Binding

This example illustrates the symbiosis between correct grammar in natural language and in formalism design.

Consider the following equations in equally familiar-looking notations:

$$S \subseteq T \equiv \forall\, x : S \,.\, x \in T, \tag{4}$$

$$S \subseteq T \equiv \forall\, x \in S \,.\, x \in T. \tag{5}$$

The right-hand side in (4) is read "For all x in S, x is in T", which is grammatically correct. Consistently, the right-hand side in (5) is read "For all x is in S, x is in T". Bad marks for grammar in grade school!

The problem in (5) is overloading $\in$ for two purposes: as the usual set membership operator in $x \in T$ and for introducing (declaring) a variable x by $x \in S$. Overloading is avoided in (4) in an obvious manner. Of course, correct grammar is not the goal but is a good "sanity check".

Technically, $\in$ is the standard set membership operator and is better reserved for that purpose. Also, $x \in T$, read "x is in T" or "x is an element (or member) of T", is a proposition (a boolean expression for computer scientists) that may be true or false depending on x and T.

By contrast, $x : S$ is a *binding* (declaration). It has no truth value, but in $\forall\, x : S \,.\, x \in T$ it has a *scope*, namely the part after the period. The binding specifies that $x \in S$ within the scope.

Before revealing the bigger picture, we consider a more illustrative example.

The expressions $\{x : \mathbb{R} \mid x > 7\}$ and $\{x^3 \mid x : \mathbb{N}\}$ directly look familiar; most readers will not even notice the colon, and all will read the expressions correctly. However, consider the general pattern. Let v be a variable, p a propositional expression, S and T sets, and e any expression. The general patterns for sets are $\{v : S \mid p\}$ and $\{e \mid v : T\}$. Many authors would write $\{v \in S \mid p\}$ and $\{e \mid v \in T\}$, but $\{x \in S \mid x \in T\}$ fits both patterns, an ambiguity! With proper distinction between binding and set membership, $\{x : S \mid x \in T\} = S \cap T$ and $\{x \in S \mid x : T\} \subseteq \mathbb{B}$ (booleans). Thinking that $\{x \in S \mid x : T\}$ is not useful is incorrect and not relevant: the issue is that the syntax in $\{x \in S \mid x \in T\}$ is ambiguous and that correction is very simple.

Intermezzo: The Bigger Picture

Here is an outline of the language design aspects of the formalism used in [10]. Expecting completeness would cause misleading oversimplification.

All abstraction is unified by *function abstraction*: $v : S \mathbin{\vartriangle} p \,.\, e$ ($\vartriangle\, p$ optional) denotes a function whose domain is the set of all v in S satisfying p and that maps v to e. So $n : \mathbb{Z} \mathbin{\vartriangle} n > 0 \,.\, 2 \cdot n$ and $n : \mathbb{N} \,.\, 2 \cdot n$ both denote the function that doubles natural numbers, and $f = x : \mathcal{D} f \,.\, f(x)$.

Predicates are boolean-valued functions. Quantifiers are predicates on predicates: if P is a predicate, $\forall\, P \equiv P = x : \mathcal{D} P \,.\, 1$ and $\exists\, P \equiv P \neq x : \mathcal{D} P \,.\, 0$. This supports pointfree expression (as for $\sum f$, $\lim_a f$, etc.) and the common pointwise style via function abstraction, as in (4).

The *function range* operator $\mathcal{R}$ is defined by $y \in \mathcal{R} f \equiv \exists\, x : \mathcal{D} f \,.\, y = f(x)$. A synonym is $\{\ \}$. Together with the "abbreviations" $v : S \mid p$ for $v : S \mathbin{\vartriangle} p \,.\, v$ and $e \mid v : S$ for $v : S \,.\, e$, this yields familiar-looking forms such as $\{v : S \mid p\}$ and $\{e \mid v : T\}$.

Warning: This outline does not cover foundational issues. The notation is intended to be "portable" across various set-theoretic foundations, from ZF [47] to NFU [29]. For instance, a simple interface to Suppes's formulation comprises two quantifiers $\forall$ and $\exists$ from [47, pp. 3–4], used for that purpose only and not needed in "working-type" calculations, as in the examples. Further explanation would lead us astray from the subject matter of this article.

Still, this notational digression demonstrates the simplicity of designing unambiguous notation while maintaining close resemblance to the "average" of the various disparate notations throughout the literature. It also ensures the feasibility of the discipline proposed below.

The Bane of Notational Negligence

Notational negligence is a major impediment against eliminating the cultural divide. At the same time, it is the easiest to remove and hence a logical first step, were it not for various human factors discussed next.

In the literature and in many professions, a text full of spelling mistakes and grammatical errors is considered disgraceful, and reviewers as well as editors would do their best to ensure the necessary corrections. Yet, in mathematical texts, sloppy conventions are tolerated uncritically. Often one notices a supercilious attitude (others have called it machismo) whereby notational conventions are considered too unimportant to deserve some care, or poor notation is defended by (false) claims of "common practice". Interestingly, such a casual attitude is typical in areas where discourse has

not yet progressed beyond the syncopation stage. Unfortunately, it creates an obvious vicious circle impeding the development of symbolic reasoning, where notation should be sufficiently trustworthy for uninterpreted use.

A pervasive form of sloppiness concerns some of the most useful forms of expression in mathematics: variables and functions. In particular, the distinction between free and bound variables and the rules governing them seem to be insufficiently appreciated.[4] For instance, in [51] the definition "Truth set of $P(x) = \{x \mid P(x)\}$" has x occurring free on the left and bound on the right of the equality sign. The incorrect appearance of (x) in "the truth set of $P(x)$" is analogous to talking about "the function $f(x)$" when actually meaning "the function f", which invoked the justified wrath of calculus professors fifty years ago!

Unfortunately, the good advice of these professors is still necessary today: it suffices to Google "function f(x)" or, even worse, "x = x(t)" to know how things stand.

In a widely disseminated paper about education in signal processing [36], Lee and Varaiya criticize evidently erroneous conventions in classical analysis:

> Most texts call the expression $x(t)$ a function. A better interpretation is that $x(t)$ is an element in the range of the function x. The difficulty with the former interpretation becomes obvious when talking about systems. Many texts pay lip service to the notion that a system is a function by introducing a notation like $y(t) = T(x(t))$. This makes no distinction between the value of the function at t and the function y itself.
>
> Why does this matter? Consider our favorite type of system, an LTI system. We[5] write $y(t) = x(t) * h(t)$ to indicate convolution. Under any reasonable interpretation of mathematics, this would seem to imply that $y(t-\tau) = x(t-\tau) * h(t-\tau)$. But it is not so! How is a student supposed to conclude that $y(t-2\tau) = x(t-\tau) * h(t-\tau)$? This sort of sloppy notation could easily undermine the students' confidence in mathematics.

Still, some later texts in the same field stick to defective conventions, often pointing out the errors but perpetuating them with the (fortunately false) "common practice" argument.

Gold [21] reports that many students do not know the difference between a function and an expression until some computer algebra system forces them to recognize it.

One of the consequences observed in my own students is the difficulty in understanding higher-order functions, for instance, the fact that $f(x)$ *can* be a function, e.g., if f is a higher-order function defined by $(f(x))(y) = x + y$ (all parentheses optional, but written for emphasis).

Similar misuse of notation is writing $x[n]$ instead of just x for a sequence. Even worse, as noted in [36], is writing $x[n]$ for the sampled version of a continuous signal written $x(t)$ in the same context. The mathematics literature does not do any better: many authors write a sequence x as x_j or even as $\{x_j\}$: one wonders how they would write the singleton set with x_j.

Reviewers and editors of mathematical texts let all kinds of junk notation pass routinely, although this attitude is fully equivalent to accepting pidgin in literary texts.

A few authors even suggest that sloppy formalism is in the interest of the students, preparing them for worse to come. This is a dangerous misconception. The best basis for students to learn to cope with the wide diversity of often defective formalisms is a flawless formalism. Not only does this provide a reliable reference frame but it also instills a sense of discrimination [22], [36]. Defects create a disparity between the image of mathematics as being very exact and the actual practice of condoning sloppy and misleading use of notation. No wonder many students feel that symbolism is primarily obfuscation or even that mathematics is yet another hoax!

Conclusion

As for many issues, education is a crucial element in bridging a methodological gap. Ample evidence has been given that learning to practice faithful formalization and true symbolism (not syncopation) should be an integral part of mathematics education. Most effective for dealing with logical arguments is the ability to calculate with predicates and quantifiers as fluently as is customary for derivatives and integrals.

Suitable study material can be found on the Web, and there is also the excellent textbook by Gries and Schneider [23]. One reason why [23] should be complemented by material from the Web is that it does not cover pointfree expression (in particular for quantifiers) and a systematic approach to bindings (although it contains no errors in this respect). Yet there are high expectations for a new edition that has been in preparation for some years.

A further complement is provided by the many texts taking the more traditional approach, such as [51], because they are a treasure trove of

[4] *It is significant that lambda calculus is little known in mathematics yet is basic knowledge in computer science.*

[5] *The pronoun "we" might be misleading; apparently the intended meaning is "some authors".*

$$
\begin{array}{l}
\text{Overlap}\,(s) \\
\quad \equiv \langle \text{Basic form} \rangle \quad \underline{\exists\, k,k' : \mathbb{Z}_{\neq}^{2} \,.\, \exists\, g : \mathbb{R} \,.}\; g \in R_k(s) \cap R_{k'}(s) \\
\quad \equiv \langle \text{Insert defs.} \rangle \quad \text{— Head calculation step; details:} \\
\qquad \left| \begin{array}{ll}
\equiv \langle \text{Definition } \cap \rangle & \triangleq\; g \in R_k(s) \wedge g \in R_{k'}(s) \\
\equiv \langle \text{Defin. } R_k(s) \rangle & \triangleq\; g \in \{f - ks \mid f : S\} \wedge g \in \{f - k's \mid f : S\} \\
\equiv \langle \text{Definition } \{\} \rangle & \triangleq\; (\exists\, f : S \,.\, g = f - ks) \wedge (\exists\, f' : S \,.\, g = f' - k's)
\end{array} \right. \\
\quad \equiv \langle \text{Rearrange } \exists \rangle \quad \text{— Head calculation step; details:} \\
\qquad \left| \begin{array}{ll}
\equiv \langle \text{Distr. } \wedge/\exists,\ 2\times \rangle & \triangleq\; \exists\, f : S \,.\, \exists\, f' : S \,.\, g = f - ks \wedge g = f' - k's \\
\equiv \langle \text{Swap } \exists \rangle & \underline{\exists\, f : S \,.\, \exists\, f' : S \,.\, \exists\, k,k' : \mathbb{Z}_{\neq}^{2} \,.}\; \exists\, g : \mathbb{R} \,.\, g = f - ks \wedge g = f' - k's
\end{array} \right. \\
\quad \equiv \langle \text{Eliminate } g \rangle \quad \text{— Head calculation step; details:} \\
\qquad \left| \begin{array}{ll}
\equiv \langle \text{Rearrange } = \rangle & \triangleq\; \exists\, g : \mathbb{R} \,.\, g = f - ks \wedge f - ks = f' - k's \\
\equiv \langle \text{Distrib. } \wedge/\exists \rangle & \triangleq\; (\exists\, g : \mathbb{R} \,.\, g = f - ks) \wedge f - ks = f' - k's \\
\equiv \langle \text{One-pt. rule } \exists \rangle & \triangleq\; f - ks \in \mathbb{R} \wedge f - ks = f' - k's \\
\equiv \langle f - ks \in \mathbb{R} \rangle & \triangleq\; f - ks = f' - k's
\end{array} \right. \\
\quad \equiv \langle \text{Definition } S \rangle \; \exists\, f : P \cup N \,.\, \exists\, f' : P \cup N \,.\, \exists\, k,k' : \mathbb{Z}_{\neq}^{2} \,.\, f - ks = f' - k's \qquad \text{(A)}
\end{array}
$$

Figure 14. Bandpass sampling: (a′) details of concretizing the support overlap formula.

interesting example problems that can improve any introductory course on constructing proofs. For the time being, such a combination of various sources is the best basis for helping students to develop a sense of discrimination and quality, together with proficiency in formulating problems and theories, in problem solving and in mathematical reasoning and discourse.

From a traditional perspective, these concerns might be considered university level, but thorough studies and contributions by others [3] have shown the desirability and feasibility of preparing symbolic reasoning already in high school.

Acknowledgement

The author wishes to thank Leslie Lamport for many useful suggestions.

Appendix: Organizing and Presenting Sizeable Calculations

For various reasons, large symbolic calculations are preferably neither elaborated nor presented in one stretch but are split into parts more convenient for both writer and reader. Obviously, this is common practice throughout the mathematics, physics, and engineering literature.

A few more remarks may be useful, using Figures 14 and 15 as examples. These figures directly illustrate how a larger calculation can be cut into parts at convenient places.

Even then, the calculations would not normally be presented with all details "in plain view". In a hypertext environment, only top level ⟨*justification*⟩s and resulting formulas would appear, to be expanded only when the reader wishes. In a regular text environment, the familiar forms of presentation have shown their value: splitting lower-level calculations as well into "chunks", interspersing textual comments, and highlighting the main intermediate results as numbered formulas. For the example of Figure 15, the textbook variant would be Figure 13.

References

[1] Vladimir I. Arnold, *On Teaching Mathematics*, 1997. http://pauli.uni-muenster.de/~munsteg/arnold.html

[2] Ralph Back, Jim Grundy, and Joakim von Wright, Structured calculational proof, *Formal Aspects of Computing* **9** (1997), 469–483. http://crest.abo.fi/publications/public/1997/StructuredCalculationalProofA.pdf

[3] Ralph-Johan Back and Joakim von Wright, *Mathematics with a Little Bit of Logic: Structured Derivations in High-School Mathematics*, Åbo Akademi University, 2006.

[4] Isabella G. Bashmakova and Galina S. Smirnova, The birth of literal algebra, *The American Mathematical Monthly* **106** (1999), no. 1, 57–66.

[5] ———, The literal algebra of Viète and Descartes, *The American Mathematical Monthly* **106** (1999), no. 3, 260–263.

[6] Richard Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, 1987.

[7] Jon Bentley, *Programming Pearls*, 2nd edition, Addison-Wesley, 2000.

[8] Raymond Boute, The Geometry of Bandpass Sampling: A Simple and Safe Approach", *IEEE Signal Processing Magazine 29*, **4**, pp. 90–96 (July 2012).

Given $P := [L, U]$, $N := [-U, -L]$, $s : \mathbb{R}_{>0}$, we simplify (A) as follows.

$\exists\, f : P \cup N \,.\, \exists\, f' : P \cup N \,.\, \exists\, k, k' : \mathbb{Z}_{\neq}^2 \,.\, f - ks = f' - k's'$ (A)

$\equiv$ ⟨Simplify inner $\exists$⟩

- $\equiv$ ⟨Arithmetic⟩ $\triangleq \exists\, k, k' : \mathbb{Z}_{\neq}^2 \,.\, f - f' = (k - k')s$
- $\equiv$ ⟨Change var⟩ $\triangleq \exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, f - f' = \ell s$
- $\equiv$ ⟨Shorthand⟩ $\underline{\textbf{let } Q_x \equiv \exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, x = \ell s \textbf{ in } \exists\, f : P \cup N \,.\, \exists\, f' : P \cup N \,.\, Q_{f-f'}}$

$\equiv$ ⟨Exploit symmetry⟩

- $\equiv$ ⟨Dom. split $\exists$⟩ $\triangleq (\exists\, f : P \,.\, \exists\, f' : P \cup N \,.\, Q_{f-f'}) \vee (\exists\, f : N \,.\, \exists\, f' : P \cup N \,.\, Q_{f-f'})$
- $\equiv$ ⟨Change var⟩ $\triangleq (\exists\, f : P \,.\, \exists\, f' : P \cup N \,.\, Q_{f-f'}) \vee (\exists\, f : P \,.\, \exists\, f' : N \cup P \,.\, Q_{f'-f})$
- $\equiv$ ⟨$Q_{-x} \equiv Q_x$⟩ $\triangleq \exists\, f : P \,.\, \exists\, f' : P \cup N \,.\, Q_{f-f'}$

$\equiv$ ⟨Dom. split $\exists$⟩ $\triangleq (\exists\, f : P \,.\, \exists\, f' : P \,.\, Q_{f-f'}) \vee (\exists\, f : P \,.\, \exists\, f' : N \,.\, Q_{f-f'})$ (B)

$\equiv$ ⟨Simplify 2nd term⟩ — Priority for what seems the more interesting term

$\lozenge\, \exists\, f : P \,.\, \exists\, f' : N \,.\, \exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, x = \ell s$

- $\equiv$ ⟨Arithmetic $\leq$⟩ $\exists\, f : P \,.\, \exists\, f' : N \,.\, \exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, f - f' = \ell s \wedge f - f' \in [2L, 2U]$
  - $\lozenge$ Loc: $L \leq f \leq U \wedge -U \leq f' \leq -L \Rightarrow 2L \leq f - f' \leq 2U$
- $\equiv$ ⟨Leibniz⟩ $\exists\, f : P \,.\, \exists\, f' : N \,.\, \exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, f - f' = \ell s \wedge \ell s \in [2L, 2U]$
- $\equiv$ ⟨Eliminate f'⟩
  - $\equiv$ ⟨Swap $\exists$⟩ $\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \exists\, f : P \,.\, \exists\, f' : N \,.\, f - f' = \ell s \wedge \ell s \in [2L, 2U]$
  - $\equiv$ ⟨Distrib. $\wedge/\exists$⟩ $\triangleq\triangleq \underline{\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [2L, 2U] \wedge} \ \exists\, f : P \,.\, \exists\, f' : N \,.\, f - f' = \ell s$
  - $\equiv$ ⟨Arithmetic⟩ $\triangleq \exists\, f : P \,.\, \exists\, f' : N \,.\, f' = f - \ell s$
  - $\equiv$ ⟨One-pt. rule $\exists$⟩ $\triangleq \exists\, f : P \,.\, f - \ell s \in N$
- $\equiv$ ⟨Eliminate f⟩ $\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [2L, 2U]$
  - $\lozenge$ Loc: $\forall\, v : [2L, 2U] \,.\, \exists\, f : [L, U] \,.\, f - v \in [-U, -L]$. Proof: witness $f := v/2$

⟨Simplify 2nd term⟩ $\triangleq (\exists\, f : P \,.\, \exists\, f' : P \,.\, Q_{f-f'}) \vee (\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [2L, 2U])$

$\equiv$ ⟨Similarly 1st term⟩ $(\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [-B, B]) \vee (\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [2L, 2U])$

$\equiv$ ⟨Prune negative ℓ⟩

- $\equiv$ ⟨Symmetry⟩ $(\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [0, B]) \vee (\exists\, \ell : \mathbb{Z}_{\neq 0} \,.\, \ell s \in [2L, 2U])$
- $\equiv$ ⟨$s \geq 0 \wedge \ell s \geq 0 \Rightarrow \ell \geq 0$⟩ $(\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [0, B]) \vee (\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [2L, 2U])$

$\equiv$ ⟨Absorb 1st term⟩ $\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [2L, 2U]$

$\lozenge$ Loc: $\underline{(\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [0, B])} \Rightarrow (\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [2L, 2U])$

- $\triangleq \Rightarrow$ ⟨Eliminate ℓ⟩ $s \leq B$
  - $\lozenge$ Loc: $(\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \leq B) \Rightarrow s \leq B$ (proof: $\forall\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \leq B \Rightarrow s \leq B/\ell \leq B$)
- $\Rightarrow$ ⟨Weaken: $B \leq 2B$⟩ $s \leq 2B$
- $\equiv$ ⟨Euclid IDT⟩ $0 \leq 2U - (2U \div s)s < s \wedge s \leq 2B$
- $\equiv$ ⟨Arithm. $\leq$⟩ $2L < (2U \div s)s \leq 2U \wedge s \leq 2B$
- $\Rightarrow$ ⟨$2U \div s \in \mathbb{N}_{\neq 0}$⟩ $\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, \ell s \in [2L, 2U]$

$\equiv$ ⟨Def. [ ], arithm.⟩ $\exists\, \ell : \mathbb{N}_{\neq 0} \,.\, 2L/\ell \leq s \leq 2U/\ell$ (C)

Figure 15. Bandpass sampling: (b′) stepwise simplification to obtain a practical formula.

[9] Raymond Boute, The Euclidean definition of the functions div and mod, *ACM TOPLAS* **14** (1992), no. 2, 127–144.

[10] ———, Functional declarative language design and predicate calculus: A practical approach, *ACM TOPLAS* **27** (2005), no. 5, 988–1047.

[11] Carl B. Boyer and Uta C. Merzbach, *A History of Mathematics*, second edition, Wiley, 1991.

[12] Ronald N. Bracewell, *The Fourier Transform and Its Applications*, second edition, McGraw-Hill, 1978.

[13] Peter J. Cameron, *Sets, Logic and Categories*, Springer, 1999.

[14] Jeffrey Clark, Derivative sign patterns, *The College Mathematics Journal* **42** (2011), no. 5, 379–383.
[15] Mary Joan Collison, The unique factorization theorem: From Euclid to Gauss, *Mathematics Magazine* **53** (1980), no. 2, 96–100.
[16] Edsger W. Dijkstra, How computing science created a new mathematical style, *EWD* **1073** (1990). http://www.cs.utexas.edu/users/EWD/ewd10xx/EWD1073.PDF
[17] Paulo J. S. G. Ferreira and Roeland Higgins, The establishment of sampling as a scientific principle, *Notices of the AMS* **58** (2011), no. 10, 1446–1450.
[18] Wim H. J. Feyen, Antoinetta J. M. Van Gasteren, David Gries, and Jayadev Misra, eds., *Beauty Is Our Business*, Springer, 1990.
[19] Mohan Ganesalingam, *The Language of Mathematics*, Ph.D. thesis, University of Cambridge, 2009. http://people.pwf.cam.ac.uk/mg262/GanesalingamMdis.pdf
[20] ———, *The Language of Mathematics*, presentation, 2010. http://www.srcf.ucam.org/principia/files/ganesalingam.pdf
[21] Bonnie Gold, How your philosophy of mathematics impacts your teaching, *The College Mathematics Journal* **42** (2011), no. 3, 174–182.
[22] David Gries, Improving the curriculum through the teaching of calculation and discrimination, *Communications of the ACM* **34** (1991), no. 3, 45–55.
[23] David Gries and Fred B. Schneider, *A Logical Approach to Discrete Math*, Springer, 1993.
[24] ———, Teaching math more effectively, through calculational proofs, *The American Mathematical Monthly* **102** (1995), no. 8, 691–697.
[25] Paul Halmos and Steven Givant, *Logic as Algebra*, volume 21 of The Dolciani Mathematical Expositions, The Math. Assoc. of America, 1998.
[26] John Harrison, Formal proof—theory and practice, *Notices of the AMS* **55** (2008), no. 11, 1395–1406.
[27] ———, *Handbook of Practical Logic and Automated Reasoning*, Cambridge, 2009.
[28] Israel N. Herstein, *Topics in Algebra*, Xerox College Publishing, 1964.
[29] Randall Holmes, *Elementary Set Theory with a Universal Set*, http://math.boisestate.edu/~holmes/holmes/head.pdf
[30] John L. Kelley, *General Topology*, Van Nostrand, 1955.
[31] Morris Kline, *Mathematics in Western Culture*, Oxford University Press, 1953.
[32] Steven G. Krantz, *Real Analysis and Foundations*, second ed., Chapman & Hall/CRC, 2005.
[33] Leslie Lamport, How to write a proof, *The American Mathematical Monthly* **102** (1995), no. 7, 600–608. http://www.jstor.org/pss/2974556, or the 1993 manuscript at http://research.microsoft.com/en-us/um/people/lamport/pubs/lamport-how-to-write.pdf
[34] ———, *Specifying Systems: The TLA$^+$ Language and Tools for Hardware and Software Engineers*, Pearson Education Inc., 2002.
[35] ———, *How to Write a 21st Century Proof*, technical note, Microsoft, November 2011.
[36] Edward A. Lee and Pravin Varaiya, Introducing signals and systems—the Berkeley approach, in *First Signal Processing Education Workshop*, Hunt, Texas, October 2000. http://ptolemy.eecs.berkeley.edu/publications/papers/00/spe1
[37] ———, *Structure and Interpretation of Signals and Systems*, Addison-Wesley, 2003.
[38] S. K. Lui, An interview with Vladimir Arnol'd, *Notices of the AMS* **44** (1997), no. 4, 432–438.
[39] Richard G. Lyons, *Understanding Digital Signal Processing*, Prentice Hall, 2004.
[40] John H. Mason, On the use and abuse of word problems for moving from arithmetic to algebra, in *The Future of the Teaching and Learning of Algebra, Proc. 12th ICMI Study Conference*, Helen Chick, Kaye Stacey, Jill Vincent, and John Vincent (eds.), December 2001, pp. 430–437.
[41] Reviel Netz and William Noel, *The Archimedes Codex*, Phoenix, London, 2008.
[42] Dag Prawitz, *Natural Deduction, a Proof-Theoretical Study*, Dover, 2006.
[43] Walter Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 1964.
[44] Lucio Russo, *The Forgotten Revolution*, Springer, 2004.
[45] Yoram Sagher, What Pythagoras could have done, *The American Mathematical Monthly* **95** (1988), no. 2, 117.
[46] Chris Sangwin, Modelling the journey from elementary word problems to mathematical research, *Notices of the AMS* **58** (2011), no. 10, 1436–1445.
[47] Patrick Suppes, *Axiomatic Set Theory*, Dover, 1972.
[48] Alfred Tarski and Steven Givant, *A Formalization of Set Theory without Variables*, AMS Colloquium Publications, volume 41, American Mathematical Society, Providence, Rhode Island, 1987.
[49] Paul Taylor, *Practical Foundations of Mathematics*, no. 59 in Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1999. http://www.cs.man.ac.uk/~pt/Practical_Foundations/html/s10.html.
[50] Antoinetta J. M. Van Gasteren, *On the Shape of Mathematical Arguments*, Springer Lecture Notes in Computer Science, no. 445, Springer, 1990. http://www.springerlink.com/content/978-3-540-52849-4
[51] Daniel J. Velleman, *How to Prove It: A Structured Approach*, second ed., Cambridge, 2006.