PROOF AS A CLUSTER CONCEPT

Keith Weber

Rutgers University

Proof is a central concept in mathematics education, yet mathematics educators have failed to reach a consensus on how proof should be conceptualized. I advocate defining proof as a clustered concept, in the sense of Lakoff (1987). I contend that this offers a better account of mathematicians' practice with respect to proof than previous accounts that attempted to define a proof as an argument possessing an essential property, such as being convincing or deductive. I also argue that it leads to useful pedagogical consequences.

PROOF CONCEPTUALIZATION IN MATHEMATICS EDUCATION

It is widely accepted that having students successfully engage in the activity of proving is a central goal of mathematics education (e.g., Harel & Sowder, 1998). Yet mathematics educators cannot agree on a shared definition of proof (Balecheff, 2002; Reid & Knipping, 2010; Weber, 2009). This is recognized as problematic: without a shared definition, it is difficult for mathematics educators to meaningfully build upon each another's research and it is impossible to judge if pedagogical goals related to proof are achieved (e.g., Balacheff, 2002; Weber, 2009). Until now, most mathematics educators have sought to define proof as an argument that possesses one or more desirable properties, such as employing deductive reasoning (Hoyles & Kuchemann, 2002) or being convincing to oneself (Harel & Sowder, 1998) or community (Balacheff, 1987). However, there is not a consensus on which property or properties capture the essence of proof. The main thesis of this paper is that, in mathematical practice, there are no properties that are the essence of proof and viewing proof as a clustered model in the sense of Lakoff (1987) offers a better account of how proof is practiced by mathematicians.

Two approaches to defining proof

There are two approaches that philosophers and mathematics educators have used to define proof (CadwalladerOlsker, 2011). In the analytic philosophical tradition, some have sought to define a proof as a formal object, usually as a strictly syntactic object within a formal theory. Unfortunately, there is little intersection between the objects satisfying definitions of these types and the arguments that mathematicians refer to as proofs. Consequently, such a definition cannot provide a reasonable account of how proofs are produced or how they advance our mathematical knowledge (cf., Pelc, 2009). Further, from an instructional perspective, this can imply the pedagogically dubious suggestion of focusing on the form of proof rather than its meaning.

A second approach to proof is to define proofs as the proofs that mathematicians actually read and write or as the arguments that mathematicians label as proofs.

2014. In Nicol, C., Oesterle, S., Liljedahl, P., & Allan, D. (Eds.) Proceedings of the Joint Meeting of PME 38 and PME-NA 36,Vol. 5, pp. 353-360. Vancouver, Canada: PME.

However, such a characterization is too broad to do useful philosophical or pedagogical work. What is needed is a sense *what* types of arguments mathematicians recognize as proof. Further, this sense should be philosophically and pedagogically pertinent. For instance, the observation that mathematicians usually publish their proofs is LaTEX will not inform instructional practice. If we accept Larvor's (2012) observation that, "the field [the philosophy of mathematical practice] lacks an explication of 'informal proof' as it appears in expressions such as 'the informal proofs that mathematicians actually read and write'" (p. 716), then it is clear that there is more work to do in this area.

DIFFICULTIES IN FINDING AN ESSENCE OF PROOF

A common approach to defining proof is to locate a characteristic (or set of characteristic) that is shared by all arguments that mathematicians consider to be proofs and not present in all other arguments. If successful, this approach would yield a clear way of characterizing proof. Unfortunately, this approach has not been successful. For instance, a proof has sometimes been defined as an argument that convinces oneself (or one's community) that an assertion is true (e.g., Harel & Sowder, 1998). However, Tall (1989) noted that there are convincing arguments that would not qualify as proofs. For instance, Eccheveria (1996) claims that the empirical evidence in support of Goldbach's Conjecture is so overwhelming that the mathematical community is certain of its truth, but the claim is not proven. Proofs are sometimes defined to be *a priori* deductive arguments that do not depend on one's observations or experience, but Fallis (1997) noted that computer-assisted arguments would not satisfy this description.

It is natural to try to define proof as a category of objects sharing some properties. After all, this is how mathematical concepts are defined (Alcock & Simpson, 2002). However, I argue that proofs are not mathematical concepts, they are *discursive concepts*. And I further argue that there is no property that distinguishes proofs from non-proofs.

Three proofs

To highlight the difficulties of characterizing proofs, consider these three proofs as they appear in the mathematics literature.

Theorem 1: If n is a number of the form $6k$-1, then n is not perfect.

Proof 1: Assume n is a positive integer of the form $6k$-1. Then $n \equiv -1 (\text{mod } 3)$ and hence n is not a square. Note also that for any divisor d of n, $n = d\left(\frac{n}{d}\right) \equiv -1(\text{mod } 3)$ implies that $d \equiv -1(\text{mod } 3)$ and $\left(\frac{n}{d}\right) \equiv 1(\text{mod } 3)$ or $d \equiv 1(\text{mod } 3)$ and $\left(\frac{n}{d}\right) \equiv -\mathbf{1}(\text{mod } 3)$. Either way, $d + \left(\frac{n}{d}\right) \equiv 0$ (mod 3) and $\sigma(n) = \sum_{d|n, d<\sqrt{n}} d + \frac{n}{d} \equiv 0$ (mod 3). Computing $2n$=2($6k$-1) $\equiv 1(\text{mod } 3)$, we see that n cannot be perfect. (from Holdener, 2002)

Theorem 2: $\pi = \sum_{k=0}^{\infty} \frac{1}{16^k}\left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6}\right)$.

Proof 2: Here is a proof using *Mathematica* to perform the summation.

FullSimplify[TrigtoExp[FullSimplify[

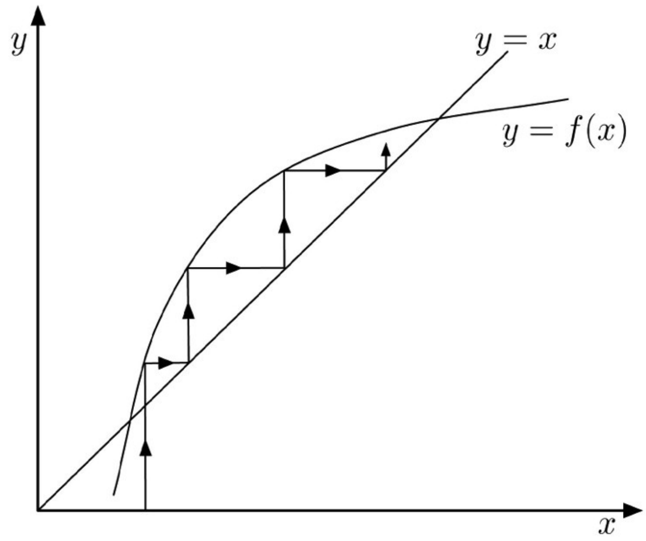
$\pi = \sum_{k=0}^{\infty} \frac{1}{16^k}\left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6}\right)$.]]/.

a_Log[b_]+a_Log[c_]:>a Log[b c]].

π (from Adamchik & Wagon, 1997)

Theorem 3: (Fixed Point Theorem) Let $f(x)$ be continuous and increasing on [0, 1] such that $f([0,1]) \subseteq [0,1]$. Let $f_2(x)=f(f(x))$ and $f_n(x)=f(f_{n-1}(x))$. Then under iteration of f, every point is either a fixed point or else converges to a fixed point.

Proof 3: The only proof needed is:



(from Littlewood, 1957)

These proofs vary widely in terms of the types of inferences that were made, the representation systems used, their level of transparency, and the level of detail they provide. At this point, the reader may want to make three objections: (1) Some of these "proofs" are not really proofs; (2) These proofs are outliers; (3) These proofs are considered controversial.

I do not think (1) is a fair objection. If we were defining what proof *ought* to be, one could say Proof 2 or Proof 3 ought not be considered as a proof. However, if we wish to describe the proofs that mathematicians actually read and write, we must account for Proof 2 and Proof 3 because they were published in the literature by mathematicians as proofs. With (2), Proof 2 and Proof 3 were deliberately chosen to be provocative, yet they are also representative of the wider categories of computer-assisted proofs and visual proofs.

With (3), these proofs *are* controversial. In Adamchik and Wagon's (1997) paper in which their proof was presented, they admitted that, "Some might even say this is not truly a proof! But in principle, such computations *can* be viewed as proofs" (p. 852). In an experimental study, Inglis and Mejia-Ramos (2009) demonstrated that mathematicians collectively find Proof 3 significantly less convincing than more conventional proofs. I accept that these proofs are controversial, but argue this controversy has important consequences for the nature of a descriptive account of proof.

Proof*

Aberdein (2009) coined the term, "proof*", as "species of alleged 'proof' where there is no consensus that the method provides proof, or there is a broad consensus that it doesn't, but a vocal minority or an historical precedent point the other way". As examples of proof*, Aberdein included "picture proofs*, probabilistic proofs*, computer-assisted proofs*, [and] textbook proofs* which are didactically useful but would not satisfy an expert practitioner". As Proof 2 is a computer-assisted proof and Proof 3 is a picture proof, these qualify as proofs*.

Proofs* do not pose a problem for analytic philosophers who attempt to pose *normative judgments* for what should be considered a proof. Recently, there have been arguments that picture proofs, such as Proof 3, are perfectly valid and ought to be on par epistemologically with the more traditional verbal-symbolic proof (e.g., Kulpa, 2009). Granted there may be some mathematicians who disagree, such as those in Inglis and Mejia-Ramos' (2009) experimental study, but the proponents of picture proofs can argue that these mathematicians are simply mistaken.

However, proofs* *do* pose a problem for philosophers and mathematics educators who, as Larvor (2012) put it, wish to describe "the proofs that mathematicians actually read and write". Take picture proofs*, for instance. A proposed criteria of proof must either admit some picture proofs* as proofs or claim that all picture proofs* are not. If the former occurred, one could challenge this claim by citing the large number of mathematicians who do not produce such proofs and reject such proofs when they read them. If the latter occurred, one could rebut the claim by citing the picture proofs in the published literature as well as the large number of mathematicians (or at least the vocal minority) who accept such proofs. Similar arguments could be made for all types of proofs*.

PROOF AS CLUSTER MODEL

Cluster concepts

Lakoff (1987) noted that "according to classical theory, categories are uniform in the following respect: they are defined by a collection of properties that the category members share" (p. 17). This perspective has dominated the way that philosophers have attempted to define proof. However, Lakoff's thesis is that most real-world categories cannot be characterized this way. In particular, he argued that some categories might be better thought of as *clustered models*, which he defined as occurring when "a number of cognitive models combine to form a complex cluster that is psychologically more basic than the models taken individually" (p. 74). I will argue that mathematical proof should be regarded in the same way.

As an illustrative example of a clustered concept, Lakoff considered the category of *mother*. According to Lakoff, there are several types of mothers, including the birth mother, the genetic mother, the nurturance mother (i.e., the adult female caretaker of the child), and the marital mother (i.e., the wife of the father). These concepts are highly correlated-- the birth mother is nearly always the genetic mother and more often than not the caretaker. In the prototypical case, these concepts will converge—that is, the birth mother will also be the genetic mother, the nurturance mother, and so on. And indeed, when one hears that the woman is the mother of a child, the default assumption is that the woman assumes all roles. However, this is not always the case.

Lakoff raised two points that will be relevant to this paper. First, there is a natural desire to pick out the "real" definition of mother, or the true *essence of motherhood*. However, Lakoff rejected this essentialist disposition. Different dictionaries list different conceptions of mother as their primary definition. Further, sentences such as, "I was adopted so I don't know who my *real mother* is" and "I am uncaring so I doubt I could be a *real mother* to my child" both are intrinsically meaningful yet define *real mother* in contradictory ways. Second, in cases where there is divergence in the clustered concept of mother (e.g., a genetic but not adoptive mother), compound words exist to qualify the use of mother. Calling one a birth mother typically indicates that she in not the nurturance mother; calling one an adoptive mother or a stepmother indicates that she is not the birth mother.

Proof as a clustered concept

The main thesis of this paper is that it would be profitable to consider proof as a clustered concept. The exact models that should form the basis of this cluster should be the matter of debate, but I will propose the following models as a working description to highlight the utility of this approach: (1) A proof is a *convincing argument* that convinces a knowledgeable mathematician that a claim is true. (2) A proof is a *deductive argument* that does not admit possible rebuttals. The lack of potential rebuttals provides the proof with the psychological perception of being *timeless*. Proven theorems remain proven. (3) A proof is a *transparent argument where a mathematician can fill in every gap* (given sufficient time and motivation), perhaps to the level of being a formal derivation. In essence, the proof is a blueprint for the mathematician to develop an argument that he or she feels is complete. This gives a proof the psychological perception of being *impersonal*. Theorems are *objectively* true. (4) A proof is a *perspicuous argument that provides the reader with an understanding of why a theorem is true*. (5) A proof is an *argument within a representation system satisfying communal norms.* That is, there are certain ways of transforming mathematical propositions to deduce statements that are accepted as unproblematic by a community and all other steps need to be justified. (6) A proof is an *argument that has been sanctioned by the mathematical community*.

Of course, the criteria above are not original. All have previously been proposed by other philosophers and mathematicians. What is original here is claiming that one cannot demarcate proofs from non-proofs by saying that proofs must satisfy some subset of the criteria above.

I argue that each of these more basic models do not, by themselves, characterize proof completely. I previously argued that (1) fails because there are convincing empirical arguments that are not proofs. Fallis (1997) notes that computer-assisted proofs fail to satisfy (2) and (3), since a possible rebuttal is that the computer software was faulty and since the proof does not give a blueprint for how a human could perform the computer checks for himself or herself. Similar arguments can be given for (4), (5), and (6).

If we accept proof to be a clustered concept as defined above, we would expect the following to occur: (a) proofs that satisfied all of these criteria should be uncontroversial, but some proofs that satisfy only a subset of these criteria might be regarded as contentious; (b) compound words exist that qualify proofs that satisfy some of these criteria but not others; (c) it would be desirable for proofs to satisfy all six criteria.

Regarding (a) and (b), Aberdein's (2009) discussion of proofs* supports these points. He explicitly highlighted compound words delimiting the sense that arguments are proofs. For instance, computer-assisted proofs* are not transparent and it is not clear how a mathematician can fill in every gap of the proof and probabilistic proofs* are not deductive. Not only do these qualifying compound words exist, but as Aberdein (2009) argued, there is not a consensus on their validity amongst mathematicians. For (c), we can consider Dawson's (2006) analysis of why mathematicians re-prove theorems. Dawson's analysis demonstrated that sanctioned proofs are reproven to avoid controversial methods, fill in perceived gaps, become more perspicuous, and increase mathematician's conviction, which correspond to the first four components of the cluster model described above.

IMPLICATIONS FOR PEDAGOGY

If we view proof as a cluster concept, like that of mother, we might expect that this concept is perhaps not best taught by direct instruction, but instead through practice in a community. For instance, Thurston (1994) described how he sought a clear definition in proof in graduate school; he did not find one but through experience, he began to "catch on". Of course, we know that mathematics majors often do not catch on and remain deeply confused about the meaning of proof when they graduate. Here the instructor might help by pointing to features of the argument that make the argument a better or worse example of proof, rather than solely presenting the argument as right or wrong.

At a broad level, the components of the clustered model of proof are correlated with one another. For instance, as an argument becomes more deductive, it often tends to become more convincing, easier to translate into a formal proof, and more likely to be sanctioned by one's peers. Hence, encouraging students to make their arguments more deductive would usually make their arguments more proof-like in other respects as well. However, this is not the case if we take some of these criteria to extremes.

For a first example, suppose we strive to present students with arguments that are as convincing as possible in geometry. In many cases, an exploration on a dynamic geometry package would be entirely convincing, both for mathematicians and for students (de Villiers, 2004). For a student, such explorations would probably be *more* convincing than a complicated deductive argument because the student may worry that he or she has overlooked an error in the argument. If we view the mode of reasoning (deductive vs. perceptual) and the representation system in which an argument is couched as irrelevant, it is difficult to argue why demonstrations on dynamic geometry software packages are not proofs.

A similar claim relates to how formal an argument is. Increasing the formality of an argument usually makes the argument more deductive and more acceptable to the mathematical community. However, it is generally accepted that there is a point where an argument is "formal enough" and making it more rigorous would be detrimental. Filling in *all* the gaps would make the proof impossibly long and unwieldy. The result would be a proof that masks its main ideas. As understanding these ideas is important for determining the validity of the proof, so increasing the rigor of the proof would lessen its persuasive power.

If we want students and teachers to present proofs that satisfy all or most of the criteria above, it would be best not to focus on a single criterion. Not only would the other criteria be ignored, a singular focus on one criterion might actually lessen the possibilities of the other criteria being achieved.

References

Aberdein, A. (2009). Mathematics and argumentation. *Foundations of Science*, *14*, 1-8.

Adamchik, V. & Wagon, S. (1997). A simple formula for π. *American Mathematical Monthly*, *104*, 852-855.

Alcock, L., & Simpson, A. (2002). Definitions: Dealing with categories mathematically. *For the Learning of Mathematics*, *22*(2), 28-34.

Balacheff, N. (1987). Processus de preuves et situations de validation. *Educational Studies in Mathematics*, *18(*2), 147-176.

Balacheff, N. (2002). The researcher epistemology: A deadlock from educational research on proof. In F.L. Lin (Ed.), *2002 International Conference on Mathematics Education: Understanding proving and proving to understand* (pp. 23-44). Taipei, Taiwan: NCS and NSUT.

CadwalladerOlsker, T. (2011). What do we mean by mathematical proof? *Journal of Humanistic Mathematics*, *1*(1), 36-60.

Dawson, J. (2006). Why do mathematicians re-prove theorems? *Philosophia Mathematica*, *14,* 269-286.

de Villiers, M. (2004). The role and function of quasi-empirical methods in mathematics. *Canadian Journal of Science, Mathematics, and Technology Education*, *4*, 397-418.

Echeverria, J. (1996). Empirical methods in mathematics. A case study: Goldbach's Conjecture. In G. Munevar (Ed.), *Spanish studies in the philosophy of science* (pp. 19-56). Dordrecht: Kluwer.

Fallis, D. (1997). The epistemic status of probabilistic proof. *The Journal of Philosophy*, *94*(4), 165-186.

Harel, G., & Sowder, L. (1998). Students proof schemes. *Research in Collegiate Mathematics Education*, *3,* 234-282.

Holdener, J. (2002). A theorem of Touchard on the form of odd perfect numbers. *American Mathematical Monthly*, *109,* 661-663.

Hoyles, C., & Kuchemann, D. (2002). Students' understanding of mathematical implication. *Educational Studies in Mathematics*, *51*, 193-223.

Inglis, M., & Mejia-Ramos, J. P. (2009). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction*, *27,* 25-50.

Kulpa, Z. (2009). Main problems with diagrammatic reasoning. Part 1: The generalization problem. *Foundations of Science*, *14*, 75-96.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Cambridge: Cambridge University Press.

Larvor, B. (2012). How to think about informal proofs. *Synthese*, *187*, 715-730.

Littlewood, J. E. (1953). *A mathematician's miscellany*. London: Methuen.

Pelc, A. (2009). Why do we believe theorems? *Philosophia Mathematica*, *17*, 84-94.

Reid, D., & Knipping, C. (2010). *Proof in mathematics education*. Rotterdam: Sense Publishers.

Tall, D. (1989). The nature of mathematical proof. *Mathematics Teaching*, *127*, 28-32.

Weber, K. (2009). Review of Boero's Theorems in schools. *Mathematical Thinking and Learning*, *11*, 289-294.