Formal Proofs in Geometry

Thomas C. Hales

October 6, 2006

Formal Proofs in Geometry

Thomas C. Hales

October 6, 2006

On the one hand, mathematicians often think of mathematics as a perfect, infallible realm . . . On the other hand, they make a great issue of the imperfections of computers.

The imperfect world of computers

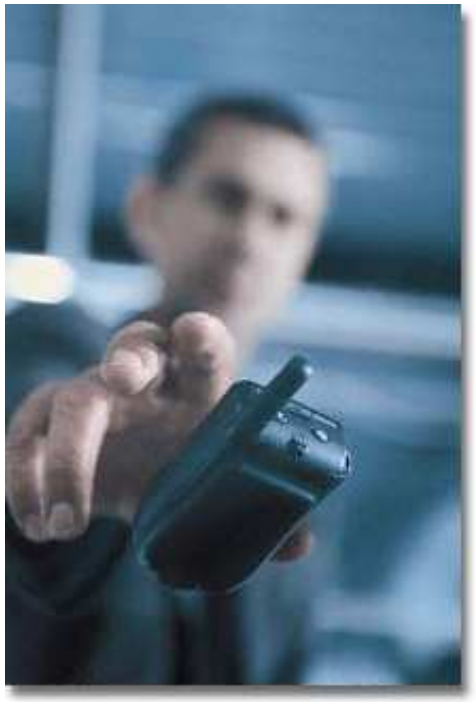
Chuck Barr received a library fine for 40 trillion dollars in 2001.

Severe Bugs

- Intel Pentium Chip division error.
- Ariane 5 explosion.

Bug Rates

- 1 bug per 100 statements.
- 600 bugs in a typical cell phone.
- Software testers do not try to get rid of all the bugs.



Bugs: Man or Machine?

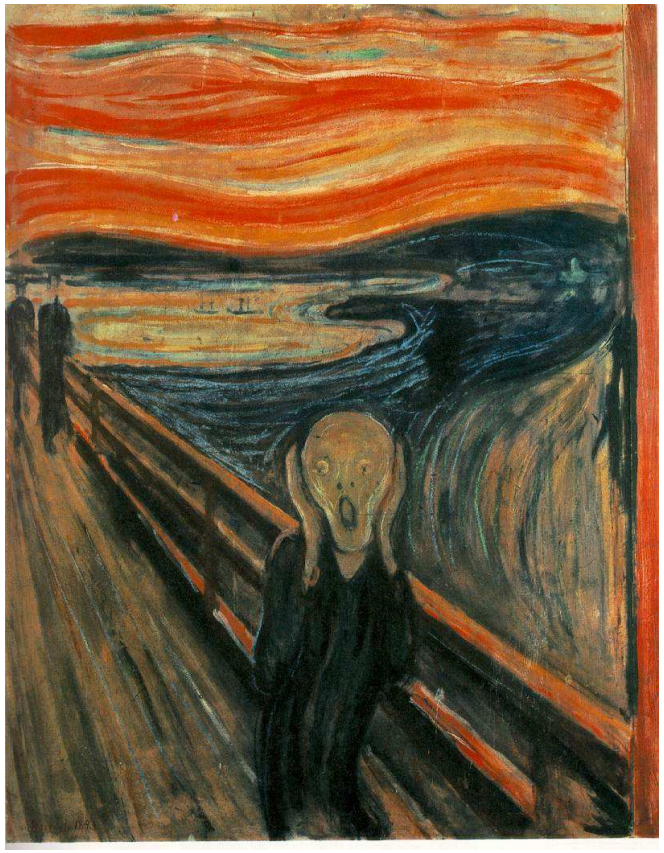
Our experience with computers is that once given a consistent set of instructions, they compute consistently. It’s just hard to give them a consistent set. – Georges Gonthier

The perfect realm of mathematics

Proof that $1 + 1 = 2$

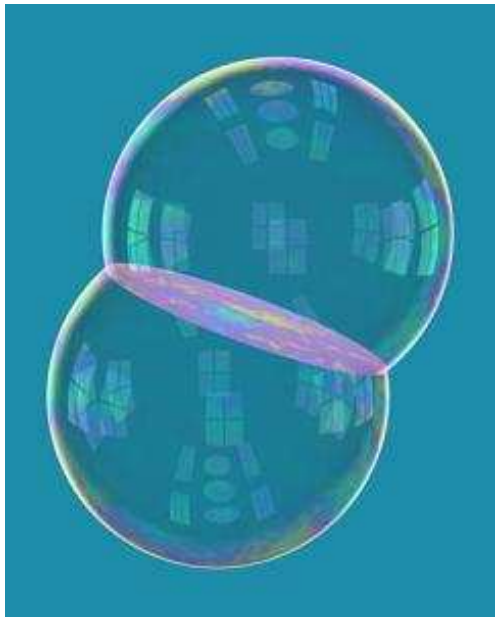
$$\begin{aligned} 1 + 1 &= 1 + \mathrm{SUC}\, 0 \\ &= \mathrm{SUC}(1 + 0) \\ &= \mathrm{SUC}\, 1 \\ &= 2 \end{aligned}$$

Mathematicians can be devastated by the slightest flaw in their work



Is math really as simple as $1 + 1$?

- Not all proofs are as simple as 1+1 = 2.
- Almgren's Big Paper is 1728 pages long.
- He spent 13 years writing the paper.
- It was still unpublished at the time of his death.



Proof. Using A.2.2(3)(4) and part 2(12)(13) we estimate

$$\begin{aligned}
&\mathrm{Dir}(\theta_3(\phi_{\varepsilon_1} * \phi_{\varepsilon_0} * F) + (1-\theta_3)(\phi_{\varepsilon_0} * F); \mathbf{B}^m(0, r_{12}) \sim \mathbf{U}^m(0, r_{14})) \\
\leq &\mathrm{Dir}(\phi_{\varepsilon_0} * \phi_{\varepsilon_1} * F; \mathbf{B}^m(0, r_{12}) \sim \mathbf{U}^m(0, r_{13})) \\
&+ (6 + 3\Gamma_{T.2}^2)\mathrm{Dir}(\phi_{\varepsilon_0} * F; \mathbf{B}^m(0, r_{12}) \sim \mathbf{U}^m(0, r_{15})) \\
\leq &(8512 + 4032\Gamma_{T.2}^2)\sigma[\mathrm{Lip}(\boldsymbol{\xi}(l))^2 + 2\boldsymbol{\alpha}(m)^{1/2}\Gamma_{3.10}^2 + \boldsymbol{\alpha}(m)\Gamma_{3.10}^4]E_*^{1+2Q(m+1)\alpha+4\alpha}
\end{aligned}$$

with

$$\begin{aligned}
&\mathrm{Lip}(\theta_3(\phi_{\varepsilon_1} * \phi_{\varepsilon_0} * F) + (1-\theta_3)(\phi_{\varepsilon_0} * F) | \mathbf{B}^m(0, r_{12}) \sim \mathbf{U}^m(0, r_{14})) \\
\leq &2\mathrm{Lip}(\boldsymbol{\xi}(l))\Gamma_{3.10}E_*^{\alpha}.
\end{aligned}$$

The conclusions of part 6 readily follow.

Part 7.

$$\mathcal{L}^m(W) \leq m\boldsymbol{\alpha}(m)\sigma\mathrm{Lip}(\boldsymbol{\xi}(l))[\boldsymbol{\alpha}(m)^{1/2} + \Gamma_{3.10}^2]E_*^{\alpha/4m}.$$

Proof. We estimate

$$\begin{aligned}
&\int_{x \in \mathbf{U}^m(0,1-\sigma\varepsilon_0)} \int_{y \in \mathbf{B}^m(0,\sigma\varepsilon_0)} |y|^{1-m} ||DF(x+y)|| \, d\mathcal{L}^m y \, d\mathcal{L}^m x \\
= &\int_{y \in \mathbf{B}^m(0,\sigma\varepsilon_0)} |y|^{1-m} \int_{x \in \mathbf{U}^m(0,1-\sigma\varepsilon_0)} ||DF(x+y)|| \, d\mathcal{L}^m x \, d\mathcal{L}^m y \\
\leq &\int_{y \in \mathbf{B}^m(0,\sigma\varepsilon_0)} |y|^{1-m} \int_{x \in \mathbf{U}^m(0,1)} ||DF(x)|| \, d\mathcal{L}^m x \\
= &\int_0^{\sigma\varepsilon_0} s^{1-m}(m\boldsymbol{\alpha}(m)) s^{m-1} \, ds \int_{x \in \mathbf{U}^m(0,1)} ||DF(x)|| \, d\mathcal{L}^m x \\
\leq &m\boldsymbol{\alpha}(m)\sigma\varepsilon_0 \int_{\mathbf{U}^m(0,1) \sim Z_\alpha} ||DF|| \, d\mathcal{L}^m + \int_{Z_\alpha} ||DF|| \, d\mathcal{L}^m \\
\leq &m\boldsymbol{\alpha}(m)\sigma\varepsilon_0[\boldsymbol{\alpha}(m)^{1/2} \left(\int_{\mathbf{U}^m(0,1) \sim Z_\alpha} ||DF||^2 \, d\mathcal{L}^m \right)^{1/2} + \mathcal{L}^m(Z_\alpha) \sup ||DF||] \\
\leq &m\boldsymbol{\alpha}(m)\sigma\varepsilon_0[\boldsymbol{\alpha}(m)^{1/2}\mathrm{Lip}(\boldsymbol{\xi}(l))\mathrm{Dir}(f; \mathbf{B}^m(0,1) \sim Z_\alpha)^{1/2} \\
&+ \Gamma_{3.10}E^{1-2Q(m+1)\alpha}\mathrm{Lip}(\boldsymbol{\xi}(l))\Gamma_{3.10}E^{\alpha}]
\end{aligned}$$

(by 1.4, 3.23(6), 3.22(m)(n))

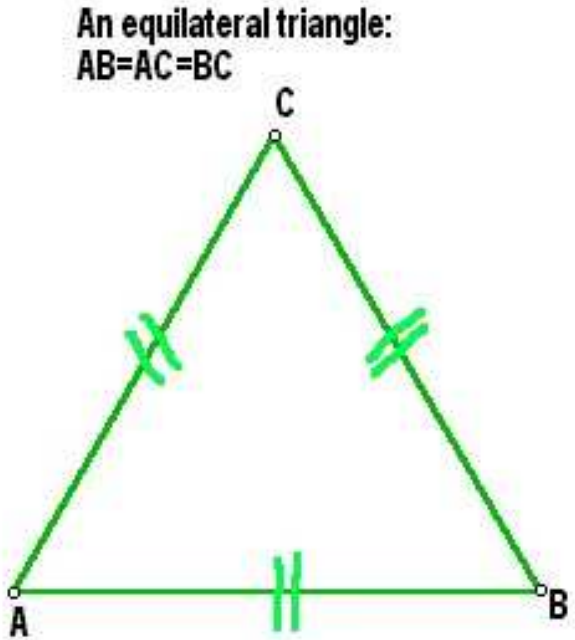
$$\leq m\boldsymbol{\alpha}(m)\sigma\mathrm{Lip}(\boldsymbol{\xi}(l))[\boldsymbol{\alpha}(m)^{1/2} + \Gamma_{3.10}^2]E_*^{1/2+\alpha/2m}.$$

We estimate further from the definition of W, either $\mathcal{L}^m(W) = 0$ or

$$\begin{aligned}
&\int_{x \in \mathbf{U}^m(0,1-\sigma\varepsilon_0) \cap W} \int_{y \in \mathbf{B}^m(0,\sigma\varepsilon_0)} |y|^{1-m} ||DF(x+y)|| \, d\mathcal{L}^m y \, d\mathcal{L}^m x \\
> &\mathcal{L}^m(W)E_*^{1/2+\alpha/4m}.
\end{aligned}$$

Euclid is our role model

- The first mistake in Euclid occurs in the very first Proposition.
- “On a given finite straight line to construct an equilateral triangle.”



HOL LIGHT (Higher Order Logic)

- It is a computer program that checks every single logical step in a proof.
- There are no shortcuts, no approximations, no handwaving.
- The program is written in a way that it becomes self-checking after the first 500 lines of computer code.
- Highly reliable computer code can reach reliability rates of less than one bug per 10,000 lines of code (Space Shuttle software)
- Dare we hope that the 500 lines of HOL light are bug free?

Why Believe HOL Light?

- John Harrison (the author of HOL Light) writes extremely clean and transparent code.
- A number of mathematicians, computer scientists, and logicians have made a careful analysis of the code.
- The underlying logic and system of axioms is weaker than ZFC.
- The kernel of the system contains only about 500 lines of code. Error rates can be as low as one bug per $10,000$ lines.
- The kernel is now certified to be free of defects. . ..

The Consistency of HOL Light

- Harrison has recently produced a proof in HOL Light of its own consistency! (2006)
- The model that he builds of the semantics is quite faithful to the actual language semantics.
- The proof of consistency can be automatically translated to other proof-checking systems.

Other reasons to be optimistic

- The semantics of ML (a functional programming language) have a complete mathematical specification (Harper,. . .).
- ML semantics have been formalized in twelf (a system that allows formal reasoning about the properties of the language), and basic formal properties such as type safety have been established (Harper,. . ., 2006).
- OCAML, the implementation language for HOL Light, is very close to ML.
- Technology has now improved to the point that a compiler has been written for C that has been formally proved correct. (X. Leroy, 2006).

Difficult proofs have been formally verified are several orders of magnitude more reliable than difficult proofs that have been verified by the conventional route (journal submission, peer review, etc.)

Ultimately (all philosophical arguments aside), the reliability of our discipline is not one bit more reliable than the processes in regular use to check the correctness of proofs.

The *Annals* Policy on Computer Proofs

Computer-assisted proofs of exceptionally important mathematical theorems will be considered by the Annals.

The human part of the proof, which reduces the original mathematical problem to one tractable by the computer, will be refereed for correctness in the traditional manner. The computer part may not be checked line-by-line, but will be examined for the methods by which the authors have eliminated or minimized possible sources of error: (e.g., round-off error eliminated by interval artihmetic, programming error minimized by transparent surveyable code and consistency checks, computer error minimized by redundant calculations, etc. [Surveyable means that an interested person can readily check that the code is essentially operating as claimed]).

HOL light – history

- LCF – original system in 1972 by Robin Milner
  - LCF = Logical for Computable Functions.
  - Based on a logic developed by Dana Scott in 1969.
  - a proof-checking program.
- Milner designed ML (meta-language) to add programming capabilities to LCF.
- Type checking guarantees that every value of type ‘theorem’ has been created by the rules of inference.

- Proofs can be developed in a backwards style.
  - Goals are declared.
  - Tactics reduce a goal to subgoals.

- Mike Gordon around 1983 developed a variant of LCF called HOL
- HOL = Higher order logic
- LCF primitive axioms and inference rules were modified for higher order logic.
- theorem proving methods were essentially the same as LCF.
- The original purpose of the system was hardware verification.

```
let union-subset = prove-by-refinement(
‘∀ Z1 Z2 A. ((Z1 ∪ Z2) ⊂ A) =
   (Z1 ⊂ A) ∧ (Z2 ⊂ A)‘,
[
DISCH-ALL-TAC;
REWRITE-TAC[UNION;SUBSET;IN;IN-ELIM-THM’];
ASM-MESON-TAC[];
]);;
```

Drop the quantifiers, rewrite using the definition of union, subset, and set-notation. Finally, apply the model elimination tactic to finish the proof.

We could have written the same proof more compactly as a single step: meson-tac with arugments (unions, subset,. . .).

Logical Axioms

The theorems of HOL-light are the same as the theorems of HOL, but the axioms are formulated in a somewhat different manner.

Basic primitive inference rules include:

- Reflexive: $x = x$
- Transitivity: $x = y$ and $y = z$ gives $x = z$.
- Combination: $f = g$ and $x = y$ gives $f(x) = g(y)$.
- Abstraction: $u = v$ gives $\lambda x.u = \lambda x.v$.
- Beta: $(\lambda x.t)x = t$.
- Assume: $\{p\} \vdash p$.
- MP: $p = q$ and p gives q.
- Deduct: $\{q\} \vdash p$ and $\{p\} \vdash q$ gives $p = q$.

Logical and Mathematical Axioms

- There are rules for term and type instantiation.
- There is a mechanism for extending the system (conservative extensions) with new definitions and new types.
- There is an axiom of infinity, an axiom of choice, and an axiom of extensionality.

Results that have been Formally Verified

The Four Color Theorem (G. Gonthier, COQ, 2004).

Results that have been Formally Verified

- The Prime Number Theorem (J. Avigad, Isabelle, 2004).
- The Jordan Curve Theorem (T.C.H., HOL Light 2005).
- The Fundamental Theorem of Calculus, basic real analysis, Gauge Integration in n-dimensions, Brouwer Fixed Point Theorem, Fundamental Theorem of Algebra (Harrison, HOL Light 1997–2006).
- Quantifier Elimination over the reals (Harrison-McLaughlin, 2006).
- Verification of one of the main computer programs in the proof of the Kepler Conjecture (Enumeration of Tame Planar Graphs) (Nipkow-Bauer-Schultz, Isabelle 2005).
- Basic algebraic topology and the fundamental group (Various contributors, MIZAR).

Current Verification Projects

- The Feit-Thompson odd-order theorem (G. Gonthier, COQ)
- The fundamental theorem of Galois theory (McLaughlin, COQ)
- The formal verification of the Kepler conjecture [FLYSPECK] (Hales, Nipkow, Zumkeller, Obua, Avigad, Harrison, McLaughlin, Bauer, etc.)

A day of theorem proving.

How much terrain can be covered with HOL-light in a day? The general guideline is a page of textbook mathematics a week.

- Friday April 2, 2004.
- 9.5 hour day (not counting breaks and time fighting a virus)
- 27 definitions, lemmas, and theorems
- 474 proof steps
- 23 proof steps per lemma on average
- 694 line objective caml file
- 2 pages and 7 lines of mathematical text

April 2 context

Results shortly before April 2.

- Continuous functions on a compact topological space achieve their maximum.
- A subset of $\mathbb{R}^n$ is compact iff it is closed and bounded.
- A topological space is compact iff it is complete and totally bounded.

April 2 results

The results were in three areas:

- Homeomorphisms of topological spaces.
- Transporting results about $\mathbb{R}^1$ to $\mathbb{R}$.
- Connected sets in topological spaces.

Note that there is a difference between $\mathbb{R}^1$ and $\mathbb{R}$:

$$\mathbb{R}^n = \{f : \mathbb{N} \to \mathbb{R} \mid f(m) = 0 \text{ for } m \geq n.\}$$

New definitions

- homeomorphism [4 minutes] (A bijective continuous function between topological spaces such that is an open mapping)
- connected (A set is connected if it cannot be divided between two disjoint open subsets.)
- connected component (Two points are equivalent if there is a connected set that contains them both.)
- the metric on the real line ($d\ x\ y = |x - y|$).
- the topology on the real line (the topology associated with the metric)

Results about homeomorphisms

- If f is a homeomorphism, then the inverse function is continuous. [45 minutes]
- If a function is bijective, continuous, and the inverse function is continuous, then the function is a homeomorphism. [18 minutes]
- For bijective maps, the image of a complement is the complement of the image. [13 minutes]
- A bijective mapping sends closed sets to closed sets iff it sends open sets to open sets
- A continuous bijective function from a compact space to a hausdorff space is always a homeomorphism [20 minutes]

Results about $\mathbb{R}^1$ and $\mathbb{R}$

- The ith coordinate function $\mathbb{R}^n \mapsto \mathbb{R}$, $(x_1, \ldots, x_n) \mapsto x_i$ is continuous.
- The vector $x\delta_0$ belongs to $\mathbb{R}^1$.
- The value of $x\delta_0$ at 0 is x. (Dirac delta function)
- An closed interval $[a, b]$ in $\mathbb{R}^1$ is a closed ball (centered at $(a + b)/2$).
- A closed interval $[a, b]$ in $\mathbb{R}^1$ is closed.
- A closed interval $[a, b]$ in $\mathbb{R}^1$ is bounded.
- A closed interval $[a, b]$ in $\mathbb{R}^1$ is compact.
- The image of the closed interval $[a, b]$ in $\mathbb{R}^1$ in $\mathbb{R}$ is $[a, b]$.
- A closed interval $[a, b]$ in $\mathbb{R}$ is compact.

Results about connectedness

- A union is a subset iff the individuals are each subsets
- The union of two connected non-disjoint sets is again connected.
- A point is a connected set.
- A point is in the same connected component as itself.
- The “connected component” relation is symmetric.
- The “connected component” relation is transitive.
- [Partial proof] (If the second hypotheses of Bolzano’s lemma holds, then) A closed interval $[a, b]$ in $\mathbb{R}$ is connected.

[Recall Bolzano’s lemma.]

Bolzano's Lemma

val it : thm =

—- $\forall$P. ($\forall$a b c. a $\leq$ b $\wedge$ b $\leq$ c $\wedge$ P (a,b) $\wedge$ P (b,c) $\Rightarrow$ P (a,c)) $\wedge$

($\forall$x. $\exists$ d. 0.0 $<$ d $\wedge$ ($\forall$a b. a $\leq$ x $\wedge$ x $\leq$ b $\wedge$ b - a $<$ d $\Rightarrow$ P (a,b)))

$\Rightarrow$ ($\forall$a b. a $\leq$ b $\Rightarrow$ P (a,b))

In words, a property holds of all intervals if the following two conditions hold.

- Whenever the property holds of two subintervals, the property holds of the full interval.
- The property holds of sufficiently small intervals around each real number.

For connectedness, take P to be the property

λ (u ,v).

$$(u < a) \vee (b < v)$$

$$\vee (\{x \mid u \leq x \wedge x \leq v\} \subset A)$$

$$\vee (\{x \mid u \leq x \wedge x \leq v\} \subset B)‘$$

Self-evaluation

- 19 minutes were lost in locating the name of Bolzano's lemma.
- 57 minutes were lost in administrative chores:
  - Loading took 10 minutes because of a stray character in a file.
  - A virus forced a reboot.
  - I documented a bug in a tactic.
- About 3 hours were lost because of a digression into properties of real numbers. (compactness of intervals)