AUGMENTING THE HUMAN MATHEMATICIAN

Henrik Kragh Sørensen & Mikkel Willum Johansen
Section for History and Philosophy of Science
Department of Science Education
University of Copenhagen
Denmark
{henrik.kragh, mwj}@ind.ku.dk

Renee Hoekzema
Mathematical Institute
The University of Oxford
Oxford, United Kingdom
hoekzema@maths.ox.ac.uk

Hester Breman
Faculty of Psychology and Neuroscience
Maastricht University, The Netherlands
hb4367@ou.ac.uk

ABSTRACT

In this article we consider important developments in artificial intelligence within automated and interactive theorem provers (ATP/ITP). Our focus is to describe and analyze key challenges for interactive theorem provers in mainstream mathematical practice. Our broader research program is motivated by studying the functions of visual internal and external representations in human mathematicians and the role of epistemic emotions. These aspects remain gorges to bridge in developing ITPs. But by seeing ITPs as augmenting the human mathematician, we stand to gain the best of two epistemic practices in the form of a hybrid — a centaur.

1 INTRODUCTION

Computers are ubiquitous in mathematical practice, yet the prospect of profound epistemological revolutions of mathematics have not yet materialized. It has been foreseen, though, often either in the form of epistemic and creative autonomy granted to automated theorem provers or in the form of outsourcing lengthy computations and relying on the results as a posteriori epistemic claims.

One productive way of framing this debate can be found in the sociologist of mathematics and AI Donald MacKenzie, who distinguished between different and largely disjoint domains: In the AI research community, MacKenzie suggested, "automatic theorem proving [was pursued] where the aim is to simulate human processes of deduction". This contrasted with the efforts in mathematical logic which pursued "automatic theorem proving where any resemblance to how humans deduce is considered to be irrelevant" and with the efforts in computer science for the verification of software and hardware using "interactive theorem proving, where the proof is directly guided by a human being" (MacKenzie, 1995). MacKenzie's typology is already 25 years old, but in the following we use his distinctions to gauge and discuss challenges facing the broader acceptance of (interactive) theorem provers in mathematical research. For this, we identify the three strands with 'artificial', 'autonomous' and 'augmented' mathematical theorem proving, respectively, and we return to this in the conclusion.

1.1 BACKGROUND

Recent advances in AI research and a new wave of computer savvy mathematicians have renewed the discussions over the role of digital augmentation of the human mathematical process. Machine learning agents have progressed to the point where they can suggest statements for mathematical inquiry (Raayoni et al., 2021), and efforts are underway to formalize large parts of undergraduate mathematics in Lean (Buzzard, 2020) or Isabelle/HOL (Koutsoukou-Argyraki, 2020) and actively use such proof assistants in teaching (Buzzard, 2019). This trend has also not escaped the broader professional press, and hopes, visions and ambitions are high (Ornes, 2020).

Yet, the history of AI urges caution when the hype is at its height. Famously, after constructing their General Problem Solver (GPS) — indeed one of the first breakthroughs in automating mathematical and logical thought — Herbert Simon and Allen Newell predicted in 1958 that "within ten years a digital computer will discover and prove an important mathematical theorem" (Simon & Newell, 1958). In fact, it took until the Appel-Haken proof of the Four-Colour Theorem announced in 1976 (Appel & Haken, 1976) for computers to really enter essentially into a mathematical proof, and this was in the form of checking a large number of cases with the construction of the proof done by humans. And the first discovery of an 'interesting' mathematical result obtained by computers is perhaps the proof of the Robbins Conjecture in 1997 (McCune, 1997). So the initial optimism was somewhat dampened — both by the delay of decades and by the contested (if not outright sceptic) reception of new computer-assisted mathematical knowledge claims: The Appel-Haken proof is still sometimes considered 'ugly' (Montaño, 2014; 2012) and William McCune's automated solution to the Robbins problem was immediately transformed into more 'accessible' (anthropomorphised) form by human mathematicians. Thus, both these famous milestones in computer-assisted mathematics have come to point to a reluctance among human mathematicians in relying on computers and now form a barrier to be overcome in communicating proofs.

In what could be called the next phase of computerized mathematics, with the advent of so-called 'experimental mathematics' in the 1990s, proponents were arguing that mathematics could rely on computers for both heuristic, exploratory and demonstrative functions (Sørensen, 2008; Borwein, 2012). Jon Borwein and David Bailey presented their arguments for drawing inspiration from digital experiments in mathematics and using these to construct human proofs (Borwein & Bailey, 2004). And at the same time a debate arose around Doron Zeilberger, who would controversially list his computer as a co-author on papers, proving identities by the Wilf-Zeilberger algorithm (Zeilberger, 1994; Ekhad & Zeilberger, 1996; Petkovšek et al., 1997). In a response to Zeilberger, George Andrews defined the gold standard: "Until Zeilberger can provide identities which are (1) discovered by his computer, (2) important to some mathematical work external to pure identity tracking, and (3) too complicated to allow an actual proof using his algorithm, then he has produced exactly no evidence that his Brave New World is on its way" (Andrews, 1994, p.17).

In the following, we combine empirical evidence in the form of qualitative interviews with working mathematicians with a conceptual analysis and overview of features of interactive theorem provers to point to some key philosophical issues about mathematical practice that lie at the nexus of a broader acceptance of interactive theorem provers into 'mainstream' mathematical practice (Sørensen, 2012).

1.2 Insights from working mathematicians

The emerging field of philosophy of mathematical practice makes use of empirical methods to get access to the mathematical research process Löwe & Kerkhove (2018). In semi-structured interviews with research mathematicians in 2012, the respondents were asked about the practices of developing and solving new mathematical research problems (Misfeldt & Johansen, 2015; Johansen & Misfeldt, 2014). These interviews revealed the surprisingly intricacy of problem choice. Thus, respondents would try to balance three criteria: 1) personal interest to the researcher, 2) perceived level of difficulty as non-trivial but also solvable, and 3) interest among peers in the research community (Misfeldt & Johansen, 2015).

The first of these criteria does not warrant much more attention except to say that choices need to be made about which problems to develop and attack. The second criterion involves metacognition in the sense of the ability to assess the limits of your own cognitive powers and the ability to reliably rank these with those of other mathematicians and approaches. The respondents would often rely on professional experience and their assessment of where they could gain a 'head start' by drawing on their work with similar problems or similar promising methods. The third criterion was perceived as being both of utmost importance and difficulty, as essentially it focuses on carving a niche for your work among your intended audience (Andersen et al., 2019; Ashton, 2020). Therefore, recognizability is a powerful value in problem choice, as peers would have to be able to identify and learn from your work. Thus, mathematical practice is a social one that also relies on more informal ways of communication for its practitioners to follow and calibrate with the interests of their peers, for instance by attending conferences and workshops, by collaborating and by using a shared pool of techniques and visualizations to aid understanding and generate new high-level concepts.

These informal and social aspects of human mathematical practice pose a real challenge to any attempt to automate mathematics. Crucial aspects of mathematical practice (such as the development and choice of suitable problems as well as the development of the representations and concepts needed to attack these problems and communicate solutions in a meaningful way) seem to lie beyond the scope of formal and mechanical reasoning. This goes some way to explain the limits to purely formal approaches such as those exposed by the GPS and the Robbins Conjecture, above. However, these limits can be approached in different ways; one can try to overcome them, or one can accept them and work within them as boundary conditions. In this short paper we will try the last approach and explore how humans and computers can work together in a constructive way (or rather: how human mathematical practice can incorporate computers as a new and powerful tool).

1.3 Interactive theorem provers

Based on the example of the GPS, a number of features are recurring in automated and interactive theorem provers: First, the computer software works towards a "goal" (hypothesis) which is described by a human mathematician and encoded in something like human-readable form. Various systems employ different formats and syntax, and learning curves for the novice mathematician can be more or less steep. The ITP can then proceed to try to break the goal into subgoals, the realization of which would inform a proof of the overarching goal. This process can draw on libraries of strategies, and it can include supervised guidance by a human mathematician. The idea is to provide a break-down of the main goal into sub-goals such that a proof of each subgoal would lead to a proof of the main objective. Some of the subgoals may have 'easy' proofs that can be filled either through a variety of mechanisms called 'resolution' or through e.g. library look-ups of theorems, formula reductions or the like. Thus, the 'interactive' part of ITPs mainly involves the human user providing goals and heuristic strategies to the computer, which in turn suggests possible sub-goals to be explored either by human, by computer, or by a combination.

Automated mathematics has already existed for decades, with Automath (Bruijn, 1970) being one of the first more generally viable systems. However, the real change has come with the introduction of *interactive* theorem provers with a machine-learning based component and suited for ordinary desktop research. These components are based on techniques used in language processing, like syntactic trees (Yang & Deng, 2019; Purgał et al., 2021), and in the form of a type of neural networks called "Transformers" (Vaswani et al., 2017). Transformers — or more precisely attention-based neural network architectures (Lindsay, 2020) — are for example used for translating and composing syntax; a well-known non-mathematical application of transformers is GPT-3 (Brown et al., 2020). But they also have shown effective in mathematics; Polu & Sutskever (2020) used GPT-f to discover a short proof, which was added to the Metamath library.

So what can we expect of future developments in ITPs? In addition to suggesting tactics, a new generation of theorem provers might even discover theorems, for example by employing "self-play". "Self-play" — a notion already mentioned at the AITP Conference 2020 — refers to an AI finding rules in a world by exploration (OpenAI et al., 2019). Thus, self-play adds a new paradigm to the automation of human cognition: Gopnik (2020) argues that human childhood is characterised by the "exploration" stage, while adults use "exploitation", which is goal-directed. Observing how children learn as a means to obtain general artificial intelligence was already suggested by Alan Turing. And whereas we have suspected for decades that embodiment plays a role in human mathematical thought (Lakoff & Núñez, 2000), by combining our knowledge of interactive computers proving theorems with new discoveries about computers searching out theorems in an exploratory way, could provide bridges to 'mainstream' mathematical practice.

1.4 Waiting for the "ATP/ITP revolution"?

Considering the potential of current ITPs, one could expect a large interest from the mathematical community. This has, however, not really been the case. When asking mathematicians why they do not use theorem provers, Bundy (2011) was frequently confronted with the following reasons: 1) Logic proofs are too detailed and long, 2) Provers are insufficiently powerful, 3) Provers are too tedious to use, 4) Provers are hard to use, and 5) Why give up the fun of proving? In order to contribute to the development of more human-friendly ITPs, we will address the most of these objections in our research program.

1.4.1 OBJECTION: TOO DETAILED, TEDIOUS AND HARD TO USE

Due in large part to probabilistic data analysis and computer vision applications, the field of AI has grown spectacularly over the past decades. Yet, mathematical problem solving by AIs has mainly been non-visual and rigid and has thus not reaped huge benefits from this renaissance. However, there are strong indicators that internal (visual thinking, mental simulation) and external visual presentations (diagrams) play important roles for human mathematicians (Anderson et al., 2015; Johansen, 2014). For instance, in an interview at the Heidelberg Laureate Forum, Fields Medal winner Terence Tao described vividly how he solved a problem by mentally navigating through space.

Although the role of spatial skills has enjoyed a lot of attention in mathematical education research (Hegarty & Kozhevnikov, 1999; Gilligan-Lee et al., 2021), the study of visual representations in mathematics experts is more limited (Cipora et al., 2016; Butterworth, 2006; Stylianou & Silver, 2004; Ma'ayan et al., 2020; Amalric & Dehaene, 2016; Giaquinto, 2007). This is perhaps correlated with the exclusion of graphical or visual arguments by the predominant disciplinary standards of the 20th century beginning with Hilbert's attempts to stop incorrect reasoning inferred from figures (Davis & Hersh, 1981). As emphasised by Whiteley (2010), however, this absence of visual elements in publications does not reflect how important visualization is in mathematics.

These observations show a discrepancy between human and AI mathematics. Human mathematicians switch between several representations, while theorem provers use a single representation (Purgał et al., 2021), which still need improvement (Kaliszyk & Rabe, 2020). To gauge the depth of this gap, our research program seeks to quantify the use of internal and external visual representations by human mathematicians (e.g. visual thinking and conceptual diagrams) and to qualitatively describe their functions (see Figure 1). This information will enable AIs to assist human mathematicians and make ITPs more intuitive at producing visual representations, which could even be formal. The program "Penrose" (Ye et al., 2020) can for example produce figures from any mathematical texts. An even more formal example in that direction is perhaps Globular (Bar et al., 2016), producing graphical proofs in category theory. This newly acquired information about visual and spatial thinking (e.g. mental imagery, mental simulation, etc) by mathematicians should then be integrated with symbolic views on mathematical reasoning.

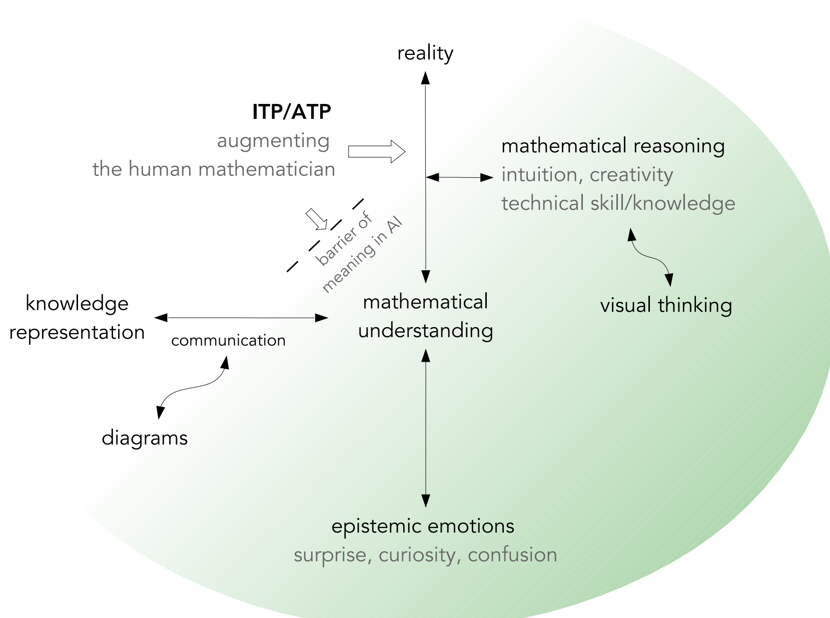

Figure 1: Possible role of ITP/ATP in human mathematical understanding in relation to external representations (in white) such as diagrams, internal representations including visual thinking (in green) and epistemic emotions, addressed by our research program. Also depicted is the relation of AI to mathematical understanding, and the barrier of meaning in AI.

1.4.2 OBJECTION: WHY GIVE UP THE FUN?

The second part of our research program focuses on how mathematicians experience practicing mathematics by investigating the epistemic emotions (Muis et al., 2015; Pekrun et al., 2016) 'curiosity', 'confusion' and 'surprise'. We predict that mathematicians are particularly resilient to confusion (Jaber & Hammer, 2015). Information about these epistemic emotions in mathematical practice might elucidate what it means to "understand" mathematics and how much joy the process actually brings to mathematicians (Satyam, 2020).

When thinking of emotions in mathematics (Borovik, 2017), we realise that understanding is also a subjective experience. As such, understanding seems a major limitation for AI. In cognitive psychology for example, neural networks were originally introduced to model human cognition (Goebel, 1991). These models are so powerful that many researchers in cognitive computational neuroscience now employ deep learning such as image recognition as a 'realistic' way to study human processes (Storrs & Kriegeskorte, 2020). However, the use of deep learning ultimately leads to the question what it means to "understand a neural process via a computational model" (Saxe et al., 2020). A human observer is always needed to assign meaning. An ITP could prove theorems, or perhaps even discover theorems, but in order to understand such AI mathematics, a human mathematician seems necessary in the loop.

This illustrates that the introduction of AI invariably results in existential questions concerning meaning and understanding. And we concur with Mitchell (2020) in concluding that interdisciplinary collaboration is important to overcome the barrier of meaning in AI.

1.5 CONCLUSION

To summarise, our research program into visual representations and epistemic emotions of professional mathematicians will not only contribute to the collaboratively-derived research agenda of challenges in mathematical cognition (Alcock et al., 2016), but also inform the design of ITPs to make them more human-friendly, joining mathematics automation usability efforts like Lean Forward (Doorn et al., 2020) and SErAPIS (Stathopoulos et al., 2020). And, rather than introducing artificial or autonomous mathematical entities, through augmenting the human mathematician like the mythological centaur, this might produce a mathematician — neither computer nor human, but more powerful than each.

REFERENCES

L. Alcock, D. Ansari, S. Batchelor, M.J. Bisson, B. De Smedt, C. Gilmore, S.M. Göbel, M. Hannula-Sormunen, J. Hodgen, M. Inglis, I. Jones, M. Mazzocco, N. McNeil, M. Schneider, V. Simms, and K. Weber. Challenges in Mathematical Cognition; A Collaboratively-Derived Research Agenda. *Journal of Numerical Cognition*, 2016.

M. Amalric and S. Dehaene. Origins of the brain networks for advanced mathematics in expert mathematicians. *PNAS*, 113(18):4909–4917, 2016.

L.E. Andersen, M.W. Johansen, and H.K. Sørensen. Mathematicians Writing for Mathematicians. *Synthese*, 2019. doi: 10.1007/s11229-019-02145-5.

G. Anderson, G. Buck, T. Coates, and A. Corti. Drawing in mathematics; from inverse vision to the liberation of form. *Leonardo*, 48(5):439–448, 2015.

G.E. Andrews. The Death of Proof? Semi-Rigorous Mathematics? You've Got to Be Kidding! *The Mathematical Intelligencer*, 16(4):16–18, 1994.

K. Appel and W. Haken. Every Planar Map is Four Colourable. *Bulletin of the American Mathematical Society*, 82(5):711–712, 1976.

Z. Ashton. Audience role in mathematical proof development. *Synthese*, 2020. doi: 10.1007/s11229-020-02619-x.

K. Bar, A. Kissinger, and J. Vicary. Globular: an online proof assistant for higher-dimensional rewriting. *arXiv*, 2016.

A. Borovik. *Understanding Emotions in Mathematical Thinking and Learning*, chapter Being in Control, pp. 77–96. Academic Press, 2017.

J.M. Borwein. *Exploratory Experimentation: Digitally-Assisted Discovery and Proof*, chapter 4, pp. 69–96. Number 15 in New ICMI Study Series. Springer, 2012.

J.M. Borwein and D. Bailey. *Mathematics by Experiment: Plausible Reasoning in the 21st Century*. A K Peters, 2004.

T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D.M. Ziegler, J. Wu, Winter. C., C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners. *arXiv*, 2020.

N.G. de Bruijn. The mathematical language AUTOMATH, its usage, and some of its extensions. In *Symposium on automatic demonstration*. Springer Berlin Heidelberg, 1970.

A. Bundy. Automated theorem provers: a practical tool for the working mathematician? *Annals of Mathematics and Artificial Intelligence*, 61(1):3–14, 2011.

B. Butterworth. *The Cambridge Handbook of Expertise and Expert Performance*, pp. 553–568. Cambridge University Press, 2006.

K. Buzzard. Computers and mathematics. *Newsletter of the LMS*, (484):32–36, 2019.

K. Buzzard. Proving Theorems with Computers. *Notices of the American Mathematical Society*, 67 (11):1, 2020. doi: 10.1090/noti2177.

K. Cipora, M. Hohol, H.C. Nuerk, and K. Willmes. Professional mathematicians differ from controls in their spatial-numerical associations. *Psychological Research*, 80(4):710–726, 2016.

P.J. Davis and R. Hersh. *The Mathematical Experience*. Penguin Books, 1981.

F. van Doorn, G. Ebner, and R.Y. Lewis. *CICM 2020: Intelligent Computer Mathematics*, volume 12236 of *Lecture Notes in Computer Science*, chapter Maintaining a Library of Formal Mathematics, pp. 251–267. Springer, 2020.

S.B. Ekhad and D. Zeilberger. The Number of Solutions of $X^2 = 0$ in Triangular Matrices over $GF(q)$. *Electronic Journal of Combinatorics*, 3(R2):1–2, 1996.

Marcus Giaquinto. *Visual Thinking in Mathematics*. Oxford University Press, Oxford, 2007. ISBN 0199285942.

K.A. Gilligan-Lee, A. Hodgkiss, M.S.C. Thomas, P.K. Patel, and E.K. Farran. Aged-based differences in spatial language skills from 6 to 10 years: Relations with spatial and mathematics skills. *Learning and Instruction*, 73, 2021.

R. Goebel. *Connectionist Models*, chapter Binding, Episodic Short-Term Memory, and Selective Attention, Or Why are PDP Models Poor at Symbol Manipulation?, pp. 253–264. Morgan Kaufmann, 1991.

A. Gopnik. Childhood as a solution to explore-exploit tensions. *Philosophical Transactions of the Royal Society (B)*, 2020.

M. Hegarty and M. Kozhevnikov. Types of Visual-Spatial Representations and Mathematical Problem Solving. *Journal of Educational Psychology*, 91(4):684–689, 1999.

L.Z. Jaber and D. Hammer. Learning to Feel Like a Scientist. *Science Education*, 100(2):189–220, 2015.

M.W. Johansen. *Model-Based Reasoning in Science and Technology*, chapter What's in a Diagram? On the Classification of Symbols, Figures and Diagrams. Springer-Verlag Berlin Heidelberg, 2014.

M.W. Johansen and M. Misfeldt. *An empirical approach to the mathematical values of problem choice and argumentation*, pp. 259–269. Trends in the history of science. Birkhäuser, 2014.

C. Kaliszyk and F. Rabe. *CICM 2020: Intelligent Computer Mathematics*, volume 12236 of *Lecture Notes in Computer Science*, chapter A Survey of Languages for Formalizing Mathematics, pp. 138–156. Springer, 2020.

A. Koutsoukou-Argyraki. Formalising mathematics — in praxis. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 123(1):3–26, 2020. doi: 10.1365/s13291-020-00221-1.

G. Lakoff and R.E. Núñez. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics Into Being*. Basic Books, New York, 2000.

G.W. Lindsay. Attention in Psychology, Neuroscience, and Machine Learning. *Frontiers in Computational Neuroscience*, 2020.

B. Löwe and B.V. Kerkhove. *Methodological Triangulation in Empirical Philosophy (of Mathematics)*, chapter 2, pp. 15–37. Advances in Experimental Philosophy. Bloomsbury, 2018.

D. Ma'ayan, W. Ni, K. Ye, C. Kulkarni, and J. Sunshine. How Domain Experts Create Conceptual Diagrams and Implications for Tool Design. In *CHI 2020, April 25–30, 2020, Honolulu, HI, USA*, 2020.

D. MacKenzie. The Automation of Proof: A Historical and Sociological Exploration. *IEEE Annals of the History of Computing*, 17(3):7–29, 1995.

W. McCune. Solution of the Robbins Problem. *Journal of Automated Reasoning*, 19:263–276, 1997.

M. Misfeldt and M.W. Johansen. Research mathematicians' practices in selecting mathematical problems. *Educational Studies in Mathematics*, 89(3):357–373, 2015.

M. Mitchell. On Crashing the Barrier of Meaning in AI. *AI Magazine*, 41(2):86–92, 2020.

U. Montaño. Ugly Mathematics: Why Do Mathematicians Dislike Computer-Assisted Proofs? *The mathematical intelligencer*, 34(4):21–28, 2012. doi: 10.1007/s00283-012-9325-9.

U. Montaño. *Explaining Beauty in Mathematics*. Number 370 in Synthese Library: Studies in Epistemology, Logic, Methodology, and Philosophy of Science. Springer, 2014. doi: 10.1007/978-3-319-03452-2.

K.R. Muis, C. Psaradellis, SP Lajoie, I. Di Leo, and M. Chevrier. The role of epistemic emotions in mathematics problem solving. *Contemporary Educational Psychology*, 42:172–185, 2015.

OpenAI, C. Berner, G. Brockman, B. Chan, V. Cheung, P. Debiak, C. Dennison, D. Farhi, Q. Fischer, S. Hashme, C. Hesse, R. Józefowicz, S. Gray, C. Olsson, J. Pachocki, M. Petrov, H.P. de Oliveira Pinto, J. Raiman, T. Salimans, J. Schlatter, J. Schneider, S. Sidor, I. Sutskever, J. Tang, F. Wolski, and S. Zhang. Dota 2 with large scale deep reinforcement learning. 2019. URL https://arxiv.org/abs/1912.06680.

S. Ornes. How close are computers to automating mathematical reasoning? *Quanta Magazine*, 2020. URL https://www.quantamagazine.org/how-close-are-computers-to-automating-mathematical-reasoning-20200827/.

R. Pekrun, E. Vogl, K.R. Muis, and G.M. Sinatra. Measuring emotions during epistemic activities: the epistemically-related emotion scales. *Cognition and Emotion*, 31(6):1268–1276, 2016.

M. Petkovšek, H.S. Wilf, and D. Zeilberger. $A = B$. 1997.

S. Polu and I. Sutskever. Generative language modeling for automated theorem proving. *arXiv*, 2020.

S. Purgał, J. Parsert, and C. Kaliszyk. A study of continuous vector representations for theorem proving. *Journal of Logic and Computation*, 2021.

G. Raayoni, S. Gottlieb, Y. Manor, G. Pisha, Y. Harris, U. Mendlovic, D. Haviv, Y. Hadad, and I. Kaminer. Generating conjectures on fundamental constants with the Ramanujan Machine. *Nature*, 590(7844):67–73, 2021. doi: 10.1038/s41586-021-03229-4.

V.R. Satyam. Satisfying moments during the transition-to-proof: Characteristics of moments of significant positive emotion. *The Journal of Mathematical Behavior*, 59, 2020.

A. Saxe, S. Nelli, and C. Summerfield. If deep learning is the answer, then what is the question? *Nat Rev Neurosci*, 22(1):55–67, 2020.

H.A. Simon and A. Newell. Heuristic Problem Solving: The Next Advance in Operations Research. *Operations Research*, 6(1):1–10, 1958.

H.K. Sørensen. *Exploratory experimentation in experimental mathematics*, pp. 341–360. Number 11 in Texts in Philosophy. College Publications, 2008. URL `http://www.lib.uni-bonn.de/PhiMSAMP/Book/`.

H.K. Sørensen. *'The End of Proof'? The integration of different mathematical cultures as experimental mathematics comes of age*, pp. 139–160. Trends in the history of science. Birkhäuser, 2012. doi: 10.1007/978-3-319-28582-5_9.

Y. Stathopoulos, A. Koutsoukou-Argyraki, and L.C. Paulson. SErAPIS : A Concept-Oriented Search Engine for the Isabelle Libraries Based on Natural Language. In *IJCAR*, 2020.

K.R. Storrs and N. Kriegeskorte. *The Cognitive Neurosciences*, chapter Deep learning for cognitive neuroscience. Boston: MIT Press, 6th edition edition, 2020.

D.A. Stylianou and E.A. Silver. The role of visual representations in advanced mathematical problem solving: An examination of expert-novice similarities and differences. *Mathematical Thinking and Learning*, 6(4):353–387, 2004.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.

W. Whiteley. Visualization in mathematics: Claims and questions towards a research program. In *The 10th International Congress on Mathematical Education*, 2010.

K. Yang and J. Deng. Learning to prove theorems via interacting with proof assistants. In *International Conference on Machine Learning*, 2019.

K. Ye, W. Ni, M. Krieger, D. Ma'ayan, J. Wise, J. Aldrich, J. Sunshine, and K. Crane. Penrose: from mathematical notation to beautiful diagrams. In *SIGGRAPH 2020*, 2020.

D. Zeilberger. Theorems for a Price: Tomorrow's Semi-Rigorous Mathematical Culture. *The mathematical intelligencer*, 16(4):11–14,76, 1994.