Proof as a cluster concept in mathematical practice

Keith Weber
Rutgers University

Approaches for defining proof

- In the philosophy of mathematics, there are two approaches to defining proof:
  - *Logical or formalist approach:* Proof can be defined as a syntactic formal object. There are rules for forming well-formed sentences. There are a collection of axioms and rules for deducing new sentences from previous ones. A sequence of sentences beginning with axioms, inferring a sequence of new statements, and concluding with the theorem is a proof of the theorem.

Approaches for defining proof

- A standard critique of this approach is that it does a poor job of characterizing mathematical practice.
  - Few proofs that are published in mathematical journals come close to matching this standard (e.g., Davis & Hersh, 1981; Rav, 1999)
  - Even if published proofs "map" to formal derivations, this is rarely done so its tough to see what benefits could be accrued from engaging in this process. More broadly, it's tough to say how derivations leads to conviction or knowledge, given their scarcity (Pelc, 2009)
  - There are some who argue that formal derivations provide considerably *less* conviction or understanding than proofs as they are normally written (e.g., Rav, 1999; Thurston, 1994)

Approaches for defining proof

- *Sociological approach*: We should define a proof to be the types of proofs that mathematicians read and write and define proof.

Problems with the sociological approach to proof

- Defining proof purely descriptively as “the types of proofs that mathematicians produce” also does little work for us.
  - As Larvor (2012) noted, “the field lacks an explication of ‘informal proof’ as it appears in expressions such as ‘the informal proofs that mathematicians actually read and write’” (p. 716).

- This is *pedagogically useless*. We need some sense to describe similarities between (desired) student proofs and actual proofs.
  - What’s to stop us from saying, “students should write proofs in LATeX”?

My approach: This should be treated as an empirical question

- If we are describing proofs "out in the world", we can look at these proofs.

- If we are describing mathematicians' views on proof, we can talk to and discuss these issues with mathematicians.

“Mathematical proof does not admit degrees. A sequence of steps in an argument is either a proof, or it is gibberish”

(Rota, 1997, p. 183).

“The concept of mathematical proof, like mathematical truth, does not admit degrees”

(Montano, 2012, p. 26).

Proof as a binary judgment

- Mathematicians all agree on whether something is a proof.
  - Azzouni (2004) attempted to explain why "mathematicians are so good at agreeing with one another on whether a proof convincingly establishes a theorem" (p. 84).
  - "All agree that something either is a proof or it is not and what makes it a proof is that every assertion in it is correct" (McKnight et al, 2000, p. 1).
  - Selden and Selden (2003) marveled at "the unusual degree of agreement about the correctness of arguments and the proof of theorems […] Mathematicians say an argument proves a theorem. Not that it proves it for Smith but possibly not for Jones" (p. 11).

Is this a proof?

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

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

FullSimplify[TrigtoExp[FullSimplify[

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

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

Is this a proof?

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

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

FullSimplify[TrigtoExp[FullSimplify[

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

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

From Adamchik and Wagon (1997), published in the *American Mathematical Monthly*.

Is this a proof?

Notice that this proof:

- Does not provide explanation
- Involves untested hidden assumptions (Mathematica is reliable)
- Gaps in the proof cannot easily be deductively verified by mathematicians (or at least it does not hint at a method other than use Mathematica)

Family resemblance

- Wittgenstein (1953, 2009) noted that philosophers desired necessary and sufficient conditions for concept membership, but this "craving for generality" was misplaced.

- Some concepts (famously *game*) may not have a feature that all its members share but overlapping similarities amongst all members of the concept.

Family resemblance

- Wittgenstein (1953, 2009) noted that philosophers desired necessary and sufficient conditions for concept membership, but this “craving for generality” was misplaced.
- Some concepts(famously *game*) may not have a feature that all its members share but overlapping similarities amongst all members of the concept.

Name	Eyes	Hair	Height	Physique
Aaron	Green	Red	Tall	Thin
Billy	Blue	Brown	Tall	Thin
Caleb	Blue	Red	Short	Thin
Dave	Blue	Red	Tall	Fat

Cluster concepts

- Lakoff (1987) said 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).
  - But like Wittgenstein, Lakoff argued that most real-world categories and many scientific categories cannot be defined in this way.
- Lakoff says some categories might be better defined 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).

Cluster concepts: Mother

- A classic example is the category of *mother*, which is an amalgam of several models:
  - *The birth mother*
  - *The genetic mother*
  - *The nurturance mother* (the female caretaker of the child)
  - *The wife of the father*
  - *The female legal guardian*

Cluster concepts: Key points

- The prototypical mother satisfies all models. Our default assumption is that a mother (probably) satisfies these models.

- There is no true *essence* of mother.
  - Different dictionaries list different primary definitions.
  - “I am uncaring so I could never be a *real mother* to my child”; “I’m adopted so I don’t know who my *real mother* is”, illustrate that “real mother” doesn’t have one definition.

- Compound words exist to qualify limited types of mothers.
  - *Stepmother* implies wife of the father but not the birth or genetic mother.
  - *Birth mother* implies not the caretaker
  - *Adoptive mother* implies not the birth or genetic mother.

Cluster concepts: Proof

Proof is:

- *A convincing argument*
- *A surveyable argument understandable by a human mathematician*
- *An a priori argument* (starting from known facts, independent of experience, deductive)
- *A transparent argument where a reader can fill in every gap*
- *An argument in a representation system*, with social norms for what constitutes an acceptable transformation or inference
- *A sanctioned argument* (accepted as valid by mathematicians by a formal review process)

Cluster concepts: Proof

Proof is:

- *A convincing argument*
- *A surveyable argument understandable by a human mathematician*
- *An a priori argument* (independent of experience, deductive)
- *A transparent argument where a reader can fill in every gap*
- *An argument in a representation system*, with social norms for what constitutes an acceptable transformation or inference
- *A sanctioned argument* (accepted as valid by mathematicians by a formal review process)

Cluster concepts: Proof

Proof is:

- *A convincing argument*
- *A surveyable argument understandable by a human mathematician*
- *An a priori argument* (independent of experience, deductive)
- *A transparent argument where a reader can fill in every gap*
- *An argument in a representation system*, with social norms for what constitutes an acceptable transformation or inference
- *A sanctioned argument* (accepted as valid by mathematicians by a formal review process)

Cluster concepts: Proof

Proof is:

- *A convincing argument*
- *A surveyable argument understandable by a human mathematician*
- *An a priori argument* (independent of experience, deductive)
- *A transparent argument where a reader can fill in every gap*
- *An argument in a representation system*, with social norms for what constitutes an acceptable transformation or inference
- *A sanctioned argument* (accepted as valid by mathematicians by a formal review process)

Cluster concepts: Predicted consequences

1. The prototypical proof satisfies all criteria. Proofs that satisfy all criteria would be better representatives of proof than those that satisfy some criteria and would not be controversial.
2. Proofs that only satisfy some would be controversial and spark disagreement.
3. There are compound words that qualify “proofs” that satisfy some criteria but not all criteria.
4. There are *default judgments* when you hear an argument is a proof– properties you think the argument is likely to have but are not necessarily sure of.
5. There is no single essence of proof.

Cluster concepts: Predicted consequences

1. The prototypical proof satisfies all criteria. Proofs that satisfy all criteria would be better representatives of proof than those that satisfy some criteria and would not be controversial.
2. Proofs that only satisfy some would be controversial and spark disagreement.
3. There are compound words that qualify "proofs" that satisfy some criteria but not all criteria.
4. There are *default judgments* when you hear an argument is a proof– properties you think the argument is likely to have but are not necessarily sure of.
5. There is no single essence of proof.

Cluster concepts: Predicted consequences

- Proofs that only satisfy some would be controversial and spark disagreement.
- There are compound words that qualify “proofs” that satisfy some criteria but not all criteria.

- Picture proofs* are not in standard representation system of proof.
- Probabilistic proofs* are not deductive or *a priori*.
- Computer assisted proofs* are not transparent.
- One might add unpublished proofs*, or proofs* with gaps/ incomplete proofs*, etc.

The need for empirical studies: Status of computer-assisted proofs

"The glamorous instance of a verification that falls short of being accepted as a proof- despite its undeniable correctness- is the computer verification of the four color conjecture".

(Rota, 1997, p. 186).

The need for empirical studies: Status of computer-assisted proofs

"We are entering into a grey area: computer-assisted proofs. They are not proofs in the standard sense in that they can be checked by a line-by-line verification. They are especially unreliable when they claim to make a complete list of something or another"

(Jean-Pierre Serre, as cited in Raussen & Skau, 2004).

The need for empirical studies: Status of computer-assisted proofs

"When the critics spoke of [a computer assisted proof as] an ugly proof, they were conceding it was a *genuine* proof, for the concept of mathematical proof, like mathematical truth, does not admit degrees"

(Montano, 2012, p. 25).

The need for empirical studies: Status of computer-assisted proofs

“I now need to argue that a computer proof can legitimately stand in for a mathematical proof. However, for the following reasons, I will not do so. First, I really have nothing to add to the debate that has already been carried out on this issue. (See Tymoczko, Teller, and Detlefsen) […] Most importantly, this debate is rather anachronistic. The prevailing sentiment among mathematicians is that a computer proof is a legitimate way to establish the truth of a mathematical claim”.

(Fallis, 1996, p. 494).

The study

- A survey was completed by 108 mathematicians.
- Mathematicians were shown four proofs in a randomized order.
  - They were told not to focus on correctness. They could assume each statement in the proof was true and each calculation was carried out correctly.
  - They were told where each proof was published.
  - The goal was to have them focus on the types of reasoning that were used.

The study

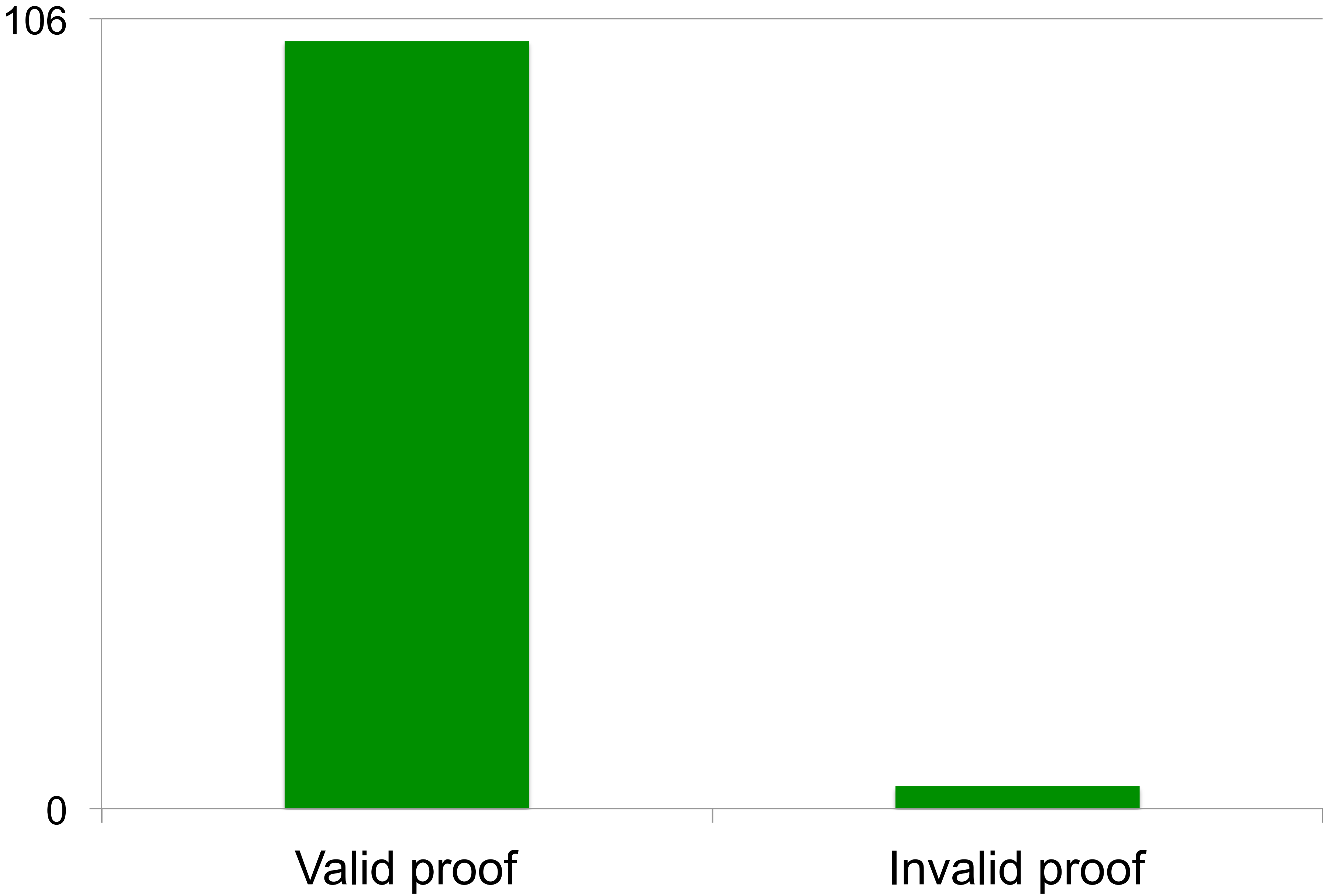
- They were asked questions.
  - On a scale of 1 through 10, how typical was the reasoning used in this proof of the proofs they read and wrote.
  - Was the proof valid? (Yes/No)
  - What percentage of mathematicians did they think would agree with them?
  - Was the argument valid (invalid) in nearly all math contexts or was it generally valid (invalid) but there were contexts in which it invalid (valid).

Theorem: If *n* is a number of the form 4*k*+3, then *n* is a not perfect number.

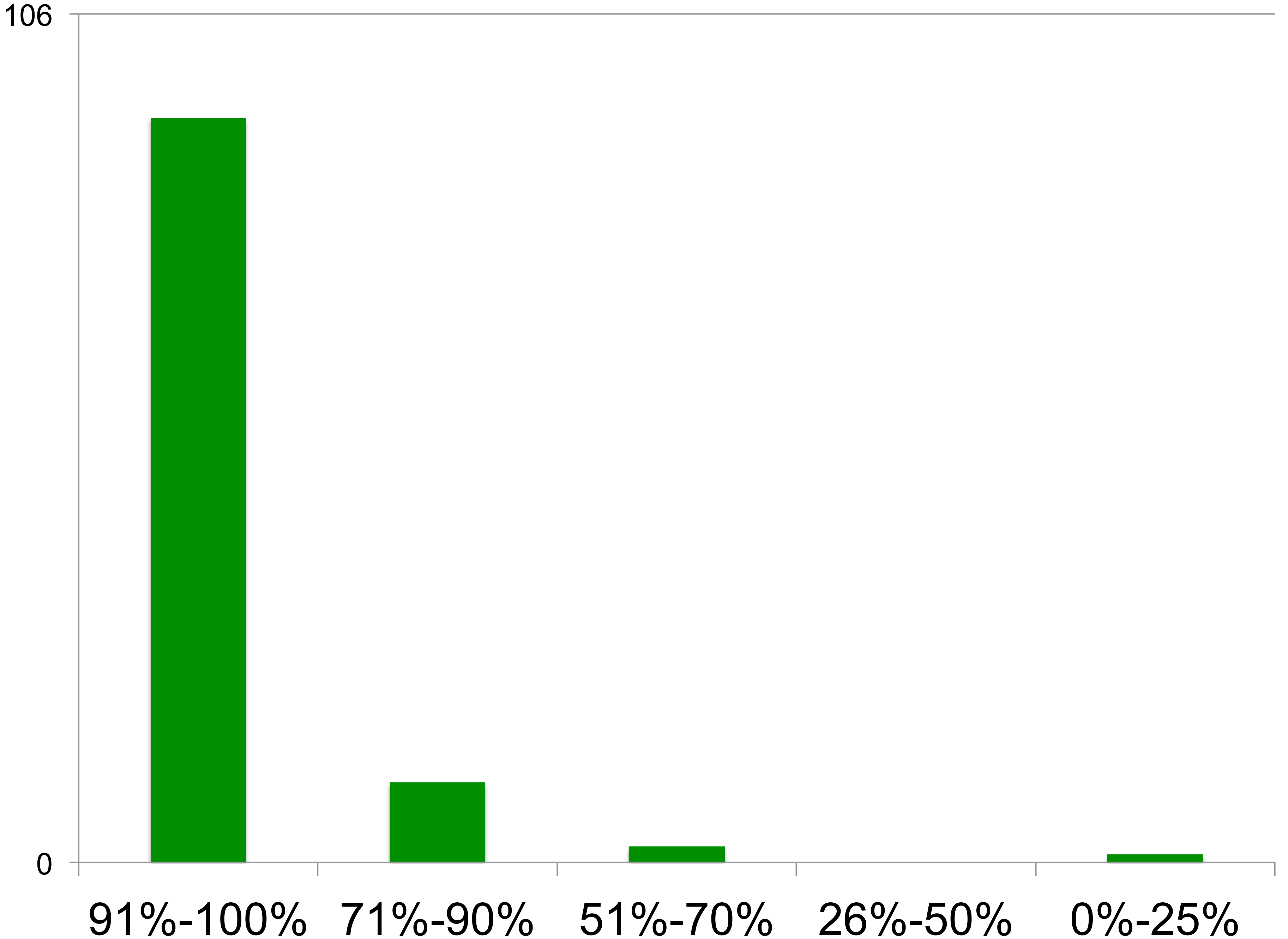
Proof: Assume *n* is a positive integer of the form 4*k*+3. Then *n* is not a square. If *d* and $\left(\frac{n}{d}\right)$ are divisors of *n*, then either $d<\sqrt{n}$ or $\left(\frac{n}{d}\right) < \sqrt{n}$. Without loss of generality, assume $d<\sqrt{n}$. Since $n = d\left(\frac{n}{d}\right) \equiv 3 (\text{mod } 4)$, either $d \equiv 3(\text{mod } 4)$ and $\left(\frac{n}{d}\right) \equiv 1(\text{mod } 4)$ or $d \equiv 1(\text{mod } 4)$ and $\left(\frac{n}{d}\right) \equiv 3(\text{mod } 4)$. Either way, $d + \left(\frac{n}{d}\right) \equiv 0(\text{mod } 4)$ and $\sigma(n) = \sum_{d|n, d<\sqrt{n}} d + \frac{n}{d} \equiv 0(\text{mod } 4)$. Computing $2n=2(4k+3) \equiv 2(\text{mod } 4)$, we see that *n* cannot be perfect.

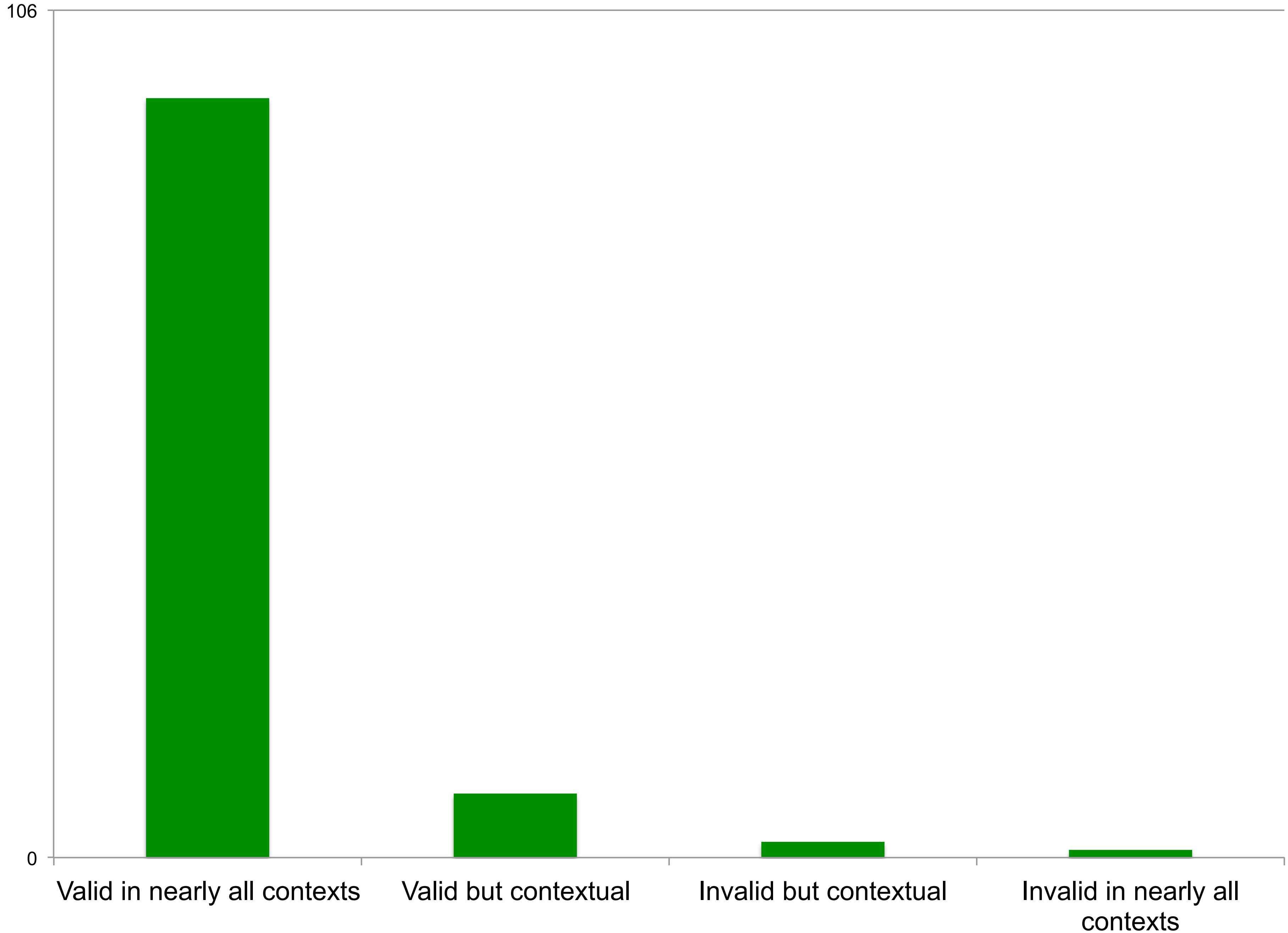
Published in: J. Holdener (2002). A theorem of Touchard on the form of odd perfect numbers. *American Mathematical Monthly*, 109, 661-663.

(Note: $\sigma(n)$ is the sum of the divisors of *n*. *n* is a perfect number if and only if $\sigma(n) = 2n$)



106
0
Valid proof
Invalid proof


106
0
91%-100%
71%-90%
51%-70%
26%-50%
0%-25%

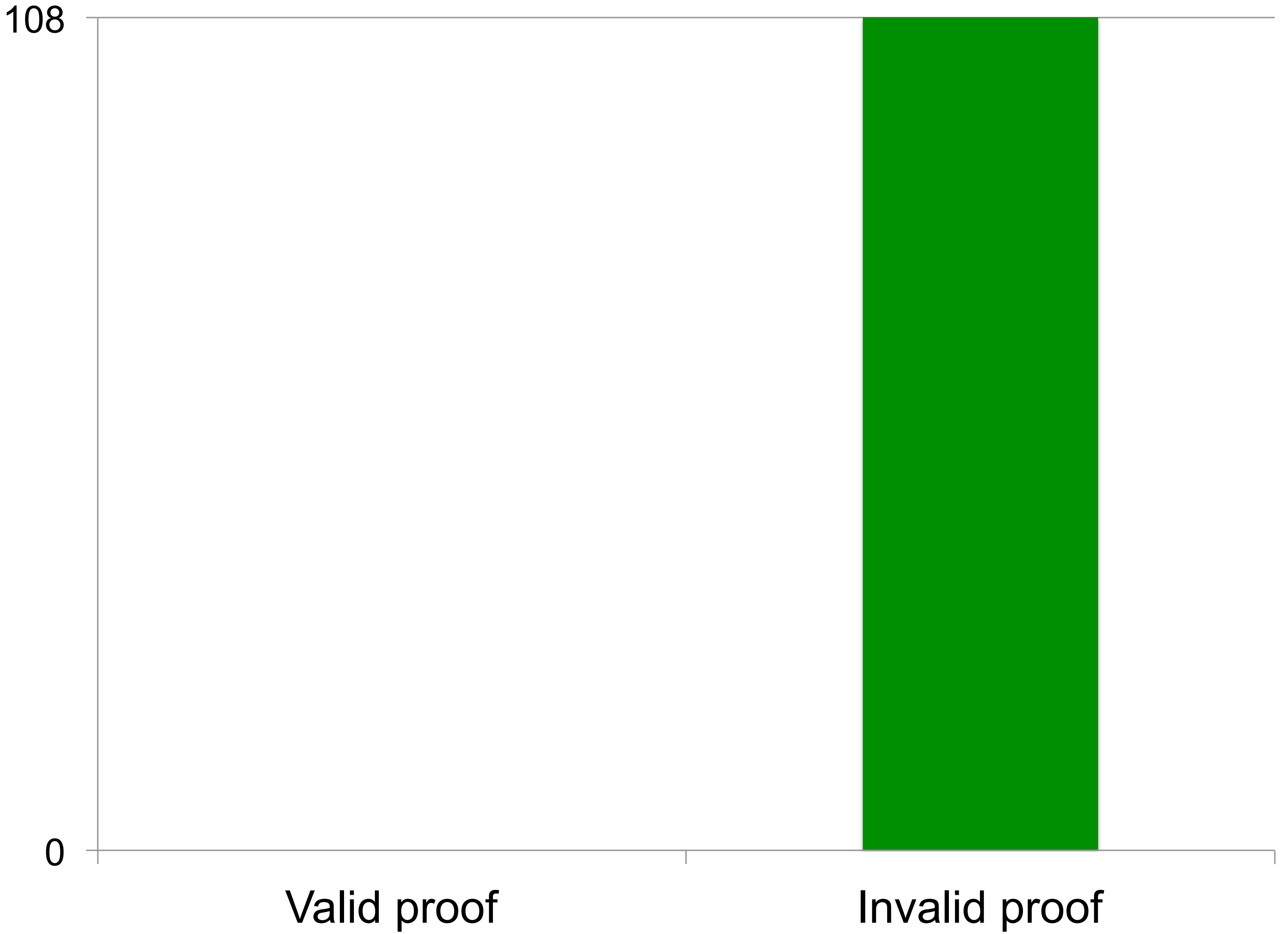

106
0
Valid in nearly all contexts
Valid but contextual
Invalid but contextual
Invalid in nearly all contexts

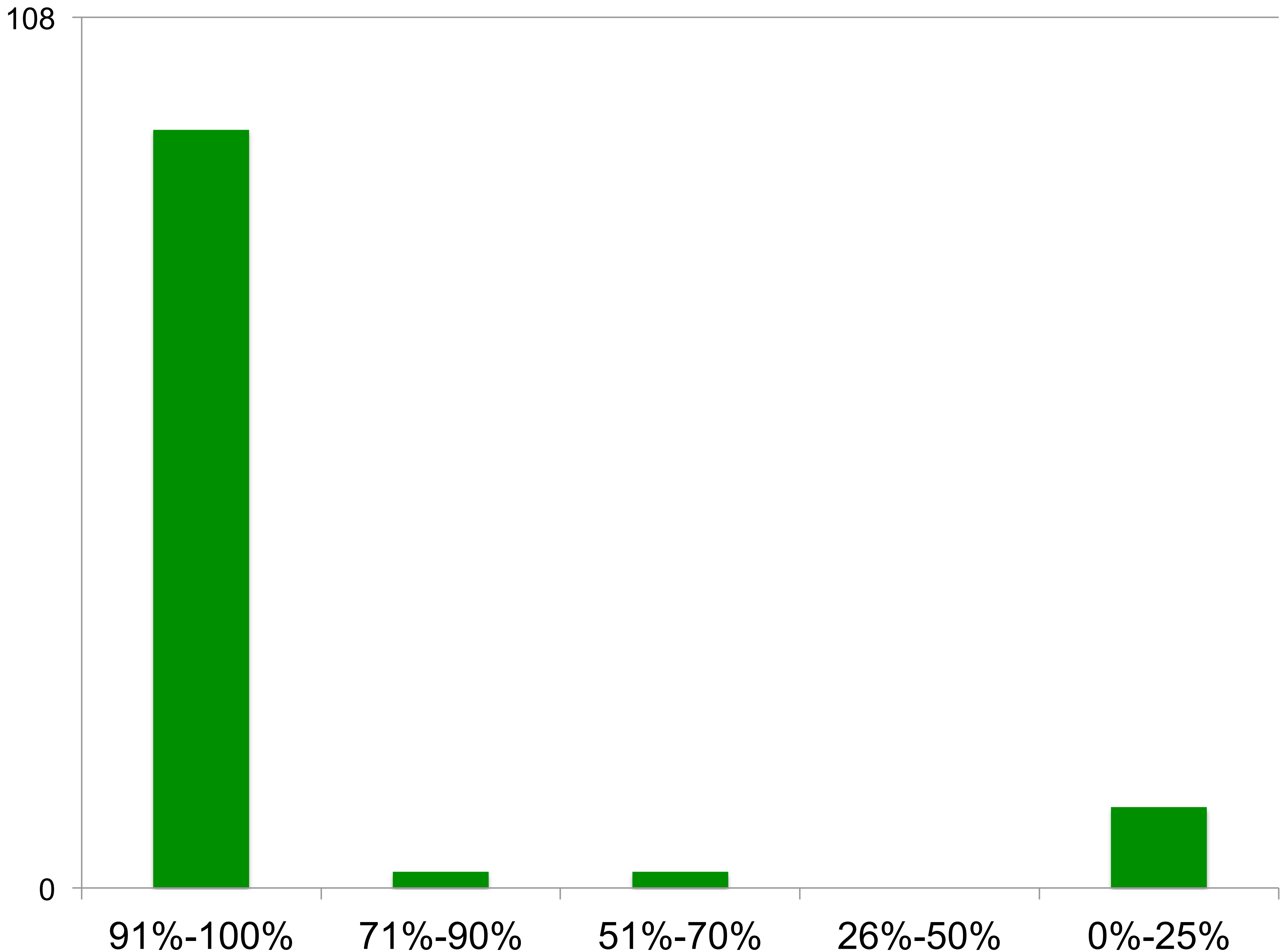
The following proof was generated by a mathematics major in an introduction to proof course.

Theorem: If n is an odd natural number, then n^2 is odd.

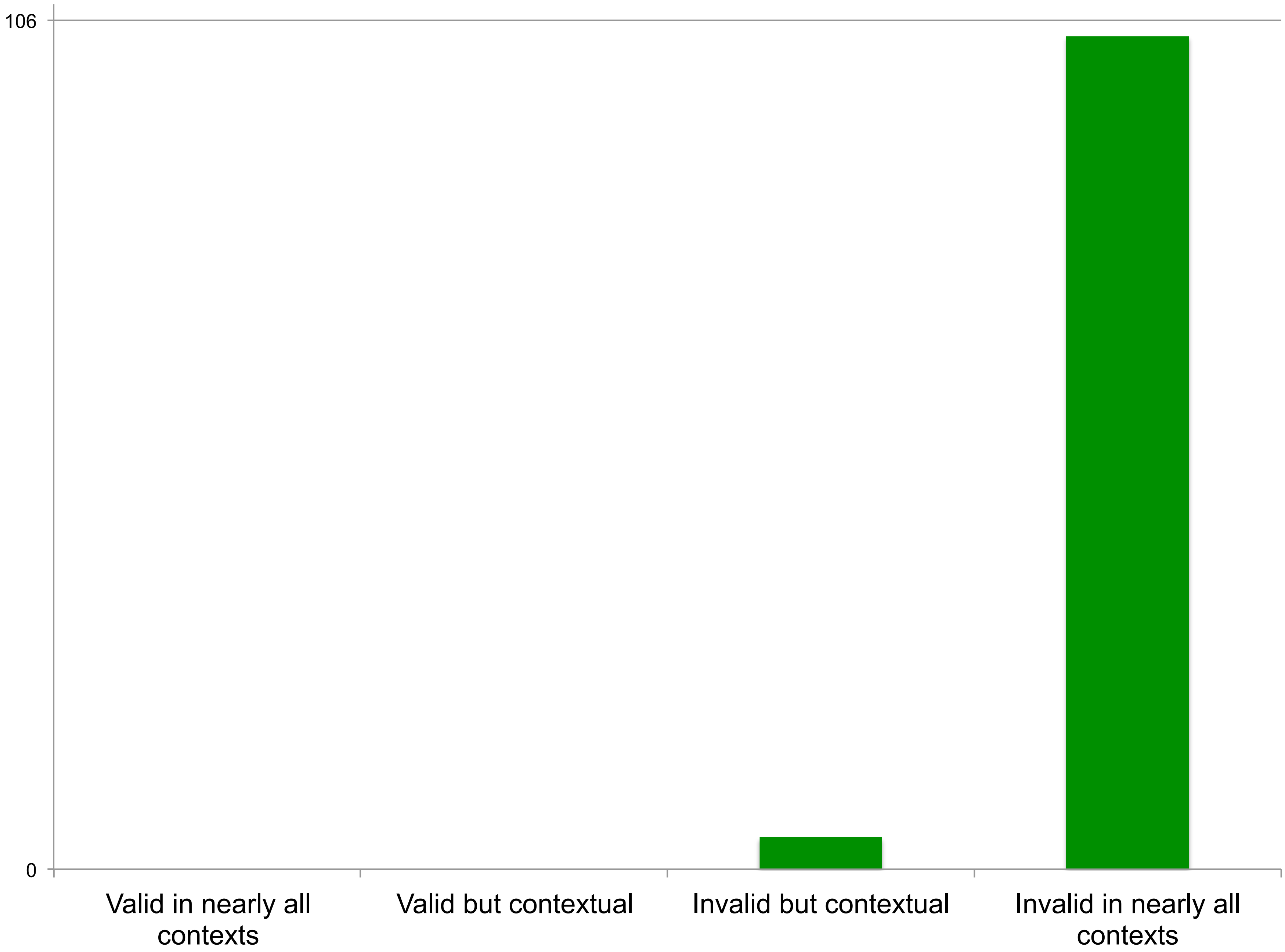
Student-generated proof: $1^2 = 1$, which is odd. $3^2 = 9$, which is odd. $5^2 = 25$, which is odd. I am convinced that this pattern will hold and the result will always be true. Therefore, whenever n is odd, n^2 is odd.

Example of student proof cited in K. Weber (2003), *Research Sampler on Undergraduate Mathematics Education*, published on-line by the MAA.


108
0
Valid proof
Invalid proof



108
0
91%-100%
71%-90%
51%-70%
26%-50%
0%-25%

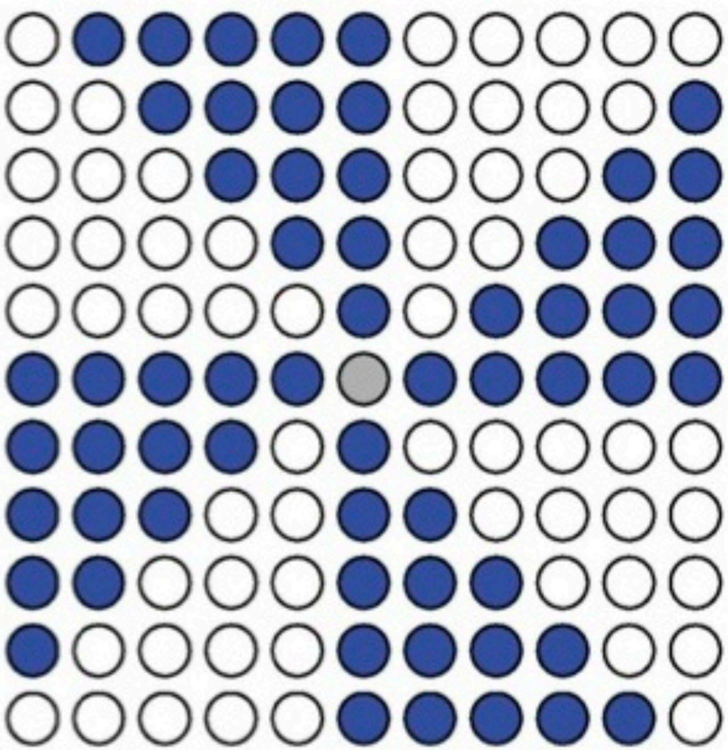

106
0
Valid in nearly all contexts
Valid but contextual
Invalid but contextual
Invalid in nearly all contexts

Cluster concepts: Predicted consequences

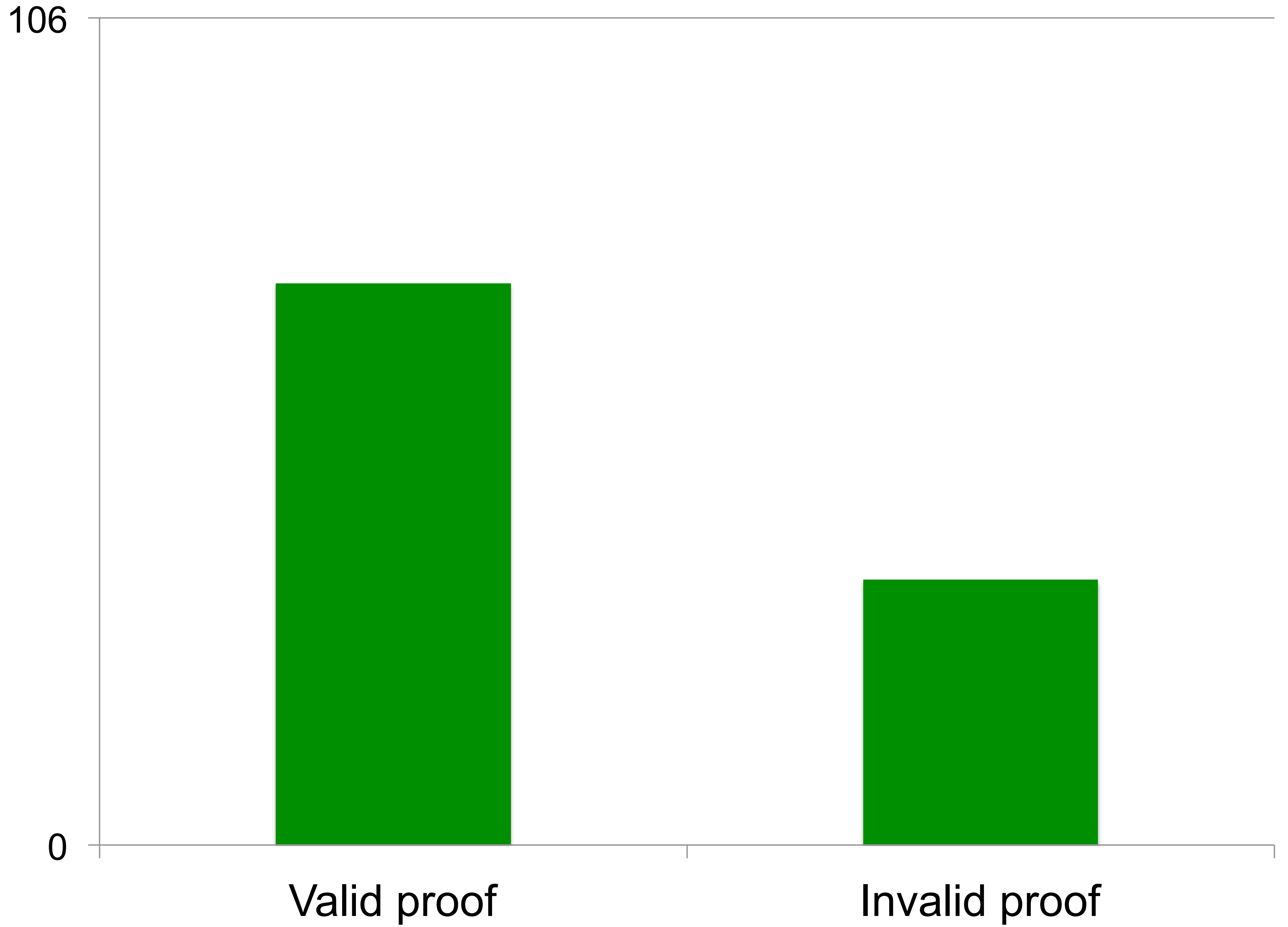
1. The prototypical proof satisfies all criteria. Proofs that satisfy all criteria would be better representatives of proof than those that satisfy some criteria and would not be controversial.
2. Proofs that only satisfy some would be controversial and spark disagreement.
3. There are compound words that qualify “proofs” that satisfy some criteria but not all criteria.
4. There are *default judgments* when you hear an argument is a proof– properties you think the argument is likely to have but are not necessarily sure of.
5. There is no single essence of proof.

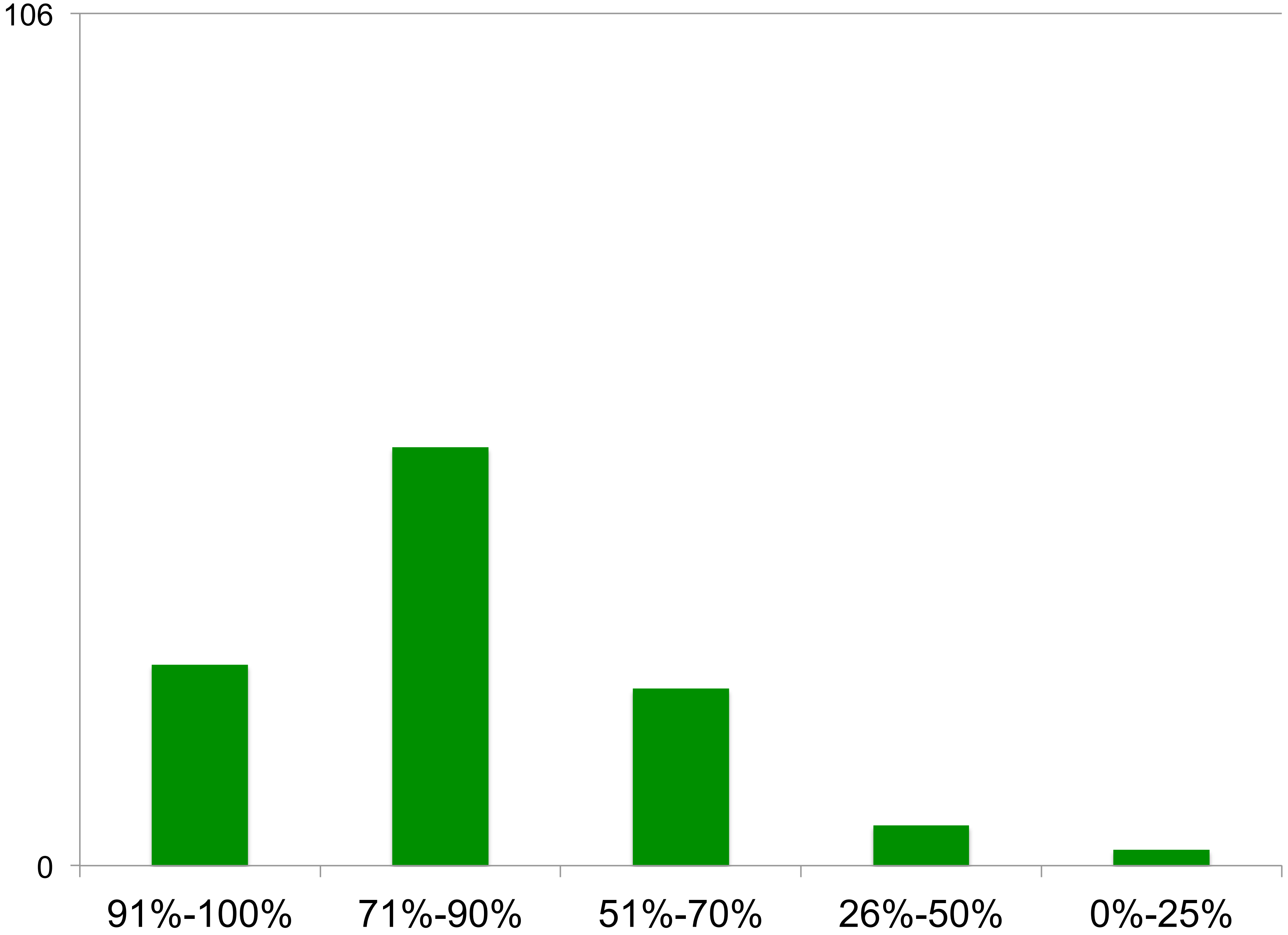
Theorem: If n is an odd natural number, then $n^2 \equiv 1 \pmod 8$

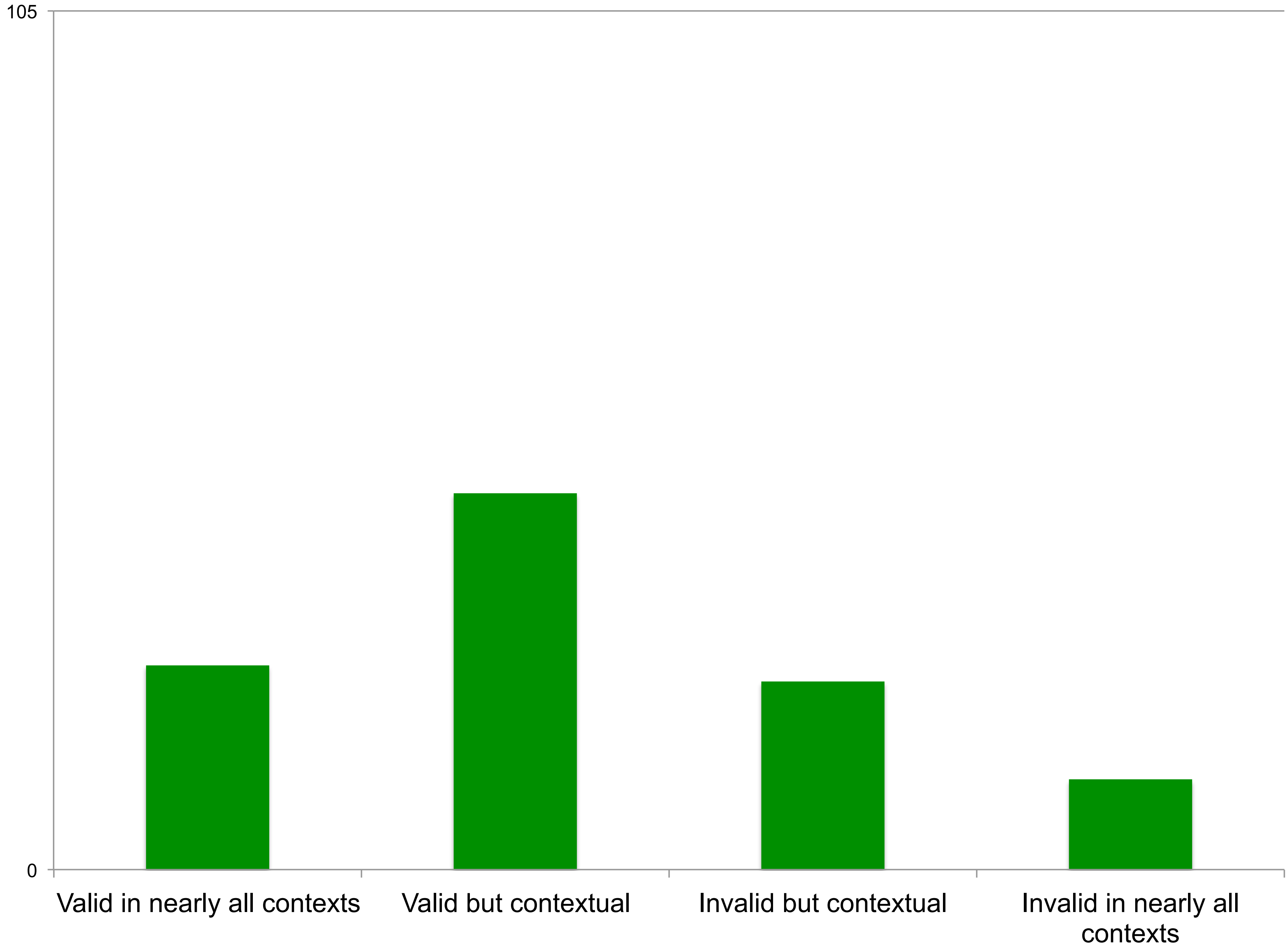
Proof: The proof is given in the picture below.



From R. Nelsen (2008). Visual gems in number theory. *Math Horizons*, 31, 7-9.


106
0
Valid proof
Invalid proof


106
0
91%-100%
71%-90%
51%-70%
26%-50%
0%-25%


105
0
Valid in nearly all contexts
Valid but contextual
Invalid but contextual
Invalid in nearly all contexts

Theorem: $\sum_{k=0}^{\infty}\left(\frac{-1}{4}\right)^k\left(\frac{2}{4k+1}+\frac{2}{4k+2}+\frac{1}{4k+3}\right)=\pi$

Proof: In this proof, we use *Mathematica* to assist with our computations.

Using *Mathematica,* we can verify that

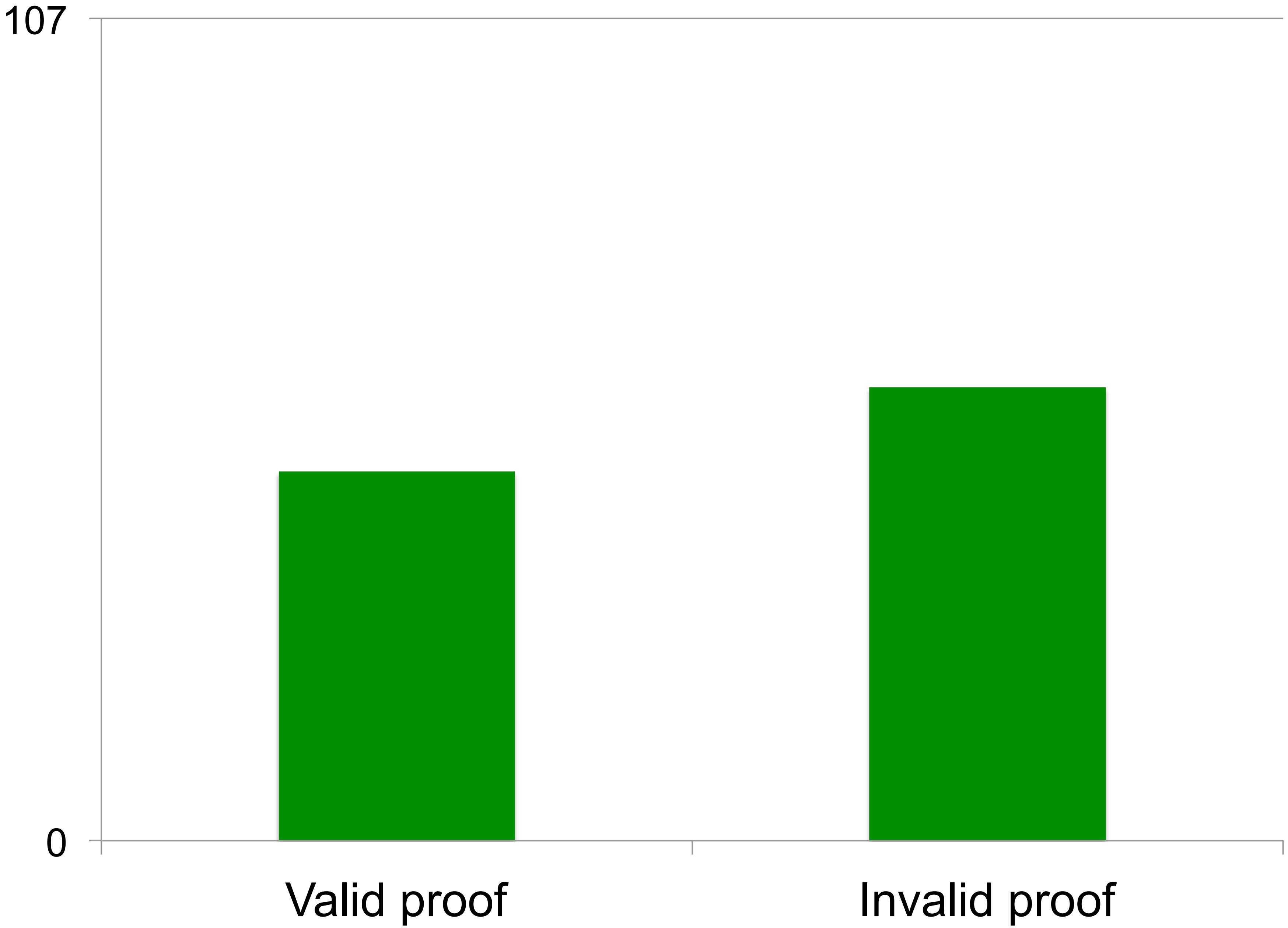
$$\sum_{k=0}^{\infty}\left(\frac{-1}{4}\right)^k\left(\frac{a_1}{4k+1}+\frac{a_2}{4k+2}+\frac{a_3}{4k+3}\right)=$$

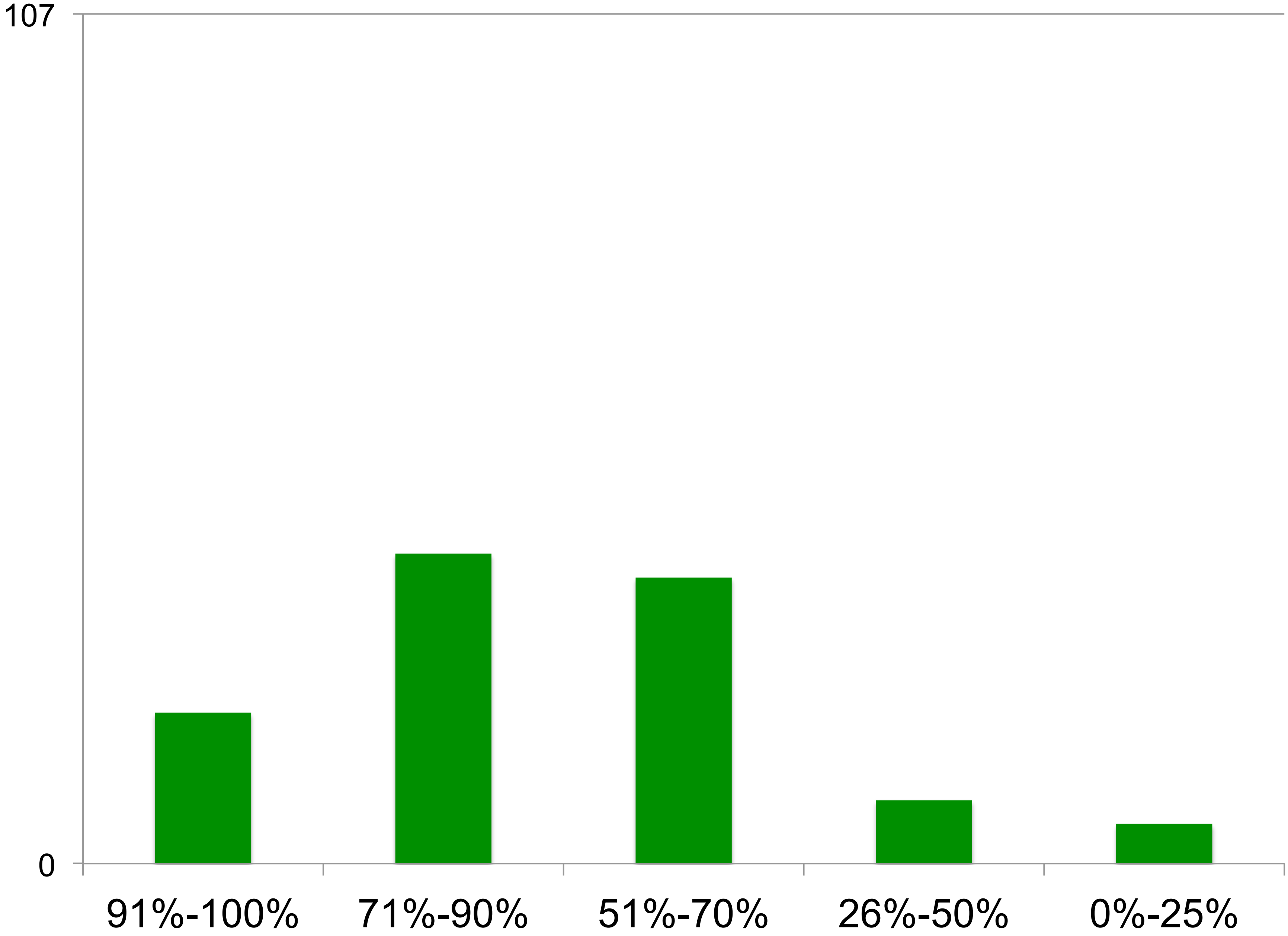
$$\frac{a_2}{2}\pi+\left(\frac{a_1}{2}-a_2+a_3\right)\arctan(2)+\left(\frac{a_1}{4}-\frac{a_3}{2}\right)\ln(5).$$

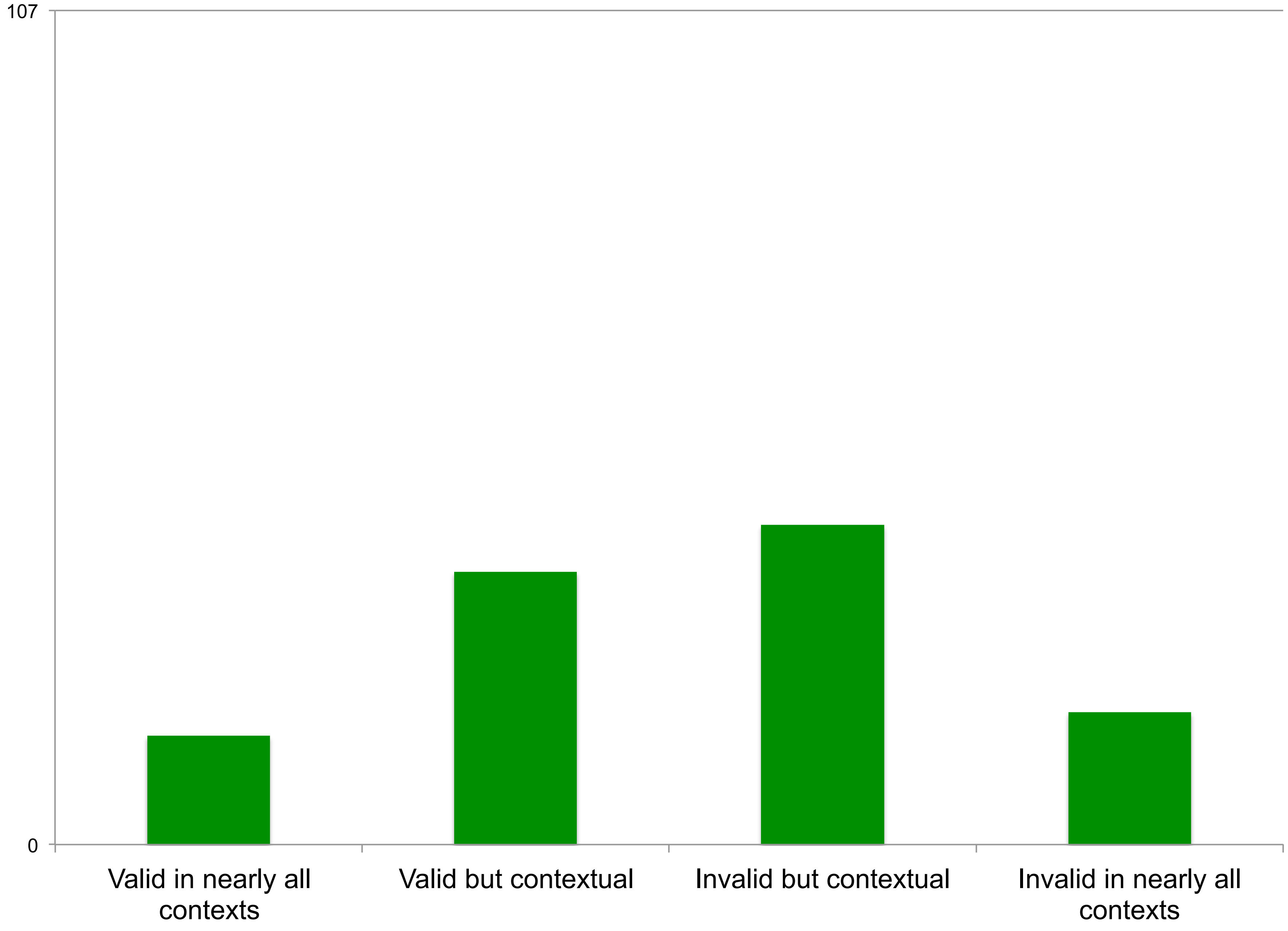
Setting a_1=2, a_2=2, and a_3=1 yields $\sum_{k=0}^{\infty}\left(\frac{-1}{4}\right)^k\left(\frac{2}{4k+1}+\frac{2}{4k+2}+\frac{1}{4k+3}\right)=\pi$, completing the proof.

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

(Note: The authors noted that *Mathematica* could provide a verification for the correctness of the computations above).


107
0
Valid proof
Invalid proof


107
0
91%-100%
71%-90%
51%-70%
26%-50%
0%-25%


107
0
Valid in nearly all contexts
Valid but contextual
Invalid but contextual
Invalid in nearly all contexts

Cluster concepts: Predicted consequences

1. The prototypical proof satisfies all criteria. Proofs that satisfy all criteria would be better representatives of proof than those that satisfy some criteria and would not be controversial.
2. Proofs that only satisfy some would be controversial and spark disagreement.
3. There are compound words that qualify “proofs” that satisfy some criteria but not all criteria.
4. There are *default judgments* when you hear an argument is a proof– properties you think the argument is likely to have but are not necessarily sure of.
5. There is no single essence of proof.

Default frames

- Minsky (1975) introduced the idea of *frames* as default assumptions about a new situation.
  - If we go to a fine restaurant, we *expect* to be seated and order are food, although we realize that it could be a buffet, prepare your own food, etc.

- The notion of cluster concept predicts the elements of the cluster are default options.

Default frames

- Participants were told that they came across a conjecture X in an old paper in your field, they asked a respected colleague about the status of the conjecture, and the colleague said, “X was proved by Smith”. What probability would they give to the following?
  - If they read the proof, they’d be certain X was true
  - If they read the proof, they’d have high confidence that X was true
  - They would be capable of filling in any gap in Smith’s proof of X
  - They could, in principle, remove any diagram from X without affecting its validity
  - Smith’s proof of X would be published in a respected outlet

Default frames

	Avg	100%	51-99%	0-50%
Certain of X	81	20%	66%	14%
Confident in X	92	43	55	2
Fill in all gaps	78	13	72	14
No diagram inf.	64	14	49	36
Published	73	8	75	17

Default frames

	Avg	100%	51-99%	0-50%
Certain of X	81	20%	66%	14%
Confident in X	92	43	55	2
Fill in all gaps	78	13	72	14
No diagram inf.	64	14	49	36
Published	73	8	75	17

Default frames

	Avg	100%	51-99%	0-50%
Certain of X	81	20%	66%	14%
Confident in X	92	43	55	2
Fill in all gaps	78	13	72	14
No diagram inf.	64	14	49	36
Published	73	8	75	17

Default frames

	Avg	100%	51-99%	0-50%
Certain of X	81	20%	66%	14%
Confident in X	92	43	55	2
Fill in all gaps	78	13	72	14
No diagram inf.	64	14	49	36
Published	73	8	75	17

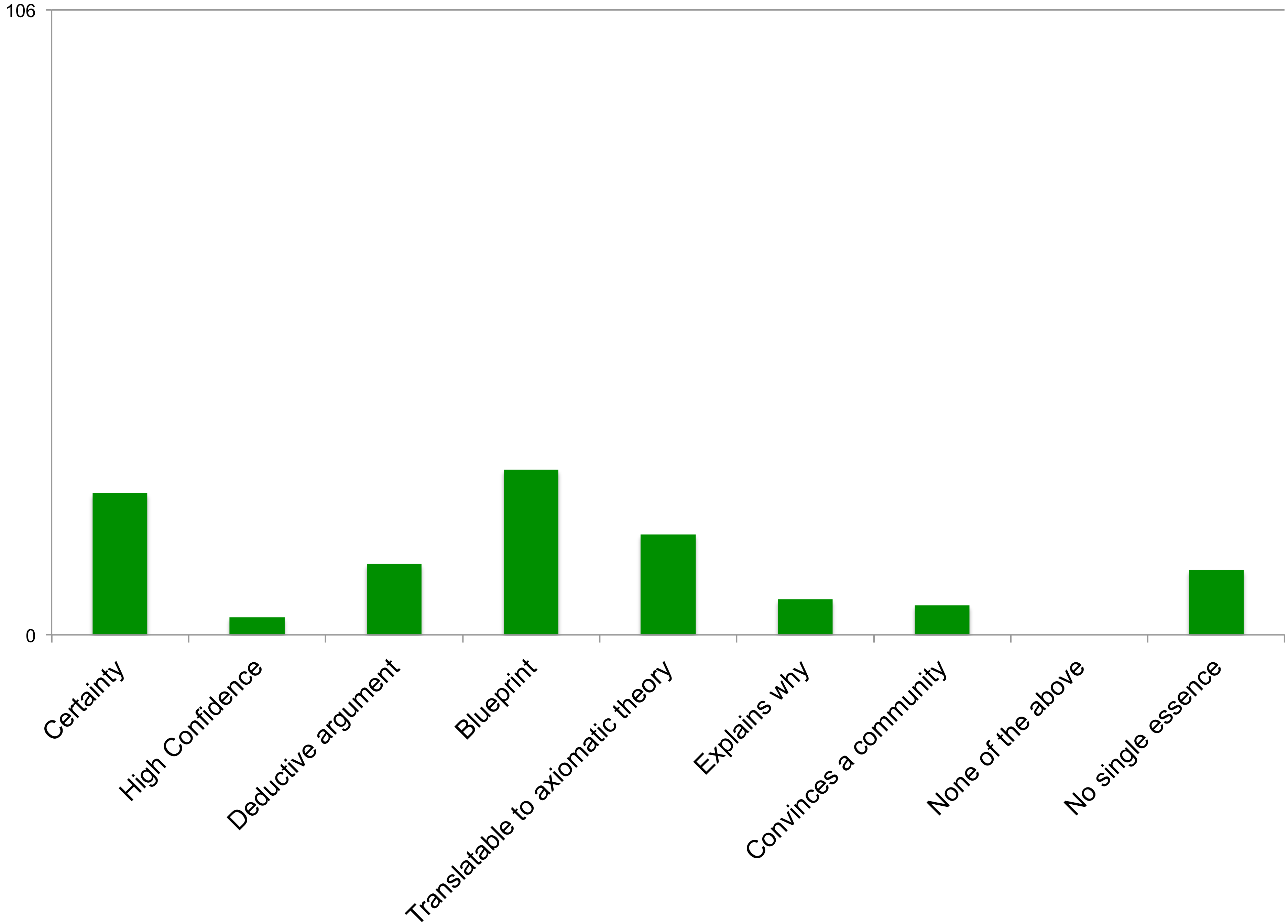
Cluster concepts: Predicted consequences

1. The prototypical proof satisfies all criteria. Proofs that satisfy all criteria would be better representatives of proof than those that satisfy some criteria and would not be controversial.
2. Proofs that only satisfy some would be controversial and spark disagreement.
3. There are compound words that qualify “proofs” that satisfy some criteria but not all criteria.
4. There are *default judgments* when you hear an argument is a proof– properties you think the argument is likely to have but are not necessarily sure of.
5. There is no single essence of proof.

A single essence of proof?

Which of the following best captures the essence of proof?

- Proof provides certainty in a theorem
- Proof provides high degree of confidence
- Deductive argument w each step being logical consequence of previous ones
- Proof is a blueprint I can use to write a complete formal proof myself
- A proof, in principle, can be translated into a formal argument in an axiomatized theory
- Proof explains why a theorem is true
- Proof convinces a particular math community a result is true
- None of the above captures the essence of proof
- There is no single essence of proof



106
0
Certainty
High Confidence
Deductive argument
Blueprint
Translatable to axiomatic theory
Explains why
Convinces a community
None of the above
No single essence

Summary

- I proposed that proof can be thought of as Lakoff's cluster concept.

- Empirical verified consequences of this:
  - A remarkable level of agreement among a prototypical and a clearly problematic proof
  - Disagreement amongst proofs that satisfied some aspects of cluster concept but not others; proof is viewed as contextual and individually-based.
  - The majority of mathematicians viewed the cluster concept as defaults
  - A single essence of proof was not agreed upon

- It might be validity of proof is agreed upon in *prototypical* cases but disagreement is made on cases where arguments satisfy some, but not all, aspects of the cluster.

Consequences of a cluster concept

- Fallis (1997, 2002) argued that probabilistic proofs (Rabin's primality tests) should be epistemologically on par with proofs.
  - Probabilistic proofs do not provide certainty, but proofs with errors are accepted so they cannot provide certainty either.
  - Probabilistic proofs do not explain why, but exhaustive proofs do not provide explanation either.
  - Probabilistic proofs are not *a priori* but neither are computer-assisted proofs

"I have been considering epistemic objectives one at a time. However, it is conceivable that some disjunction of these epistemic objectives might explain the rejection of probabilistic proofs. Unfortunately, it is not immediately clear what this disjunction could be or that such a disjunction would provide a satisfying explanation" (Fallis, 2002).

Consequences of a cluster concept

- For math education, perhaps proof is *not* a useful construct for teaching younger children
  - For mathematicians it is, because the elements of the cluster correlate highly, they share default expectations, and they have technical competence. Hence, proof* are unusual.
  - For students, there is no reason to suppose the elements are correlated, expectations are being learned, and they have naturally error prone. Nearly *every* student proof will be a proof*.
  - The individual components of the model may be more basic to students.
  - Consequently, might it be better to speak of "convincing arguments", "comprehensible (clear) arguments", "deductive arguments", "algebraic arguments", etc.?

Thank you

Contact me:

keith.weber@gse.rutgers.edu