Theorem proving in

Cayden Codel
15-414 Bug Catching
Prof. Ruben Martins

April 25, 2024

Mathematical proof is hard.

Mathematical proof is hard.

Can we get computers to do it?

Mathematical proof is hard.

Can we get computers to do it?

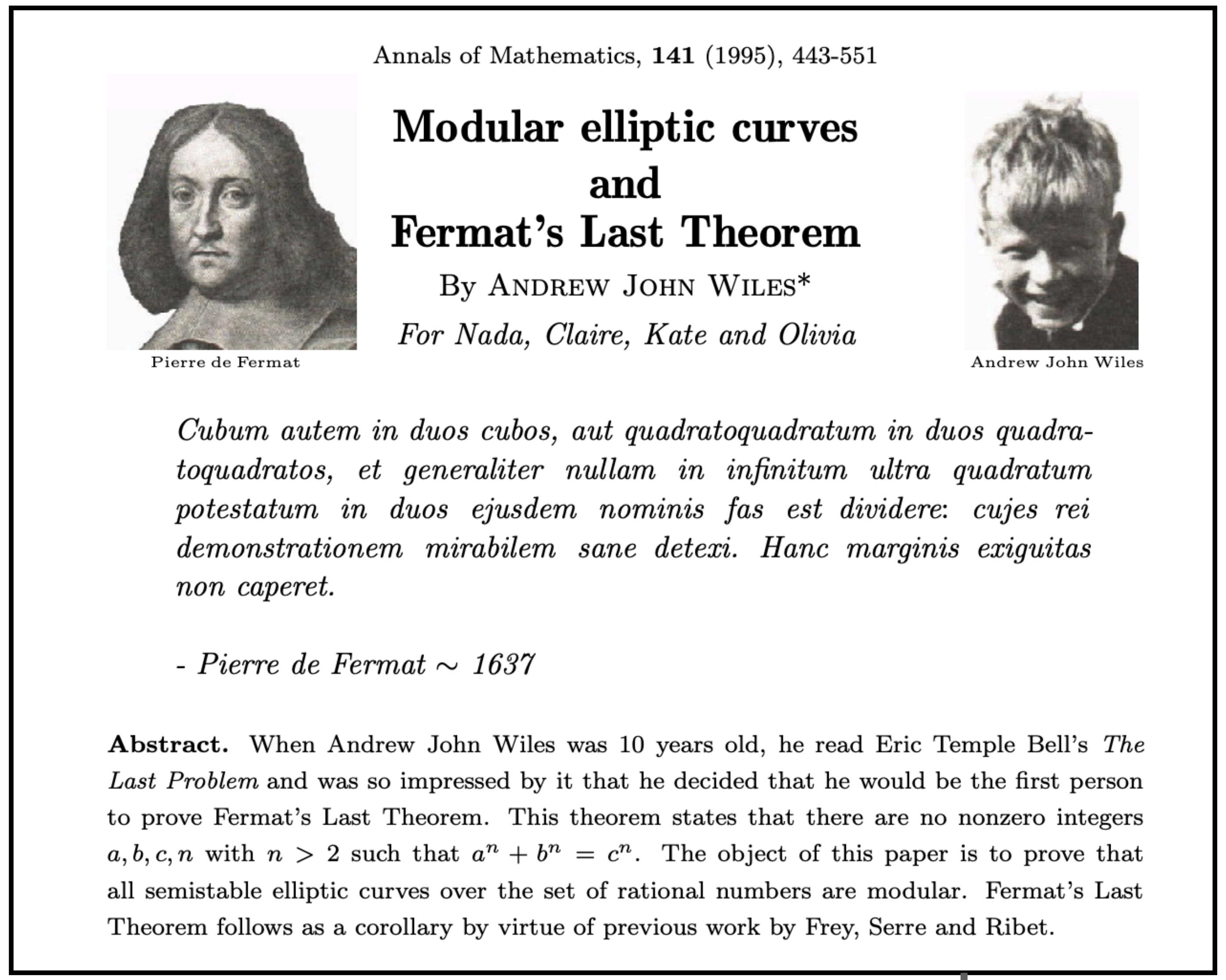
Answer: it depends.

Who does the proving?

Who does the proving?

Computers Humans

Who does the proving?


Annals of Mathematics, **141** (1995), 443-551

Modular elliptic curves
and
Fermat's Last Theorem

By Andrew John Wiles*

For Nada, Claire, Kate and Olivia

Pierre de Fermat

Andrew John Wiles

Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatum in duos ejusdem nominis fas est dividere: cujes rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

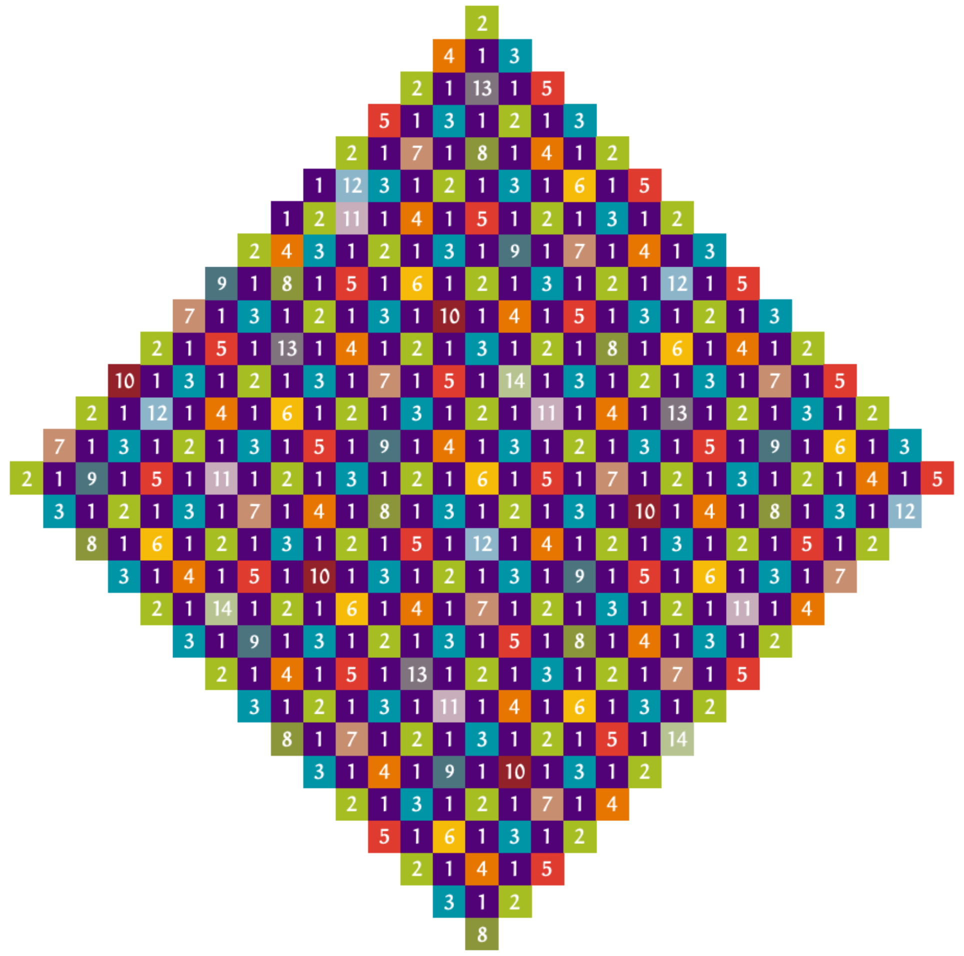
- Pierre de Fermat $\sim$ *1637*

Abstract. When Andrew John Wiles was 10 years old, he read Eric Temple Bell's *The Last Problem* and was so impressed by it that he decided that he would be the first person to prove Fermat's Last Theorem. This theorem states that there are no nonzero integers a, b, c, n with $n > 2$ such that $a^n + b^n = c^n$. The object of this paper is to prove that all semistable elliptic curves over the set of rational numbers are modular. Fermat's Last Theorem follows as a corollary by virtue of previous work by Frey, Serre and Ribet.

Computers

Humans

Who does the proving?



The Packing Chromatic Number of the Infinite Square Grid is 15

Bernardo Subercaseaux ✉ and Marijn J.H. Heule

Carnegie Mellon University, Pittsburgh PA 15203, USA
{bsuberca, mheule}@cs.cmu.edu

Annals of Mathematics, **141** (1995), 443-551

Pierre de Fermat

Modular elliptic curves and Fermat's Last Theorem

By Andrew John Wiles*

For Nada, Claire, Kate and Olivia

Andrew John Wiles

Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatum in duos ejusdem nominis fas est dividere: cujes rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

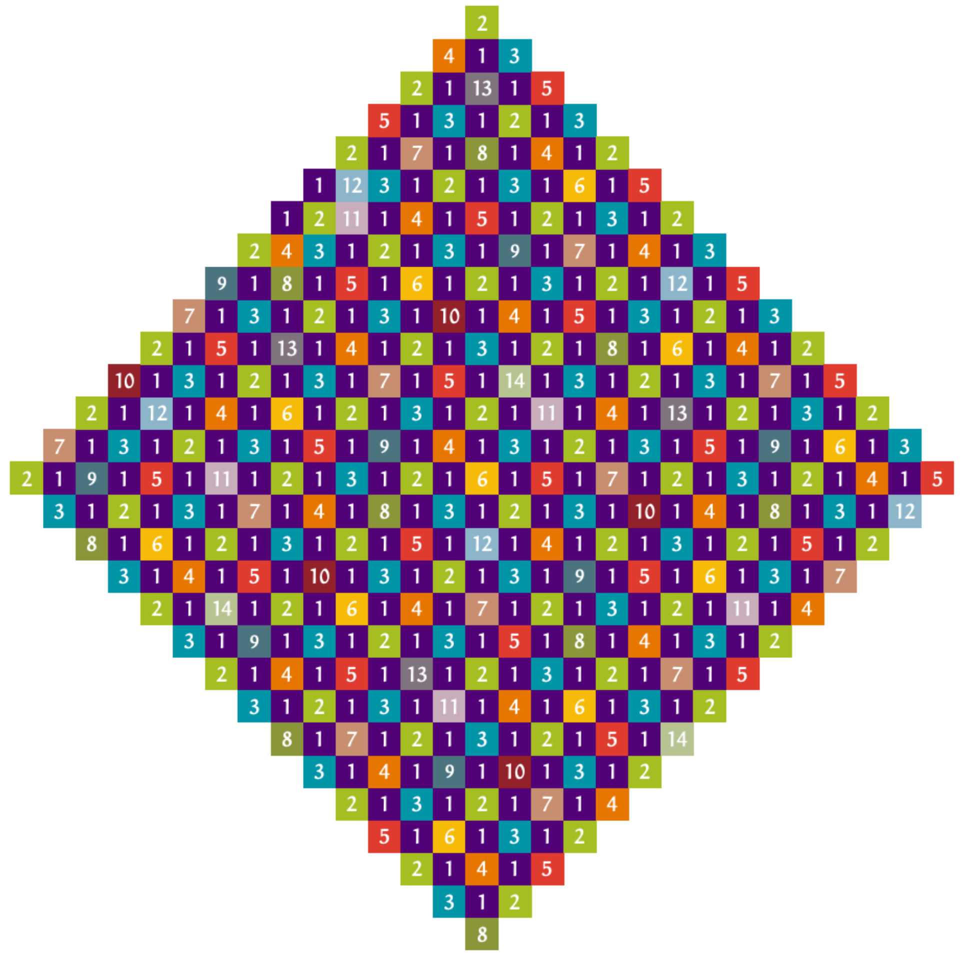
- Pierre de Fermat ~ 1637

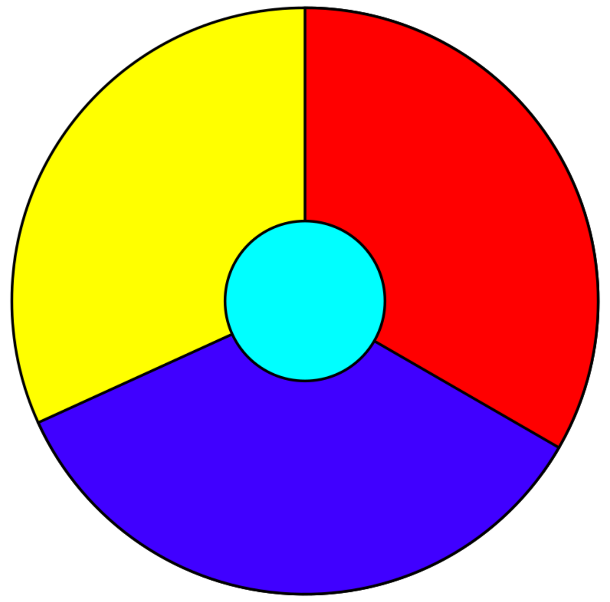
Abstract. When Andrew John Wiles was 10 years old, he read Eric Temple Bell's *The Last Problem* and was so impressed by it that he decided that he would be the first person to prove Fermat's Last Theorem. This theorem states that there are no nonzero integers a, b, c, n with $n > 2$ such that $a^n + b^n = c^n$. The object of this paper is to prove that all semistable elliptic curves over the set of rational numbers are modular. Fermat's Last Theorem follows as a corollary by virtue of previous work by Frey, Serre and Ribet.

Computers

Humans

Who does the proving?





The Packing Chromatic Number of the Infinite Square Grid is 15

Bernardo Subercaseaux and Marijn J.H. Heule

Carnegie Mellon University, Pittsburgh PA 15203, USA
{bsuberca, mheule}@cs.cmu.edu

A computer-checked proof of the Four Colour Theorem

Georges Gonthier
Microsoft Research Cambridge

This report gives an account of a successful formalization of the proof of the Four Colour Theorem, which was fully checked by the Coq v7.3.1 proof assistant [13].

Annals of Mathematics, **141** (1995), 443-551

Pierre de Fermat

Modular elliptic curves and Fermat's Last Theorem

By Andrew John Wiles*

For Nada, Claire, Kate and Olivia

Andrew John Wiles

Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatum in duos ejusdem nominis fas est dividere: cujes rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

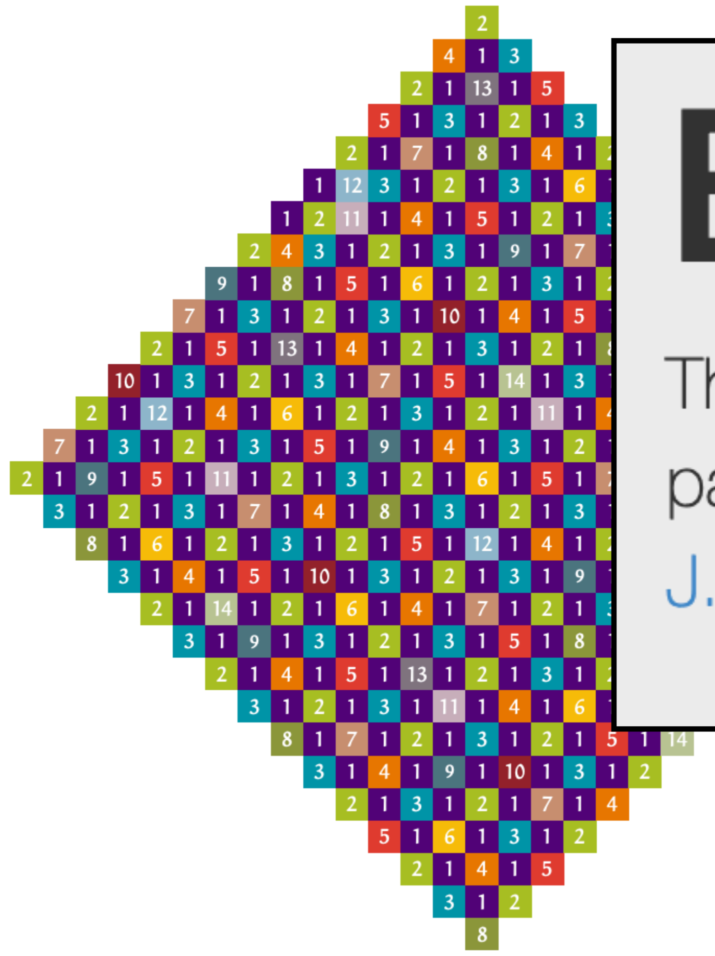
- Pierre de Fermat ~ 1637

Abstract. When Andrew John Wiles was 10 years old, he read Eric Temple Bell's *The Last Problem* and was so impressed by it that he decided that he would be the first person to prove Fermat's Last Theorem. This theorem states that there are no nonzero integers a, b, c, n with $n > 2$ such that $a^n + b^n = c^n$. The object of this paper is to prove that all semistable elliptic curves over the set of rational numbers are modular. Fermat's Last Theorem follows as a corollary by virtue of previous work by Frey, Serre and Ribet.

Computers

Humans

Who does the proving?



Everything's Bigger in Texas

This page provides access to results, proofs and tools on "the largest math proof ever" presented in the SAT 2016 paper **Solving and Verifying the boolean Pythagorean Triples problem via Cube-and-Conquer** by Marijn J.H. Heule, Oliver Kullmann, and Victor Marek (best paper award). A preprint is available on arXiv.

Fortunately there comes SAT solving to the rescue, which actually is r
approach solved the problem and resulted into a 200 terabytes proof

ematics, **141** (1995), 443-551

r elliptic curves
and
s Last Theorem

By Andrew John Wiles*

For Nada, Claire, Kate and Olivia

Fermat

Andrew John Wiles

um autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatum in duos ejusdem nominis fas est dividere: cujes rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

- Pierre de Fermat ∼ 1637

Abstract. When Andrew John Wiles was 10 years old, he read Eric Temple Bell's *The Last Problem* and was so impressed by it that he decided that he would be the first person to prove Fermat's Last Theorem. This theorem states that there are no nonzero integers a, b, c, n with $n > 2$ such that $a^n + b^n = c^n$. The object of this paper is to prove that all semistable elliptic curves over the set of rational numbers are modular. Fermat's Last Theorem follows as a corollary by virtue of previous work by Frey, Serre and Ribet.

The Packing Chromatic Number of the Infinite Square Grid is 15

Bernardo Subercaseaux and Marijn J.H. Heule

Carnegie Mellon University, Pittsburgh PA 15203, USA
{bsuberca, mheule}@cs.cmu.edu

A computer-checked proof of the Four Colour Theorem

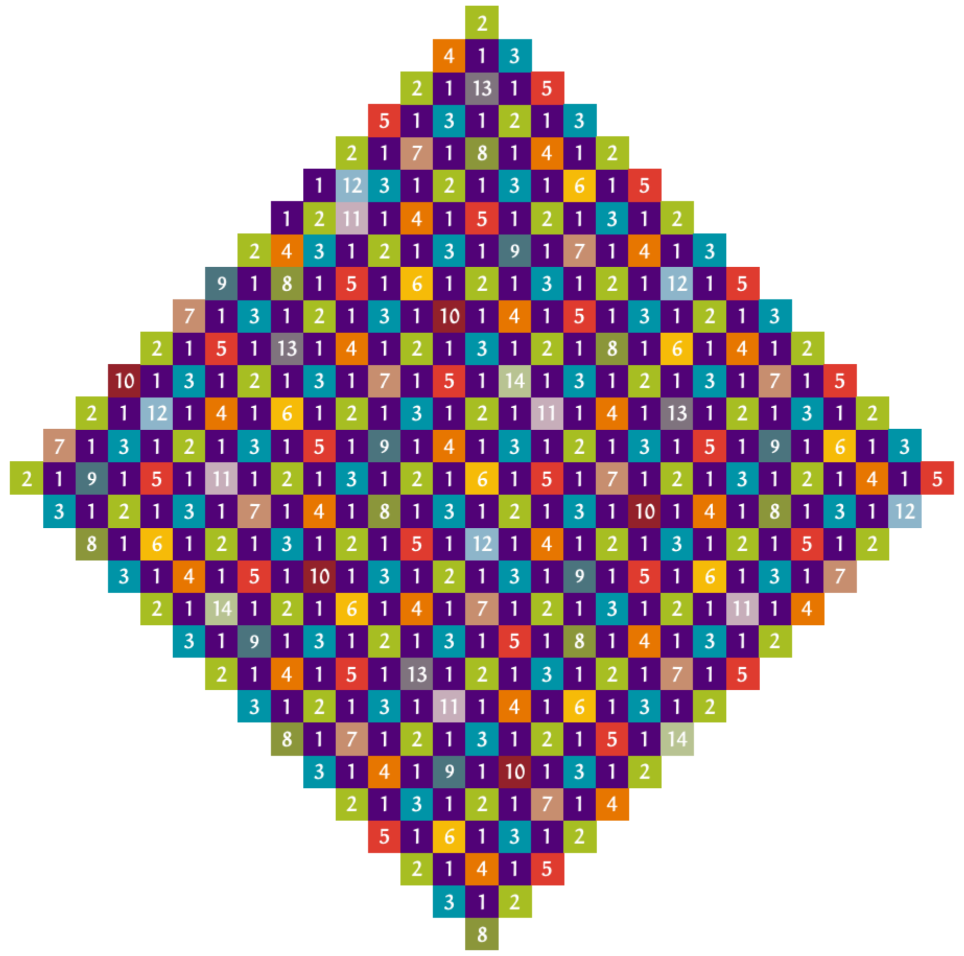
Georges Gonthier
Microsoft Research Cambridge

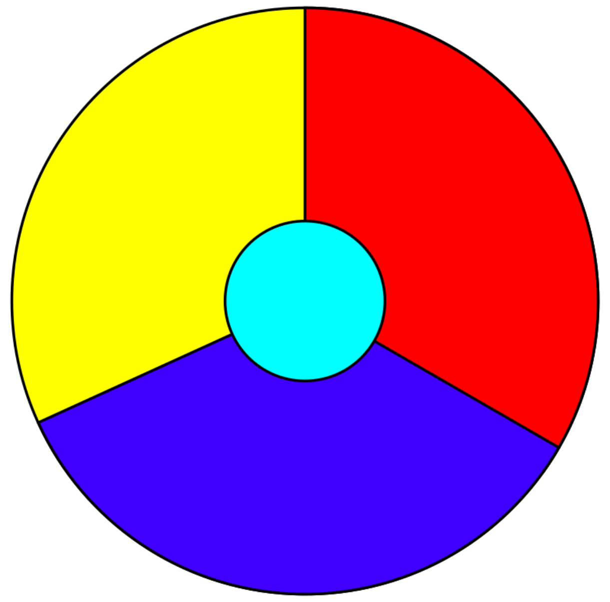
This report gives an account of a successful formalization of the proof of the Four Colour Theorem, which was fully checked by the Coq v7.3.1 proof assistant [13].

Computers

Humans

Who does the proving?





The Packing Chromatic Number of the Infinite Square Grid is 15

Bernardo Subercaseaux and Marijn J.H. Heule

Carnegie Mellon University, Pittsburgh PA 15203, USA
{bsuberca, mheule}@cs.cmu.edu

A computer-checked proof of the Four Colour Theorem

Georges Gonthier
Microsoft Research Cambridge

This report gives an account of a successful formalization of the proof of the Four Colour Theorem, which was fully checked by the Coq v7.3.1 proof assistant [13].

Annals of Mathematics, **141** (1995), 443-551

Pierre de Fermat

Modular elliptic curves and Fermat's Last Theorem

By Andrew John Wiles*

For Nada, Claire, Kate and Olivia

Andrew John Wiles

Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatum in duos ejusdem nominis fas est dividere: cujes rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

- Pierre de Fermat ∼ 1637

Abstract. When Andrew John Wiles was 10 years old, he read Eric Temple Bell's *The Last Problem* and was so impressed by it that he decided that he would be the first person to prove Fermat's Last Theorem. This theorem states that there are no nonzero integers a, b, c, n with $n > 2$ such that $a^n + b^n = c^n$. The object of this paper is to prove that all semistable elliptic curves over the set of rational numbers are modular. Fermat's Last Theorem follows as a corollary by virtue of previous work by Frey, Serre and Ribet.

Computers

Humans

Interactive theorem provers

A human-computer compromise

- Write code (functional, with inductive types)
- Write theorems
- Prove theorems with tactics
- Libraries of mathematical proofs

Interactive theorem provers

A human-computer compromise

- Write code (functional, with inductive types)
- Write theorems
- Prove theorems with tactics
- Libraries of mathematical proofs

Lean

The best theorem prover!

Lean

The best theorem prover!

- A functional programming language with theorem-proving capabilities

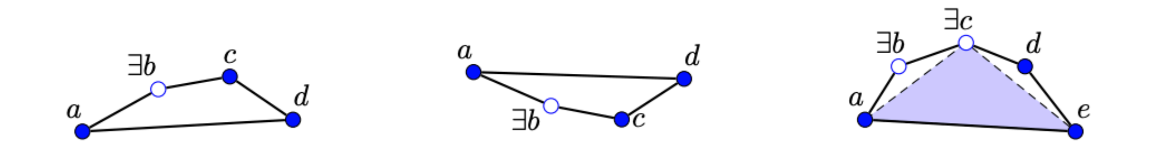
Lean

The best theorem prover!

- A functional programming language with theorem-proving capabilities
- Critical for several formalization efforts

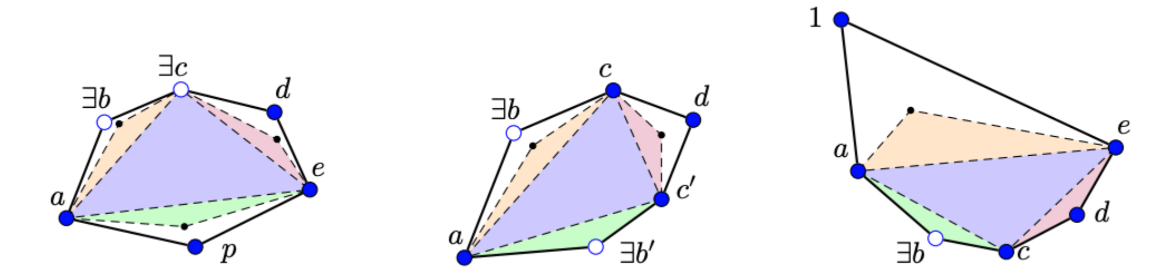
Formal Verification of the Empty Hexagon Number

Bernardo Subercaseaux
Carnegie Mellon University
Wojciech Nawrocki
Carnegie Mellon University
James Gallicchio
Carnegie Mellon University
Cayden Codel
Carnegie Mellon University
Mario Carneiro
Carnegie Mellon University
Marijn J. H. Heule
Carnegie Mellon University



(a) $u^4_{a,c,d}$ (b) $v^4_{a,c,d}$ (c) $u^5_{a,d,e}$

Figure 7 Illustration of the 4-cap (7a), 4-cup (7b), and 5-cap (7c) variables. The highlighted region denotes an empty triangle.



(a) $(u^5_{a,d,e} \wedge o_{a,p,e})$ (b) $(u^4_{a,c,d} \wedge v^4_{a,c',d} \wedge h_{a,c,c'})$ (c) $(v^4_{a,c,d} \wedge o_{c,d,e} \wedge h_{a,c,e})$

Figure 8 Illustration of some *forbidden configurations* that imply 6-holes. Figure 8a corresponds to the configuration forbidden by clause (13), Figure 8b to the one forbidden by clause (15), and Figure 8c to clause (17). All highlighted regions denote empty triangles.

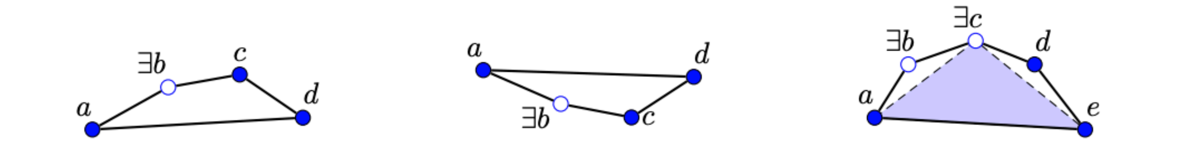
Lean

The best theorem prover!

- A functional programming language with theorem-proving capabilities
- Critical for several formalization efforts
- A way for computer scientists to verify code

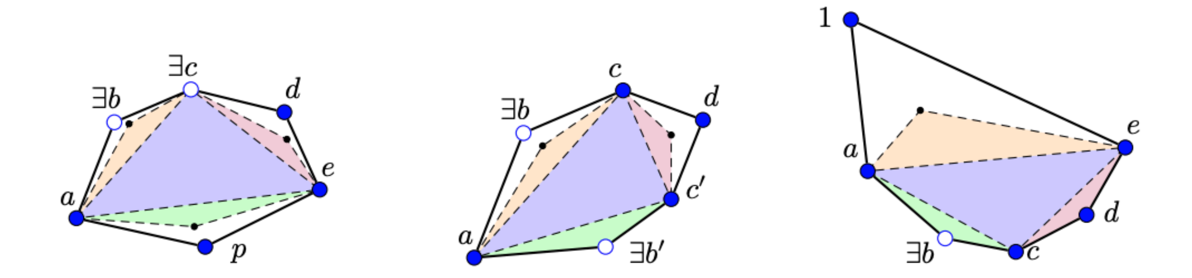
Formal Verification of the Empty Hexagon Number

Bernardo Subercaseaux
Carnegie Mellon University
Wojciech Nawrocki
Carnegie Mellon University
James Gallicchio
Carnegie Mellon University
Cayden Codel
Carnegie Mellon University
Mario Carneiro
Carnegie Mellon University
Marijn J. H. Heule
Carnegie Mellon University



(a) $\mathsf{u}^4_{a,c,d}$ **(b)** $\mathsf{v}^4_{a,c,d}$ **(c)** $\mathsf{u}^5_{a,d,e}$

Figure 7 Illustration of the 4-cap (7a), 4-cup (7b), and 5-cap (7c) variables. The highlighted region denotes an empty triangle.



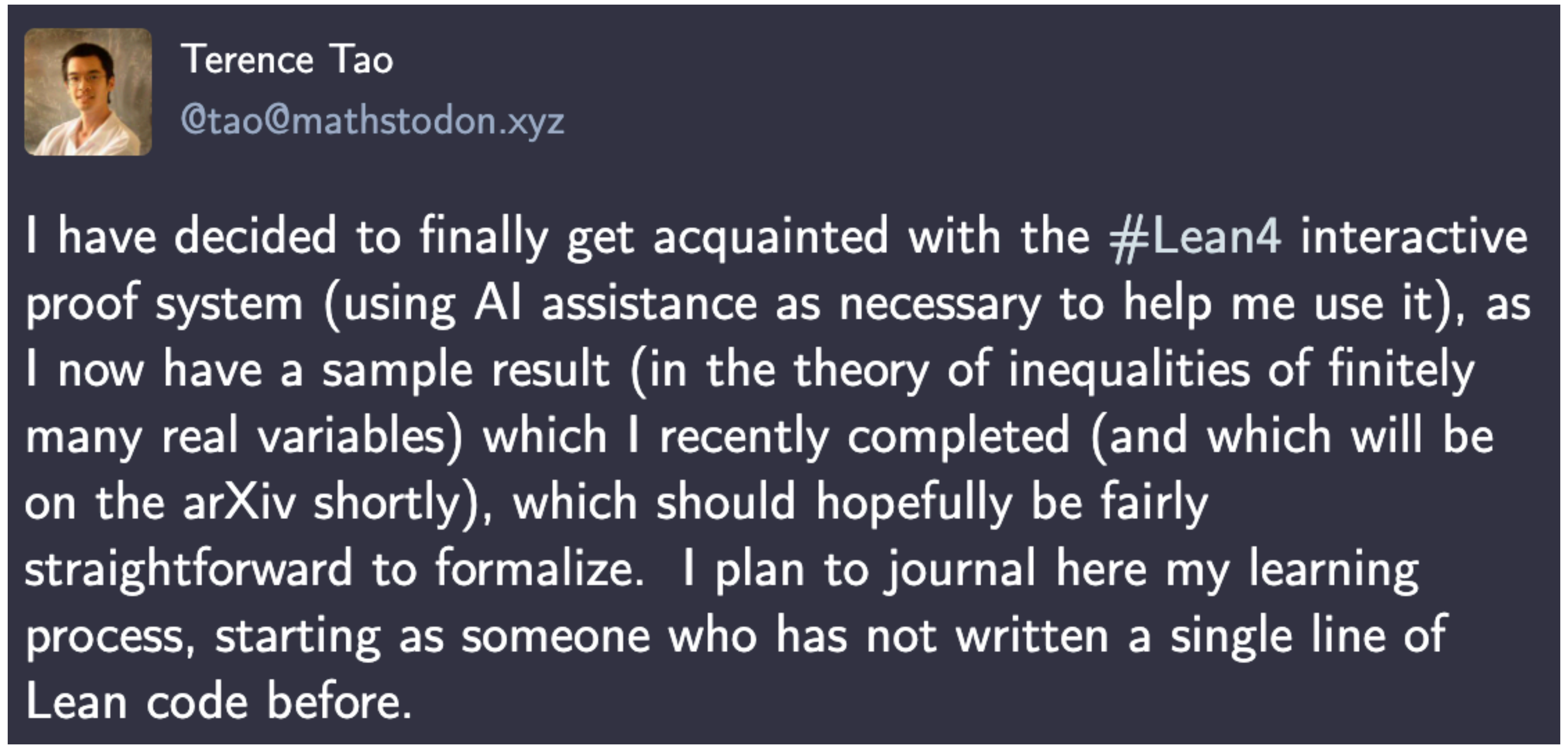
(a) $(\mathsf{u}^5_{a,d,e} \wedge \mathsf{o}_{a,p,e})$ **(b)** $(\mathsf{u}^4_{a,c,d} \wedge \mathsf{v}^4_{a,c',d} \wedge \mathsf{h}_{a,c,c'})$ **(c)** $(\mathsf{v}^4_{a,c,d} \wedge \mathsf{o}_{c,d,e} \wedge \mathsf{h}_{a,c,e})$

Figure 8 Illustration of some *forbidden configurations* that imply 6-holes. Figure 8a corresponds to the configuration forbidden by clause (13), Figure 8b to the one forbidden by clause (15), and Figure 8c to clause (17). All highlighted regions denote empty triangles.

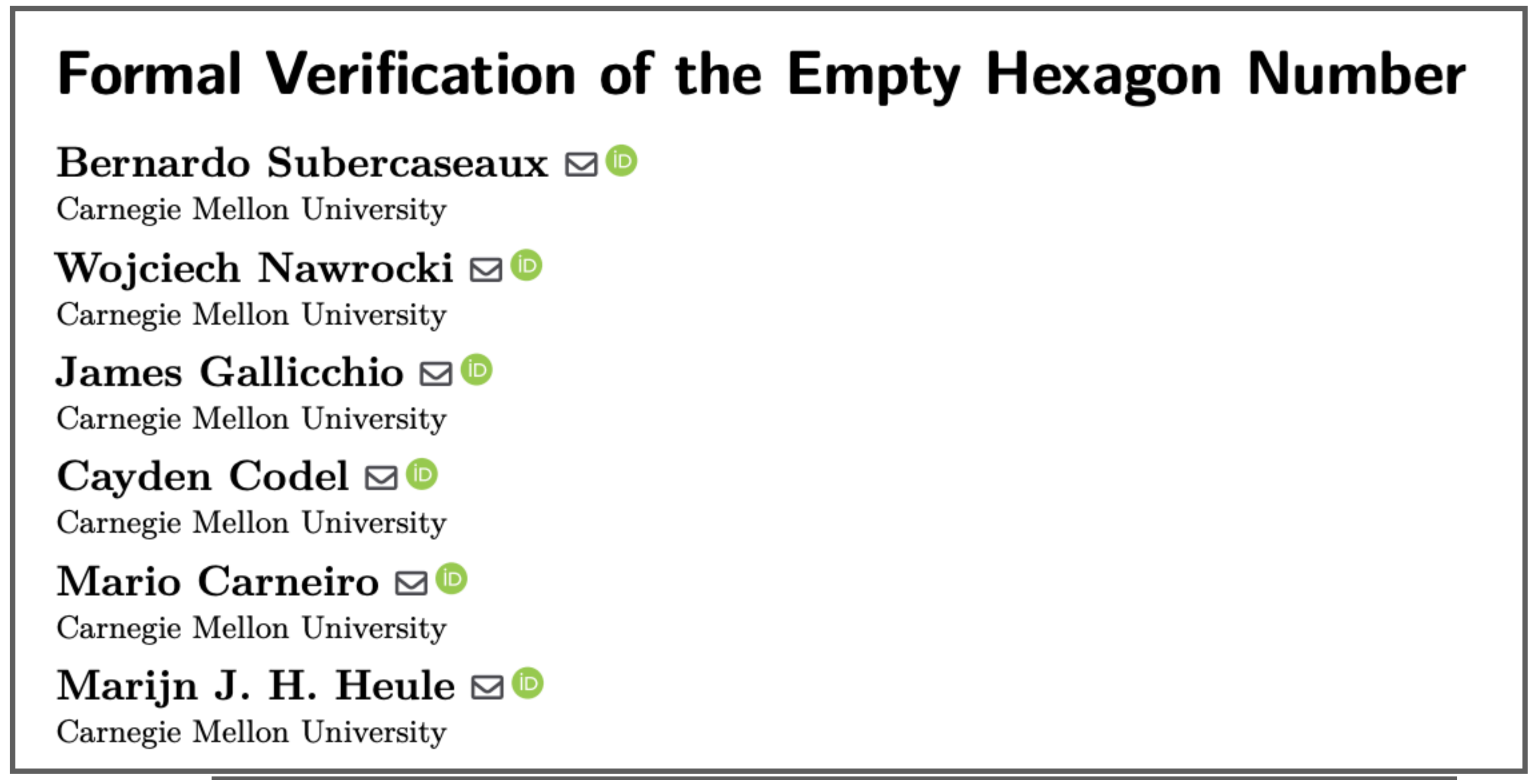
Lean

The best theorem prover!

- A functional programming language with theorem-proving capabilities
- Critical for several formalization efforts
- A way for computer scientists to verify code
- A way for mathematicians to verify proofs


Terence Tao
@tao@mathstodon.xyz

I have decided to finally get acquainted with the #Lean4 interactive proof system (using AI assistance as necessary to help me use it), as I now have a sample result (in the theory of inequalities of finitely many real variables) which I recently completed (and which will be on the arXiv shortly), which should hopefully be fairly straightforward to formalize. I plan to journal here my learning process, starting as someone who has not written a single line of Lean code before.


Formal Verification of the Empty Hexagon Number

Bernardo Subercaseaux
Carnegie Mellon University
Wojciech Nawrocki
Carnegie Mellon University
James Gallicchio
Carnegie Mellon University
Cayden Codel
Carnegie Mellon University
Mario Carneiro
Carnegie Mellon University
Marijn J. H. Heule
Carnegie Mellon University

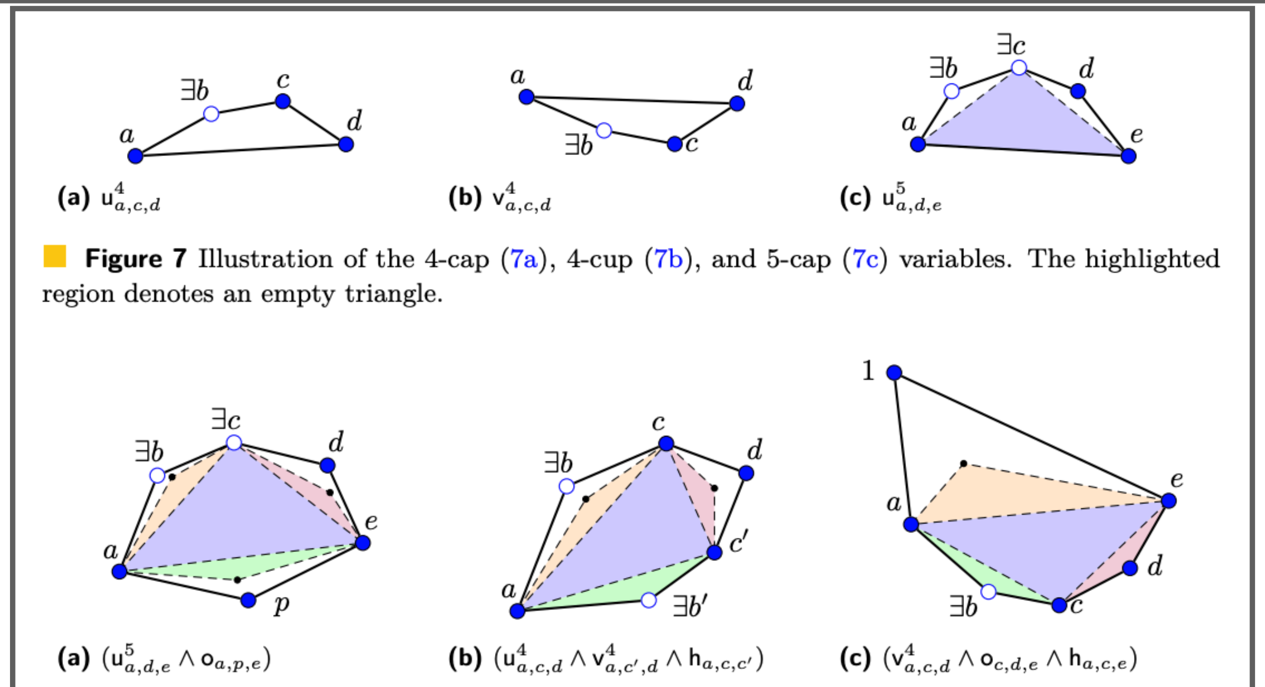


Figure 7 Illustration of the 4-cap (7a), 4-cup (7b), and 5-cap (7c) variables. The highlighted region denotes an empty triangle.

Figure 8 Illustration of some *forbidden configurations* that imply 6-holes. Figure 8a corresponds to the configuration forbidden by clause (13), Figure 8b to the one forbidden by clause (15), and Figure 8c to clause (17). All highlighted regions denote empty triangles.

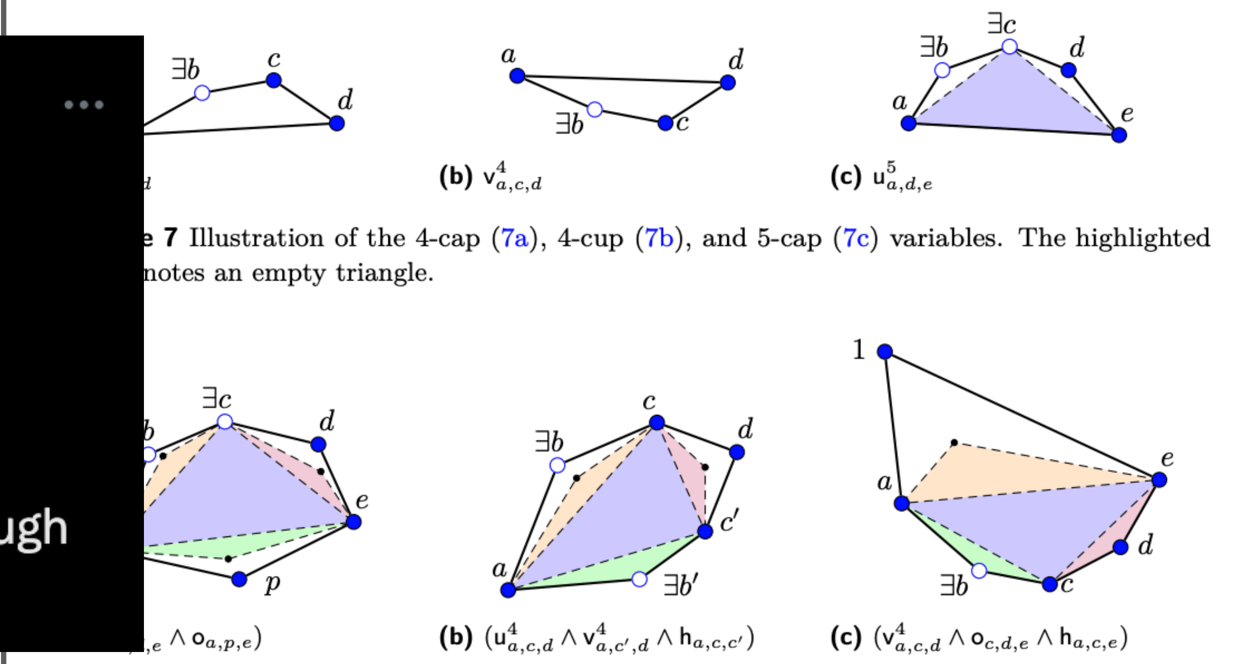
Lean

The best theorem prover!

- A functional programming language with theorem-proving capabilities
- Critical for several formalization efforts
- A way for computer scientists to verify code
- A way for mathematicians to verify proofs

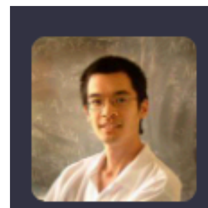
Formal Verification of the Empty Hexagon Number

Bernardo Subercaseaux
Carnegie Mellon University
Wojciech Nawrocki
Carnegie Mellon University
James Gallicchio
Carnegie Mellon University
Cayden Codel
Carnegie Mellon University
Mario Carneiro
Carnegie Mellon University
Marijn J. H. Heule
Carnegie Mellon University



…e 7 Illustration of the 4-cap (7a), 4-cup (7b), and 5-cap (7c) variables. The highlighted …notes an empty triangle.

Figure 8 Illustration of some *forbidden configurations* that imply 6-holes. Figure 8a corresponds to the configuration forbidden by clause (13), Figure 8b to the one forbidden by clause (15), and Figure 8c to clause (17). All highlighted regions denote empty triangles.



Terence Tao
@tao@mathstodon.xyz

I have decided to finally get acquainted with the #Lean4 interactive proof system (using AI assistance as necessary to help me use it), as I now have a sample result (in the theory of inequalities of finitely many real variables) which I recently completed (and which will be on the arXiv shortly), which should hopefully be fairly straightforward to formalize. I plan to journal here my learning process, starting as someone who has not written a single line of Lean code before.

Kevin Buzzard
@XenaProject

I got a research grant to begin the proof of formalising Fermat's Last Theorem in Lean! gow.epsrc.ukri.org/NGBOViewGrant....

The research buys out my teaching and admin for 5 years, which I suspect will not be enough to get it done, but it will certainly be enough to make a big dent in it.

Let’s play around with Lean!