Notes by J.H.Davenport

28 May 2019

Contents

1 27 May 2019 **3**

2 28 May 2019 **4**
2.1 Mechanization of Mathematics: Avigad 4
2.1.1 Digital Infrastructure 5
2.1.2 What I really think 5
2.2 Modeling human proof checking in the Naproche-SAD system: Koepke 6
2.3 Online proof and elementary mathematics from an educational perspective: Sangwin 7
2.4 String theory, mathematical databases and formal methods: Douglas 8
2.4.1 String Theory 9
2.5 Mathematical Knowledge Representation and Reasoning in the Wolfram Language: Ford 9
2.6 Janus in the Mathematics Library: Baramy 9
2.7 Verification of Neural Networks: Komendantskaya 10
2.8 How to be a High-Frequency Trader: Donald MacKenzie 11
2.8.1 High Frequency Trading 11
2.8.2 Cryptocurrencies 12

3 29 May 2019 **13**
3.1 Mathematical objects in dependent type theory: Buzzard 13
3.2 Diagrammatic Notations in Mathematical Proofs: De Toffoli . . . 14
3.3 Formalising Mathematics-In Praxis: First Experiences with Isabelle/HOL: Koutsoukou-Argyraki 15
3.3.1 Challenges 16
3.4 Teaching machines to do mathematics: Saxton 16
3.4.1 Problems 17
3.5 The HOL Light Mathematical Libraries: Harrison 17
3.5.1 Decision Procedures 18
3.5.2 Keep your theorems sharp 18
3.5.3 $A = B$ is too glib 18
3.6 Group Knowledge and Mathematical Collaboration: Tanswell . . 18

3.6.1 4CT 19
3.6.2 CFSG 19
3.7 Automating "human-like" example-use in mathematics: Pease . . 20
3.7.1 ?? 20

4 30 May 2019 **21**
4.1 Kohlhase 21
4.2 Learning and Reasoning over Big Proof Corpora: Urban 22
4.3 A formal classical prof of Hahn–Banach in Coq: Kerjean 23
4.4 Why did mathematicians not embrace the vision of the QED Manifesto: Löwe 24
4.5 On the Impact of Big Proofs on Provers: Gonthier 25
4.6 Logipedia:a system-independent encyclopedia of formal proofs: Dowek 26
4.7 Why won't they use my stuff, and what can I do about it: Martin 27

5 31 May 2019 **28**
5.1 The "NASA Effect" of a Global Digital Math Library: Watt . . . 28
5.2 Panel 29

A JHD's reflections on Machine Learning **31**

Chapter 1

27 May 2019

JHD alas missed this half-day due to illness.

Chapter 2

28 May 2019

2.1 Mechanization of Mathematics: Avigad

[Avi18] came out of Big Proof 2017. Written as a survey, but published under "Opinion Piece".

Hales (1998) proved Kepler's Conjecture, but [Hal05] showed the problems with the refereeing process.

[HKM16] can colour up to but not 7825, by SAT solving. We can see SAT(7824), but to see UNSAT(7825) we have a 200TB UNSAT core.

Hence "formal methods". Llull (13th century) used a set of wheels to generate combinations, and this inspired Leibniz, [Lei66] proposed *calculus ratiocinator*. Note the flagship results are noteworthy, but the underlying libraries, which formalise most of undergraduate mathematics these days. 150kLOC, 4000 definitions, 13000 results for [GT12]'s proof of the Odd Order Theorem [FT63]. But showed 2-line formal statement of Prime Number Theorem[1], 7-line formal statement of Jordan Curve Theorem, etc. Formal statement of Central Limit Theorem needs whole slide. Also recent proof that ZFC$\not\Rightarrow$Continuum Hypothesis. Also Ellenberg–Gijswijt Cap Set Theorem.

So what does verified computation mean?

- Rewrite the computations to construct proofs (Flyspeck bounds)
- verify certificates (proof sketches, duality in linear programming etc.)
- verify the algorithm and execute it with trusted evaluation (4CT)
- verify the algorithm and extract code and run it (Flyspeck enumeration)

[CMSPT14] verification of irrationality of $\zeta(3)$: Salvy's Maple worksheets converted to Coq. Tucker computed the existence of the Lorenz attractor, which Fabien Immler converted to Isabelle and used code extraction.

[1] But needs function π, definition not shown.

McCune showed that $(w((x^{-1}w)^{-1}z))((yz)^{-1}x) = e$ formalises groups, and Kunnen showed this is the shortest such single axiom. There are lots of other combinatorial/algebraic results proved this way.

2.1.1 Digital Infrastructure

Used computers for computation, searching etc. for a long time. Since 2017, more mathematicians are using Lean: list. But why? It's a nice expressive system, but the experts were great at answering questions and building a community.

Floor Jeremy's book is also very important.

Thanks, but we are still not ready for prime time.

- Applications to date are isolate extreme examples
- Tools require training and effort.

2.1.2 What I really think

Mathematicians think it's really cool that CS is developing these tools, but CS doesn't necessarily care about mathematics as such. Computer scientists don't win promotion by making mathematicians happy.

There are still questions about "how do we have reliable mathematical knowledge"? etc.

Q–FW Why is Lean successful and Coq not?

A Possibly the interface.

Q–Silvia Examples of errors in combinations?

A Various examples of ultimately immaterial errors.

Q Maybe there is a sharp distinction: CS deals with objects are already digital, mathematics isn't.

A I am not sure this is the biggest difference, but its there.

Q Patrick and I are introducing Lean in our graduate course.

Q What about challenges to the formalisation?

A There are doubts: "formality means death" etc.

Q You commented that we should accept incremental results — look at Moore's Law, which is codification of incremental results.

A I tend to agree.

Q–MK Add to message to CS: "There is life outside deep learning".

A I agree that we need to say this. I am not worried abut the future of logic.

Q–UHM But Moore's Law became a self-fulfilling prophecy in the industry.

A True, but this became a perception of utility.

2.2 Modeling human proof checking in the Naproche-SAD system: Koepke

Entailment $\Phi \vdash \psi$ is the formal way we work, but hard. Contrasts a statement in [Hal05] with the Flyspeck code. A mathematical story is a story which is (more) formal, and usually every variable is typed and statements must be type-correct. Compares "is at most" in the text with $\leq$ in a formula. Note that we analyse a text in phases: reading, understanding and reasoning (the main process), which often involves identifying smaller tasks, heuristics etc.

Naproche is "Natural proof checking". Based on Glushkovs SA: System for Automated Deduction. With MW, embedding into Isabelle's system. Based on Formula Theory Language, an enriched version of FOL.

```
Signature. A real number is a notion
Ley x,y,z stand for real numbers.
Definition. R is the set fo real numbers.
Signature x.y is a real number.
Axiom. x.y = y.x
```

It is a typed language: variables belong to notions ($\sim$ types).

"Signature. `\Prod{m}{n}{f}` is a real number" is an example of the introduction of LaTeX type-setting. Note that we need type-correctness reasoning. Note that $\frac{x}{y}$ requires a proof that $y \neq 0$. We do not require notions to be non-empty, which is a small difference from Mizar, but otherwise very similar.

We now (unlike SAD) have term rewriting, which may spawn proof obligations.

So what can we do. More of Rudin's *analysis*.

Theorem 1 $\forall x, y > 0 \exists nn \cdot x > y$.

Kelly's Topology Appendix on Kelly–Morse Set Theory (p. 259 is ordered pairs, for example, and our version looks pretty natural), etc.

Does this "natural language formal mathematics" work? It seems to. Can we get as far as other ITPs? Open questions over the organisation of libraries etc.

Note that, as a set theorist, I hold that set theory and type theory are not incompatible.

Q Why textbooks? This theorem (Theorem 1) is a didactic example.

A Always challenge e.g. Automath for Grundlagen. But textbooks are the equivalent of libraries.

Q We can see textbook writing as a first step towards formalisation.

A Indeed.

Q–Sarah There are levels of textbooks,

A Indeed.

2.3 Online proof and elementary mathematics from an educational perspective: Sangwin

Can we automatically assess students' proofs. I am looking at SQA Advanced Higher Mathematics, similar to English FM or IB's HL mathematics. This affects far more people than Kepler, and everyone starts here. Reminds us of Stack [San13], and its wide use (all Finland, 35 Germany). Example of differentiation, with a CAS as his back end. But I intended this to be formative. [KoecherSangwin201a]. In IB, 18% were precisely STACK, rising to 37% with method marks. Reasoning by equivalence is also a major component (37% in IB). I regard the entire proof as one entity. Note that there is equivalence of expressions and equivalence of equations. Equations coefficients. Support for Boolean connectives. Simple systems of inequalities. Automated deduction of calculus operations. Evaluation of previous lines.

Note that $f = g$ might be an equation or a statement of equivalence of expressions. Claims that "Reasoning by Equivalence" is important in mathematics education. This is now (after the death of geometry) the start of formal logic. Accounts for 1/3 of IBM marks, etc.

Note that we let students work line by line with no explanation, and therefore we have to support this.

SQA Advanced Higher, taken by about 5% of cohort in Scotland. One 3-hour paper worth 100 marks. STACK v4.3 can award 61% of the marks. RE is 31%. Calculus moves (6) contribute to 15%. It is possible to regard these results as a condemnation of the examination system, which I do not dispute.

I currently cannot assess the proof questions. "prove that the sum of any three consecutive integers is divisible by 3", "explain why the matrix P is not associated with a rotation about the origin".

Cajori claims that De Morgan 1838 is the first proof by induction. Induction is in SQA. I can't do this, but can prove the correctness of the algebra. Note also the importance of [Pól62].

See also `https://fourferries.com` and gradarius. Slide on textbooks. Note that the basic rule $ca = cb \Rightarrow a = b$ is wrong, we need $ca = cb \Rightarrow a = b \vee c = 0$.

Note that Babbage set out to produce log tables, Knuth set out to replicate movable types, and I am setting out to reproduce what teachers do. I tried to reproduce the columnar layout of Pell's algebra, but it turned out tot be a pedagogic disaster.

Q Could you assess things as in Project Euler?

A

Q How are you envisaging students writing solutions?

A Good question. Mixing symbolic and natural languages is an issue.

Q Formative?

A Students need templates, e.g. proof by induction.

2.4 String theory, mathematical databases and formal methods: Douglas

Various equations etc.

Quantum Mechanics for a rigorous basis: Schrödinger equation and functional analysis.

Classical Mechanics Can be treated rigorously

QED Everything is an expansion in $\alpha \approx \frac{1}{137}$. Recent rigorous treatment in Costello's book.

YM and other QFD coming

String theory Nothing

But a physicist's job is to predict the real world, not to look at formalisms. Contact between the model and the real world is based on an interpretation. Can we formalise this, so that "do these measurements contradict my model" can be answered formally. For example, observing Mercury shows perihelion progression incompatible with Newton. Need a basic model of astronomy, model of observers, model of the surface of the earth, model of an observation, and an error model.

Then we need to map everything into fourvectors in an earth-centred frame etc. Since N-body is not soluble, need a perturbation approach.

Example 1 ($g-2$ of a muon) *The lifetime is approximately $2\mu s$. can get to within 10^{-9}, and the Standard Model can compute this to similar precision. The two seem to be differing. Some time ago we also thought they differed by 3σ, but there was a mistake in the theory, in particular "weak hadronic light scattering". See hep-ph/011???. This was a surprise, as many groups worked on this. Note that this arose when adding up different theories, so would need a "grand theory" to catch.*

2.4.1 String Theory

A quantum unified theory of gravity and Yang–Mills theory coupled to matter a candidate fundamental theory. But what is M. It might be a six-torus, Calabi–Yau manifold, etc. There's Kreuzer–Skarke database of reflexive polytopes. This gives M as 3D CY. This dataset has an unexpectedly smooth boundary. This is a *platonic* rather than experimental dataset. But a very large one.

Q We teach "working out by hand".

A All theoretical physicists use computer algebra. But the problem is the number of conventions. There are 20 different Mathematica packages for Ricci tensors etc.

2.5 Mathematical Knowledge Representation and Reasoning in the Wolfram Language: Ford

"Wolfram Language" used to be Mathematica, but it's now more than that.

Note that, as well as numbers, strings, we also have images, sounds and Symbol, `plot`, `integrate`, or `foo`.

Q You said a lot of this was manually created.

A Yes, but there is a lot of tool support. Looked at an ML parser for arXiv papers.

Q People accuse Google of being parasitic on Wikipedia. Do you look at such ecosystem effects?

A Above my pay grade?

***** Stock cop out!

Q How does the link with Lean work?

A There are translation rules that work in one or the other direction (or both) e.g. polynomials $\Leftrightarrow$. But it's extensible

Q Real world is open-ended and messy. What about controversial questions?

A Perennial problem.

2.6 Janus in the Mathematics Library: Baramy

Start with Weil's anger at the Mathematics Library at IAS threat, and Oppenheimer's reprimand to Weil. Quotes T.S. Kuhn "For reasons and in ways that remain obscure to me, the sciences destroy their past more thoroughly than do mathematics or the arts". Hankel: "In the sciences, each generation destroys

what the previous one built, whereas in mathematics each generation builds on the previous". Latour: "Ready made Science versus Science in the Making".

Claims that mathematicians have a good sense of how they work, but this rarely comes across when they are asked to characterise their work. Mathematics occurs in blogs, social media, etc. Claims that there is Stabilized (written up) versus Situated (read down) media: e.g. the blackboard. "Writing up" is often viewed as secondary labour, but look at the number of issues that crop up.

1950 ICM was "the largest ever", but no Soviet block, many excluded by visas etc. Laurent Schwartz's theory of distributions was worked on in every inhabited continent"[2]. André Weil was imprisoned for avoiding national service, and was very productive there!

Note the importance of bibliographic literature: Zbl and MR. Claims this organised work. Note the postcards "I saw your citation, could I have a copy".

So what about arXiv.org? Claims that this is very different from the "static repository" framework that is the starting point for much "big proof".

No contrast isn't good/bad: they are co-produced. Both views are important.

2.7 Verification of Neural Networks: Komendantskaya

HW has just launched a "Lab for AI and Verification". I never got students to do verification projects until I added "for NN"!. AI is pervasive, but in need of verification: robustness, etc. Also worried about verification of AI planning languages, etc. This is "Big Proof" as the objects, and #parameters are in 10^6. Is this different? Not in principle. Note that we may be interested in verifying the net, or the algorithm that produced the net. The outputs tend to be statistical/probabilistic. Another issue is $\mathbf{R}$, but in practice $\mathbf{R}_{IEEE}$ at best[3]. Actually quite a lot of work, see [BS19].

Adversarial attacks Panda/Gibbon image. Discussion on how these are done. considers a Python/Z3 example, using the rich Python infrastructure. But inks are weak. Or could use Coq but there's less automation , and computing with defined objects is awkward. Neither group has any real infrastructure.

Artificial neuron, weights $w_{i,j}$ on inputs, compute $\sum_i w_{i,j}p_i - \Theta_j$ and output is ψ_j applied to this. Special case is perceptron, a linear neuron. Showed us the Python code (question: why does this matter) for perceptron for Boolean "and". Need to prove that a finite set of tests (the "ladder") is sufficient (pen/paper).

Problems of multiple languages: F* might be better

Q–Bundy Questions of ethics, sentencing etc.

[2] JHD: but Africa was represented by Algeria, then French.

[3] JHD: there's a lot of low-precision work: [Fel17] for 8-bit.

A Not really my area.

Q–Shankar This is really not the hard question: look at automated differentiation etc.

A You would be surprised how good some of the literature is (but also how bad some is!).

Q You talked about languages, but maybe the architectures of the NN are more critical.

A Daniel is working on this.

2.8 How to be a High-Frequency Trader: Donald MacKenzie

Public lecture at Informatics.

Speaker has a BSc (Maths) from Edinburgh, then moved in sociology, especially of science and mathematics. Recently turned his attention to financial algorithms.

1381 St Alban's townspeople stormed the Abbey and smashed the stones of the abbey, made from confiscated hand mills: see [Blo35]. The abbot's confiscation of these (to the benefit of his water mill) was an act of "material political economy"

Now Chicago Mercantile Exchange can quote times of 40ns for orders. But European Sovereign bonds forbid such high-frequency trading. Quotes that CME futures affect share prices in NJ immediately thereafter. But the speed of light in fibre optic is $\frac{2}{3}$ that of air, hence installing microwave links from 2010 onwards to Chicago from NJ. The firm Jump paid $40M for the field next to CME to park a microwave uplink there. There are discernible effects in patterns of US stock prices when it rains in Ohio and Pennsylvania, disrupting the 6GHz transmission.

2.8.1 High Frequency Trading

Similarly, project Express laid a fresh cable NY–London as close to the geodesic as possible, even though that was much shallower by US continental shelf. A geodesic link from London to Frankfurt would need masts $> 300m$ high — apparently planning permission was refused.

A U.K. farmer tweeted a picture of a box with aerials found in his field (sloping west) — apparently an attempt to do short-wave transmission from US.

Algorithms are thought of as market-making, entering passive bid/offer into order books. There are also "taking" algorithms try to grab these bids from order books before they are cancelled as the latest news arrives from Chicago.

This is essentially an arms race. Possible suggestions — asymmetric speed-bumps; periodic auctions (say 1/ms) rather than continuous trading.

2.8.2 Cryptocurrencies

Bitcoin mining: amazing photo of a mining farm. Probably uses "Bitmain Antminer S8 Hydro": 189 ASICs each with 100 hashing units. We believe bitcoin uses at least 4.77GW of power (Ireland 3.1GW). [dV18]. ASIC-resistant hashes are the equivalent of St Alban's hand-milling.

Q I am amazed that you have been able to go up to these locations and photograph them. Think of the damage you could do.

A I haven't given exact locations. Also note that rain is already a disruptor. Note that the markets wouldn't grind to a halt.

Q But aren't you worried?

A I am careful. That photograph was taken from a public cycleway, with an iPhone.

Q Where does the money come from?

A These firms that own the high-speed networks tend to be set up by retired traders who made their fortunes in the previous generation.

Chapter 3

29 May 2019

A random photograph: https://www.flickr.com/photos/icmsnews/47951277506/in/album-72157708786979088/.

3.1 Mathematical objects in dependent type theory: Buzzard

This talk is on video at https://www.youtube.com/watch?v=PSQq0RbgWH8.

"I saw Tom's talk live-streamed at INI 2017[1], and it changed my life". [Avi18] said "mathematicians need to put some skin in the game", and that got under my skin! I aim to digitise the IC Math curriculum. This is my midlife crisis, and, as Harrison said, this change is cheaper than buying a Ferrari. IC have funded a postdoc to observe the students watching me use Lean in my lectures. There isn't a set theorist in my Department: if we thought it was an important part of mathematics, we'd have one.

This talk is precisely about what mathematicians find difficult when using theorem provers, especially Lean. "Trying to do my own problem sheets in Lean taught me a lot about my own first-year course. If you'd told me a few years ago that I would learn new things about this course, I'd have laughed in your face."

Example 2 (What is a group) *We tend to see "G is a set with a multiplication", and e and $^{-1}$ tend not to be mentioned. Is it a pair $\langle G, \cdot \rangle$, or a 4-tuple $\langle G, \cdot, e, ^{-1} \rangle$*

Example 3 (My Q1.1) *"T/F: $x^2 - 3x + 2 \Rightarrow x = 1$?" Needed to get Mario to hack something into the Lean library, and now it works. Similarly needed Mario to add a* `Ring` *tactic.*

Example 4 ($\mathbf{N} \subset \mathbf{Z}$) *JHD didn't really understand this: it seemed to be about exact division. LCP noted Isabelle doesn't have his problem.*

[1] See §1.1 of https://staff.bath.ac.uk/masjhd/Meetings/JHDonBigProof.pdf.

Example 5 (First Sheet) *"Let $X = \{1, 2, \{2\}, \{1, 2\}\}$. Is $2 \in X$? Is $\{2\} \in X$? Is $\{2\} \subset X$." Lean's* `Set` *won't cope, because it's heterogeneous. I now realise that sets like this keep cropping up in example sheets, but not in actual mathematics.*

Example 6 ($F - E + V = 2$) *I couldn't do this, as the book says*

Example 7 *Let S be a set with two elements , and let $*$ be a reflective binary relation on S, Prove S is reflexive.*

The standard proof begins "w.l.o.g. $S = \{0, 1\}$", but this assumes a hidden result.

Lemma 1 *Let S and T be two sets and $f : S \to T$ is a bijection. Let $*$ be a relation on S. Then if $*$ is reflective (resp. transitive), then $f(*)$ is reflective (resp. transitive).*

Mathematicians never need to even state this result.

Example 8 $R[1/r][1/s] = R[1/rs]$. *This usage of "=" is totally common in mathematics, even EGA ("the most formal"). Grothendieck will write $R[1/r] = R[1/r^2]$ for example. This issue multiplies and multiplies. We could prove a whole family of "similar" results Neil had a completely different definition of $R[1/r]$ which were consequences of the universality of $R[1/r]$, and all this worked.*

3.2 Diagrammatic Notations in Mathematical Proofs: De Toffoli

1. Euclid = diagrams + text
2. Knot diagrams
3. Commutative diagrams

Will need a working definition of "diagram". Some diagrams enable an "acceptable" use of spatial-temporal intuition (Knot, topological, Venn, monoidal categories), but others don't involve this at all.

[Hal08] "Proofs are written in a way to make them easily understood by mathematicians. Routine logical steps are omitted. An enormous amount of context is assumed on the part of the reader." ... Quotes the "crisis in intuition" [Hah33]. [Hal08] "Proofs, especially in topology and geometry, rely on intuitive arguments in situations where a trained mathematician would be capable of translating those intuitive arguments."

Definition 1 *A* mathematical diagram *is a two dimension interpreted display which is an element of mathematical notation. As with elements of notation in general, a mathematics diagram is deployed ... Note that a linear e.g. LaTeX) encoding doesn't prevent it being inherently two-dimensional.*

Definition 2 *The* operative dimension *of a diagram ...*

Examples are Reidemeister moves in knot theory. Because of the singular points convention, it is easy to envisage a space know from the 2D diagram. These arguments relying on intuition are reliable proxies for formal arguments. But two 10-crossings in [little1899] were only found to be the same in 1974. Note that there are good knot codes.

Commutative diagrams are very different. They can be formalised (example in Lean) but are important for our understanding.

Therefore the "crisis in intuition" should not lead to a ban on diagrams.

Q–MK We've been looking at 3D, and think that the choice of cognitive operations is key, not the dimensionality.

A A major question for me is what we can do beyond linear sequences.

Q An argument about "all triangles" illustrated with a triangle, implies that we won't use special properties.

A

3.3 Formalising Mathematics-In Praxis: First Experiences with Isabelle/HOL: Koutsoukou-Argyraki

Quotes [BW05] Proof mining and proof interpretations.. Various logical metatheorems apparently justify this, but they serve as a guidelines for the extractability.. Doesn't guarantee how the quantitative information cane be extracted.

Define one-parameter non-expansive semigroups. So what are the common fixed points of $\{T(t) : t \geq 0\}$?

What makes a good proof?

- Shorter
- More elegant
- Simpler — Hilbert's 24th (JHD: see [Thi03])
- Reverse mathematics — a proof in a weaker subsystem.
- Interdisciplinary
- Easier to combine or reuse
- better computational content (complexity, numerically, "more elegant")

Hence LCP's project ALEXANDRIA. Want to

- expand libraries of formal proofs

- improve automation
- organise/consolidate libraries
- improving search
- verification of research-level mathematics
- assisting mathematicians writing research-level proofs.

Formalised a fairly recent paper in irrationality of numbers defined by series. We discovered an error in second paper, for which the original authors provided a fix. These are conditional on assuming Roth's Theorem.

3.3.1 Challenges

- Isar/jedit easy to use, but many syntax features are strange to newcomer
- "search theorems" doesn't help — many key names such as Borel do not feature.
- Manual search is very time-consuming. names/concepts is not a bijection! Asking on Math Overflow got "an overwhelming" response.
- Also searching for proof patterns, or algorithms.
- Automation (or its failures)

A major problem is *ex falso sequitur quodlibet*.

3.4 Teaching machines to do mathematics: Saxton

Team of authors from DeepMind. Examples of how powerful machine learning is. So what counts as Mathematics: arithmetic; proof; conjecture?

What is an NN: a highly parameterised function. Shows a graph of #parameters versus accuracy, with #parameters going up to 600M. Alternatively, message passing on graphs.

We regard mathematical structures as graphs, e.g. polynomials, Boolean structures. Looked at predicting entailment in propositional logic. Then there's extracting problems from algebraic word problems.

"algebra systems etc." are very powerful, but the lesson from machine learning is that if the system can learn by itself, it will grow in power.

3.4.1 Problems

- So what graph is needed to state the Riemann hypothesis
- Module+environment (e.g. ITP) are slightly more powerful, but can the environment ever be sufficient

* general NN problems.

- Data hungry
- These models are brittle.

Usual model is large dataset (e.g.Imagenet) goes to progress. Hence we collected/released a dataset of school-level problems. Synthetically generated with 50 types of questions (modules) and machine-generated 2M examples/module. We were hoping for transfer learning, e.g. addition helps with learning. Our structure was sequences of characters, eschewing the "hint" that structured trees gave. Because it's a synthetic dataset, we can control the various levels/axes of generalisation.

Tried two models.

- LSTM with Attention (2014) [BCB14]
- Transformer (2017) [Vaswanietal2017a] — state of the art in machine translation.

Give plausible looking (but often wrong) answers: e.g. integer factorisation, where they get small factors but get the rest wrong, so $N \neq \prod p_i!$

Both models trained on sequences of > 7 numbers of 1–3 digits, but given 1+1+1+1+1+1+1 (seven '1'), both models gave 6.

Q I thought you needed exponential data in #parameters, but you had "only" 10^8 examples.

A "Unreasonable success of machine learning"

Q What have you learned.

A They can't compose procedural knowledge.

3.5 The HOL Light Mathematical Libraries: Harrison

HOL-Light descends from MJCG's HOL, an LCF checked for classical HOL. Written in OCAML. Family tree of such "where the arrows denote plagiarism". This library was designed to support the D of Flyspeck, but grew. Many contributors. c200kLOC in basic, then 50kLOC for complex analysis and more topology, 20kLoC misc.

3.5.1 Decision Procedures

Very important. They can rarely solve problems on their own, but give a nice natural steps from "big lines". Examples are Gröbner bases (does solve theorems on quartics). Also covering congruences via Presburger.

I have some new ones, e.g. divisibility reasons over **Z**. Also normed spaces [SAH12].

3.5.2 Keep your theorems sharp

Minimise assumptions. Consciously test each hypothesis. Also formulate definitions as weakly as possible. Controversially, consider totalizing functions that take convenient values.

3.5.3 $A = B$ is too glib

You have to some overall integers, but what about factorials of negatives etc. $\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$ is only valid for $0 \leq k \leq n$ etc. We explain WZ via limits (problems with $(-1)^n$. Need a proof that can approach the limit avoiding bad values.

Apéry's $\zeta(3)$ proof involves recurrences a_n, which I can do in HOL Light.

Q How do people find these?

A They attend my talk. But seriously, this is similar to the theorem-finding problem.

Q Term rewriting etc.?

A I tend to use only bounded ones, which may be a false dichotomy.

Q Are you using these "bad values", rather like people thought about negatives, etc.

A Not personally. There is the structure of a "meadow", a field with total division.

3.6 Group Knowledge and Mathematical Collaboration: Tanswell

Paper 'The group theory of group theory". Yesterday, JA challenges philosophers to understand issues of reliability, knowledge and understanding in "big proof". Social epistemology seeks o investigate the epistemic effects of social interactions and social systems. What does it mean for a group to believe something. [MS95] "The Los Alamos team knows how to build an atomic bomb.

3.6.1 4CT

[Tym79] — "if we accept this, we are changing the underlying concept of 'proof'." He has a concept of "surveyable". The use of the computer is at first glance oracular, telling you something is true without telling you why.

Speaker I think GG referred to a "folklore theorem" he couldn't prove.

GG I made some different choices, so the theorem was trivial here.

3.6.2 CFSG

[Ste12]: 300–500 articles, 5–10,000 journal pages. So what is/isn't part of the proof? To what extend can mathematical knowledge be obtained by an individual. "Because the proof is long and scattered across multiple journals etc. . . . ".

Floor Not true today.

In Tymoczko's sense, the proof isn't surveyable.

Bundy "Surveyable" also means you can learn from the proof technique.

[Gor83, pp. 7-8] "The prevailing opinion is that every significant configuration has been considered". Mathematicians don't check every proof in detail, and standards vary.

Hence the claim is that the proof is understood at a collective level.

CJS I am not sure: Feynman commented that a problem with Challenger was the failure by the design team to understand shapes of constant width, a 19th-century problem.

So should we believe that CFSG is understood by a collective machine.

Q Can't we talk about justification at the individual level: both mathematical and ?structural.

A But structural doesn't cover the mathematical.

Q Trust is an important part of the equation.

A Yes, and the collective needs collective trust.

Q The trick is to turn something that's difficult to validate into something that's easy to validate.

Q What does checking mean? Consider [Gon08] on 4CT. This is the same sort of proof as before, except that we believe Coq, and in principle we could write a checker for this proof.

Q Logic is a model of human thought? So what about multi-agent systems?

A But there are also issues of credit. There is no credit for a second proof. There is some (smaller) for a formalised proof.

SMW Also zero-knowledge proofs.

3.7 Automating "human-like" example-use in mathematics: Pease

"How do people do mathematics?" — there isn't one answer, either across communities or across time. Examples are important, and the more concrete the better. This is especially in "back-stage" reasoning. 1/3 of conversations in MathOverflow have comments. Hence a typology of examples.

3.7.1 ??

We applied NLP to learning patterns in data sets. We are "early explorers". NLP generation has progressed immensely over time. Dataset from MathStackExchange with 1M questions, 1.4M answers, 450K questions with one accepted answer.

1. Markov Model n-gram. Standard NLP technique five years ago.
2. Case-based: look for most similar past question, take its best answer, and massage (regex). Implementation based on ElasticSearch — very simple.
3. LTSM (Long Term Short Memory) NN. Slow to train — 1month for a fairly small NN on a consumer-grade GPU. Was state-of-the-art until recently, now surpassed by "attention models". Contrasts £300 unit we had with £10M TensorFlow component. However, one can use transfer learning to build on such products.
4. GPT2 — a project from OpenAI, which is a 345M weights deep network. Learns P(next word—previous test). It's an Attention-based Transformer. Initially trained on 8M web pages, with no specific maths training.

But how to evaluate? So we evaluate "recognising the correct answer from candidates". But we only gave them correct answers to choose from. Random: 20%, Markov 45%, LTSM 40%, Case-based 60%. GPT2 is "further work". The Markov model kept producing $\$^2$. Case-based is high-grade regurgitation. GPT2 produces complete gibberish, but extremely high-quality gibberish.

Q How did the ML answer Alison's RQs.

A One question was whether we can build an answering machine.

Q Way forward?

A GPT2 + transfer learning, then filter out the gibberish!

[2] JHD: LaTeX markup? Yes.

Chapter 4

30 May 2019

4.1 Kohlhase

"Big Math", not just "Big Proof". Note that there's a big push on research data, especially FAIR, and I think Math should be part of this.

Q KB said "If you build it, they will come", but is this "it"

A Good question. CFSG is an example of mathematics tackling ever larger questions. It is estimated that the "condensed" second-generation proof will be 5000 pages. Hence we are hitting the "one brain barrier". This might be the "small group barrier", but see [Bro75].

I claim that "Big Proof" doesn't work: rather "Big Library+Medium Proof". Note that "Library" includes abstraction. There's also an implicit library of (counter-)examples. Claims Ontology is the centre of Narration/Computation/ Models/Inference tetrapod.

Q What do you mean by "ontology"?

A Not OWL? A formal representation of objects and how they relate to each other.

LCP: "I'd like to inherit this from there ...". Notes that there are several one-aspect systems, a few two-aspect systems, but the only complete system is Mathematica.

Note that "research data" is a big wave that is coming. This caused some debate in the audience, some saying that "Pure Mathematicians don't do data" is sufficient.

EOSC is a great new federated something-or-other. There is only one mathematical data set there, and no service. Hence the FAIRMath proposal, which was rejected as "not sufficiently metadata". Both Wikipedia and arXiv are oriented towards human brains, as is OEIS/ LMFDB.

There is a strong open-source spirit (IMU resolutions etc.) in the mathematical community. But accessible is a problem due to internal structure. Reusable — no copy/paste: what is D_4 etc. Findable: formula search engines are in their infancy.

Symbolic Data Very context-sensitive

Concrete Data OEIS, LMFDB etc. But users have to know (and have an implementation of) the representation theorems.

Linked Data Still not FAIR, as you don't really have the semantics of the object.

LMFDB is about the best there is, but in practice one e-mails John Cremona to get an export one can handle in a tool. And even then the semantics is being conveyed informally. See problems with semantics of units [Nat99] and genetics/Excel [Got16].

There is "shallow FAIR" at the data set level (provenance is mostly here), and "deep FAIR" at the item level.

Floor Biologists have been doing Deep FAIR. Astronomy is Shallow FAIR,

MK The Astronomy is mostly array data: there is little linkage with papers, symbolic data etc.

We can have symbolic data in our system, with "little theories" via various codecs. MathDataHub needs semantics, but semantic services tend to be data-specific. We have a MDDL, and there's an underpinning database which does extensibility etc.

4.2 Learning and Reasoning over Big Proof Corpora: Urban

Mathematical (scientific) thinking is pattern-matching, analogy, induction from examples. Deductive reasoning, and complicated feedback loops. Poincaré versus Hilbert. My MSc (1998) Try ILP to learn rules and heuristics from IMPS/Mizar. Symbolic methods at the time didn't work.

High-level Choose lemmas, pre-select ATP strategy/portfolio; preselect hints

low-level guide every inference (tableau, superposition); guide kernel steps of LCF-style ITPs;

mid-level guide application of tactics, invent suitable ATP strategies, invent suitable conjectures, invent suitable concepts/models

Proof sketches explore stronger/related theories to get roof ideas

Theory exploration

1. Hammering Mizar. Take a Mizar proof, delete justifications, and get Hammer to reinstate

2. Tactics in HOL4. TacticToe research. [GKU17].

3. 45% of top-level lemmas can be proved this way.

Looking at various new techniques, e.g. distance-weighted k-NNN, LI boosted trees. Work on finite (counter)-models. On MTPT278, KNN proves 900, bets current models nearly 1200. 260 theorems have one proof, 160 with 2, 140 with 3 etc.

A good heuristic

$$\frac{w_i}{n_i}\,{}_c\, p_i \sqrt{\frac{\ln N}{n_i}}$$

is a good heuristic for exploring new versus exploiting old. While his initial rlCoP is significantly worse that leanCoP, after eight iterations it is 42% better.

With "Curriculum learning" (version of Reinforcement Learning)we can learn addition and multiplication perfectly. But this doesn't yet extract to something he can reason about.

Also very good results on recurrent NNs with attention. Very good at inf2formal. Can do rewriting and normalisation (he means canonicalisation) of polynomials.

Feedback loop for ENIGMA on Mizar data sets. Went from 14933 solves to 25k.

Also ProofWatch: Statistical/Semantic guidance of E [GJSU18]. Uses Veroff's "hints" method for Prover9/AIM.

Combining these gives us ENIGMAWatch.

4.3 A formal classical prof of Hahn–Banach in Coq: Kerjean

"I've been doing formal proof for six months". Based on MathComp and MathCompAnalysis libraries. So I am new to ssreflect: a few months ago reading these lines was very painful. `case: z{zmax} gP => [c [_ _ bp _]] /= gP; apply/bp/gP.`

States Hahn–Banach in three lines of Coq.

Textbook: extending f to a linear function $F \oplus \mathbf{R}v$ bounded by p follows from the convexity of p and the linearity required for the extension. Then extend to whole space "by Zorn's Lemma".

There is also work on a constructive version by Coquand et al., but not my concern. [Gon12] for MathComp.

Because we want to talk about partial linear functions, we really need linear relations.

MathComp proofs are often written in an imperative minimal style: claims this is easier to maintain (debate here).

Note that Coq doesn't have to be used for constructive mathematics — we can do classical analysis. Mathematical Components Analysis : CIC+ + AxiomOfChoice + Excluded Middle + functional Extensionality + Propositional Equality. This reinterprets and extends Coquelicot [BLM15]. Alas there are (at least) three treatments of **R** in Coq: currently various transfer lemmas are needed.

Reasoning on the graphs of linear functions which are bounded by a convex function and which extends f.

Searching by patterns doesn't find a lemma about continuous functions, and by name doesn't find Hahn–Banach. There is a naming convention, with useful suffices, e.g. A for associativity. So `locally_normE` is a definition elimination result.

But we need really to use our theorem, to prove that it is stated in a useful way. Showed one.

We use Coq's sort `Prop`,which allows propositional extensionality. Our proof used Zorn's Lemma, hence `Choice`. The part without normed spaces can be proved in MathComp. MathComp is `bool`/`Prop`, whereas MCA is purely `Prop`. This causes some issues. Also V is finite dimensional vector spaces, which can cause confusion.

4.4 Why did mathematicians not embrace the vision of the QED Manifesto: Löwe

See [Boy94]. [Wie07]: "The future envisaged by the QED Manifesto hasn't happened. The original manifesto gives 11 reasons why it might fail.

1. "Too few people are working on formal mathematics". This is probably not quite true, and certainly not the main reason.

?. "Formal mathematics doesn't look like mathematics". True at the surface level (see previous talk), but again probably not the reason. More deeply, the formal structure is very exposed.

FW To modify [Wie07], my problem isn't logicians, it's constructive logicians.

The traditional narrative is that mathematicians are Luddites: writing proofs this way would require them to write programs, and understand axiomatic methods: both are anathema. In order to make formal proofs acceptable, we have to rewrite them to conceal this.

But this is too simplistic. What are the real reasons? Mathematicians aren't hostile to logic, but it's not quite what they do. Note the Committee to consider merging the Logic and Mathematics institutes at Amsterdam: I proposed ATP, and the mathematicians said that the logicians could do that, but there was no linkage.

Long debate about the role of programming in math. education. But the "mathematicians hate programming languages" doesn't really hold up.

The typesetting issue is more relevant. I was a student when LaTeX was around, but not expected. I have therefore seen the complete revolution. It's "only about typesetting", but the conservative mathematicians of the 1990s were worried that LaTeX would change the way we write mathematics.

[Har08] support for correctness, support for refereeing. This raises questions of whether it is the formal or the informal proof that is the "real" mathematics.

> In Bourbaki's view, the foundations of mathematics are roped-off museum pieces to be silently appreciated, but not handled directly. There is an opposing view that regards the foundational enterprise as unfinished until it is realised in practice and written down in full. [Hal08]

There is a strong narrative that the role of peer review is to guarantee correctness. But this isn't true in (informal) practice [GLVK10]. But we have no statistical data on the practice of refereeing. Empirically we can see that mathematicians do not value refereeing enough to write formal proofs to help the referee.

But note that the standard of proof has changed in the past. What we see is a reflection of the system of the (late) 20th century.

4.5 On the Impact of Big Proofs on Provers: Gonthier

The practice of theorem proving has much progressed from batch in the 1960s to interactive systems. Much is technological, but some is from tackling progressively larger proofs. Holds for the proof language, but especially library design. Very personal account.

Started by verifying the OCAML concurrent GC in Lamport's TLA+. This was a shell on top of ?REVE on top of LP. Essentially proof by guided rewriting.

Six years later bitten by 4CT. Revised proof has a reasonably compact decision procedure. Need a set of configurations with is both unavoidable and reducible (colourable by induction). The first part was 10^4 cases, the latter 10^9. Turned out BDD caching wasn't needed, hence it's purely functional (2000). Colouring only was done by small-scale reflection (hence SSReflect; 2001–2). The graph theory needed a fair bit of word (hypergraphs etc.), then a major discovery, and finished in 2004.

Making each group into a separate type causes a great deal of injections. So we place everything in a universal group, and really do subgroup theory. The $G/H := N_G(H)\langle H\rangle/\langle H\rangle$.

Formalising characters was a challenge. You never mix characters from different groups, so make `class_fun G` into a type `CF(G)`.

There were integer norm problems. Naïve SMT failed, as did SAT bit-blast. I wrote my own decision procedure.

Use Group Functions (Garillot) to map characteristic subgroups like $Z(G)$.

What next: support for (re)structuring proofs, and the ability to handle open/incomplete proofs. The latter would help me handle Landau notation.

Q When you say groups are sets, what are they?

A Essentially a list of Booleans, saying whether the element is present.

4.6 Logipedia:a system-independent encyclopedia of formal proofs: Dowek

There have been reference to Wikipedia, but these were errors! In pre-history, we wrote software, and that implicitly defined a format. How we define formats first: ASCII, HTML etc. But the formal proof community has remained in prehistory, e.g. "a Coq proof of 4CT". This has problems of interoperability and sustainability.

Of course, there are limitations, e.g. ZF≠ZFC. But in fact one direction is trivial, and the other is partially trivial.

Example 9 ($\exists$ a basis of $\mathbf{R}^2$) *One proof is via incomplete basis theorem (and AC), or, e.g.,* $\langle 0,1\rangle, \langle, 1, 0\rangle$.

This leads to "reverse mathematics". Let's focus on formal proofs, expressive theories (set theory etc.) and in the analysis of proofs. The advantage is that ZF and ZFC are in the same framework, predicate logic. Invented 1928 [HA28]. But Church's type theory (1940) is outside predicate logic. 1970; 1985 we have Martin-Löf etc. So do we give up?

I propose to extend predicate logic. What needs fixing? And what do we know?

1. No bound variables. λProlog, Isabelle, $\lambda\Pi$calculus etc.
2. No syntax for proofs. $\lambda\Pi$calculus.
3. No notion of computation. Deduction modulo theory.
4. No good notion of cut. Deduction modulo theory.
5. Classical and not constructive. Ecumenical logic (which has different symbols for $\exists$ and $\exists_{\text{constructive}}$ etc.

Hence we need $\lambda\Pi$calculus modulo theory — Dedukti is our implementation.

Simple type theory has eight declarations and three arrows in Dedukti. CoC has one more of each, and a few changes; arrow is dependent, for example. The new symbol is π (proof to term). We can see the differences clearly as the two are in the same framework. Hence every Simpletype theory translated directly to CoC. For example, we could translate the *entire* MATITA arithmetic library back into Simple Type Theory, which gives us the first constructive proof in

Simple Type Theory of Fermat's Little Theorem. Can therefore export this to many systems. So we put this in a database.

Next candidate is the standard library of HOL-Light. This uses inductive types/Q_0, which I view as a big step backwards. Hence we need to do concept alignment. This also needs to make formal the statement "Cauchy sequences: Dedekind cuts — doesn't matter". Because of ecumenical logic, we can use both disjunctions.

QED [Boy94] wanted a simple system. We understand theories much better. We can produce a new logical framework to express these theories.

Interoperability is not about standards committees: it's a research problem.

4.7 Why won't they use my stuff, and what can I do about it: Martin

Actually change my/our and I/we. Showed a word cloud of the abstracts. Blissfully free of management speak. Shows agenda of latest HoDoMS meeting, which is completely free of proof.

Vannevar Bush 1945: "a rant".

Discussed her paper about research impact based on REF 2014. Social Sciences distinguish various kinds of impact.

Instrumental Impact New products

Conceptual Impact "you can do it this way"

Capacity building

Attitude/cultural change

Enduring connectivity

Note the role of "knowledge intermediary" — often silent.

Quotes Gowers' "relaxed" attitude to computer proofs. KB's talk (section 3.1) is attitude building; TH's team in Vietnam is capacity building. But how do we get the message across, especially to Deans.

Floor Maybe the right word is "proofsmith".

FW Henk B. always says "Computer Mathematics".

Chapter 5

31 May 2019

5.1 The "NASA Effect" of a Global Digital Math Library: Watt

Come from computer algebra via languages to MKM. Good talk by UHM (section 4.7), but she wasn't perhaps direct enough. In computer algebra, some domains (polynomial algebra etc.) are "baked in" and correct. More generally, though, we then more (often imperceptibly) into looser semantics, partly because of [Ric68]. In Axiom, we had formal properties, but they were strictly documentation — in the 1980s formal proof was considered infeasible.

Document analysis also exists, mostly interested in metadata. Also the whole formal side. These three join in MKM. Presentation from 2006 quoted 2M articles, growing at 80K/year. Hence soon (if not now) half of all mathematics is born digital. Note [Nat10] as an example of a modern view of "born digital" mathematics, but still not formal.

Even basic ideas such as equation can vary [MW12]: note that equation/identity is undecidable. Conversely J_ν can be a Bessel function or an angular momentum or ....

Example of derivation of Dirac's equation — what do the symbols mean? So how does one do flexiformalism? Modula 3 had the `unsafe` declaration.

IMKT.

1957 Sputnik.

1958 NASA

1962 We choose to go to the Moon (Kennedy).

Pushback costs too much.

1969 Apollo 11.

1979 Majority of Americans felt Apollo didn't justify its costs.

1999 Majority of Americans felt Apollo did justify its costs.

\+ Increased PUS, technological advances (remote sensing, sea floor mapping, electronics), commercial spinoffs, even political benefit (quoting Sakharov's open letter). Intel, Surrey Satellite Technology Limited. Bezos was president of Students for Exploration and Development of Space, hence Amazon!

− Wikipedia points out that Tang, Teflon and Velcro were *not*.

Q–FW IPR.

A This was much more true ten years ago than today. Publishers are more frightened. Some databases are more open than others.

Q Gray literature?

A You have to start somewhere. IMKT SC has different views here. But we're happy to hear ideas.

Q–JU I was involved in lobbying EU for AI and in France the argument was "France is good at Maths". "Verification of AI" is a key area.

Floor Discussion on nationalism versus science.

5.2 Panel

Natarajan Shankar Chair.

Freek Wiedijk I seen various verifications of some systems in others, which is good. I am an old man, and have been in this field for a long time. But at the Lean workshop I saw many people I hadn't seen before. FAbstracts is fundamental, as it let's us talk real mathematics. Floor: not convinced — the libraries are nowhere near being able to formalise my statements. Floor: I am writing a paper, part of which has some fiddly combinatorial arguments, and these we are doing in Lean, and will place with the Lean on the arXiv. Floor': a real problem is when the computer scientist says "it's only an engineering problem", meaning we won't do it, and things like "interfacing to SAGEmath rather than Mathematica" won't happen.

John Harrison Interested in scaling, both the proofs but also the number of people who can do this. This workshop is formally top-down, but I've been impressed by the bottom-up energy. This meeting has brought people together, but has also demonstrated that there is no real mathematician who has actually built a theorem-prover. Agree with AB about the need to involve students. Note that Cauchy formalised calculus to ensure his students wouldn't make mistakes. JU: it is scary to hear people at POPL say "I could formalise this in Coq" despite not knowing the semantics. Floor: Bourbaki was fundamental to perfectoid spaces: hence first four chapters of General Topology (i.e. Vol 1) is not formalised.

Alan Bundy My FRANK (NK= New Knowledge) project is about reasoning and moving information retrieval beyond fact retrieval. With regard to FAbstracts: these things succeed if they improve the current system for the individuals. They key thing about "Big Proof" is concerns over surveyability, not about correctness. We also find ways to improve our students in the construction and development of knowledge.

Alison Pease Follow UHM (Section 4.7). Hue philosophical issues of trust, more complex with machines. Additional information, such as "why" and "how".

Floor Look at the "supporting comments" on MathStackOverflow etc. We mustn't ignore the human factor.

Silvia De Toffoli Organising a conference on Philosophy of Mathematics next year. Is it's dying? We should focus on the dynamic aspect of mathematics, as well as the static side that FAbstracts seem to be serving.

Floor With a computer, you have the ability to delete a supposition and see what breaks.

Ian Ford (Wolfram) I spend a lot of my time at Wolfram on documentation etc., so very keen on this aspect, which seems to be one of Lean's strength.

Stephen Watt Just spoken.

Appendix A

JHD's reflections on Machine Learning

Inspired by the Deepmind presentation (§3.4).

JHD commented: Look at Todai Robot [AMIA14], which can almost get you into Tokyo University to read mathematics, and can get you into most other Japanese universities, versus Deepmind's failure to get anywhere with GCSE Maths (the examination taken by practically everyone aged 16 in England). [New19].

Todai Robot uses ML to read the exam paper, then ML+linguistics to understand the text (distinctly non-trivial: see [AMIA14, §3]), then domain-specific reasoners. The commonest one is "Real Algebraic Geometry". The Deepmind system is presented as being purely ML.

Bibliography

[AMIA14] N.H. Arai, T. Matsuzaki, H. Iwane, and H. Anai. Mathematics by Machine. In K. Nabeshima, editor, *Proceedings ISSAC 2014*, pages 1–8, 2014.

[Avi18] J. Avigad. The Mechanization of Mathematics. *Notices AMS*, 65:681–690, 2018.

[BCB14] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. `http://arxiv.org/abs/1409.0473`, 2014.

[BLM15] S. Boldo, C. Lelay, and G. Melquiond. Coquelicot: A user-friendly library of real analysis for Coq. *Mathematics in Computer Science*, 9:41–62, 2015.

[Blo35] M. Bloch. Avènement et Conquêtes du Moulin à Eau 1. *Annales d'histoire économique et sociale*, 7:538–563, 1935.

[Boy94] R. Boyer. The QED Manifesto. In A. Bundy, editor, *Proceedings CADE 12*, pages 238–251, 1994.

[Bro75] F.P. Brooks. The Mythical Man Month. *Addison-Wesley*, 1975.

[BS19] A. Bagnall and G. Stewart. Certifying the True Error: Machine Learning in Coq with Verified Generalization Guarantees. *To appear in AAAI 2019*, 2019.

[BW05] H. Barendregt and F. Wiedijk. The challenge of computer mathematics. *Philosophical Transactions of the Royal Society A: Mathematical*, 363:2351–2375, 2005.

[CMSPT14] F. Chyzak, A. Mahboubi, T. Sibut-Pinote, and E. Tassi. A Computer-Algebra-Based Formal Proof of the Irrationality of $\zeta(3)$. *International Conference on Interactive Theorem Proving*, pages 160–176, 2014.

[dV18] A. de Vries. Bitcoin's Growing Energy Problem. *Joule*, 2:801–805, 2018.

[Fel17] M. Feldman. Exabyte Measures Linpack Performance Across Major Cloud Vendors. https://www.top500.org/news/exabyte-measures-linpack-performance-across-major-cloud-vendors/, 2017.

[FT63] W. Feit and J.G. Thompson. Solvability of Groups of Odd Order. *Pacific J. Math.*, 13:775–1029, 1963.

[GJSU18] Zarathustra Goertzel, Jan Jakubův, Stephan Schulz, and Josef Urban. Proofwatch: Watchlist guidance for large theories in e. In *International Conference on Interactive Theorem Proving*, pages 270–288. Springer, 2018.

[GKU17] Thibault Gauthier, Cezary Kaliszyk, and Josef Urban. Tactictoe: Learning to reason with hol4 tactics. In *LPAR*, pages 125–143, 2017.

[GLVK10] C. Geist, B. Löwe, and B. Van Kerkhove. Peer review and knowledge by testimony in mathematics. *PhiMSAMP: Philosophy of mathematics: Sociological aspects and mathematical practice*, pages 155–178, 2010.

[Gon08] G. Gonthier. Formal Proof — The Four-Color Theorem. *Notices A.M.S.*, 55:1382–1393, 2008.

[Gon12] Gonthier,G. *et al.* The formalization of the Odd Order theorem has been completed on September 20th, 2012. http://www.msr-inria.inria.fr/Projects/math-components/feit-thompson, 2012.

[Gor83] D.M. Gorenstein. The Classification of Finite Simple Groups. *Plenum Press*, 1983.

[Got16] P. Gothard. Microsoft Excel errors mar one-fifth of science papers on gene research. http://www.computing.co.uk/ctg/news/2468859/microsoft-excel-errors-mar-one-fifth-of-science-papers-on-gene-research, 2016.

[GT12] G. Gonthier and L. Théry. Formal Proof — The Feit–Thompson Theorem. http://www.msr-inria.inria.fr/events-news/feit-thompson-proved-in-coq, 2012.

[HA28] D. Hilbert and W. Ackermann. Grundzüge der theoretischen Logik. *Grundlehren der math. Wiss.*, 27, 1928.

[Hah33] H. Hahn. The Crisis in Intuition. In *Krise und Neuaufbau in den exakten Wissenschaften*. Fünf Wiener Vorträge, 1933.

[Hal05] T.C. Hales. A proof of the Kepler conjecture. *Ann. Math.*, 162:1065–1185, 2005.

[Hal08] T.C. Hales. Formal Proof. *Notices A.M.S.*, 55:1370–1380, 2008.

[Har08] J. Harrison. Formal Proof — Theory and Practice. *Notices A.M.S.*, 55:1395–1406, 2008.

[HKM16] M.J. Heule, O. Kullmann, and V.W. Marek. Solving and verifying the boolean pythagorean triples problem via cube-and-conquer. In *Proceedings International Conference on Theory and Applications of Satisfiability Testing*, pages 228–245, 2016.

[Lei66] G.W. Leibniz. *Dissertatio de arte combinatoria.* Fickius et Seuboldus, 1666.

[MS95] Donald MacKenzie and Graham Spinardi. Tacit knowledge, weapons design, and the uninvention of nuclear weapons. *American journal of sociology*, 101(1):44–99, 1995.

[MW12] S. Marcus and S.M. Watt. What is an Equation? In *Proceedings SYNASC 2012*, pages 23–29, 2012.

[Nat99] National Aeronautical and Space Administration. Mars Climate Orbiter: Phase I Report. `ftp://ftp.hq.nasa.gov/pub/pao/reports/1999/MCO_report.pdf`, 1999.

[Nat10] National Institute for Standards and Technology. The NIST Digital Library of Mathematical Functions. `http://dlmf.nist.gov`, 2010.

[New19] New Scientist. DeepMind taught an AI to take a school maths exam — but it failed. `https://www.newscientist.com/article/2198761-deepmind-taught-an-ai-to-take-a-school-maths-exam-but-it-failed/`, 2019.

[Pól62] G. Pólya. *Mathematical Discovery: on understanding, learning and teaching problem solving.* Wiley, 1962.

[Ric68] D. Richardson. Some Unsolvable Problems Involving Elementary Functions of a Real Variable. *Journal of Symbolic Logic*, 33:514–520, 1968.

[SAH12] R.M. Solovay, R.D. Arthan, and J. Harrison. Some new results on decidability for elementary algebra and geometry. *Annals of Pure and Applied Logic*, 163:1765–1802, 2012.

[San13] C.J. Sangwin. *Computer-Aided Assessment of Mathematics.* Oxford University Press, 2013.

[Ste12] A. Steingart. A group theory of group theory: Collaborative mathematics and the 'uninvention' of a 1000-page proof. *Social Studies of Science*, 42:185–213, 2012.

[Thi03] R. Thiele. Hilbert's Twenty-Fourth Problem. *Am. Math. Monthly*, 100:1–24, 2003.

[Tym79] T. Tymoczko. The four-color problem and its philosophical significance. *The Journal of Philosophy*, 76:57–83, 1979.

[Wie07] F. Wiedijk. The QED Manifesto Revisited. *Studies in Logic Grammar and Rhetoric*, 10.23:121–133, 2007.