Journal of Humanistic Mathematics

Volume 13 | Issue 1 January 2023

Human-Machine Collaboration in the Teaching of Proof

Gila Hanna
University of Toronto

Brendan P. Larvor
University of Hertfordshire

Xiaoheng (Kitty) Yan
University of Toronto

Recommended Citation

Gila Hanna, Brendan P. Larvor & Xiaoheng (Kitty) Yan, "Human-Machine Collaboration in the Teaching of Proof," *Journal of Humanistic Mathematics*, Volume 13 Issue 1 (January 2023), pages 99-117. . Available at: https://scholarship.claremont.edu/jhm/vol13/iss1/7

Human-Machine Collaboration in the Teaching of Proof

Cover Page Footnote

Acknowledgment We are grateful to the anonymous referees for their helpful comments. We wish to acknowledge the generous support of the Social Sciences and Humanities Research Council of Canada.

Human-Machine Collaboration in the Teaching of Proof

Gila Hanna

Ontario Institute for Studies in Education, University of Toronto, CANADA
gila.hanna@utoronto.ca

Brendan P. Larvor

School of Social Sciences, Humanities & Education, University of Hertfordshire, UK
b.p.larvor@herts.ac.uk

Xiaoheng (Kitty) Yan

Ontario Institute for Studies in Education, University of Toronto, CANADA
xh.yan@utoronto.ca

Abstract

This paper argues that interactive theorem provers (ITPs) could play an important role in fostering students' appreciation and understanding of proof and of mathematics in general. It shows that the ITP *Lean* has three features that mitigate existing difficulties in teaching and learning mathematical proof. One is that it requires students to identify a proof strategy at the start. The second is that it gives students instant feedback while still allowing them to explore with maximum autonomy. The third is that elementary formal logic finds a natural place in the activity of creating proofs. The challenge in using Lean is that students have to learn its command language, in addition to mathematics course content and elementary logic.

Keywords: proof; logic; interactive theorem prover; Lean (theorem prover); undergraduate mathematics

1. Introduction

Proof is a central concept in mathematics and also an important tool for learning mathematics. Mathematicians and mathematic educators both have repeatedly emphasized the importance of teaching proof, yet research consistently demonstrates that students continue to have difficulty with it, starting in secondary school and continuing into post-secondary education. How best to teach students to understand and construct proofs remains one of the most important open questions in mathematics education.

1.1. Computer-assisted proofs

In the last few years there have been exciting developments in mathematical practice in the area of proof and verification, occasioned by the growing use of computers in mathematical research. A major outcome has been the acceptance of what might be referred to as computer-assisted theorem development, in which mathematicians make use of computer-based proof assistants. In their early stages, proof assistants were limited in function and used mainly to verify the correctness of existing proofs. The major spur to their use, in fact, was that mathematicians have long been aware that humans are not always successful at detecting errors in proofs [12, 29].

Today's proof assistants can do more than verification. Their expanded capabilities, notably the ability to work in interactive mode, allow mathematicians to take the next step and use them to create a new proof from scratch. To cite Avigad [3, page 684], "Interactive theorem proving involves the use of computational proof assistants to construct formal proofs of mathematical claims using the axioms and rules of a formal foundation that is implemented by the system."

The programme introduction to a 2017 workshop on computer-aided proof stated that:

> Proof technology can be used to perform large calculations reliably, solve systems of constraints, discover and visualise examples and counterexamples, simplify expressions, explore hypotheses, navigate large libraries of mathematical knowledge, capture abstractions and patterns of reasoning, and interactively construct proofs. The scale and sophistication of proof technology is approaching a point where it can effectively aid human mathematical creativity at all levels of expertise. (Computer-aided mathematical proof - Isaac Newton Institute, 2017; a six-week workshop, Big Proof.[1])

[1] https://sms.cam.ac.uk/collection/2520307

In looking at the history of the use of computers in mathematics, Dick [9] noted the various reactions that their use provoked:

> Practitioners of this field sought to program computers to prove mathematical theorems or to assist human users in doing so. Everyone working in the field agreed that computers had the potential to make novel contributions to the production of mathematical knowledge. They disagreed about almost everything else. Automated theorem-proving practitioners subscribed to complicated and conflicting visions of what ought to count and not count as a mathematical proof.

1.2. What is mathematical proof?

It's not surprising that people developing automatic theorem provers had complicated and conflicting views about mathematical proof, because the concept of proof was already complex and ambiguous before they started.[2] In deciding what we mean by 'proof', the simplest option is to identify mathematical proof with the objects of the branch of formal logic called 'proof theory'. However, these proofs are mathematical objects and not bound by material constraints. There could be a proof of a given theorem within some formal system, and we might know it exists (perhaps by a completeness result about the system), but the proof might have more steps in it than there are atoms in the solar system. If we are interested in proofs as the means for establishing and communicating mathematical knowledge, then such proofs are not relevant. Even when the formal proof of a theorem is short enough to write out, mathematicians rarely bother to set down all the details.

Perhaps, therefore, the texts that mathematicians publish in journals ought to be our paradigmatic proofs. After all, the standard for accepting a theorem is not that someone claims to have a proof, but rather that a proof has been published after competent peer review. This proposal has its own difficulties. Printed proofs are notoriously incomplete, and require expert readers who are able to fill in the gaps and perform the inferences to get from one stage to the next. Exactly how gappy can a text be before it ceases to be a proof? This matter was the core of a famous dispute in the early 1990s between Arthur Jaffe and Frank Quinn [18] on one hand and

[2] For an overview of the relevant literature, see [6].

William Thurston [27] on the other. Jaffe and Quinn accused Thurston of publishing results without fully worked out proofs. Part of Thurston's answer was that mathematical understanding is in the community of working mathematicians, and this is supported by conversations and informal notes as well as proofs published in journals. Effectively, Thurston seemed to say that a proof lies in the intersubjective common understanding of the relevant mathematical community. Here too, there are problems. A proof is valid or not independently of whether the mathematical community understands it or whether it has the approval of journal referees.

So, there are at least three unsatisfactory answers to 'what is a proof?' – it might be a mathematical object, a text, or a social object. Recent work in the philosophy of mathematical practice seeks to put these simplistic options in relation to each other (see [2, 4, 14, 19, 24, 25]). Tensions arise because the first answer takes us too far from practice and the other answers take us too far from logic. In short, we do not really know what a mathematical proof is. For example, there is a range of respectable views on whether diagrams have a proper role in mathematical proofs and if so what it might be (see [8, 13]). Perhaps it is, as Czocher and Weber [7] suggest, a cluster concept.

Part of what motivated Jaffe and Quinn, writing in 1993, was a worry that advances in computing and communication technology would change mathematics for the worse, because the gatekeeping and curating functions of journal editors would be undermined. Dick [9], writing a quarter century later, perceived a change in the character of mathematical activity:

> Automated theorem-proving practitioners took their visions of mathematicians, minds, computers, and proof, and built them right in to their theorem-proving programs. Their efforts did indeed precipitate transformations in the character of mathematical activity but in varied and often surprising ways. They crafted new formal and material tools and practices for wielding them that reshaped the work of proof. They also reimagined what "reasoning" itself might be and what logics capture or prescribe it.

Jaffe and Quinn were right to think that computers and the internet would change mathematics. In fact, if Dick is right, they underestimated the effect. However, they failed to anticipate the possibility that interactive theorem provers might solve the very problems that worried them.

They accused Thurston of not offering fully worked out proofs. Thurston replied that he didn't have time to work them out because he was so busy writing lecture notes to help the mathematical community understand his work, and this had to happen before he could publish because otherwise there would be no independent referees competent to review anything he might submit for publication. Much of the debate was about the poor incentives on offer for checking the rigour of difficult proofs in new areas. Why would someone capable of checking Thurston's work bother to do so when they could be making a career doing their own original research? How differently that dispute might have gone, if Thurston could have done what Peter Scholze did, that is, ask the interactive theorem proving community to check his work on a computer. In that case, the verification became an interesting piece of research in its own right.

Dick [9] is surely correct that computer proof assistants will change what we understand by 'proof' and 'mathematics' in unpredictable ways. To see how difficult it is to predict the effects of technological change, note that engineers put SMS (Short Message Service) capability into mobile phones in order to move small packets of data around. They never imagined that teenagers would use it to write to each other, relentlessly.[3] Returning to mathematical proof, here is one advocate of computer assisted proof, echoing Dick:

> But as new methods arise and fashions change, our views of mathematics will change as well. And so our current preoccupation with conceptual methods could well give way to a more expansive view of mathematical understanding, one in which the computer plays a more central role. This isn't to say that we will then abandon the conceptual viewpoint entirely [1, page 114] .

Avigad [1] characterizes 'the conceptual approach' thus:

> a focus on axiomatically described structures; the use of certain tools, such as limit and quotient constructions, for building new structures from old ones; and an emphasis on characterizing structures and their properties in relation to other structures,

[3] The future as imagined in popular culture (*Star Trek, Dr Who,* etc.) in the late twentieth century involved very little writing because it was supposed that computers would be voice-activated. As it turned out, e-mail, social media, and messaging have supplanted the telephone so completely that we may well be doing more writing than ever.

> for example, by studying the morphisms between them and by viewing the structures themselves as elements of even more abstract structures. [1, page 106]

He associates it with Noether, van der Waerden, Bourbaki and Grothendieck. As he says, this is the dominant outlook among professional mathematicians now. Perhaps Avigad is correct that computer assistants will enable other views of mathematics to flourish. On the other hand, the conceptual approach has been one of the main drivers towards ever longer and more difficult proofs. The problem of finding competent reviewers for such proofs has worsened in the three decades since Thurston [27] pointed it out. It's possible that automation may save the conceptual approach from stagnation by solving its most pressing limiting problem.[4]

2. Approaches to the teaching of proof

For future research mathematicians and future mathematics teachers, a good command of proof and proving is essential. But mathematics educators have long ascribed considerable importance to teaching proof to a broader audience as well, as an integral part of basic mathematical literacy. The USA's National Council of Teachers of Mathematics (NCTM), for example, has recommended that secondary-school **i**nstructional programs should enable every student to:

- Recognize reasoning and proof as fundamental aspects of mathematics
- Make and investigate mathematical conjectures
- Develop and evaluate mathematical arguments and proofs
- Select and use various types of reasoning and methods of proof [21]

[4] If popular music is any guide, the future of mathematics, shaped by technology, will be surprising and diverse. Autotune has gone from being a cheat to an aesthetic choice, hit recordings are made by artists who cannot play an instrument but know how to use digital technology to shape sound, and yet the biggest selling album at the time of writing (Adele's *30*) is a collection of ballads recorded without autotune, accompanied by an acoustic piano played by a human pianist.

NCTM did not refer explicitly to deductive reasoning in its recommendations, perhaps assuming that it is generally understood to be part of "reasoning and proof." Indeed, secondary school curricula do as a rule include instruction on how to make deductive inferences from premises accepted as true (often in the context of geometry or simple proofs by induction).

It is widely recognized, nevertheless, that many students entering university have difficulty constructing valid proofs. This is hardly surprising, given that the concept of mathematical proof is ambiguous between the logical, textual, and psycho-social aspects discussed above. Creating a new proof requires a subtle, iterated movement between an intuitive sense of what the theorem is about and the formal definitions and rules of inference that code those meanings rigorously. The art of shifting between more and less formal versions of the same mathematical thought is not easy for instructors to demonstrate, and is certainly not taught simply by presenting students with formal definitions and telling them that these are what university-level mathematics is about (see [22]).

2.1. The variety of current approaches

Mathematics educators have valued and embraced a great variety of classroom approaches to proof and proving, if for a moment we take these terms in the broadest sense of finding a convincing path from one or more mathematical statements to another. Several reports and two recent international studies on proof and proving, the nineteenth International Commission on Mathematical Instruction (ICMI) study [10, 22] and European Society for Research in Mathematics Education (ERME) study [20], provide ample evidence for the many different types of proof and ways of proving, again in the broadest sense, that have been found useful in mathematics education. These two comprehensive studies in particular also summarize the issues associated with the effective teaching of proof from the historical, epistemological, and pedagogical perspectives.

In addition to portraying the many forms of proof in use, they discuss the variety of proving activities, the role of logic in teaching proof, and the experimental approaches to proving (with particular mention of dynamic geometry software). Indeed, the ICMI study devoted an entire chapter to a description and analysis of many different approaches to proving, in the broad sense, that have been seen to help foster mathematical understanding (see [10]).

The ICMI and ERME studies employ, for greater clarity, the useful distinction between proof and argumentation, in which proof is associated with mathematical rigour (admittedly to varying degrees), while argumentation encompasses the many teaching approaches in which the student learns to construct a link that is convincing without necessarily being unassailable from a strict logical point of view. In a teaching situation, as in mathematical practice, these two modes are not mutually exclusive, of course, in the sense that an argument can often lead to a proof or make an existing proof more convincing.

In their chapter, Dreyfus, Nardi, and Leikin [10] discuss the entire gamut of reasoning modes and point out in particular the following:

- different representations, including visual, verbal and dynamic, that may be used in the course of proof production;
- different ways of arguing mathematically, such as inductive example-based arguments, example-based generic arguments and general arguments, as well as individually versus socially produced arguments;
- different degrees of rigour and of detail in proving – including different degrees of pointing out assumptions, whether in terms of first principles or previously proven statements – and where and how these are used;
- multiple proofs; that is, different proofs for the same mathematical statement, which may be used in parallel or sequentially, by a single person or a group. (page 191)

Here, we see some of the same ambiguities in the concept of proof already noted above. In their chapter on argumentation[5] and proof in the 2018 ERME report, the authors [20] acknowledge that both these modes of reasoning have been subject to heated debate, reflecting the wide variety of

[5] Many mathematics educators have found it very useful to keep Toulmin's model in mind when assessing various modes of argumentation [28]. Toulmin claimed that all plausible arguments necessarily follow the same pattern. He has offered a very general model that has three main elements: 1) A claim (C), the statement to be proved (or otherwise justified); 2) the data (D) used to justify the claim; 3) the warrant (W) the inference rule that connects the data to the claim. His model also allows for qualifying its elements as to their strength, and makes explicit the role of rebuttal.

viewpoints held by the conference participants and mathematics educators in general. Though the heuristic methods characteristic of argumentation, such as visual demonstrations, have often proved useful in the classroom, many educators have not accepted them as valid methods for teaching proof or as replacements for it. The primary controversy is over the epistemological gap between heuristics and proof, with proof being seen as the only route to sure knowledge. Many educators do not want to forgo teaching proof, which they see as an essential characteristic of mathematics. But then, anyone who insists that all students should gain a practical grasp of the concept of proof is under pressure to explain how it is to be taught.

Another controversial issue discussed at these two conferences, without reaching consensus, is the explicit teaching of rules of logic in argumentation and proof. We list the main points that emerged, because they are particularly relevant to the use of ITPs, the focus of this paper:

- identification, in the relationship between logic and language, of aspects that are likely to be an obstacle for developing proof and proving skills, and of aspects that are likely to favour it;
- the value of teaching logic for fostering proof and proving competencies;
- the usefulness for teachers of logical analysis in mathematical discourse, and how to do it; and
- the relationships between logic and formalization [20, page 81].

The ICMI study included an entire chapter on the role of logic in teaching proof [11]. In this context the word "logic" refers only to the basic operators of propositional logic ("and", "or", "not", "if-then") and of predicate logic ("for all", "there exists"). One question discussed was the extent to which such basic principles of logic should be taught in the context of teaching proof, as being reflective of mathematical practice. The other question was whether explicit instruction in logic would foster students' competence in proving. The authors of the chapter concluded by stating that the principles of logic are an essential aspect of proving, and by recommending that instructors make explicit to students the logical aspects of a proof for didactic reasons as well. On the other hand, even the most elementary formal logic is additional curriculum content if it is to be taught explicitly.

Teachers are unlikely to welcome it unless students can find it in the practical activity of creating and assessing their own proofs. If it is not a spontaneous part of the activity of proving, the danger is that teaching formal logic in the context of mathematical proof will leave students trying to learn two topics at once (logic and whatever the proof is about) [15, 23] .

In the next section of this paper, we argue that it is desirable to adopt interactive theorem provers as an additional instructional tool in undergraduate mathematics teaching (see [17]). We show that such proof technology, once difficult to exploit because it required specialized computer skills, has become realistically accessible to new learners. In particular, we focus on undergraduate mathematics and consider ways in which a specific interactive theorem prover, Lean, can help students in their reasoning, connect formal logic with proof-making in other areas of mathematics and enhance their ability to construct a valid proof. In addition, we look at the challenges that the use of Lean ITP might present to the teaching of proof and assess the prerequisites to success.

2.2. An additional approach: the use of the ITP "Lean"

Lean can guide students in constructing a proof in several ways.

First of all, it forces students to come up with an overall strategy for the proof. Students would need such a proof plan with or without Lean, of course, but Lean confronts them with that fact at the earliest stage.

Secondly, Lean immediately examines each step that students type in and tells them whether it is valid (syntactically correct) or not. It is able to do this because students are required to frame their arguments as a sequence of simple steps expressed in the Lean language. The process is similar to that of computer coding, a task many students are familiar with. One of the problems that educators are trying to solve with technology is to give instant feedback without compromising student autonomy.[6] Lean allows students to

[6] It's easy to create exercises that students can do in their own time and get instant feedback: multiple-choice quizzes do that and they are easy to set up online now. However, such quizzes are teacher-led. The teacher chooses the questions and frames the answers, which deprives the students of almost all their autonomy. It would be better for students to get instant feedback on work that they have chosen and directed. For example, in modern languages, gamified learning apps such as Duolingo provide instant feedback, but only on sentences that it supplies, whereas a word processor can tell you instantly whether

work with high levels of autonomy; they could in principle decide what to prove and in any case decide on strategy and choice of steps, yet Lean will tell them straight away whether their latest step is correct. Thus, students have instant running feedback with high levels of autonomy.

Thirdly, Lean tells students whether the steps they enter do or do not move them toward their final goal (the proposition to be proved), and why. To enable this impressive capability, the students will have earlier entered Lean meta-commands known as "tactics" that give Lean information on the path they propose to follow in building their arguments. As the Lean manual puts it, these tactics "… naturally support an incremental style of writing proofs, in which users decompose a proof and work on goals one step at a time."

Lastly, the iterative proof construction enabled by Lean gives students greater comfort and confidence in the very process of proving. Far from operating as a forbidding black box, Lean reveals itself to students as a transparent assistant that responds to human input with immediate and explanatory feedback.

The following examples give a flavour of what it is like to use Lean.

2.3. Proofs using Lean

The Lean interface has a *Tactic state* that changes interactively as one proceeds with a proof.[7] In other words, the *Tactic state* communicates with students about the goals of proof to indicate where the students are at and where they are going in the proving process. Tactics are commands that Lean understands at any stage of a proof and that instruct Lean how to build a proof. A tactic operates on a proof-goal by either proving it or creating new sub-goals. Lean performs tactics that the user enters, and subsequently modifies the goals. A *Tactic state* will track the open sub-goals and will stop changing sub-goals only if Lean considers a proof as complete by showing "goal accomplished" with a party popper emoji. Using a sequence of tactics enables proving in a way that is quite similar to conventional paper-and-pencil reasoning, while automating parts of a proof. As Buzzard [5] put it, Lean's tactic state can often be easily understood without any specialist

---

you've made any gross grammatical errors regardless of what you choose to write about. Lean offers gamification and instant feedback.

[7] Lean has also ways of writing proofs in "term-style," which we will not discuss.

knowledge of ITPs, because the notation used is close to standard mathematical notation.

The changes in the *Tactic state* result from the *tactics* used in Lean proofs. For example, the tactic "rw" ("rw" is an abbreviation of "rewrite") tells Lean to do a replacement. Therefore, given a hypothesis "h" that is a proof of "A = B", the tactic "rw h", tells Lean to replace all As with Bs.

The following are two examples of proofs that feature some representative Lean tactics, such as "intro h" which introduces a hypothesis or a variable and "cases" which decomposes a conjunction or disjunction.

Example 1: Proof using the commutativity law of multiplication

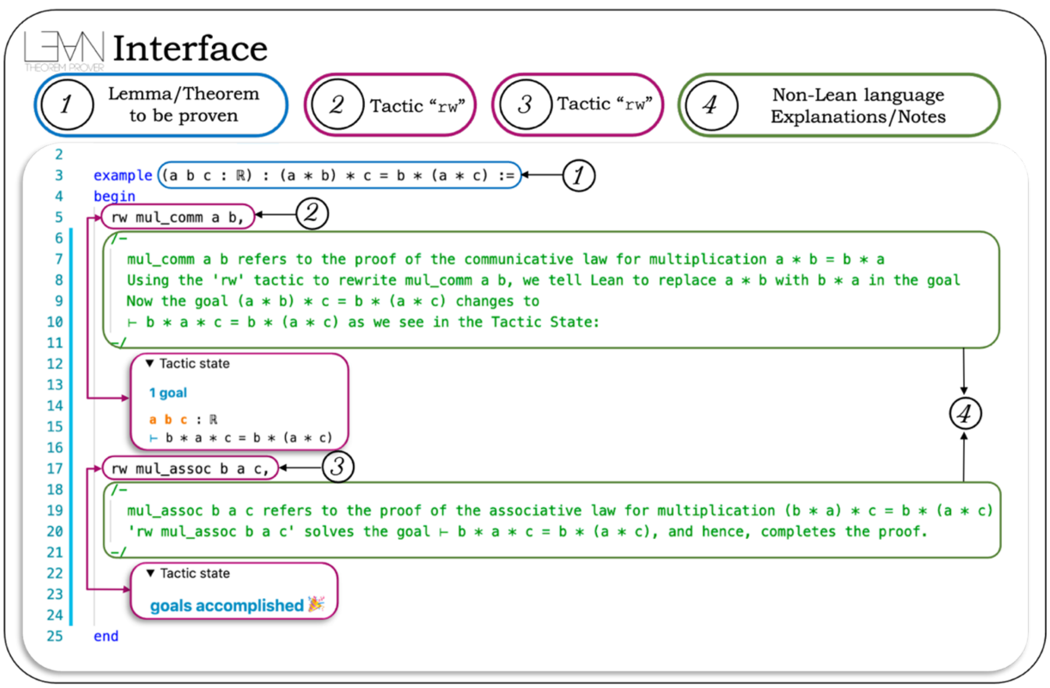

Figure 1: A Lean proof of $(a \times b) \times c = b \times (a \times c)$.

In Figure 1, to prove $(a \times b) \times c = b \times (a \times c)$ where a, b, and c are real numbers, the "rw" tactic itself suffices to solve the goal. The proof begins with rw mul_comm a b, where the left-hand side of the commutativity law for multiplication $a \times b = b \times a$ is replaced with the right-hand side $b \times a$.

Subsequently, the goal in the *Tactic state* changes from $(a \times b) \times c = b \times (a \times c)$ to $(b \times a) \times c = b \times (a \times c)$. Now using the "rw" tactic again with the associative law for multiplication with b, a, and c - rw mul_assoc b a c, it closes the goal.

As shown in Figure 1, it is worth noting that one can insert non-Lean language between the tactics to communicate with the reader about the proof. If one types "/-", Lean responds "-/" to it and recognizes that one wishes to add some text that Lean can ignore. This space is particularly useful to instructors who want to add explanations or share notes with students.

Before diving into another example of a Lean proof, we shall first take a close look at two commonly used tactics: intro(s) and cases. The intro tactic introduces a hypothesis or a variable that is a member of a set. A plural form of intro, intros, introduces a set of variables or hypotheses. For example, If P and Q are sets, intros p q means "let p be an arbitrary element of P and let q be an arbitrary element of Q." If P and Q are propositions, then intro p says "assume P is true" and turns a goal P → Q into a hypothesis p : P and goal Q.

The cases tactic decomposes a conjunction $P \wedge Q$ or disjunction $P \vee Q$ by changing one main goal to two goals. For example, to prove P Q, one needs to give either a proof of P *or* a proof of Q, so if h : P Q then cases h with p q will change one goal into two, one with p : P and the other with q : Q.

Example 2 Proof of ¬¬P → P

Figure 2 shows a Lean proof of ¬¬P → P. The proof in Lean starts with the intro tactic, introducing hypothesis h : ¬¬P to indicate that ¬¬P is assumed for conditional reasoning. Since ¬¬P is assumed, the proof proceeds with the goal "P is true." Taking into account classical.em P ("em" stands for excluded middle) that for any proposition P, P ¬P is true, using the cases tactic classical.em P efficiently splits the disjunction into two cases: h_1 : P and h_1 : ¬P. Now we use the assumption tactic to tell Lean to search hypotheses that will close the current goals. Lean finds h_1 : P successfully and closes the first case P. For the remaining case, it is worth noting that h : ¬¬P and h_1 : ¬P coexist. Now we use the contradiction tactic to tell Lean to search contradictory hypotheses to close the second goal.

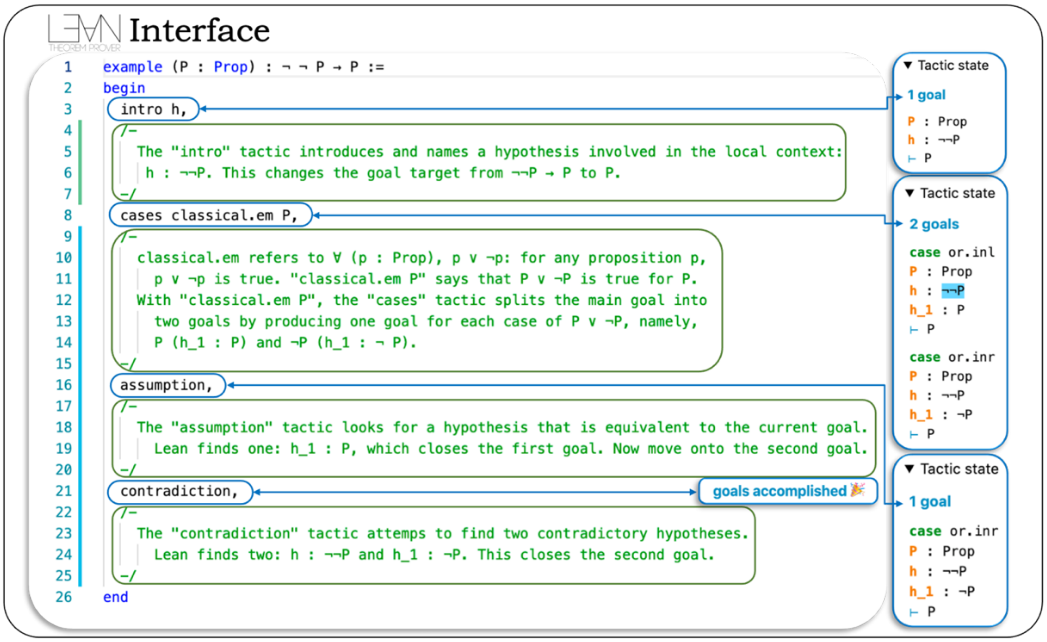

Figure 2: One way to prove $\neg\neg P \rightarrow P$ in Lean.

Clearly, students could easily do these two proofs without the help of an interactive theorem prover. However, we have shown these proofs to indicate that Lean can provide proofs in collaboration with a student, to illustrate how it is done, and how it allows keeping track of what is fixed and what is moving. As in the second proof, principles of basic formal logic appear as tactics such as assumption or contradiction. Rather than watching a teacher pointing to a proof on a blackboard and saying "by the way, this is an example of conditional reasoning," the students find logic plumbed in to their proving activity. In fact, Lean allows students to apply basic formal logic to a proof at hand in a creative manner. Surprisingly enough, it is often the case that multiple solutions to the same theorem can be produced in Lean.

Figure 3 shows a different Lean proof of the double negation, in which the "by_contra" tactic makes the method of proof by contradiction explicit, and brings this version of Lean proof closer to a paper-and-pencil proof. Whether this proof is introduced by an instructor or, better yet, produced by a student, a discussion about different Lean proofs and distinctions between Lean proofs and standard written proofs would arise organically.

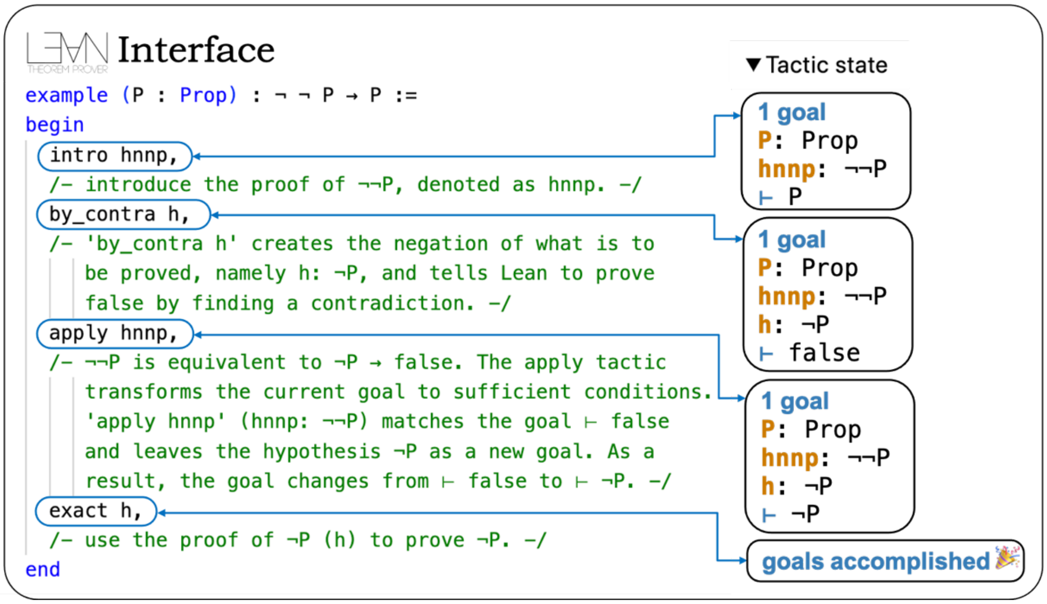

Figure 3: Another approach to the proof of $\neg\neg P \rightarrow P$ in Lean.

As far as we know, Lean has been used to advantage in a few undergraduate classes in the US and in Europe. Nevertheless, we concede that it is *not easy* to learn to use Lean proficiently. Currently the learning curve is somewhat steep.

There is not enough evidence yet to know whether the benefits of using Lean in the classroom outweigh the challenges. Hanna and Yan [16] asked nine professors who had chosen to use Lean in their teaching for their insights and suggestions. All responded that Lean was helpful in their teaching, made the course material less dry, and benefited all the students. All of them plan to continue using Lean despite its challenges, such as having to learn the necessarily rigid Lean syntax. The only systematic evaluation of which we are aware is the study carried out by Thoma and Iannone [26] at a single university; the authors offered this tentative evaluation of the benefits of Lean for the undergraduate curriculum:

> …it may help students with developing proving habits which are conducive to successful proofs, it may introduce a programming aspect to modules often taught very traditionally and it may help bridging the gap between the way in which mathematics is taught and the way in which modern mathematics evolves by allowing

students to become familiar with some of the tools used in this discipline.

We hope that before long, Lean will be made more user friendly.

3. Conclusion

In this paper we have argued that the concept of mathematical proof is ambiguous and contested. Therefore, it is not entirely clear what it means to say that mathematics students should learn to construct their own proofs. For example, it is not obvious what level of formality is appropriate at a given educational level. This is in part because the teaching of proof has multiple educational goals: to support learning of mathematical content, to teach the specific nature of mathematical reasoning, and to help students become better at reasoning on any subject, mathematical or not.

Nevertheless, a proof, however formal or informal, has a logical structure. Learning to discern and exploit the logical structures of proofs supports all the educational goals associated with learning to construct proofs. This suggests that training in elementary formal logic might help students learn to construct proofs, but such training introduces its own costs and challenges. Students' previous experience has been with informal proof construction, using argumentation and heuristics. Instructors will need to learn how to help students see the connection between this previous experience and the new more formal way of proving using Lean. In particular, instructors will need to bring students to understand that the use of symbols and tactics is not a thoughtless mechanical procedure and that thinking with symbols is no less reflective than thinking with argumentation and heuristics.

The Lean system has three features that mitigate existing difficulties in teaching and learning mathematical proof. One is that it requires students to identify a proof strategy at the start. The second is that it gives students instant feedback while still allowing them to explore with maximum autonomy. The third is that students formulate their instructions to Lean using the notation of elementary formal logic, so that their knowledge of logic finds a natural place in the activity of creating proofs. The cost is that students have to learn the system's command language, in addition to their mathematics course content and elementary logic.

We can reasonably expect two things. One is that the Lean interface will improve and become more intuitive for anyone familiar with ordinary mathematical notations and terminology. The other is that automated and interactive proof assistants will become more widely used in mathematical research and teaching, and will transform practice. What and how deep those changes will be can only be guessed at. In the meantime, there is scope for a study on the ways in which Lean can be made more helpful in teaching.

References

[1] Jeremy Avigad, "Varieties of mathematical understanding," *Bulletin of the American Mathematical Society*, Volume **59** Number 1 (2022), pages 99–117.

[2] Jeremy Avigad, "Reliability of mathematical inference," *Synthese*, Volume **198** (2021), pages 7377–7399. doi:10.1007/s11229-019-02524-y

[3] Jeremy Avigad, "The mechanization of mathematics," *Notices of the American Mathematical Society*, Volume **65** Number 6 (2018), pages 681–690.

[4] Jody Azzouni, "Why do informal proofs conform to formal norms?" *Foundations of Science*, Volume **14** Number 1–2 (2009), pages 9–26.

[5] Kevin Buzzard, "Proving theorems with computers," *Notices of the American Mathematical Society*, Volume **67** Number 11 (2020), pages 1791–1799.

[6] Todd CadwalladerOlsker, "What Do We Mean by Mathematical Proof?," *Journal of Humanistic Mathematics*, Volume **1** Issue 1 (January 2011), pages 33–60. doi:10.5642/jhummath.201101.04

[7] Jennifer Czocher and Keith Weber, "Proof as a cluster category," *Journal for Research in Mathematics Education*, Volume **51** Number 1 (2020), pages 50–74.

[8] Silvia De Toffoli, "Chasing the diagram – the use of visualizations in algebraic reasoning," *The Review of Symbolic Logic*, Volume **10** Number 1 (2017), pages 158–186.

[9] Stephanie Dick, "Making up minds: Computing and proof in the postwar United States," colloquium talk, 2019. Abstract available at `https://www.chstm.org/events/colloquium-stephanie-dick`, last accessed on January 31, 2023.

[10] Tommy Dreyfus, Elena Nardi, and Roza Leikin, "Forms of proof and proving," pages 191–214 in *Proof and Proving in Mathematics Education: The 19th ICMI Study*, edited by G. Hanna & M. deVilliers (Springer, 2012).

[11] Viviane Durand-Guerrier, Paolo Boero, Nadia Douek, Susanna Epp, and Denis Tanguay, "Examining the role of logic in teaching proof," pages 369–389 in *Proof and Proving in Mathematics Education: The 19th ICMI Study*, edited by G. Hanna & M. deVilliers (Springer, 2012).

[12] Mohan Ganesalingam and Timothy Gowers, "A fully automatic theorem prover with human-style output," *Journal of Automated Reasoning*, Volume **58** Number 2 (2017), pages 253–291.

[13] Valeria Giardino, "L'imagination manipulatoire en mathématique," *Bulletin d'Analyse Phénoménologique* [En lingne], Volume **13** Number 2 (2017), pages 252–268.

[14] Yacin Hamami, "Mathematical rigor and proof," *The Review of Symbolic Logic*, Volume **15** Issue 2 (June 2022), pages 409–449.

[15] Gila Hanna and Michael de Villiers, "ICMI Study 19: Proof and proving in mathematics education," *ZDM: the international journal on mathematics education*, Volume **40** Number 2 (2008), pages 329–336.

[16] Gila Hanna and Xiaoheng Yan, "Teaching with computer-based proof assistants: Perspectives from instructors of mathematics," *Canadian Mathematics Society Education Notes*, 2022. Available at https://notes.math.ca/en/article/teaching-with-computer-based-proof-assistants-perspectives-from-instructors-of-mathematics/, last accessed on January 31, 2023.

[17] Gila Hanna and Xiaoheng Yan, "Opening a discussion on teaching proof with automated theorem provers," *For the Learning of Mathematics*, Volume **41** Number 3 (2021), pages 42–46.

[18] Arthur Jaffe and Frank Quinn, "Theoretical mathematics: Toward a cultural synthesis of mathematics and theoretical physics," *Bulletin of the American Mathematical Society*, Volume **29** Number 1 (1993), pages 1–13.

[19] Brendan Larvor, "How to think about informal proofs," *Synthese*, Volume **187** Number 2 (2012), pages 715–730.

[20] Maria Alessandra Mariotti, Viviane Durand-Guerrier, and Gabriel Stylianides, "Argumentation and proof," pages 75–89 in *Developing research in mathematics education: twenty years of communication, cooperation, and collaboration in Europe*, edited by T. Dreyfus, M. Artigue, D. Potari, S. Prediger, & K. Ruthven (Routledge, 2018).

[21] National Council for Teachers of Mathematics (NCTM), *Principles and Standards for School Mathematics*, NCTM, Reston VA, 2000. Available at https://www.nctm.org/Standards-and-Positions/Principles-and-Standards/Process/, last accessed on January 31, 2023.

[22] Annie Selden, "Transitions and proof and proving at tertiary level," pages 391–422 in *Proof and Proving in Mathematics Education: The 19th ICMI Study*, edited by G. Hanna & M. deVilliers (Springer, 2012).

[23] John Selden and Annie Selden, "Unpacking the logic of mathematical statements," *Educational Studies in Mathematics*, Volume **29** Number 2 (1995), pages 123–151.

[24] Fenner Tanswell, "A problem with the dependence of informal proofs on formal proofs," *Philosophia Mathematica* (III), Volume **23** Number 3 (2015), pages 295–310.

[25] Oliver Tatton-Brown, "Rigour and proof," *The Review of Symbolic Logic*, *First View*, 2020, pages 1-29. doi:10.1017/S1755020320000398

[26] Athina Thoma and Paola Iannone, "Learning about proof with the theorem prover LEAN: the abundant numbers task," *International Journal of Research in Undergraduate Mathematics Education*, Volume **8** (2022), pages 64–93. doi:10.1007/s40753-021-00140-1

[27] William Thurston, "On proof and progress in mathematics," *Bulletin of the American Mathematical Society*, Volume **30** Number 2 (1994), pages 161–177.

[28] Stephen Toulmin, *The Uses of Argument*, 8th edition, Cambridge University Press, 2008.

[29] Vladimir Voevodsky, "The origins and motivations of univalent foundations," *IAS, The Institute Letter*, 2014, pages 8–9. Available at https://www.ias.edu/sites/default/files/documents/publications/ILsummer14.pdf, last accessed on January 31, 2023.