What's Special about Mathematical Proofs?

Remarks for the Williams Symposium on Proof

University of Pennsylvania, Nov. 9, 2012

Solomon Feferman

There are some great math and science cartoons by Sidney Harris that you can find on the web. Three of my favorites feature two professors, one bearded and the other with glasses, standing by a blackboard filled with long, complicated mathematical formulas. In the first of these cartoons, the bearded professor is pointing to a break in the formulas where it is written, "Here a miracle occurs!" as he says, "I think you should be more explicit in step two." In the second one, the same professor is pointing to the very beginning of the list of formulas and saying, "Here's where you made your mistake." In the third cartoon, now the bearded one is landing a haymaker on the chin of the other professor, knocking off his glasses while saying, "You want proof? I'll show you proof!" Those illustrate some of the extremes of what can go wrong with proofs, even in mathematics.

But I imagine most people consider mathematics to provide the paradigm of what constitutes a proof, to which all other subjects aspire. For, what's more certain and permanent? We can still follow Euclid's proof of Pythagoras' theorem 2400 years later and be led step by step from the assumptions to the incontrovertible conclusion. Actually, it took 2200 of those years for mathematicians to realize that not everything in Euclid's development of geometry was justified by his axioms, but instead made use of what appeared to be evident properties of the diagrams accompanying his proofs. And though mathematics nowadays has arrived at an unprecedented general degree of rigor and reliability, problems with the validity of proofs can arise even at the most advanced parts of the subject. What I want mainly to do here is explain what makes mathematical proofs so special and why the ideal can be achieved when it is achieved, and I'll be doing this from the perspective of mathematical logic. But I want, first, to say some words about the very different view of our subject from the perspective of practice.

Presentations of proofs take place in a variety of social contexts such as in the classroom, or with one or a few colleagues in an office, or in a lecture for a seminar, colloquium or a meeting; standards vary accordingly. It also takes place more permanently in writing and subject to greater scrutiny in a publication in a journal or in a book. Journal publications vary in length from a few pages to hundreds of pages in possibly several parts. The longest proof thus far, the so-called classification of finite simple groups runs over 10,000 pages, scattered over hundreds of papers written by hundreds of mathematicians. In 2005, one of those mathematicians, Michael Aschbacher, wrote that:

> By consensus of the community of group theorists, the Classification has been accepted as a theorem for roughly 25 years, despite the fact that, for at least part of that period, gaps in the proof were known to exist. At this point in time, all known gaps have been filled.

One of those gaps alone took two volumes and 1200 pages to fill. Who is to say whether there may not still be some unidentified gaps? The book by Steven Krantz, *The Proof is in the Pudding. The changing nature of mathematical proof*, written for a general audience, has a number of other interesting examples of famous problematic proofs.

At the everyday publication level, articles in journals are supposed to be refereed, and some referees are very conscientious and check every detail. Other referees may just skim, relying on their general experience of what works in the specific subject at hand and/or the reputation of the author. Errors may come out later when other mathematicians try to use the claimed results. As Krantz writes,

> It should be stressed that all mathematicians make mistakes. Any first-rate mathematician is going to take some risks, work on some hard problems, shoot for the moon. In doing so, this person may become convinced that he/she has solved a major problem. And thus mistakes get made. Almost any good mathematician has published a paper with mistakes in it…And the referee did not catch the mistake either, so it must have been a pretty good mistake.

Then there is the question who can be called on to referee. To begin with, the field breaks down into between five and ten general subfields, depending on how you count: geometry, algebra, analysis and so on are some of the usual ones. But the Mathematical Classification Scheme (MCS) used by the two main review journals has a three-level break down. Its top level has 64 distinct categories; at the second level, each of those is further subdivided into something like 8-20 subspecialties, and on the third level, we get to the fine nitty-gritty of areas of work. Altogether, this works out to over 5000 third-level classifications. I count myself as generally knowledgeable across the board in the top level of my own field, Mathematical Logic and the Foundations of Mathematics, but I would not be able to referee anything at the cutting edge of research outside of my specialty of Proof Theory and Constructive Mathematics at the next level and even much of that would take a real effort. All of this points to the exceptional degree of specialization in mathematics and why confirmation of the correctness of proofs is usually limited to a small number of experts.

Still another problem is raised in recent years by the massive use of computers in several proofs. Two famous examples are the verification of the 4 Color conjecture, according to which 4 colors suffice for any map, and Kepler's sphere packing conjecture, according to which, roughly speaking, the most economical way of stacking spheres of the same size is in a pyramid, the way a grocer stacks his oranges. Thomas Hales submitted his 200 page proof of the Kepler conjecture, plus computer program, to the prestigious *Annals of Mathematics* in 1998. About this, Steven Krantz writes that:

> The refereeing process was so protracted and arduous that there was some attrition among the referees: some quit and others retired or died. At the end they said that they were able to check the mathematical part of the paper but it was impossible to check the computer work…[but] they were 99% sure it was right.

It took seven years for Hales' paper to be finally accepted by the *Annals*, and that was only in an outline form; it took another five years for the full details to be published in a series of papers elsewhere.

Despite this fragmentation and these problems, we experience mathematics as a distinctive unity, and there is substantial agreement in the mathematical community as to what sort of thing counts as mathematical knowledge and—on the other side—what can be dismissed out of hand. I think we need to turn to the theoretical account of mathematics offered by mathematical logic to explain how and why this is so. Curiously, even though the current standards of rigor in mathematics require closely reasoned arguments, most mathematicians make no explicit reference to the role of logic in their proofs, and few of them have studied logic in any systematic way. That just shows that the study of logic is not necessary for the ability to use it correctly, no more than the study of linguistics is necessary to speak one's language correctly, nor of physiology to digest our food properly, and so on. Nevertheless, I think we have to look at the logical analysis of mathematics to see what's special about mathematical proofs. I beg your patience while I go into some of the details of this picture.

That analysis breaks into two parts, first concerning the vocabulary of mathematics and then the principles of reasoning that are formulated in that vocabulary. The vocabulary itself breaks into two parts: the first deals with *mathematical objects* of various kinds and is thus *subject specific*, while the second is the *logical part* of the vocabulary and is *subject independent*. Mathematical objects are conceived to be abstract and immutable, with no location in space or time (though mathematical theories are of course applied to concrete objects and forces that *do* exist in space and time). Among the basic kinds of objects we would count the integers, the real numbers (i.e. measurement numbers), functions and sets. Also in the subject specific side of our vocabulary we have such simple primitive notions as that of equality and order, of sum and product of two numbers, of application of a function to an object, and of membership of an object in a set. These are taken to be undefined, though explanations of them are given informally. All further mathematical notions are supposed to be defined in terms of the basic ones, and for that purpose we need only use the logical part of our vocabulary. That is simply given by the words "not", "and", "or", "if…then…" (or "implies"), "if and only if", and "all" (or "every") and "some" ("or there exists"). The operations "all" and "some" are said to be the *universal* and *existential quantifiers*, because they tell how many elements of the domain have a given property. What I have just described is called the *language of*

first order logic (FOL). (There are languages for what are called second order and higher order logics, but we don't need to consider them here.) It is a remarkable empirical fact that every concept that has come to be accepted as a precise mathematical notion can be defined within FOL given the primitive notions at the base. It is the use of the universal and existential quantifiers that has proved to be essential for this representation of mathematical concepts. Some simple examples are those of being a prime number, or an irrational real number, or a continuous function, or an infinite set, and so on. I don't have any idea how many distinct precisely defined mathematical notions there are in use, but if we take our three-level MCS as a basis, with some 5000 or so terminal nodes, and if each of those sub-sub-specialties makes use, say, of a hundred basic definitions, we are talking of a half-million or so mathematical notions in use. (Of course that's not very much when people are throwing around a billion-this and a billion-that, a trillion-this and a trillion-that.)

Up until the mid-19th century, logic was largely the province of philosophers, building on the work of Aristotle. But their language and principles of reasoning were completely inadequate to represent Euclidean geometry, let alone all further mathematical developments. Mathematicians entered the picture in the mid-19th century, beginning with the work of George Boole, but the language of FOL did not emerge until early in the 20th century. The great mathematician David Hilbert took an increasing interest in logic and the foundations of mathematics, and in the 1920s he isolated a simple basic system of reasoning for *classical* FOL. The word "classical" here is used for logics in which every statement is regarded as true or false; there is no in-between. And that is reflected in what is called the *Law of Excluded Middle*, i.e. that A or not-A holds for each statement A. Classical FOL is a system of pure logic, applicable to all domains of discourse with any given primitive notions. In 1928, Kurt Gödel showed that Hilbert's system for this logic is *complete* in the sense that if A is valid in every domain of discourse and for every choice of the primitive notions, then A is provable from the logical axioms by closing under the rules of inference, and vice versa. Classical FOL is just one among a plethora of logics that have been considered in modern times, but none of them has as satisfactory a completeness theorem (if it has one at all), and most of them are not relevant to mathematics.

Coming back to Gödel's completeness theorem, it readily extends to particular mathematical axiom systems, like those for geometry, number theory, analysis and set theory in the following way. Suppose S is any system of axioms—such as one of these—for given basic mathematical notions; by a *proof from S* is meant any sequence of statements, each of which is either an axiom of S or an axiom of classical FOL or follows from earlier statements by one of the rules of inference of that logic. Then a statement A is said to be *provable from S* (or to be a *theorem of S*) if there is a proof from S that ends with A. Gödel's completeness theorem in general tells us that if A is true in every interpretation of the vocabulary of S that makes each axiom of S true, then A is provable from S, and vice versa. Note well, though, that this notion of proof is a *relative* one, not an *absolute* one as mathematicians think of their work in practice. What the proof demonstrates is not necessarily true in some absolute sense, but only true in any world satisfying the underlying axioms. Which of those axioms ought to be accepted then becomes a separate issue. The mathematical study, initiated by Hilbert, of what exactly can and can't be proved from given axiom systems is called *metamathematics.*

It is again a remarkable empirical fact that every proof found in mathematical practice can be formalized, that is it can be represented as a formal proof from a suitable system of axioms for that part of practice. That is a more controversial claim, because it is not immediately obvious how to do this for many proofs. To begin with, the underlying axioms are not usually mentioned and there is in general a long chain from such axioms through a large background of previously established theorems that is taken for granted, all of whose proofs must be formalized before the given one is treated. Secondly, simple logical steps are usually omitted, and the formal version has to supply these. Also the proof may contain phrases like: "it is obvious that", or "it may be seen that", or "one argues by symmetry", or "the other cases may be treated in a similar way", and so on. In some proofs there are large jumps that are only readily filled by experts. Finally, the proof may make essential use of diagrams that are not immediately converted into symbolic form. Nevertheless, there is a comprehensive body of work, much of it that has been carried out using computers, that verifies the formalizability of substantial,

paradigmatic proofs in practice from appropriate background axiom systems based on classical FOL.[1]

What can we say about the role of logic and proofs outside of mathematics? Looking, to begin with, at the role of quantifiers in everyday discourse, we find that one is faced with a much richer collection than is expressed within FOL. Here are a few examples, taken from the work of linguists on quantifiers in natural language.

1. *Most linguists* are bilingual.

2. *More science students than humanities students* work hard.

3. *Not as many boys as girls* did well on the exam.

Note that in each case, the domain of discourse such as "linguists", "science students", and so on, is understood to be finite, unlike FOL where we also allow infinite domains. The semantics of "most", "more …than…", "not as many as", etc. can each be given by comparing the numbers of members of the relevant sets. But they are not definable in FOL and again do not carry any fully worked out system of logic. Also, the semantics that linguists take for "*Most* As are Bs" makes it true just in case more than half of the As have the property B. According to this, if there are, say, 100 linguists in our universe of discourse, it only takes 51 of them to be bilingual in order to assert that most linguists are bilingual. But that doesn't accord to our usual sense of "most"; we'd think it would take a lot more to draw that conclusion. But how much more? 60%? 70%? Any such proposed alternative semantics for "most" seems arbitrary; indeed in actual use "most" is a vague quantifier. That is clearly the case for the quantifier "many", as in

4. *Many people* have stopped smoking,

[1] See my discussion of the Formalizability Thesis in the article, "And so on… Reasoning with infinite diagrams", *Synthese* 186 (2012), 371-386 (also on my home page at http://math.stanford.edu/~feferman/papers/And_so_on(pub).pdf .)

for which no exact definition at all can be proposed. This last also shows that when we use quantifiers in everyday language, the domain of discourse to which they apply (e.g., living people) is often not fixed in any definite way.

Aside from quantifiers, our daily exchanges and in the media are replete with *vague concepts*, witness those that have relentlessly occupied the recent political campaigns: "middle class", "inequality", "job creators", "taxes", "small businesses," etc. And look back at what I said a minute ago and you'll see that practically every general noun, adjective, verb and adverb I've used is vague: "domain of discourse", "understood", "allow", "semantics", "relevant", "linguist", "bilingual", etc., etc.

So the presumption underlying the use of FOL that we are dealing with definite domains of discourse and definite concepts that are true or false of the members of those domains is simply not satisfied in our day-to-day use of language. Some have suggested that we need to modify our logic accordingly, that we should use some kind of *logic of vague concepts*. The first problem with that is the famous *Sorites Paradox*, typified by the *Paradox of the Heap*: consider a heap of sand; if one removes a single grain from it, the result is still a heap of sand. Continuing step by step in this way, we are brought to the conclusion that a single grain of sand is still a heap. So if there is to be a logic of vague concepts, it must somehow avoid this paradox. One way that has been suggested is in the use of what is called *fuzzy logic*. In that, propositions do not have just the two truth values, *True* and *False*, but instead lie in a continuum of truth values ranging from *Absolutely True* (or degree 1) down to *Absolutely False* (or degree 0), with everything in between as just *true to a degree*. In this logic, a heap is not just a heap without qualification, but rather a heap to a certain degree. But the assignment of that degree seems to be somewhat arbitrary—what determines when a collection of grains of sand is a heap to degree 1, but is less so when one removes one grain? Where in the visual spectrum must something occur to be absolutely red, but from which the slightest change in frequency makes it less red? Fuzzy logic itself happens to be a coherent (and precise!) formal system, but it does not help in making day-to-day reasoning more exact. Interestingly, fuzzy logic has had some practical applications, for example in control systems. By the way, once, when my wife and I were on the market for a new

dishwasher, the salesman showed us one that he said is governed by fuzzy logic: it supposedly adjusts its program according to how dirty are the dishes that have been stacked in it. We didn't buy it.

Despite the ubiquity of vague concepts, vague domains of discourse and vague quantifiers in everyday use, the ubiquitous application of classical reasoning seems to be inescapable in our daily lives. The reason is that we are continually faced with choices, some minor and routine, some major. The simplest form of a choice is whether or not to take some action. For example, if one has some chronic physical condition like a very bad knee, one is faced with the decision whether or not to have surgery; there is no in-between. We then start thinking: the surgery has a non-trivial risk of failure which is worrisome, and even if it works well, it will take six months to fully recover; on the other hand, if I don't do the surgery, I will continue to be in pain and be seriously hampered in my walking. It is only when contemplating the possible consequences of each side of such a choice in terms of possible risks and benefits that we go beyond classical logic into issues of probability and the subjective value we place on various possible consequences.

Beyond our personal lives we have a constant need for careful, critical reasoning in both our public lives and our professional pursuits, but the problems of vagueness limit how well suited to that task is the logic underlying mathematics. On the other hand, there is no evidence that any of the logics that have been proposed to deal with vague concepts can help us carry the requisite reasoning in any way better than we do by the light of day. Some claim that an approach called *Argumentation Theory*, or *Informal Logic*, is much more suited to those needs in the case of public discussions. That analyzes various general forms of both good arguments and fallacious reasoning, in a way that is also aided to some extent by formal logic. But that only goes so far, especially for arguments that are carried out on the fly. And I can't say I know of any chain of reasoning in public discussions that deserves to be called a conclusive proof. By contrast, the systematic disciplines of natural science, social science, medicine, the law, and philosophy are marked by critical reasoning of quite varied kinds and it is their hope and aim to come to some definitive conclusions that would warrant being called proofs. But clearly, what

constitutes a proof in physics is quite different from what constitutes a proof in medicine or the law. We can assume that reasoning in these disciplines all respect the basic laws of logic, but we would require another story above and beyond that in each case as to what makes a proof *a proof of that sort*. In other words, for each of these, we'd need a "meta" study that does for it what metamathematics does for mathematics, and that's where the respective experts would need to take over.