ORIGINAL PAPER

Proof, Semiotics, and the Computer: On the Relevance and Limitation of Thought Experiment in Mathematics

Johannes Lenhard[1]

Abstract

This contribution defends two claims. The first is about why thought experiments are so relevant and powerful in mathematics. Heuristics and proof are not strictly and, therefore, the relevance of thought experiments is not contained to heuristics. The main argument is based on a semiotic analysis of how mathematics works with signs. Seen in this way, formal symbols do not eliminate thought experiments (replacing them by something rigorous), but rather provide a new stage for them. The formal world resembles the empirical world in that it calls for exploration and offers surprises. This presents a major reason why thought experiments occur both in empirical sciences and in mathematics. The second claim is about a looming aporia that signals the limitation of thought experiments. This aporia arises when mathematical arguments cease to be fully accessible, thus violating a precondition for experimenting in thought. The contribution focuses on the work of Vladimir Voevodsky (1966–2017, Fields medalist in 2002) who argued that even very pure branches of mathematics cannot avoid inaccessibility of proof. Furthermore, he suggested that computer verification is a feasible path forward, but only if proof is not modeled in terms of formal logic.

Keywords Accessibility · Computer verification · Exploration · Semiotics · Voevodsky

1 Introduction

Thought experiments (TEs) are performed in the sciences as well as in mathematics. In the sciences, they present an intriguing puzzle: how can something that happens merely in thought provide knowledge about the real world? This puzzle touches upon core aspects of epistemology, especially the question how operations in

✉ Johannes Lenhard
johannes.lenhard@mv.uni-kl.de

1   Thermodynamics Engineering Laboratory, TU Kaiserslautern, Schrödinger-Str. 44, 67466 Kaiserslautern, Germany

thought relate to the world.[1] Clearly, TEs can indicate in which ways such operations are limited because of the concepts and rules that are in play. Analyzing the limitations of theoretical concepts can happen (partly) independently from questions of whether and how this analysis speaks about the world. In mathematics, the situation is special because operating with signs plays such a central role. From the semiotic perspective this paper entertains, TEs are frequently performed in mathematics.[2] However, their status is controversial. A wide-spread point of view restricts TEs to a heuristic role. When a proof comes under investigation, this view maintains, TEs must give way to rigorous, formal methods.[3]

This contribution defends two claims; the first is about why TEs are so relevant and powerful in mathematics, the second is about why TEs are limited. Part one (Sect. 2, 3 and 4), argues that TEs are much more relevant in mathematics than the common view suggests. Heuristics and proof are not as strictly separable as the standard view presumes. Thus, the relevance of TEs is not contained to heuristics. The main argument is based on a semiotic analysis of how mathematics works with signs. Seen in this way, formal symbols do not eliminate TEs (replacing them by something rigorous), but rather provide a new stage for them. The formal world resembles the empirical world in that it calls for exploration and offers surprises. This presents a major reason why TEs occur both in empirical sciences and in mathematics.

Section 2 introduces the common point of view according to which TEs are tools for creating insight, but in a somewhat heuristic manner. The context of justification requires to shift gears, i.e., to check with formal means whether some TE holds its promise. The exuberant productivity in early analysis fits well into this picture. Cauchy was not only an ingenious generator of theorems, but some of his claims were erroneous, like that the limit of continuous functions is again a continuous function.[4] This period was followed by a movement of rigor in which Weierstrass and other proponents wanted to provide guardrails to thinking, introducing the (epsilon, delta) definitions that allowed a more arithmetical way of operation with such concepts as continuity.

Such episodes are generic motivations for the common point of view on TEs in mathematics. This paper focuses on one expression of this view, namely the essay by Jaffe and Quinn (1993, then editors of the AMS Bulletin) that warned mathematics were in danger of incorporating speculative theorems without rigorous proofs. Their warning can be read as a plea to limit the use of TEs. The Jaffe/Quinn essay created high waves in the mathematical community. Section 3 discusses the reply by William Thurston (another Fields medalist). He was one of those mathematicians

---

[1] The recent Routledge Companion (Stuart et al. 2018) and entry to the Stanford Encyclopedia (Brown and Fehige 2022) provide a well-informed overview of the recent discussions on TEs.

[2] The semiotic perspective is not without alternative. From an empirically oriented perspective, one can question whether TEs happen in mathematics at all. See Buzzoni (2021) for a discussion on whether speaking of TEs in mathematics is terminologically adequate.

[3] Cf. van Bendegem (2003) for an example.

[4] Cauchy also introduced the topological proof of Euler's theorem on polygons by an ingenious TE that Lakatos (1976) made the starting point of his "Proofs and Refutations".

critically mentioned by Jaffe/Quinn. Although Thurston does not explicitly mention TEs, the substance of his argument effectively is about the use of TEs. His point is that mathematics is about understanding, i.e., about making something accessible to thinking—preparing the terrain for TEs—, not about exercising formal rules.

Section 4 sums up the discussion about why TEs achieve a particular relevance for mathematics. For semiotic reasons, it is very hard to eliminate TEs. Formalization of one context prepares the stage for TEs on a new level. The emerging new levels give mathematical epistemology a semiotic nature: The formal world can achieve a life of its own (paraphrasing Hacking's (1983) verdict about experiments). Entities might arise from construction, but this does not render them arbitrary nor does fully determine them. Thus, a major component of mathematization is exploration. This grants the relevance for TEs: they enable exploration.

Part two (Sect. 5) addresses the second main claim of this paper. A looming aporia signals the limitation of TEs. This aporia—there is no path forward—arises when mathematical arguments cease to be fully accessible. On the one side, TEs cannot deal with what is not fully accessible. Whereas laboratory experiment can fail because the experimenter had not thought of an important factor, the situation in thought experiment is different: Factors not thought of simply do not exist in a TE (even if their implications may be surprising). On the other side, the reliability of proof rests on being accessible for critical inspection. This condition is strongly dependent on the capabilities of human minds. The question then arises to what extent mathematical theorems and proofs do respect these capabilities. Instead of arguing about principles, the paper focuses on the work of Vladimir Voevodsky (1966–2017, Fields medalist in 2002) and what is called here the Voevodsky problem, namely finding ways for coping with inaccessibility that some branches (even of very pure) mathematics cannot avoid. In a recent attempt, Vladimir Voevodsky tried to incorporate the computer into mathematical practice. However, he was fully aware that older approaches to computer-based formal verification do not promise a way forward. Instead, he was going to devise new foundations for mathematics in a way that allows computer verification of proofs, liberating them from the limiting human capabilities.

2 Shifting Gears—Away from Thought Experiment

The common picture of TEs in mathematics appreciates their role in finding convincing and elegant proofs. However, when a TE is not convincing enough to some mathematicians, then the validity of the proof is in question. In this situation, so the common view, one has to shift gears into a more rigorous and formal mode of proving. Introduced in this way, the notion of rigorous proof is an umbrella term under which different variants of what counts as rigorous exist.[5] Nevertheless, TEs do not fall into this class, because a TE must be convincing by itself. Any attempt to

---

[5] Cf. the insightful discussion by Brown (2022) in this volume where he examines the qualities of TEs against a perceived view that ties rigor to formal logic.

disassemble the TE and to find a more formal and controlled way of articulating the proof would automatically destroy the character as TE.

This common view is well entrenched in the history of mathematics and also in the history of thinking about mathematics. One typical instance is the essay by Jaffe and Quinn (1993), then the editors of the Bulletin of the American Mathematical Society. They identify a dangerous tendency in mathematics, or in a part of mathematics that they call "theoretical mathematics". This name shall indicate the similarity to theoretical physics where mathematical methods play a crucial role whereas not all of the assumed theories are rigorously justified. Or, more precisely: they sometimes lack full mathematical justification—but physics is not in danger because there is a test against empirical data that will weed out invalid hypotheses. Jaffe and Quinn argue that physics therefore can afford a certain lack of mathematical rigor. Simultaneously, they warn that mathematics cannot afford this kind of lacking reliability because there is no other line of justification than by proof. For them, something like theoretical mathematics, a cultural synthesis of mathematics and theoretical physics (so Jaffe and Quinn) is a dangerous concept because there is no check against invalid hypotheses.

When Jaffe and Quinn issue their warning that "today in certain areas there is again a trend toward basing mathematics on intuitive reasoning without proof" (p. 1), they without doubt include TEs on the side of intuitive reasoning.

Jaffe and Quinn parallel the function of proof in mathematics to that of experiment in physics. "Theoretical work requires correction, refinement, and validation through experiment or proof." (p. 2) This is exactly what TEs do not allow. Surely, TEs can be refined, but they cannot be tested against something else. Only such test can grant the reliability of knowledge. Accordingly, the reliability of mathematical knowledge hinges on rigorous proof. What makes proof rigorous is the perceived universal accessibility of formal mathematical argument.

> "Mathematicians may have even better experimental access to mathematical reality than the laboratory sciences have to physical reality. This is the point of modeling: a physical phenomenon is approximated by a mathematical model; then the model is studied precisely because it is more accessible." (p. 2).

Doesn't this coincide with TE? What can be more accessible than experiments that run in thought? This is an important point. A mathematical model is not only some abstract or ideal entity. Crucially, this entity can be characterized and analyzed in formal language. In this way, the model becomes accessible to a community of researchers. This marks a seminal step: Individual minds conduct a thought experiment—and hopefully coincide in their findings what this experiment results in. A formal rigorous proof shall put this hope for coincidence on a more secure footing. A formally defined object together with a formally defined dynamic allows examination by many researchers. In short, they have access in a different mode.

The crucial property that formal proofs have and TEs lack is their scalability. One can disassemble a contested proof into smaller steps so that everybody agrees with every step and then re-assemble the entire proof. Thus, it turns out that two factors are in play, accessibility and scalability. TEs can provide wonderful proofs—as long as they work. Transferring an argument into a different semiotic context can make

it accessible in different ways. Rigor then means that nobody has doubts. In case of doubt, one can make the proof more rigorous by splitting up the steps into sub-steps. Now it becomes evident that formal proofs are never completely formal but rather are formal on a certain level. In mathematics, proofs are never carried to an extremely formal level. One simple reason is that when some proof appears valid on some level, going further is pointless. However, to the extent that formal steps are incompletely specified, a proof contains a dose of TE. This is true even of proofs that count as rigorous.

3 Thought Experiment as Paradigm of Proof—Thurston's Reply

William Thurston (1926–2012, Fields medal in 1982), mainly known for his work in low-dimensional topology, was one of those mathematicians who were explicitly named by Jaffe and Quinn for doing work more involving speculation than rigorous proof. What Jaffe and Quinn criticized was that it remained unclear to many mathematicians, which of the published —and interesting—theorems should count as proven and which should count as only plausibly suggested. Thurston was famous for relying on intuitive, often geometrical reasoning. He was an exemplar of preferring TE to formal derivation. Thus, for Jaffe and Quinn, Thurston's work exemplified the situation discussed above: some mathematicians are not fully convinced by certain TEs that other mathematicians (including Thurston) count as proof.

Thurston replied to Jaffe and Quinn with an essay (Thurston 1994) that locates proof in a wider rationale for doing mathematics. Thurston holds that mathematical practice is not about enlarging the reservoir of proven theorems but rather advancing human understanding of mathematics. This is not unrelated to proof, but significantly different. According to Thurston (1994, p. 161), the leading question is not "How do mathematicians prove theorems?", but "How do mathematicians advance human understanding of mathematics?" (ibid., p. 162). Consequently, proofs are outside the leading motivation when they do not contribute to human understanding. However, some proofs do contribute to understanding, and TEs are among them. When mathematicians are able to conduct TEs, this proves that they have sufficient understanding—else they would not be able to execute the TE.

Thurston gives an instructive example, the proof of the Four-Color-Theorem (4CT) by Appel and Haken (1977), a famously contested early computer-based proof.[6] The 4CT itself is strikingly simple. Consider a map with arbitrarily many countries and arbitrarily complicated borders between them. The task is to color all countries in a way that no neighboring countries have the same color. How many colors are needed? The 4CT states that four colors are sufficient. The proof by Appel and Haken rests on the systematic analysis of very many cases—too many to be checked by a human mathematician. Apple and Haken deploy a computer program that verifies each case. Of course, they had to argue that their program captured every case. Thurston points out that the 4CT computer proof was controversial not

---

[6] Tymocko (1979) discusses details of the proof in a very accessible way.

because its validity was questioned. On the contrary, the correctness of the proof was quickly accepted. The reason was that the proof, in particular the enormous number of single cases, frustrated the desire for understanding. For him, mathematics primarily is not about reliable knowledge, but about human understanding.

Thurston points out that the computer proof is way more formal than any normal mathematical proof. Otherwise, a computer could not deal with it. But being formal and being reliable is not the crucial point. Understanding requires a sort of communication, a resonance with engaged mathematicians. TE is a communication tool between humans. In a way, this tool exploits human biology, i.e., the machinery of thinking—as far as it is common to humans since TEs have to work in a community. Importantly, how humans achieve understanding is a process whose characteristics are open to change. Thurston (1994, p. 162) brings in a dynamical perspective: "As our thinking becomes more sophisticated, we generate new mathematical concepts and new mathematical structures: the subject matter of mathematics changes to reflect how we think." This observation is apt also in regard of computer use and TE. Working with evolving structures and concepts changes the space for TEs.

Thurston opposes Jaffe and Quinn regarding the significance of formal proof. High reliability is important, but the practice of mathematics teaches that this reliability does not need formal "rigorous" proof.

"Our system is quite good at producing reliable theorems that can be solidly backed up. It's just that the reliability does not primarily come from mathematicians formally checking formal arguments; it comes from mathematicians thinking carefully and critically about mathematical ideas."(170).

Thurston brings out that reliability is created through practice—and practice is a team practice, like soccer. Furthermore, the reliability of the shared practice is far greater than that of a single theory, like about the foundations of mathematics. Reliability comes from how mathematical activity unfolds in practice, not from the force of a particular theory.[7]

4 The Complementarity/Autonomy View Confirmed: Rigor is Rigor in Context

Even if Jaffe/Quinn and Thurston take very different standpoints on the role TEs play in mathematics, they agree on the basic setting. Mathematical knowledge should preserve its well-deserved status as extraordinarily reliable knowledge. The controversy is about how and on what basis this sort of knowledge is actually attained. Jaffe and Quinn hold that the only way is formalization, oriented at symbolic notation and formal rules for manipulating the symbols. Thurston replies that knowledge is actually checked by critical thinking, i.e., by a process that includes thought experimentation. The main point of this paper, or rather of the first part on

---

[7] This reflects Lakatos' perspective that informal theories (trusted in mathematical practice) can falsify formal theories when the latter deal with "the foundations" (Lakatos 1967).

the relevance of TEs, is that from a semiotic perspective the seemingly contradicting standpoints are connected in a complementary way.

How can humans do mathematics at all? This question has accompanied and often provoked philosophy of mathematics from Plato onward.[8] Put simply, mathematicians engage in a semiotic activity that simultaneously takes place in the formal, symbolic realm and in that time and space the mathematicians work in. In a sense, this type of activity is at the center of modern epistemology. Foucault, for instance, identifies a major turn from the age of similarity to the age of representation where things and words have lost—or are liberated from—their internal relationship. At the beginning of the seventeenth century:

> "writing has ceased to be the prose of the world, resemblances and signs have dissolved their former alliance, similitudes have become deceptive. … Thought ceases to move in the element of resemblance. Similitude is no longer the form of knowledge, but rather the occasion of error. … 'It is a frequent habit', says Descartes, in the first lines of his *Regulae*, 'when we discover several resemblances between things, to attribute to both equally, even on points in which they are really different, that which we have recognized to be true of only one of them'. The age of resemblance is drawing to a close. … And just as interpretation in the sixteenth century … was essentially a knowledge based upon similitude, so the ordering of things by means of signs constitutes all empirical forms of knowledge as knowledge based upon identity and difference" (Foucault 1973, pp. 47–51 and pp. 56–57).

In the modern age, words and things inhabit two different worlds, each with their own and relatively independent dynamic, that human episteme has to mediate. Kant, one of the iconic proponents of this age, saw constructive activity at the root of human knowledge. The current paper does not aim at a deeper discussion of these issues. The point is that mathematical epistemology is not special regarding the mediating activity between complementary realms.

However, mathematical activity unfolds in a special context because it is concerned with the realm of symbols itself. Of course, what mathematicians do and how they do it is related to the real world and to intended applications in a variety of ways. Many philosophers have expressed that knowledge is facing an unruly world. Laws, for instance, come with conditions and exceptions. Consequently, empirical checks of theoretical claims can be very powerful (and surprising). All this is relevant for applied mathematics. However, mathematics seems to be special because much of its domain is constituted by semiotic activity. Mathematical theorems about groups depend on the very definition of groups. Or, to present another well-known example, the angle-sum of a triangle in Euclidean space can be seen from a manipulation of diagrams.[9] One of the most important facts about mathematics is how dynamically the formal realm develops.

---

[8] The problem goes back beyond Plato, at least to Pythagoras and Heraklit. For a sample of recent and significantly different formulations see Hersh (1997), or Hacking (2014).

[9] C. S. Peirce has conceived mathematics as diagrammatic reasoning.

Mathematization is a means to explore the unknown. Lakatos (1967), for instance, has analyzed mathematics as quasi-empirical theories that begin while "subjects are still indeterminate". That means mathematization is a way to explore the unknown. MacLane has expressed a similar point when explaining Atiyah's approach: "For Atiyah (1994, p. 191), it meant thinking hard about a somewhat vague and uncertain situation, trying to guess what might be found out, and only then finally reaching definitions and the definitive theorems and proofs." Recently, Lenhard and Otte (2018) addressed the topic head on and described mathematization as exploration.

Exploration is momentous (in a philosophical sense) because the semiotic activity does not merely uncover what has been there all the time but constitutes the not yet determinate. What is really special in mathematics is the richness of the formal side. There are layers of abstractions, inviting to make full use of semiotic powers. Calculating with numbers is a very old and reliable activity, but mathematics has not only formulated various foundations for numbers, it has enlarged the types of numbers in highly imaginary ways, and also has added more layers of abstraction on top of the numbers. Overall, mathematical activity is a deeply semiotic activity that offers—and is going on to offer—new objects and relationships to human minds. In this way, new material is generated for potential TEs. The relevance of TEs lies in the flexible ways in which they can support mathematicians in exploring and understanding the complex terrain of formal entities and their relationships.[10]

5 The Voevodsky Problem

Jaffe and Quinn started with an observation that was common to all participants in the debate, namely the reliability of mathematical knowledge. According to them, this reliability is in danger. Full reliability is especially vulnerable because it does not allow degrees, no watering down of conditions. "Only a small pollution of serious errors would force mathematicians to invest a great deal more time and energy in checking published material than they do now." (Jaffe and Quinn 1993, p. 9) Consequently, even small pollutions are practically intolerable. As a measure that prevents the danger, Jaffe and Quinn propose an adjustment of community norms that they call "truth in advertising". In their publications, researchers should openly declare their level of uncertainty, the degree of open problems, and where opportunities to gain credit lie for other researchers. Thurston does not argue against this measure. He, among others, just points out that the problem (of lacking advertisement) does not arise from the non-formality of proofs nor can it be eliminated via formalization.

What is here called Voevodsky's (Vv) problem is the situation where the Jaffe/Quinn solution does not work. When the pollution by errors cannot be cleaned through more time and energy because there is a lack of accessibility, a prohibitive complexity. Any critical thinking, any declaration of uncertainty does depend on accessibility. Under such condition reliability is in danger but neither the Jaffe and

---

[10] In this respect, the use of TEs in sciences like physics has much in common with the use in mathematics. See Buzzoni (2022, this issue) for a discussion of how the uses are different.

Quinn nor the Thurston approach are of much help. The following passages describe the Vv problem in some more detail. Main sources are Voevodsky's own accounts. He is often refreshingly self-critical, but of course, this cannot replace validation by historical methods. Nevertheless, even if the question of historical accuracy remains open, the Vv case works well as an illustration.

The Vv problem is associated with a limitation of thought experimentation. Almost trivially, a TE has to proceed in thinking.[11] Even if mathematical processes happen in a formal realm, this does not shield them from becoming too complicated. Mathematics has found amazing ways to restore accessibility when things become too complicated. A striking example is the young Évariste Galois who complained that nobody can keep pace with Euler's long-winded calculations—just to come up with an algebraic abstraction (Galois theory) that shifted the action to a new and open level. Another, very common, example is the use of the letter "x" for some unknown and potentially complicated entity. Calculating with x often simplifies things greatly. In short, semiotic inventiveness and accessibility are friends. Accessibility (to thinking) and rigor entertain a dynamic relationship in which both parts can change.

The Vv problem troubles this picture. It is concerned with the following gap. What if there is no way (or no known way) to "shifting gears" and making things accessible? Voevodsky describes how his own career as a mathematician was a sort of uphill battle against inaccessibility and eventually led him to realize the importance of this gap. One episode is from the 1980s when he was pioneering "motivic cohomology", a field on category theory. "The field of motivic cohomology was considered at that time to be highly speculative and lacking firm foundation." Vv worked toward establishing this foundation. However, it was not easy to decide whether a given proof is correct. In his paper, Vv established the "Block lemma" in a one-paragraph proof that looked like a sufficiently complete TE, but was not. Since Vv himself had no doubt, he could hardly advertise the actually existing uncertainty. Only years later, he realized that the proof contained a mistake. Worse, the mistake could not be repaired. However, this time things had still gone well insofar Vv was able to develop a new proof of the lemma, published in 1993 and discouraging 30 pages long. This proof needed several years to be accepted in the (specialist) community.

This episode is not the end of the story. In 1999/2000 Vv held a series of lectures on cohomology theory, with Pierre Deligne, another celebrity and Abel prize winner 2013, taking notes and checking the manuscript. However, this could not prevent the problem from popping up again. Vv discovered a mistake in a key lemma of his "Cohomological Theory of Presheaves with Transfers" that neither himself nor the prominent reviewer had detected during the lectures. Worse, Vv found no way to salvage the lemma. At least, the theory did not break down since a weaker—though

---

[11] The workings of a thought experiment have to be sufficiently clear. Lenhard (2018) argues that the epistemically opaque nature of simulation experiments marks an important difference between thought and computer experiments. For a comprehensive discussion of how computer simulation and TEs are related, see also Buzzoni (2016).

more complicated—lemma proved sufficient (Mazza et al. 2006). In short, even under close scrutiny from experts, the mistake went undetected. Seen as a thought experiment, the proof went deceptively smoothly. Voevodsky was shocked when he became aware of how permanent this sort of deception can be.

> "This story got me scared. Starting from 1993, multiple groups of mathematicians studied my paper at seminars and used it in their work and none of them noticed the mistake. And it clearly was not an accident. A technical argument by a trusted author, which is hard to check and looks similar to arguments known to be correct, is hardly ever checked in detail." (Voevodsky 2014).

Vv makes a good point here. The social organization of expertise plays a role in how and when proofs are accepted. And these experts trust in TEs as long as they look sufficiently similar to already known ones. Thus, TEs are not general to humans or mathematicians, but rather conditioned on a small group of established experts. When Abel prize winners and Fields medalists struggle, accessibility is really in danger.

The storyline of the Vv problem has another branch. In 1998, Carlos Simpson, another prominent mathematician and a CNRS research director in France, published a paper that discussed an example, or rather the existence of an example, which would be a counterexample to a theorem in Vv's work with Kapranov (1990). "It claimed to provide an argument that implied that the main result of the "∞-groupoids" paper, which Kapranov and I had published (...), cannot be true." (Voevodsky 2014) Kapranov and Vv checked their theorem carefully and convinced themselves that it is correct. Thus, a sort of a draw occurred: Two contradicting papers but, despite serious efforts, no mistake found in either one. This perceived draw persisted one and a half decades until eventually, in 2013, Vv found out that he and Kapranov were wrong. The situation was complicated by the fact that none of the authors had been able to put his finger on a potential mistake. The proofs in both papers seemed to be correct. Again, thought experiments tend to run well as long as the experts say they do so. Since the reputation of Kapranov and Voevodsky is high, nobody had challenged the (in fact wrong) result. The accessibility of a proof indeed hangs on a thread.

Vv diagnoses that the repeated problems and undetected mistakes indicate a general problem. This kind of situation is typical whenever mathematical theories and arguments are of greater complexity (higher-dimensionall mathematics he calls it). A commuting diagram with four arrows looks like a manageable object to exercise critical thinking. A commuting diagram with 40 arrows becomes meaningless in the sense that mathematicians cannot handle it. And this is what Vv says he typically would have to deal with. In these circumstances, TEs cease to be meaningful and mathematicians get lost in argumentation.

At this point, the problem lends itself to the Jaffe and Quinn approach, but now with the computer as pivotal auxiliary support. Automated verification is possible if the argument is sufficiently formal. Mathematical proof is an exemplar of computer verification. Computers pose an accessibility condition that is very different from the condition for (human) mathematicians. Consequently, mathematicians would

have to bite the bullet and formalize their favorite mix of thought experiments and symbolic manipulations to such a degree that the computer can overtake.

However, in Vv's case automated verification is a non-starter. The point is not on the computer but on the human side. The notion of formalization encompasses formal symbols in a semiotically very generic way. The examples of algebra and its further family members like matrix algebra and commuting diagrams in category theory show the breadth of the concept. When it comes to computer verification of proofs, formalization plays a crucial role because computers require a highly formalized input. However, in this context formal means formal in a logic calculus. This is a much narrower notion of being formal. Actually, squeezing a mathematical argument into a string of formal logic can be a bottleneck. Much of the dynamics of mathematics rests on avoiding this bottleneck and working with entirely different symbolic systems.

So why is computer verification a non-starter regarding the Vv problem? Vv argues that cohomology theory is so alien to formal logic (like ZFC) that a translation into this language would require like 1000 pages of formulas. This makes it a non-starter. The computer would surely be able to formally check these 1000 pages and even to do heuristic numerical testing. However, the machine cannot help in deciding whether these 1000 pages contain exactly what is at stake, i.e., express that problem that should be checked on the side of cohomology theory. Vv describes a dramatic change in his own work.

> "But to do the work at the level of rigor and precision I felt was necessary would take an enormous amount of effort and would produce a text that would be very hard to read. And who would ensure that I did not forget something and did not make a mistake, if even the mistakes in much more simple arguments take years to uncover? I think it was at this moment that I largely stopped doing what is called "curiosity-driven research" and started to think seriously about the future. I didn't have the tools to explore the areas where curiosity was leading me and the areas that I considered to be of value and of interest and of beauty." (Voevodsky 2014).

In other words, Vv perceived mathematics was approaching an impasse. Thought experiments fell short of the rigor and precision needed in complex situations. But computers, whose calculating capabilities make them promising tools, require a forbidding formal bottleneck. The pressing question is whether there is a way out? Is the computer good for escaping the limitation of TEs? But how can the social fabric be maintained when the process of verification has to go through an indigestible formal text?[12]

---

[12] A proof in formal logic is not necessarily matching the needs. A telling example for the insight is Thomas Hales' computer-based proof of the Kepler conjecture. It rests on a sort of labyrinthian formalization where the code is very large and computer runtime is astronomical, making it look like a tragicomical replay of the 4-color-theorem and its computer-based proof.

The remainder of this paper describes Vv's ambitious new program (see also reference "Homotopy..."). He agreed with Thurston that the social fabric is primary, not the (formal) text.

> "Mathematical knowledge and understanding were embedded in the minds and in the social fabric of the community of people thinking about a particular topic. This knowledge was supported by written documents, but the written documents were not really primary." (Thurston 1994, p. 169).

In other words, TE are the better and more important part. Additionally, Vv saw the danger that TEs fail and that social practice lands in an aporia. According to him, this danger could be avoided with the help of computers. Even for Thurston, the use of computers for verification is the way mathematics will have to go:

> "In not too many years, I expect that we will have interactive computer programs that can help people compile significant chunks of formally complete and correct mathematics (based on a few perhaps shaky but at least explicit assumptions), and that they will become part of the standard mathematician's working environment.
> However, we should recognize that the humanly understandable and humanly checkable proofs that we actually do are what is most important to us, and that they are quite different from formal proofs. For the present, formal proofs are out of reach and mostly irrelevant: we have good human processes for checking mathematical validity." (Thurston 1994, p. 171).

It is the last sentence where Vv disagrees. The human processes are not good—this had motivated Vv to abandon his research in cohomology theory and to enter the field of computer-based methods in proof verification. What was needed was a new concept of computer proof verification, including the software that makes computer-verification practical. He worked toward providing both parts: The software (AUT-68, Automath) and the underlying concept of "univalent foundations". In a way, this new approach replaces predicate logic by category theory, while univalent foundation still provides a complete system like ZFC. Vv argues that ZFC is so strong because it is based on hierarchies and humans are very capable of dealing with hierarchies. According to Vv, the major virtue of univalent foundation is that it fits to human mathematicians *and* to the computer. Vv names three conditions that every system shared by computers and humans must fulfill simultaneously. Firstly, it must be a formal deduction system (not necessarily ZFC nor at all built on predicate logic). "The second component is a structure that provides a meaning to the sentences of this language in terms of mental objects intuitively comprehensible to humans." This keeps TEs in play. Furthermore, and thirdly, this structure shall enable humans to encode mathematical ideas in terms of the objects directly associated with the language. This is the practical-technical side, crucially depending on a software.

With univalent foundations, Vv wanted to satisfy all three conditions. Surely, it has to prove its quality in the practice test. Whether it will be able to maintain the social fabric strongly depends on the uptake by mathematicians, on whether formal

verification and intuitive argumentation can be combined merely in theory (or promise) or rather in practice.

Independent of the future success of univalent foundations as a remedy to the Vv problem, this problem brings out a serious limitation of TEs in mathematics. Thinking and accessibility to intuition must be in tune. And being in tune refers to a community, not a single person. A person can find a proof, but cannot constitute a practice. Whether a TE works or not does not depend on the one who proposes the TE. Thus, TEs are communal entities. Individual persons have to learn how to do them, honing their intuition and being initiated on what counts as fully clear. Whether univalent foundations will become a common link to computer assistance—who knows? Anyway, the limitation of TEs seems inevitable. Finding ways for granting the reliability of mathematical knowledge likely will require the computer as a tool and, importantly, require to explore how the computer might allow to shift the boundaries of accessibility. Finally, the semiotic insight stands: mathematics is operating with signs—the question is in what sense and to what extent the operator might be a computer.

References

Appel K, Haken W, assisted by Koch J (1977) Every planar map is four colorable. Ill J Math 21(84):429–567

Atiyah M et al (1994) Responses to "theoretical mathematics: toward a cultural synthesis of mathematics and theoretical physics" by A. Jaffe and F. Quinn. J Bull Am Math Soc 30(2):178–207

Brown, James R, Yiftach F (2022) "Thought experiments", The stanford encyclopedia of philosophy, Spring 2022 edn. Zalta EN (ed.), https://plato.stanford.edu/archives/spr2022/entries/thought-experiment/

Brown, James Robert (2022) This issue

Buzzoni, Marco. 2016. Thought experiments and computer simulations. In: Magnani L, Casadio C (eds.) Model-based reasoning in science and technology. logical, epistemological, and cognitive issues. Springer, Heidelberg, New York, pp 57–78

Buzzoni M (2022) This issue

Buzzoni M (2021) Mathematical versus empirical thought experiments. In: Calosi C, Graziani P, Pietrini D, Tarozzi G (eds), Experience, abstraction, and the scientific of the world. Festschrift for Vincenzo Fano on his 60th birthday, FrancoA EM 2021, pp 215–228

Foucault M (1973) The order of things. Vintage Books, NY

Hacking I (1983) Representing and intervening, introductory topics in the philosophy of natural science. Cambridge University Press, Cambridge, UK

Hacking I (2014) Why is there philosophy of mathematics at all? Cambridge University Press

Hersh R (1997) What is mathematics, really? Oxford Univerity Press

Homotopy type theory: univalent foundations of mathematics. The Univalent Foundations Program. Institute for Advanced Study

Jaffe A, Quinn F (1993) "Theoretical mathematics": toward a cultural synthesis of mathematics and theoretical physics. Bull Am Math Soc 29(1):1–13

Lakatos I (1976) Proofs and refutations. Cambridge University Press, Cambridge

Lakatos, Imre. 1967. "A renaissance of empiricism in the recent philosophy of mathematics", in Lakatos (ed). Problems in the philosophy of mathematics, Amsterdam, North-Holland

Lenhard J, Otte M (2018) The applicability of mathematics as a philosophical problem—mathematization as exploration". Found Sci 23(4): 719–737. Online version http://rdcu.be/GS84

Lenhard J (2018) Thought experiments and simulation experiments. Exploring hypothetical worlds, in Stuart M, Fehige Y, Brown JR (eds): The Routledge Companion to Thought Experiments. Chapter 27, pp 484–497

Mazza C, Voevodsky V, Weibel C (2006) "Lecture notes on motivic cohomology", clay mathematics monographs. American Mathematical Society, Providence, RI, vol 2; Clay Mathematics Institute, Cambridge, MA

Stuart MT, Fehige Y, Brown JR (eds) (2018) The Routledge companion to thought experiments

Thurston WP (1994) On proof and progress in mathematics. Bull Am Math Soc 30(2):161–177

Tymocko T (1979) The four color problem and its philosophical significance. J Philos 76(2):57–83

Van Bendegem JP (2003) Thought experiments in mathematics: anything but proof. Philosophica 72:9–33

Voevodsky VA, Kapranov MM (1990) ∞-groupoids as a model for a homotopy category". Uspekhi Matematicheskikh Nauk 45(5):183–184

Voevodsky V (2014) The origins and motivations of univalent foundations. A personal mission to develop computer proof verification to avoid mathematical mistakes. The Institute Letter Summer 2014. Institute for Advanced Study, Princeton.