How to Write a Proof

Leslie Lamport

Microsoft Research

“And now for something completely different.”

Making it Easier to Write Proofs

Harder
Making it ~~Easier~~ to Write Proofs

Harder
Making it ~~Easier~~ to Write Proofs

It's already too easy.

Theorem We have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} |\widehat{\mu}(j)|^2 = \left(\int_M k(x)\, dx \right)^{-1} \sum_j k(x_j) |a_j|^2$$

where $k(x) = \mathrm{vol}\{\xi \in T^*_x(M) : a_1(x,\xi) \leq 1\}$.

Proof Obvious by inspection. Q.E.D.

YOU WANT PROOF? I'LL GIVE YOU PROOF!
$+\frac{K}{8.1}\,c'\sqrt{\frac{D^5}{9}}+H6$
S. Harris

Making it Easier to ~~Write~~ Read Proofs

Making it Easier to ~~Write~~ Read Proofs

And Find Mistakes

A Brief History of Notation

17th century

There do not exist four positive integers, the last being greater than two, such that the sum of the first two, each raised to the power of the fourth, equals the third raised to that same power.

A Brief History of Notation

17th century

There do not exist four positive integers, the last being greater than two, such that the sum of the first two, each raised to the power of the fourth, equals the third raised to that same power.

20th century

There do not exist positive integers x, y, z, and n, with $n > 2$, such that $x^n + y^n = z^n$.

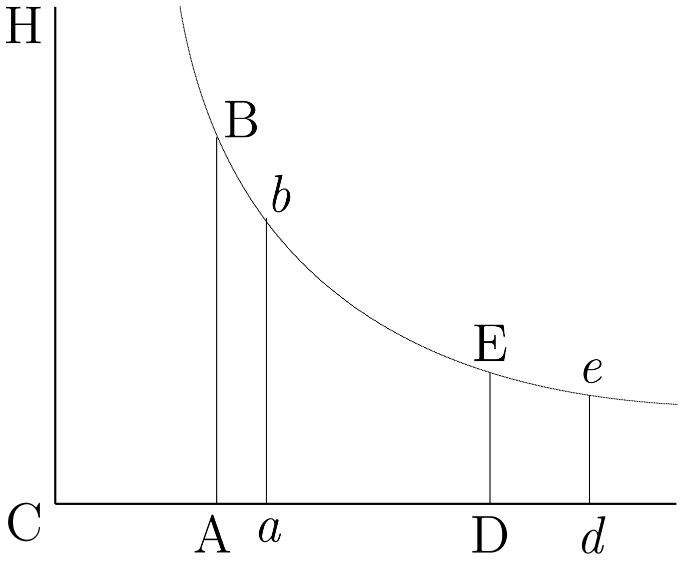
PROPOSITION VI. THEOREM IV

Homogeneous and equal spherical bodies, opposed by resistances that are as the square of the velocities, and moving on by their innate force only, will, in times which are inversely as the velocities at the beginning, describe equal spaces, and lose parts of their velocities proportional to the wholes.

Isaac Newton

Sir Isaac Newton's Mathematical Principles of Natural Philosophy. Translated by Andrew Motte, revised by Florian Cajori. Volume One, page 246 Greenwood Press (1962)

To the rectangular asymptotes CD, CH describe any hyperbola B*b*E*e*, cutting the perpendiculars AB, *ab*, DE, *de* in B, *b*, E, *e*; let the initial velocities be represented by the perpendiculars AB, DE, and the times by the lines A*a*, D*d*. Therefore as A*a* is to D*d*, do (by the hypothesis) is DE to AB, and so (from the nature of the hyperbola) is CA to CD; and, by composition, so is C*a* to C*d*. Therefore, the areas AB*ba*, DE*ed*, that is, the spaces described, are equal among themselves, and the first velocities AB, DE are proportional to the last *ab*, *de*; and therefore, by subtraction, proportional to the parts of the velocities lost, AB–*ab*, DE–*de*. Q.E.D.



20th century

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

John Kelley

Theorem 20 on page 28 of *General Topology* by John L. Kelley, van Nostrand (1955). Reprinted by Springer Verlag.

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

Why are formulas easier to read than prose?

The sum of the first two numbers, each raised to the power of the fourth, equals the third raised to that same power

The sum of the first two numbers, each raised to the power of the fourth, equals the third raised to that same power

+ naming

The sum of x and y, each raised to the power n, equals z raised to the power n.

The sum of x and y, each raised to the power n, equals z raised to the power n.

$+$ **structure**

x to the power n

plus **EQUALS** z to the power n

y to the power n

In addition to developing the students' intuition about the beautiful concepts of analysis, it is surely equally important to persuade them that precision and rigor are neither deterrents to intuition, nor ends in themselves, but the natural medium in which to formulate and think about mathematical questions.

Michael Spivak

Calculus by Michael Spivak. W. A. Benjamin (1967), page vii.

Corollary If $f'(x) > 0$ for all x in an interval, then f is increasing on the interval.

Proof Let a and b be two points in the interval with $a < b$. Then there is some x in (a, b) with

$$f'(x) = \frac{f(b) - f(a)}{b - a}.$$

But $f'(x) > 0$ for all x in (a, b), so

$$\frac{f(b) - f(a)}{b - a} > 0.$$

Since $b - a > 0$ it follows that $f(b) > f(a)$. ▌

Adapted from Corollary 3 on page 170 of *Calculus* by Michael Spivak, W. A. Benjamin (1967)

Corollary If $f'(x) > 0$ for all x in an interval, then f is increasing on the interval.

Hypotheses 1. $f'(x) > 0$ for all x in the interval.

2. a, b in the interval and $a < b$.

To prove $f(a) < f(b)$

Corollary If $f'(x) > 0$ for all x in an interval, then f is increasing on the interval.

Hypotheses 1. $f'(x) > 0$ for all x in the interval.
2. a, b in the interval and $a < b$.

To prove $f(a) < f(b)$

1.	Choose x in (a, b) with $f'(x) = \dfrac{f(b) - f(a)}{b - a}$.	1.	The Mean Value Theorem and the hypotheses imply x exists.
2.	$f'(x) > 0$	2.	Step 1 and the hypotheses.
3.	$\dfrac{f(b) - f(a)}{b - a} > 0$	3.	Steps 1 and 2.
4.	$b - a > 0$	4.	Hypothesis 2.
5.	$f(a) < f(b)$	5.	Steps 3 and 4.

Theorem There does not exist r in $\mathbf{Q}$ such that $r^2 = 2$.

Proof sketch: We assume $r^2 = 2$ for $r \in \mathbf{Q}$ and obtain a contradiction. Writing $r = m/n$, where m and n have no common divisors, we deduce from $(m/n)^2 = 2$ and the lemma that both m and n must be divisible by 2.

Theorem There does not exist r in $\mathbf{Q}$ such that $r^2 = 2$.

Assume: 1. $r \in \mathbf{Q}$
2. $r^2 = 2$

Prove: False

1. Choose m, n in $\mathbf{Z}$ such that
   (a) $\gcd(m, n) = 1$
   (b) $r = (m/n)$

2. 2 divides m.

3. 2 divides n.

4. Q.E.D.

3. 2 divides n.

3.1. Choose p in **Z** such that $m = 2p$.

Proof: By 2.

3.2. $n^2 = 2p^2$

Proof:
$$\begin{aligned} 2 &= (m/n)^2 && \text{[1(a) and hypoth. 0:2]} \\ &= (2p/n)^2 && \text{[3.1]} \\ &= 4p^2/n^2 && \text{[Algebra]} \end{aligned}$$

from which the result follows easily by algebra.

3.3. Q.E.D.

Proof: By 3.2 and the lemma.

3. 2 divides n.

. . .

3.2. $n^2 = 2p^2$

3.2.1. $(m/n)^2 = 2$

Proof: By 1(a) and hypothesis 0:2.

3.2.2. $(m/n)^2 = (2p/n)^2$

Proof: By 3.1.

3.2.3. $(2p/n)^2 = 4p^2/n^2$

Proof: By simple algebra.

3.2.4. Q.E.D.

By 3.2.1–3.2.3.

. . .

3. 2 divides n.

. . .

3.2. $n^2 = 2p^2$

. . .

3.2.3. $(2p/n)^2 = 4p^2/n^2$

3.2.3.1. $(2p/n)^2 = (2p/n) \cdot (2p/n)$

Proof: Definition of $(\ldots)^2$.

3.2.3.2. $(2p/n) = (2p) \cdot (1/n)$

Proof: Definition of /.

3.2.3.3. $(2p/n)^2 =$
$((2p) \cdot (1/n)) \cdot ((2p) \cdot (1/n))$

Proof: By 3.2.3.1 and 3.2.3.2.

3.2.3.4. $((2p) \cdot (1/n)) \cdot ((2p) \cdot (1/n)) =$
$(2 \cdot 2) \cdot (p \cdot p) \cdot ((1/n) \cdot (1/n))$

Proof: By 3.2.3.3 and . . .

. . .

3. 2 divides n.

. . .

3.2. $n^2 = 2p^2$

. . .

3.2.3. $(2p/n)^2 = 4p^2/n^2$

3.2.3.3. $(2p/n)^2 =$
$((2p) \cdot (1/n)) \cdot ((2p) \cdot (1/n))$

3.2.3.3.1. Help! I'm running out of room.

. . .

3. 2 divides n.

. . .

3.2. $n^2 = 2p^2$

. . .

3.2.3. $(2p/n)^2 = 4p^2/n^2$

3.2.3.3. $(2p/n)^2 =$
$((2p) \cdot (1/n)) \cdot ((2p) \cdot (1/n))$

3.2.3.3.1. Was that 3.2.3.3.1 or 3.2.3.3.3.1?

. . .

⟨1⟩3. 2 divides n.

...

⟨2⟩2. $n^2 = 2p^2$

...

⟨3⟩3. $(2p/n)^2 = 4p^2/n^2$

...

⟨4⟩3. $(2p/n)^2 =$
$((2p) \cdot (1/n)) \cdot ((2p) \cdot (1/n))$

⟨5⟩1. Whew, that's better.

⟨6⟩1. Now I have lots...

⟨7⟩1. and lots of room.

Proof: ...

...

Anatomy of a Proof Step

$\langle 4\rangle$2. Assume: 1. Assumption $\langle 4\rangle$:1
2. Assumption $\langle 4\rangle$:2
Prove: Goal

statement

$\langle 5\rangle$1. 1st statement in proof of $\langle 4\rangle$2
Proof of $\langle 5\rangle$1

$\langle 5\rangle$2. 2nd statement in proof of $\langle 4\rangle$2
Proof of $\langle 5\rangle$2

⋮

$\langle 5\rangle$7. Q.E.D.
Proves the Goal of $\langle 4\rangle$2

proof

Mathematical Induction

⟨4⟩2. For all n in **N**: $P(n)$

⟨5⟩1. $P(0)$

Proof: . . .

⟨5⟩2. Assume: 1. $n \in \mathbf{N}$
2. $P(n)$
Prove: $P(n+1)$

Proof: . . .

⟨5⟩3. Q.E.D.

Proof: ⟨5⟩1, ⟨5⟩2, and induction.

Proof by Cases

⟨4⟩2. Case: Assumption

is an abbreviation for

⟨4⟩2. Assume: Assumption
Prove: Q.E.D.

$\langle 3 \rangle 3$. Assume: $n \in \mathbf{Z}$

Prove: ...

$\vdots$

$\langle 4 \rangle 6$. Case: $n \geq 0$

Proof: ...

$\langle 4 \rangle 7$. Case: $n < 0$

Proof: ...

$\langle 4 \rangle 8$. Q.E.D.

Proof: $\langle 4 \rangle 6$, $\langle 4 \rangle 7$, and assumption $\langle 3 \rangle$.

How Deep?

How carefully should you write your proof?

How Deep?

How carefully should you write your proof?

As carefully as necessary.

How Deep?

How carefully should you write your proof?

As carefully as necessary.

No long paragraphs.

How Deep?

How carefully should you write your proof?

As carefully as necessary.

No long paragraphs.

Deeper than you want to.

Assume: 1. (Π, L_1) is machine closed.

2. $\Pi \wedge L_1$ implies L_2.

Prove: (Π, L_2) is machine closed.

Proof:

$$\begin{array}{rlr} \Pi & = \mathcal{C}(\Pi \wedge L_1) & [\langle 0 \rangle.1] \\ & \subseteq \mathcal{C}(\Pi \wedge L_2) & [\langle 0 \rangle.2 \text{ and monotonicity of closure}] \\ & \subseteq \mathcal{C}(\Pi) & [\text{monotonicity of closure}] \\ & = \Pi & [\langle 0 \rangle.1, \text{ which implies } \Pi \text{ closed}] \end{array}$$

This proves that $\Pi = \mathcal{C}(\Pi \wedge L_2)$. □

Assume: 1. Φ and Π are safety properties.
2. $(\Phi \mathrel{-\!\!\triangleright} \Pi, L_1)$ is μ-machine realizable.
3. (true, L_2) is μ-machine realizable.
4. $\Phi \wedge \Pi \wedge L_1$ implies L_2.

Prove: $(\Phi \mathrel{-\!\!\triangleright} \Pi, L_2)$ is μ-machine realizable.

⟨1⟩1. Assume: ρ is a finite behavior such that $\rho \models (\Phi \mathrel{-\!\!\triangleright} \Pi)$.
Prove: There exists a μ-strategy f such that

$$\mathcal{O}_\mu(f, \rho) \subseteq (\Phi \mathrel{-\!\!\triangleright} \Pi) \wedge L_2.$$

⟨2⟩1. Choose μ-strategy h such that $\mathcal{O}_\mu(h, \rho) \subseteq (\Phi \mathrel{-\!\!\triangleright} \Pi) \wedge L_1$.

Proof: Assumption ⟨0⟩.2, assumption ⟨1⟩, and Lemma 2.

⟨2⟩2. For all τ, choose a μ-strategy g_τ such that $\mathcal{O}_\mu(g_\tau, \tau) \subseteq L_2$.

Proof: Assumption ⟨0⟩.3 and Lemma 2.

⟨2⟩3. Let $f(\tau) = h(\tau)$ for $\tau \models \Phi$, and otherwise let $f(\tau) = g_\eta(\tau)$ where η is the shortest prefix of τ such that $\eta \not\models \Phi$. Then f is a μ-strategy.

Proof: By steps ⟨2⟩1 and ⟨2⟩2 (which say that h and g_η are μ-strategies).

⟨2⟩4. $\mathcal{O}_\mu(f, \rho) \subseteq (\Phi \mathrel{-\!\!\triangleright} \Pi) \wedge L_2$

Assume: $\tau \in \mathcal{O}_\mu(f, \rho)$
Prove: $\tau \in (\Phi \mathrel{-\!\!\triangleright} \Pi) \wedge L_2$

⟨3⟩1. Case: $\tau \in \Phi$

⟨4⟩1. $\tau \in \mathcal{O}_\mu(h, \rho)$

Proof: By assumption ⟨2⟩, since the case assumption ⟨3⟩ and the definition of f (step ⟨2⟩3) imply $f(\tau|_n) = h(\tau|_n)$ for all n.

$\langle 4 \rangle$2. $\tau \in (\Phi \mathrel{-\!\!\rhd} \Pi) \wedge L_1$

Proof: By step $\langle 4 \rangle$1 and the choice of h (step $\langle 2 \rangle$1).

$\langle 4 \rangle$3. $\tau \in \Phi \wedge \Pi \wedge L_1$

Proof: By step $\langle 4 \rangle$2 and case assumption $\langle 3 \rangle$.

$\langle 4 \rangle$4. Q.E.D.

Proof: By step $\langle 4 \rangle$3 and assumption $\langle 0 \rangle$.4.

$\langle 3 \rangle$2. Case: $\tau \notin \Phi$

$\langle 4 \rangle$1. Let n be the least integer such that $\tau|_n \notin \Phi$.

Proof: Such an n exists by case assumption $\langle 3 \rangle$ and assumption $\langle 0 \rangle$.1.

$\langle 4 \rangle$2. $\tau \in \mathcal{O}_\mu(g_{\tau|_n}, \tau|_n)$

Proof: By step $\langle 4 \rangle$1, assumption $\langle 2 \rangle$, and the definition of f (step $\langle 2 \rangle$3).

$\langle 4 \rangle$3. $\tau \in L_2$

Proof: By steps $\langle 4 \rangle$2 and the choice of $g_{\tau|_n}$ (step $\langle 2 \rangle$2).

$\langle 4 \rangle$4. Q.E.D.

Proof: By steps $\langle 4 \rangle$3 and case assumption $\langle 3 \rangle$.

$\langle 3 \rangle$3. Q.E.D.

Proof: By steps $\langle 3 \rangle$1 and $\langle 3 \rangle$2.

$\langle 2 \rangle$5. Q.E.D.

Proof: By steps $\langle 2 \rangle$3 and $\langle 2 \rangle$4.

$\langle 1 \rangle$2. Q.E.D.

Proof: The result follows immediately from step $\langle 1 \rangle$1 and Lemma 2.

. . .

$\langle 4\rangle 1.$ $(\mathit{Inv} \wedge \langle \mathit{NowT}\rangle_{now} \wedge \mathcal{M}) \Rightarrow (\langle \mathit{NowT}\rangle_{now} \wedge \mathcal{N}^{t})$

$\langle 5\rangle 1.$ $\mathit{Inv} \wedge \langle \mathit{NowT}\rangle_{now} \Rightarrow \mathit{RTact}_{v}$

Proof: $\langle 1\rangle 3$ (*Inv*.1 and *Inv*.2).

$\langle 5\rangle 2.$ $\forall i \in I : \langle \mathit{NowT}\rangle_{now} \Rightarrow \mathit{MinTact}(t_i, \mathcal{A}_i, v)$

Proof: $\mathit{MinTact}(t_i, \mathcal{A}_i, v) = [\ldots]_v$ and $\langle \mathit{NowT}\rangle_{now}$ implies $v' = v$.

$\langle 5\rangle 3.$ $\forall j \in J : \mathit{Inv} \wedge \langle \mathit{NowT}\rangle_{now} \wedge \mathcal{M} \Rightarrow \mathit{MaxTact}(T_j)$

Assume: 1. $j \in J$

2. $\mathit{Inv} \wedge \langle \mathit{NowT}\rangle_{now} \wedge \mathcal{M}$

Prove: $\mathit{MaxTact}(T_j)$

$\langle 6\rangle 1.$ Case: $\mathit{Enabled}\ \langle \mathcal{A}_j\rangle_v$

$\langle 7\rangle 1.$ $\mathit{now}' \leq T_j$

Proof: *Inv*.1, *Inv*.2, $\langle \mathit{NowT}\rangle_{now}$, and the definitions of *NowT*, since case assumption $\langle 6\rangle$ and the definition of T imply $T_j \geq T$.

$\langle 7\rangle 2.$ Case: $j \in J_P$

$\langle 8\rangle 1.$ Case: $T_j' = T_j$

Proof: $\langle 7\rangle 1$ and the definition of $\mathit{MaxTact}(T_j)$.

$\langle 8\rangle 2.$ Case: $T_j' = \mathit{now} + \Delta_j$

By *Inv*.2, $\langle \mathit{NowT}\rangle_{now}$, and the definition of $\mathit{MaxTact}(T_j)$.

$\langle 8\rangle 3.$ Q.E.D.

$\langle 8\rangle 1$, $\langle 8\rangle 2$, and case assumption $\langle 7\rangle$, since case assumption $\langle 6\rangle$ and the definition of $\mathit{PTact}(T_j, \mathcal{A}_j, \Delta_j, v)$ imply that these are the only possibilities.

. . .

Advantages

- Makes it hard to publish incorrect results.

Advantages

- Makes it hard to publish incorrect results.

- Proofs easier to check.

Advantages

- Makes it hard to publish incorrect results.
- Proofs easier to check.
- Proofs easier to modify.

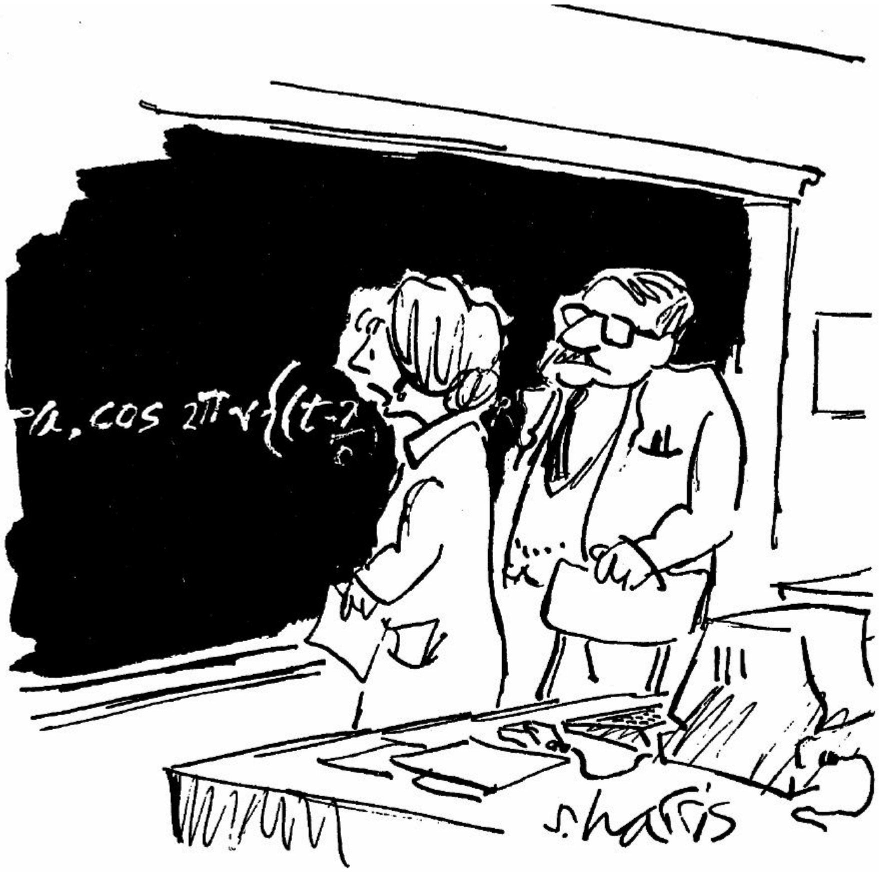
Disadvantages

- Makes it hard to publish incorrect results.

Disadvantages

- Makes it hard to publish incorrect results.

- Hard to see structure of long proof on paper.



"It's an excellent proof, but it lacks warmth and feeling."

A proof should be great literature.

art

A proof should be great ~~literature~~.

A proof should be great ~~literature~~ art.

The beauty of a proof lies in its logic, not in its prose.

The Future of Notation

The Future of Notation

17th century

There do not exist four positive integers, the last being greater than two, such that the sum of the first two, each raised to the power of the fourth, equals the third raised to that same power.

The Future of Notation

20th century

There do not exist positive integers x, y, z, and n, with $n > 2$, such that $x^n + y^n = z^n$.

The Future of Notation

21st century

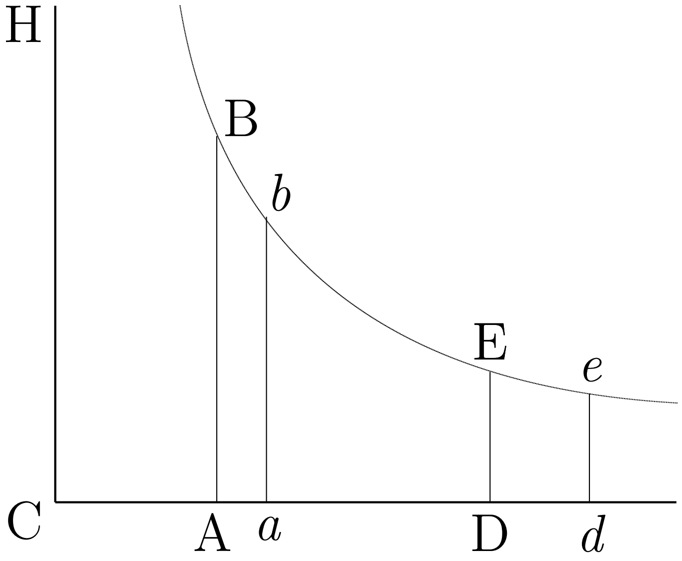
$\neg\,\exists\, x,\ y,\ z,\ n \in \mathbf{Z}^{+} : (n > 2) \wedge (x^n + y^n = z^n)$

The Future of Proofs

The Future of Proofs

17th century

To the rectangular asymptotes CD, CH describe any hyperbola B*b*E*e*, cutting the perpendiculars AB, *ab*, DE, *de* in B, *b*, E, *e*; let the initial velocities be represented by the perpendiculars AB, DE, and the times by the lines A*a*, D*d*. Therefore as A*a* is to D*d*, do (by the hypothesis) is DE to AB, and so (from the nature of the hyperbola) is CA to CD; and, by composition, so is C*a* to C*d*. Therefore, the areas AB*ba*, DE*ed*, that is, the spaces described, are equal among themselves, and the first velocities AB, DE are proportional to the last *ab*, *de*; and therefore, by subtraction, proportional to the parts of the velocities lost, AB–*ab*, DE–*de*. Q.E.D.



The Future of Proofs

20th century

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

The Future of Proofs

2002

Assume: 1. Φ and Π are safety properties.
2. $(\Phi \mathrel{-\!\!\triangleright} \Pi, L_1)$ is μ-machine realizable.
3. (true, L_2) is μ-machine realizable.
4. $\Phi \wedge \Pi \wedge L_1$ implies L_2.

Prove: $(\Phi \mathrel{-\!\!\triangleright} \Pi, L_2)$ is μ-machine realizable.

$\langle 1 \rangle$1. Assume: ρ is a finite behavior such that $\rho \models (\Phi \mathrel{-\!\!\triangleright} \Pi)$.
Prove: There exists a μ-strategy f such that

$$\mathcal{O}_\mu(f, \rho) \subseteq (\Phi \mathrel{-\!\!\triangleright} \Pi) \wedge L_2.$$

$\langle 2 \rangle$1. Choose μ-strategy h such that $\mathcal{O}_\mu(h, \rho) \subseteq (\Phi \mathrel{-\!\!\triangleright} \Pi) \wedge L_1$.
Proof: Assumption $\langle 0 \rangle$.2, assumption $\langle 1 \rangle$, and Lemma 2.

$\langle 2 \rangle$2. For all τ, choose a μ-strategy g_τ such that $\mathcal{O}_\mu(g_\tau, \tau) \subseteq L_2$.

Proof: Assumption $\langle 0 \rangle$.3 and Lemma 2.

$\langle 2 \rangle$3. Let $f(\tau) = h(\tau)$ for $\tau \models \Phi$, and otherwise let $f(\tau) = g_\eta(\tau)$ where η is the shortest prefix of τ such that $\eta \not\models \Phi$. Then f is a μ-strategy.

Proof: By steps $\langle 2 \rangle$1 and $\langle 2 \rangle$2 (which say that h and g_η are μ-strategies).

$\langle 2 \rangle$4. $\mathcal{O}_\mu(f, \rho) \subseteq (\Phi \mathrel{-\!\!\triangleright} \Pi) \wedge L_2$
Assume: $\tau \in \mathcal{O}_\mu(f, \rho)$
Prove: $\tau \in (\Phi \mathrel{-\!\!\triangleright} \Pi) \wedge L_2$

$\langle 3 \rangle$1. Case: $\tau \in \Phi$

The Future of Proofs

2010

The Future of Proofs

2010

Read on-line.

The Future of Proofs

2010

Read on-line.

Annotated with intuitive explanations.

The Future of Proofs

2010

Read on-line.

Annotated with intuitive explanations.

Must still be reduced to paper.

“That’s all folks!”