

THE NATURE OF CONTEMPORARY CORE MATHEMATICS

VERSION 0.92 MAY 2010

FRANK QUINN

ABSTRACT. The goal of this essay is a description of modern mathematical practice, with emphasis on differences between this and practices in the nineteenth century. I explain how and why these differences greatly increased the effectiveness of mathematical methods and enabled sweeping developments in the twentieth century. A particular concern is the significance for mathematics education: elementary education remains modeled on the mathematics of the nineteenth century and before, and use of modern methodologies might give advantages similar to those seen in mathematics.

This draft is about 90% complete, and comments are welcome.

1. INTRODUCTION

This essay offers a detailed qualitative description of contemporary professional practice in core mathematics. “Contemporary”, “core”, etc. are explained below. The goal is to be useful to mathematicians, users of mathematics, educators, and students. Education is a particular concern.

1.1. **About “Contemporary”.** Professional mathematics changed profoundly between about 1850 and 1930 and the concern here is with practice as it stabilized and became relatively uniform after 1950.

Roughly, for most of history mathematics was guided, validated, and explained by philosophy and by relation to the physical world. But these methodologies were shaped by preconceptions rather than adapted to the subject, and by the mid nineteenth century they were no longer reliable enough to support progress in many areas. New methodologies evolved and were crucial for the explosive development in the twentieth century.

This transition is analogous to the seventeenth-century transition in physics from Aristotle to Galileo, Newton et. al., the eighteenth-century transition from alchemy to chemistry, and similar but more gradual transitions in biology and medicine. In §9 I trace out some of the history of the transition and describe how the methodology changed. I also attempt to explain why the transformation is not more widely recognized, and related to this, in §10, why modern descriptions of mathematics are so widely divergent and ineffective.

1.2. **About “Core”.** *Core* refers to material *used within mathematics*. Core does not mean either “pure” or “useless”:

- Quite a lot of core mathematics has external use as an eventual goal, but needs mathematical development before it can actually be used.

- Similarly, quite a lot of what is identified as “applied mathematics” has the precision needed for further mathematical development, and should be considered part of the core.
- Finally, there are bits of indisputably “pure” mathematics that are—for the moment anyway—dead ends and not actually *used*. These do not always satisfy core requirements.

The point is that material that is actually used is subject to concerns and pressures that lead to uniformity of practice. Work done for outside use is not subject to these pressures and often has different goals and standards. The core/non-core distinction is crucial for a real understanding of the discipline.

In §3 I observe that practices in non-core areas are closer to those in other sciences, and suggest the term “mathematical science”.

1.3. About Mathematics. The focus here is on features of the mathematical enterprise that are special to the subject.

The first point is that science is a human activity. Scientists work hard to adapt themselves to the special demands of their subject, but even so most of what they do reflects human psychology and cognitive structure rather than the structure of the field. For example the psychology of individual discovery, the organization of research programs and agendas, and the setting of priorities all seem to me to be mainly human nature, and the rich variety of styles seen in mathematics suggests these are not even much filtered by the subject matter.

The goal, however, is to understand constraints imposed by the field itself, and how people adapt and harness their abilities to work within these constraints.

The next point is that area-specific features in other sciences tend to show up on small scales. The papers in [43], for instance, describe strong micro-adaption to different organisms in experimental biology. Lumping together fruit flies, flatworms, and a few small fish would obscure quite a bit of the subject adaption.

There is certainly micro-adaption in mathematics, and at the other end of the spectrum there is not much commonality in everything that has been called mathematics. The surprise—to me anyway—is that there are powerful commonalities at the broad level of contemporary core mathematics. The basic reason seems to be the validation criterion: criteria in other sciences are largely external (agreement with the physical world) while mathematics uses internal criteria (correct proof), see §2.2.

From this point of view the goal is to understand how people can effectively use internal validity criteria; the limits imposed by this; and how these play out in actual practice.

1.4. About the Essay. This description is not philosophy, as philosophy has defined itself. With few exceptions the philosophy of mathematics seems at best irrelevant, see §10, and it emerges as a bit of a villain in the historical account in §9, see §9.2, and in education, §8.1. This work was also not undertaken as philosophy. I never asked, for instance, ‘what is mathematics?’

In fact the essay is almost accidental: I spent twenty years trying to understand and resolve a variety of problems and—unexpectedly—the pieces fit together to give a global view of mathematics. It was developed into an essay because it provides significant explanatory power from many points of view, not just a nice story.

Many of the experiences that shaped the ideas are described here. These include publication practices, §2.1.1; sociology, §2.1.2; history, §2.1.3 and §9; physics and computing, §2.2.2; education, §2.2.3; and behind it all 40 years of toil at the research frontier, §2.2.2, 10.1, and 10.4. The points are:

- Requiring consistency and explanatory power in different contexts is a way of challenging ideas, and the more challenges the better.
- These are also the challenges that alternatives or modifications should be expected to satisfy.
- Making clear the limits of my experience may suggest where to look for new challenges that might lead to more-robust ideas.

To illustrate the last point, work of Jeremy Gray [21] and David Corfield [10] provided useful perspectives outside my own experience. Some of the ideas here were sharpened by the associated challenges; see §9.4, 10.4.

I have tried to address a wide range of concerns, and a consequence is that most readers are likely to find parts of the essay irrelevant.

For example, in §2 I describe contemporary mathematics at four successive levels of sophistication. Research mathematicians may find the first two levels plausible but want to see sensible accounts at the third and fourth levels before coming to any conclusion. Extended discussion at upper levels may be irrelevant. Mathematics educators, on the other hand, may find an extended discussion useful and the deeper levels irrelevant.

The descriptions of personal experiences mentioned above are other examples. It seemed important to include these to explain how the ideas have been shaped and tested, but it does not follow that everyone should find the stories interesting. Readers are invited to skip material that addresses other interests or concerns.

1.5. About Education. The picture that emerges is that mathematics education in the twentieth century, and so far in the twenty-first, is modeled on largely obsolete mathematical practices of the nineteenth century.

To expand on this, the math-adapted methodologies of the early twentieth century were unattractive to philosophers and many traditional mathematicians. They charged that the new methods were “devoid of meaning”, “formal symbol manipulation”, “disconnected from reality”, etc. Professional mathematics changed, as it had to, but the arguments against the new methods were never answered on the philosophical level. The connection with philosophy was abandoned rather than updated. As a result, views of people who did not have to contend with the technical failure of the old methods were also not updated. Moreover some traditional mathematicians, most notably Felix Klein, were very influential in education and strongly imprinted nineteenth-century values on early twentieth-century education reforms, see §8. The result is that the nineteenth-century philosophical description of mathematics is still the dominant one outside the core professional community, and nineteenth-century arguments against contemporary practices are deeply embedded in educational philosophy and culture.

The result is a deep divide between mathematicians and educators. One of the objectives of this essay is to promote communication across this divide.

I hope educators can come to see arguments against modern mathematics such as “devoid of meaning”, “disconnected from reality”, etc., as echos of the struggle a century ago, all the more strident and angry because they are the arguments of

the losing side. I believe modern mathematics has a lot to offer education but the antagonism must be overcome before possibilities can be explored.

On the other side of the divide, I urge mathematicians to realize that even if this account is fully correct, and even if educators accept it, the consequences for educational practice are unclear. It may be that in some cases the nineteenth-century approach is more appropriate for education even if it is mathematically “wrong”.

1.6. Acknowledgments. The inspiring collaboration [27] with Arthur Jaffe was the starting point for this work. Marjorie Grene provided early feedback and a model for careful critical thinking. Anne and John Selden deserve thanks from both the author and the reader for clarifications in the exposition. I am also grateful to Judith Grabiner and Jeremy Gray for helpful comments.

CONTENTS

1. Introduction	1
1.1. About “Contemporary”	1
1.2. About “Core”	1
1.3. About Mathematics	2
1.4. About the Essay	2
1.5. About Education	3
1.6. Acknowledgments	4
2. Mathematics and Mathematical Methods	5
2.1. Level 1: Complete reliability	5
2.2. Level 2: Methods for achieving reliability	7
2.3. Level 3: Origins of the methods	13
2.4. Level 4: Method Selection	17
2.5. Computers in Mathematics	20
3. Mathematical Science	20
4. Proof and Discovery	20
4.1. Components of Proof	21
4.2. Verifier-Specific Proofs	22
4.3. Incidental Benefits and Spectator Proofs	23
4.4. Non-Proofs	24
5. Mathematical Objects	25
5.1. Level 1: Definitions	25
5.2. Level 2: Perception	27
5.3. Level 3: Concept Formation	29
6. Mathematical Developments	34
6.1. Cognitive Difficulty	34
6.2. Cognitive Complexity	34
7. Mathematicians	34
7.1. Adapted to Mathematics	35
8. Education	35
8.1. Educational Philosophy	36
8.2. Informal Student Proofs	39
8.3. Focus on Diagnosis	39
8.4. Complicated Problems, Formal Proofs	40

8.5. Teacher Preparation	40
8.6. Applications and Word Problems	41
8.7. Counterproductive Practices	42
9. History: The Modernist Transformation and Later	43
9.1. Sketch of the Transformation	43
9.2. Confusion, Obscurity, and Philosophy	47
9.3. Later Developments	49
9.4. Gray's History	50
10. Descriptions: Other Accounts of Mathematics	53
10.1. Barry Mazur, 'When is One Thing Equal to Some Other Thing?'	53
10.2. Jonathan Borwein, 'Implications of Experimental Mathematics for the Philosophy of Mathematics'	56
10.3. Keith Devlin, 'What Will Count as Mathematics in 2100?'	56
10.4. David Corfield, 'Towards a Philosophy of Real Mathematics'	58
10.5. Michael Stöltzner; 'Theoretical Mathematics: On the Philosophical Significance of the Jaffe-Quinn Debate'	63
10.6. William Thurston; 'On proof and progress in mathematics'	66
10.7. Constructivism	66
10.8. Religion	68
11. Sociology: Culture in Mathematics and Physics	68
11.1. Reliability	68
11.2. Process versus Outcome	70
11.3. Efficiency	71
11.4. Publication	72
References	74
Index	77

2. MATHEMATICS AND MATHEMATICAL METHODS

I describe contemporary mathematics at four levels of sophistication, beginning with reliability. The fact that mathematics is reliable heavily influences mathematical social structure, publication practices, attitudes of users, and many other things. Some of this is expanded in §11.

Deeper levels concern how reliability is obtained and do not directly effect social structure etc. The second level describes how mathematical methods, particularly proofs, produce reliable results. Implications for education are explored in §8 and [63]. The third level describes the origins of reliability-producing mathematical methods, and the final level examines in more detail how the third level works.

Computer methods are not qualitatively different from this point of view and mechanisms already in place seem robust enough to deal with them. However the quantitative differences may be great enough to make this unclear, so this is discussed in §2.5.

2.1. Level 1: Complete reliability. *As a first approximation, the characteristic feature of contemporary core mathematics is that it gives completely reliable conclusions.*

Put another way, reliability seems to be the axis around which the mathematical enterprise revolves. The rest of this section relates experiences that convinced me of this.

2.1.1. *Reliability and Publication.* My first glimpse of the explanatory power of reliability came in a study of mathematical publication practices [52, 53, 50, 54, 55], in the context of the transition to electronic publication. Special features of the mathematical literature include:

- Durability. The primary literature remains useful much longer than in other sciences, and citation half-lives of individual papers are longer.
- Extent. Key papers in other sciences tend to be concentrated in a small number of elite journals while mathematics is more widely distributed. When the JSTOR archive project [29] first started, their procedure was to digitize the half-dozen or so core journals of each academic area. However they found that a reasonable approximation to the core of mathematics seemed to require at least three times this number, and they are now up to fifty (including statistics).
- Lack of repetition. Other sciences tolerate or even encourage duplication in the literature, perhaps as an extension of the idea of “replication” of experiments. Mathematics by-and-large does not tolerate duplicate publication.
- Weak secondary literature. In other sciences the primary literature is filtered and compressed in the secondary literature (monographs, surveys, texts), and most citations point to the secondary literature. In mathematics the primary literature is already reliable so it does not need to be filtered and does not compress as much. The secondary literature is less extensive and contributing to it is not as high-profile an activity as in other sciences.

All of these require consistent reliability, and reflect a dependence on it.

2.1.2. *Reliability and Sociology.* My second encounter with reliability per se was in, roughly speaking, sociology and ethics [27, 56]. The issue was whether heuristic arguments or “physical level of rigor” should qualify as finished products in mathematics¹. Two centuries ago the answer would have been “yes”, and a century ago “maybe”, but today the answer—for core mathematics—is “no”. As a human enterprise core mathematics has adapted to—and become dependent on—a degree of reliability that these methods cannot provide. This adaptation is explored in §11, through a comparison with theoretical physics.

I was led into this by experiences with physics and computing. This story, in §2.2.2 is part of the next level because eventually I learned more from it.

The behavior of users in other fields also reveals adaptation to mathematical reliability. If analysis of a model gives results not in agreement with observation, then users conclude that the wrong model is being used or the data is bad. Garbage *out* must mean garbage *in*, because they are confident that mathematical analysis will not turn good data into garbage.

There are occasional problems with misunderstandings or with material presented as mathematics that has not in fact been mathematically verified, but these are negligible compared to other sources of error. In fact, reliability of mathematics is so deeply ingrained that it has become invisible: other scientists don’t realize the

¹Heuristic conclusions are often very valuable when presented as conjectures (i.e. *unfinished*). It is the claim of completeness that is problematic, not the methods or conclusions per se.

magnitude of their dependence on it. The mathematical view, of course, is that other scientists are not nearly appreciative enough of our hard work.

2.1.3. *Reliability and History.* In a study of the history of manifolds (unpublished, see [69]) I saw the influence of reliability, and the methodological transition mentioned in the introduction and §9. Erhard Scholz has a wonderful description of this area in the period 1850–1920 [72]. It comes through clearly in his account that the methodology was pre-modern in many ways, including:

- Reliance on intuition and direct insight rather than formal definitions and proof.
- The role of philosophy in providing meaning and direction.
- The role of, in Gray’s words [21], “unspoken realist assumptions”.

However the subject was considerably too subtle for this. The intuitive approach enabled Poincaré to develop his deep and remarkable insights but it was inadequate for actual development and the methods now seem clumsy and ineffective. Real success had to wait on the development and adoption of contemporary methods.

Manifolds lie at the confluence of topology, analysis and geometry. The modernist transition in manifold theory seems to have come a little slower than in the constituent fields, no doubt due to the influence of Poincaré, Brouwer, and to some extent Klein. In any case the field came into focus in the 1930s and grew explosively for the rest of the century. There were occasional methodological lapses, with attendant reliability problems, and surgery theory in particular is still not fully consolidated.

My study of manifolds was too casual to be a reliable basis for conclusions and I did not make much of it at the time. Recently Gray[21] has provided a deep and thorough account of similar developments in many other areas of mathematics, and the manifold story turns out to be roughly representative of the whole discipline. The major differences have to do with timing (good definitions came late in the game). This essay draws extensively on Gray’s work, see §9.

2.1.4. *Summary.* Even with the clues described above (except Gray) it me took about a decade and further experiences recounted in the next section, to overcome my disbelief that a real field of knowledge could be characterized in such a simple way. Indeed, contemporary core mathematics seems to be unique in this regard, and only because it went through a period of focus and simplification in the late nineteenth and early twentieth century. However, bearing in mind that this is to be understood as a first approximation, and that “reliable” is not to be interpreted in a technical way², “complete reliability” seems to provide a robust starting point for deeper investigation.

2.2. **Level 2: Methods for achieving reliability.** Mathematics provides basic facts and methods of reasoning with the following property:

If a mathematical argument produces a false conclusion then an error can be found in it.

This is shortened to ‘mathematical methods are error–displaying’. Reliability is obtained in practice by making error–displaying arguments and checking them very carefully for errors, see Proofs, §4.

²The philosophical notion “reliabilism” [20], for instance, seems to be inappropriate here.

We will see that there are strong limits on the conclusions that can be obtained this way, and that while these methods are very robust in some ways they are fragile in others.

2.2.1. *Initial Comments.* These indicate some of the directions to be explored, and provide perspective. We add to the list in §2.2.4, after background in the next few sections.

- The error–displaying formulation is almost a restatement of ‘proof by contradiction’: an error–free argument with a false conclusion and *one* uncertain hypothesis forces that hypothesis to be false (there must be an error, and that is the only candidate).
- This formulation emphasizes that mathematics has an *internal* criterion for correctness. Mathematicians will accept things that have been proved as being correct, even when they are outrageously counterintuitive.
- This formulation identifies error detection as a crucial activity.

Proof by contradiction has had a central role in mathematics for millennia³. Considering this, and with hindsight, the path from “reliability is the characteristic feature” through “how is it achieved?” to “error–displaying” seems straightforward. However I never asked “how is it achieved?” Like most mathematicians I regard meta–questions with suspicion and more likely to lead to error than to insight⁴, so more clues were needed. My experiences are related in the next two sections: analysis of the topic resumes in §2.2.4.

2.2.2. *Adventures in Physics and Computing.* I spent a decade, roughly 1989–1999, on a computational project distantly related to physics. The goal was to detect counterexamples to the Andrews–Curtis conjecture about deformations of two–dimensional complexes. This conjecture is expected to be false and there are many specific examples for which it is expected to fail, but for fundamental reasons the standard methods of topology are unable to detect any such failure.

I was interested in this problem because it is a good model for bizarre behavior of smooth four–manifolds, and I was attracted to the latter because I had been quite successful in understanding *topological* four–manifolds, see [16]. Standard methods of topology are also powerless to deal with smooth four–manifolds. In that case, however, extremely deep and subtle analysis partly aided by clues from physics had succeeded. Furthermore there was a related development in dimension three (Reshitikhin–Turaev, [71]) and an outline of a general context related to theoretical physics, “topological quantum field theory”, was proposed by Witten [81]. There was an excellent heuristic argument that a similar approach should work for two–complexes.

The program had two components: abstract work to develop invariants and develop computational algorithms; and programming and computation. The key part was to show that prospective invariant are well–defined. It is all too easy and common to define invariants that don’t actually detect anything because a loophole

³Hardy, [23], writes “Reductio ad absurdum, which Euclid loved so much, is one of a mathematician’s finest weapons.” My thesis is essentially that it is such a fine weapon that the rest of our methodology has evolved to support confident and aggressive use of it. Ironically most of Euclid’s methodology was eventually rejected because it was *not* precise enough, see §2.2.10.

⁴See Thurston [77] where careful phrasing of questions is used to justify conclusions overwhelmingly rejected by the mathematical community.

lets the target phenomenon escape, and it is obviously a waste of time to compute such a thing.

I had serious culture shock trying to make sense of Witten’s proposal. He worked on the “physical level of rigor” using what I now see as the methodology of mathematics in the nineteenth century and before: informal definition by example or by reference, and heuristic arguments based on personal intuition. This was completely alien to a mathematician trained in the late twentieth century. After consideration I decided I was willing to work in this mode—and publish in a physics journal—if it could address my problem. Unfortunately the ideas came apart immediately when I tried to modify them. Moreover this approach is inadequate for showing something is well-defined. Witten did not address this at all, and in fact his description is not quite well-defined. This may be appropriate in physics:

- A mathematical construct that manifests itself in physical reality presumably must be well-defined. The primary goal, then, is to describe it in a way that exposes structure and permits calculation.
- It is often better to leave details ambiguous and therefore adjustable as the theory develops. For instance an extremely precise alignment of the stars is required to produce a “topological quantum field theory” satisfying the first versions of the axioms. Physically relevant theories probably won’t satisfy them, and the mathematical theories developed for smooth four-manifolds definitely do not. Too much precision too soon might lead in the wrong direction.

This approach is not effective in mathematics. Curiously, the ostensibly mathematical development [71] of a theory satisfying Witten’s description also lacked significant detail related to showing the invariants were well-defined. I now see this as an instance of residual nineteenth-century methodology in the Russian school of mathematics, see 9.3.2.

My first task in the project was therefore to develop an abstract framework that would include two-complexes and be sufficiently precise to support a dependable demonstration of well-definition.

[[unfinished]]

2.2.3. *Insight from Education.* In the past decade I have spent many hundreds of hours providing one-on-one help to students working on computer-based materials. Students asked for help when they got stuck, and my job, as I saw it, was to diagnose specific errors and correct them in a way that would keep them from recurring. I cannot count the number of times I’ve said “let’s find the mistake”. I became adept at spotting work habits that invited and hid errors, and tried to correct them. Finally, I spent a great deal of time developing materials (e.g. solution hints) designed to make errors easier to find by maximizing “error-display”. Eventually I thought carefully about all this. The fact I was relying on, that mathematical errors *can* be found, is not obvious and is actually quite remarkable. Working in a way that displays errors, combined with active error-seeking and correction, provides reliability, not just for students but for all of us. My work with students had become a model of mathematics itself.

2.2.4. *Problematic Aspects.* We return to analysis of the topic. The error-displaying formulation as given above leaves a number of issues to be clarified:

- What are the facts and methods that mathematics “provides”, and why should we believe they are error–displaying? See §2.3.
- “Error–displaying” is supposed to mean *someone* should be able to find errors. Who, exactly? See §4.
- How can we know if an argument really is error–free, and therefore has reliable conclusions? It is usually humans that do the checking and humans are not completely reliable so this is a basic source of uncertainty. See §4.

These are serious problems at deeper levels. Before tackling them we further explore the current level.

2.2.5. *The Legal Analogy.* It is as though we have stumbled onto a language for writing contracts that Nature accepts as legally binding. If we write a valid contract (proof) then we get certain goods (conclusions of the theorem). But, as with any legal system, there is a loophole problem: a contract that is not really airtight can leave us empty–handed or with unwanted goods.

In these terms the ‘error–displaying’ property is: whenever the wrong product is delivered we can find a loophole in the contract that enabled this to happen. It seems that a loophole–free contract really does guarantee delivery of the correct product. There are obvious problems with being sure any particular contract has no loopholes, but in practice *very* careful study of the language and avoidance of sloppy or wishful thinking seems to work pretty well.

Some interesting differences between mathematics and human law:

- Unlike commercial contracts we don’t have to *pay* for the goods, we just have to write a valid contract!
- The mathematical “laws” that govern these contracts aren’t recorded anywhere: we have to discover them by trial and error, see §2.3.
- We can’t legislate, enforce, or decide for ourselves what the laws are or the products to which they apply. Some things simply don’t work, and Nature generally regards our attempts to be too creative or greedy as just providing loopholes that justify non–delivery.
- Failures cannot be appealed.

But the goods acquired (mathematical conclusions) are, after all, both useful and free⁵, so we should accept what we can get and be grateful.

2.2.6. *Explicit Rules for Mathematics.* Genuinely error–displaying methods require explicit rules for argumentation: any method that is not made explicit cannot be used without appearing to be an error.

Explicit axioms describing properties of mathematical objects, and *some* of the rules of logic have been explicitly formulated for millennia. The idea that *all* of the rules should be made explicit is relatively recent⁶, and development of a really effective set of rules is even more recent. See §9 for a historical discussion. How good rules are developed (or discovered) is described in §2.3; the point for the present discussion is simply that explicit rules are required. The slavish devotion of mathematicians to rigorous methodology is required by the subject and is not a character flaw.

⁵Once the hard work of proof is done, anyway.

⁶Attributed to Frege in 1879 in [46].

2.2.7. *The ‘Correct Outcomes’ Formulation is Not Effective.* A common description of the mathematical process is *correct arguments give correct conclusions*. However this lacks logical force and invites confusion:

- Who is entitled to decide when an argument is correct, either methodologically or in the sense of being error-free? Leaving this unclear has been a perennial source of disagreement.
- This version is often interpreted as being to some extent a definition: arguments are *validated* by having correct conclusions. This is an *external* criterion in that it depends on things outside the argument itself. External validation is appropriate in other sciences but methods obtained this way are not sufficiently reliable to be satisfactory in contemporary mathematics.

2.2.8. *Mathematics vs. Science.* The observations about internal/external criteria provide a surprisingly sharp line between mathematics and the other sciences.

- Mathematicians will accept an assertion on the basis of internal criteria, even when it is outrageously counterintuitive and not supported by any independent evidence. This would be highly unprofessional in any other science.
- Mathematicians, when doing core work (i.e. for use *in mathematics*) will not accept an assertion on the basis of external criteria, even when it seems to be incredibly well tested. See the discussion of the Riemann Hypothesis in §2.4.1 for an example. This would be highly unprofessional in any other science.

The second point has the qualification “core, or “for use in mathematics” because this is the context that requires full mathematical rigor: later work depends on complete reliability.

In fact a great deal of mathematical work is done for external use. Full mathematical rigor is not needed and this work is typically done in a more scientific mode. This gives access to an much larger world, but it is important to realize that it is a different world from mathematics. Confusion about this is a source of stress and it seems possible that mathematics will divide along these lines in the near future, see §3.

This distinction between mathematics and other sciences is important but in a larger sense superficial. We are all engaged in disciplined searches for knowledge; mathematics just has a very different source of discipline. This brings me to the final point:

- Rigor plays the same role in mathematics that agreement with the physical world plays in other sciences. Relaxing rigor is like ignoring data.

2.2.9. *A Definition of Core Mathematics.* The observation in the previous section can be reformulated as a definition:

Core mathematics is the body of material that can be studied with error-displaying methods.

This is not a serious proposal to formally define the discipline; the point is to emphasize that mathematics is limited by its methods as well as empowered by them. Non-core mathematics is subject to different constraints and is much more like other sciences, see §3.

Note this is a *functional* definition in the sense that anything accessible to the methodology is fair game. Most proposed definitions have been descriptive, for example “the Science of Patterns”, c.f. Devlin, [13].

2.2.10. *Example: Euclidean Geometry.* Deficiencies in Euclid have been provoking objections for over 2,000 years quite apart from the failure of the parallel postulate, see M. Klein [32] §42.1. The perspective here provides a sharp understanding of these concerns: the methods are not error–displaying.

There are infamous proofs that all triangles are isosceles; see Klein for examples. We know the answer is wrong and after the fact we can identify the problem as unfortunate choices of diagrams. But they are identified as unfortunate using *external* criteria. They do not violate any rules and cannot be identified as erroneous using *internal* criteria. In other words, the methods do not display an error.

More generally the approach is essentially proof-by-example. It is emphasized that “generic” examples should be chosen, in other words ones that implement the universal quantifier (for every triangle . . .). But, as observed above, there is no *internal* criterion for knowing whether or not any particular example is good enough. It succeeds most of the time because the topic is very simple, but it is a bad model for almost any other topic. If we don’t want students to use proof-by-example in later work we should avoid Euclid⁷.

Another problem of Euclidean geometry is that relationships such as ‘similar’ and ‘congruent’ are essentially undefined primitives. They are described operationally: triangles are congruent if one can be “moved” to coincide with the other. Taken literally this suggests that congruence is experiential and can be checked by actually moving things around. This is the feature that has drawn the most criticism over the millennia. It provides many opportunities for invisible errors and is certainly not mathematical by current standards.

A modern version of congruence is that there exists a distance–preserving function (rigid motion) that takes one thing to the other. To show things are congruent one must produce the function or logically demonstrate its existence. An error–displaying transcript of the argument pretty much requires that the function be assigned a name. Typically the function rather than the original objects becomes the center of attention. One gets drawn into the group theory of rigid motions, crystallography and other wonderful things. Moreover the definitions are appropriate for any dimension, and when formulated carefully many of the results generalize as well. The experiential Euclidean formulation blocks the way to all this.

Thirty-five years ago I attended a lecture of Dirk Struik⁸ in which he mused on the uneven pace of mathematical development. Geometry and arithmetic were both well-known to the ancients, he observed, so why did they not come together until sixteenth century Europe? In particular he felt that Omar al-Khayyam (he of the *Rubaiyat*) had all the pieces in hand; why didn’t he put them together? Struik speculated that the synthesis was inhibited by some sort of societal influence or

⁷Actual use of Euclid is now rare in the US but still common in some other countries. There is a movement to bring it back—with flaws reinforced rather than fixed—using “dynamic geometry” software.

⁸ Mathematician and historian of mathematics, 1894–2000. See the tribute in the Notices of the AMS at <http://www.ams.org/notices/200106/mem-struik.pdf>.

mindset. Now that I see how unmathematical some of the Euclidean methods are I wonder if that might have been the problem⁹.

In §?? I analyze why Euclidean geometry has been so attractive to educators, and find another difficulty with the experiential formulation.

2.3. Level 3: Origins of the methods. Here I address some questions about the error–displaying formulation at the beginning of §2.2. Specifically: what, exactly, are the methods that “mathematics provides”, where did they come from, and why should we believe they are “error–displaying”?

Specific methods are discussed in more detail later. For the moment note that they include the assertion: *an if–then statement with a false hypothesis is always true, independently of the truth of the conclusion*. This strikes most people as artificial if not wrong–headed. If we want to assert this is the right thing to do¹⁰ then the “where” and “why” questions posed above must be answered carefully and convincingly. In particular, claims that standard logic is “natural” or “self–evident” are not convincing and much too close to “socially constructed by *my* social group”.

In a nutshell the answer is,

Mathematical methods evolve to maximize reliability, by maximizing the error–displaying property.

There are two sides to evolution: things that work are adopted, and things that don’t are eliminated. The most profound changes in the 1890–1930 transition were eliminations. This made them hard to see, particularly because the “eliminated methods” (especially intuitive or heuristic argument) are alive and well in the development process and “eliminated” only from final products. See §9 for detail on the transition.

The larger point is that mathematics has evolved to exploit a very special niche in the world of knowledge, one both empowered and constrained by complete reliability.

2.3.1. Error Magnification. At the first level I explained that excellent reliability is so commonplace in core mathematics that it is even reflected in the sociology and publication practices of the discipline. The implicit claim here is stronger: not only is reliability a fact but it is not an accident or byproduct of something else. There is something so powerful about complete reliability—or so damaging about the lack of it—that it has shaped the subject.

Here is an explanation: Delicate and elaborate arguments are undertaken to magnify and exploit subtle features. But they magnify errors as well as correct features. Deeper investigations give higher magnification, with the consequence that tinier errors in the original material will lead to painful collapses. Even “really excellent” reliability is eventually unsatisfactory: only *complete* reliability provides a stable basis for unlimited exploration.

The link between elaborate arguments and error magnification applies to any science, not just mathematics. In other sciences complete reliability is impossible so the conclusion is “avoid elaborate arguments”. Wisdom and experience play important roles in knowing which arguments could be useful and which are too

⁹Khayyam knew the Euclidean texts and identified the use of motion as a problem, but this was evidently not enough to overcome the blockage.

¹⁰There have been attempts to develop alternatives, see [47], but mathematicians have not found them useful.

elaborate to trust. Mathematics followed this pattern until the mid nineteenth century when extremely reliable methodologies began to emerge. The advantages were so powerful that they drove most of the subject toward complete reliability.

2.3.2. *Methods Evolve.* Basic mathematical methods and ingredients (natural numbers, logic, quantifiers, set theory, etc.) have been evolving for well over two thousand years. The contemporary approach is, roughly:

- Vigorously crash-test everything. In particular this includes seeking and attacking weak points; and
- Discard or modify anything that can be made to crash.

Another way to put this is that the methods of mathematics and their properties are *experimental* results. The tests applied are far harsher than in other sciences, see §2.4, but the principle is the same.

Ruthless application of this approach, and the resulting clarity and precision, is a relatively recent phenomenon. Basic materials stabilized around 1930 and were generally accepted by 1950. Examples of modern practice can be found much earlier, but examples of pre-modern practice are also common until about that time.

Methods are still evolving, but basic methodology is so effective and well established that compatibility with the base is the principal criterion. This point is explored in the fourth level, §2.4.

2.3.3. *Short-Term Success of Erroneous Formulations.* Every formulation of mathematics, at least until very recently, has been technically wrong but successful in a limited setting. It is important to understand how this happens because it is a source of confusion, and it clarifies the sense in which mathematical methods can evolve.

A formulation of mathematics is wrong if it is not error-displaying, or in other words if there is an argument without methodological errors that has a false conclusion. However this is only a problem if one actually *makes* such an argument, and it may be that lots of arguments don't run afoul of the error. In the terminology of §2.3.1, errors have not yet been magnified enough to be problematic. A related point is that believing something false does not cause problems unless the belief is used in a way that essentially uses its falseness. Some examples:

- It took the Greeks quite a while to get into trouble with the assumption that all numbers are rational.
- The example that shocked the Greeks, and that everyone knows about, is the irrationality of $\sqrt{2}$. Can this can be fixed by adding algebraic numbers (roots of polynomial with integer coefficients) to the number system? No, because π is not algebraic, but believing it won't harm calculus students. To get into trouble one must first of all say something like "let P be a polynomial with integer coefficients such that $P(\pi) = 0$ " and this is far outside the usual scope of calculus¹¹.
- Mathematicians in the eighteenth and early nineteenth centuries were successful with naïve infinitesimals, believing that bounded sequences converge, and that continuous functions are almost-everywhere differentiable.

¹¹It is hard to get in trouble even if one says this. The transcendence of π was first shown by Lindemann in 1882, and existence of *any* transcendental numbers was only firmly established by Liouville in 1844.

None of these are correct, but it took a long time to reach a depth where they became problematic.

- N ave set theory is known to be wrong by Russell’s paradox: the set–formation axiom is too strong. It is nonetheless useful to “working mathematicians” because it is extremely rare that anyone uses the axiom in a way that could not be justified by a correct weaker version¹².

The historical pattern was that mathematicians worked with a formulation until it broke down. They then revised the formulation to avoid known problems, and proceeded until the next breakdown. This is an evolutionary process, and it continued until an effectively error–displaying formulation was developed. Evolution of a somewhat different sort is still going on, see §2.4.

2.3.4. *Methods are not Proved.* I expand on sense in which it is an *experimental* conclusion that current methods have the error–displaying property.

From the historical perspective of the previous section the contemporary formulation is just the latest in a long series. The developers had hoped that there would be a *proof* that the formulation is completely reliable, but this did not happen and is no longer expected to happen¹³. The best we can say is that it is extremely well tested.

- For core methods this is humanity’s best-tested experimental conclusion by far. Insanely complicated proofs by contradiction routinely hold up.
- Since consistency is not proved it may be wrong. Note, however, that this possibility concerns the *whole system* of mathematics, not individual conclusions. Individual conclusions are uncertain because our methods for checking for errors are uncertain. The possibility that the whole system might collapse is negligible by comparison, and worrying about it is a waste of time.

Or referring again to the historical pattern, it will probably be a long time before anyone finds a way to make it fail even if it is wrong.

2.3.5. *Natural Selection and Community Norms.* In the very long run standards are enforced by natural selection in the sense that dysfunctional methods confer disadvantages on their users and eventually die out, while effective methods confer advantages and are eventually accepted by the whole community.

A complication is that the selection process acts primarily at the community level and is expressed through social norms. Individual circumstances usually include pressures more powerful than the background influence of mathematics itself.

Vulnerability to flawed methods has encouraged development of strong and quickly–enforced norms, at least in core areas of mathematics. In particular norms provide a communal memory of past failures and can quickly and efficiently reject attempts to reintroduce flaws previously rejected.

On the other hand, it is a reasonable concern that strong and conservative norms may block introduction of valuable new methods. Infinite objects, nonconstructive methods, and proof by contradiction all had acceptance impeded by norms, often inherited from the ambient society. However, it is more accurate to say that resistance from norms forced development and testing of very careful versions (and

¹²It did happen to me in my thesis work, see §10.1 for a discussion.

¹³By me, anyway.

careful disclaimers) that did not trigger strong reactions. Norms are doing their job correctly when they block sloppy versions of potentially good ideas.

The method–selection process is discussed in more detail in §2.4.

2.3.6. *Who Benefits from Methodological Norms?* The assertion above is that the “community” benefits from strong norms. The community consists of people who have worked hard to adapt themselves to the structure of mathematics so by now one might say that it is mathematics itself that benefits. Benefits for individuals are varied.

The primary beneficiaries are rank-and-file working mathematicians. Reliability and utility of the literature gives them effective tools and a solid base on which to build, see §4.3, 11.1. This enables them to make useful contributions much more routinely than in other sciences. It also makes possible steady, if often slow, progress in a much larger range of topics relative to the size of the active community, see §11.3.2. It is this professional middle class that defends norms most strongly.

2.3.7. *Problems With the Elite.* Elite, or highly–adapted mathematicians benefit less directly from norms. Some play by their own rules, and their conduct is not a good guide to professional practice.

Elite mathematicians benefit indirectly:

- through the stability and clarity in previous work resulting from their predecessors being held to high standards;
- through careful work done by those around them; and
- through a precise and highly disciplined presentation of mathematics during the training needed to become elite.

Nonetheless once a mathematician becomes elite in this sense, his or her intuition may be so good that intuitive work alone can provide very impressive accomplishments, see §7. Having to fill in details to provide error–displaying arguments slows them down, and almost always (the argument goes) just shows they were right in the first place. As a result some elite mathematicians object to being held to high standards [3, 77]. Some use the power of their ideas to force the community to accept behavior that would result in virtual expulsion of a middle–class mathematician.

Objections to high standards frequently take the form “requiring rigor is a threat to exploration and development”. This is silly: rigor is expected in the final product, and the quest for rigor guides and provides discipline for exploration, but no-one expects exploratory work itself to be rigorous. See §?? for the use of semi–rigorous argument. This objection also confuses the general issue of human creativity with questions about how to make it effective in mathematics, see §10.6 for further discussion.

For most of history the mathematical community was dominated by elite practitioners. Conflicts between their ambitions and the needs of the subject led to uncertainty about norms and uneven standards. The development of a large professional “middle class” of active mathematicians helped stabilize the situation by about 1930, see [21] and §9. By the end of the twentieth century norms were so firmly established that proposals like that of Thurston [77], §10.6 to relax them were rejected as not even interesting. The alarm expressed by Jaffe and myself in [27] about threats to mathematical norms turned out to be largely unnecessary.

The point is that the behavior of elite mathematicians is often a poor guide to well-adapted mathematical practice. It is not taken as a guide in the core community, or more precisely, young people who try to emulate casual elite approaches are unsuccessful and do not survive to perpetuate the practice. It is much more influential in education, history and philosophy, where the dysfunctionality is less clear.

2.3.8. *Insight From Editorial Experience.* The conclusions in [27] were also shaped by my experiences as editor of the Research Announcements section of the *Bulletin of the American Mathematical Society* in the early 1990s.

I came to the job believing that publication of claims speeded dissemination of knowledge in the community. After several years and around a thousand submissions, I realized there were widely divergent understandings of the editorial process. The general presumption was that complete formal proofs were in hand and would be published shortly. As often as not, rank-and-file mathematicians already had complete manuscripts and wanted to publicize their results while refereeing and long journal backlogs ran their course. Others were much more casual, and in particular some elite mathematicians saw research announcements as an outlet for their more intuitive conclusions. At one point a University of California Berkeley professor, presumably in a race for priority, sent significantly different versions every few weeks for several months.

Over time I came to worry that presenting casual announcements as citable publications came close to betraying the trust and conscientious work of the large majority of the community. Working with Jaffe on [27] sharpened my concerns and made it impossible for me to continue. Therefore I resigned as editor and argued that announcements should be dropped from the *Bulletin*.

Predictably, the argument against announcements outraged people who had benefitted from the practice. Unexpectedly, it convinced enough people that announcements were in fact discontinued. Disgruntled users organized an electronic outlet for announcements through the AMS, but eventually that was discontinued as well.

2.3.9. *Other Norms.* The norms considered here concern methodology, or in other words what sort of argumentation is, and is not, acceptable for use in a proof. The level of precision expected in the use of these methods is a separate issue and is discussed in the section on Proofs, §??.

2.4. **Level 4: Method Selection.** I have described core mathematics as characterized by complete reliability; complete reliability as obtained by use of error-displaying methods; and methods with this property as products of a long evolutionary process.

In this section I describe the selection process in more detail. Development of qualitatively new methods is described in §§2.4.1–2.4.3 and summarized in §2.4.4. In a nutshell, core methodology seems to be complete and new methods have to be shown to be compatible with it.

So far there has been no mention of computer methods because, from this point of view, they are not qualitatively new. However the quantitative differences are profound enough that the same basic considerations apply. Some of the issues are discussed in §2.5.

2.4.1. *Methods vs. Facts.* Logically speaking, facts are just short methods. A fact is error–displaying (as a method) if using it is not an error, i.e. if it really is correct. However the difficulty of demonstrating that things really are correct, and the vulnerability of mathematical methods to mistakes, leads to a distinction in practice. I discuss this here and draw a conclusion in the next section.

The mathematical world seems not to come just from logic or the methodological structure of set theory. It is necessary¹⁴ to provide a “fact” as input for the methods to work on: usually existence of the natural numbers (in the stripped–down Peano form), or just existence of an infinite set.

Mathematicians take considerable comfort that no further “facts” are needed:

- The additive and multiplicative structure of the integers, the erratic behavior of primes especially, are not new facts but consequences of the existence of the natural numbers.
- The subtle analytic structure of the real numbers and the deep magic of the complexes are not new facts but consequences of natural numbers¹⁵.
- Number theory, modern abstract algebra, topology, geometry, analysis, differential equations, and even the mathematical foundations of string theory all come from the natural numbers. No additional input is needed.

Contrast this with the Riemann Hypothesis. Everyone expects that it is correct; there are impressive heuristic arguments that it should be correct; a century of tentative use has failed to lead to an error; and it has been shown to be correct “with probability one”. It has been checked for the first several billion zeros and spot–checked incredibly further out. This is more than enough to establish it as a scientific fact but it would be unthinkable to modern mathematicians to accept it as a fact on these grounds. Why not?

The first problem is the general point that long and delicate mathematical arguments magnify errors. An argument that deduces something remarkable from the Riemann Hypothesis therefore also greatly increases exposure to any flaw it might have. The reason the bar for accepting the Hypothesis is so high is the same reason we want to know it: it has remarkable consequences.

The second concern is a more subtle version of the first. Our experience is that rare and unlikely things have special properties. Roughly speaking the constraints that make them rare also give them structure. If one writes the structure conditions for a Lie group or a category as systems of equations then they are hugely overdetermined and for many purposes it would be quite safe to assume solutions don’t exist. However the search for special structures leads us to them. If there is a failure of the Riemann Hypothesis then it’s a pretty good bet that there is an associated structure with remarkable properties, and sooner or later we will be led to it.

The Hypothesis may be correct with probability one, but if there is even a single failure then our search for significant structures will lead us to it with probability one.

¹⁴This is an oversimplification. There are formulations of set theory, c.f. [17], that seem to do everything, but this hides the issue in greater complexity of logic or set axioms. We take a naïve view here to avoid having to identify concerns about different versions of logic.

¹⁵Development of the reals from the natural numbers was undertaken in the mid to late nineteenth century to resolve problems in analysis, and has been part of the fundamentals since then.

2.4.2. *Extreme Testing.* The previous section raises the question: if mathematics is so sensitive to errors as to still be uncertain about the Riemann Hypothesis, how can we be sure of anything? In fact a few people do still worry. However to get anywhere we have to assume something, and the current core methodology is so well tested that most of the community is completely comfortable with it.

The key is the degree of testing. Turning pages in a book provides at least a weak test of the properties of the natural numbers because the pages are numbered. Even the simplest argument uses, and therefore provides a test of, basic logic. And there have been absolutely no failures that could not be traced back to methodological errors. In other words we have extraordinarily good evidence that these methods are error–displaying. By comparison the Riemann Hypothesis is quite raw.

This raises another question: is it conceivable that a genuinely new method could be used and tested so extensively to be accepted? Or is the subject closed? We will see that the core methodology probably is closed but there are still opportunities for development.

2.4.3. *Doubts into Tools.* A trend in the last seventy years or so is that some concerns about basic methodology have been resolved in interesting ways: the methodology emerges unchanged, but when carefully formulated the concern provides a new tool. We give several examples.

The “constructivist school” developed in response to concerns about validity of non–constructive existence proofs, and proofs by contradiction more generally¹⁶. It is often much harder to construct something than to show it exists. When it can be done one often learns something useful from the construction, but after a time it became clear that it is unsatisfactory as the only approach. Contradiction works fine and rejecting it is a self–inflicted disadvantage. A severe disadvantage in fact, and strict practitioners got selected out.

On the other hand when “constructable” is carefully defined, e.g., in terms of terminating algorithms, it turns out that some things are and others aren’t. A turning point was the Boone–Novikov proof that there is no (terminating) algorithm to determine from a presentation of a group whether or not the group is trivial. Algorithmic decidability is now a tool and has been used in profound and remarkable ways, c.f. [79].

Nonstandard analysis and model theory provide another example. The infinitesimal approach to calculus was rejected as insufficiently reliable for mathematics centuries ago but works so well as a heuristic setting that educational use—particularly in engineering—persists to this day. Eventually this was explained¹⁷ in a remarkable way: it works in education because nobody was trying to do anything fancy with it, and “fancy” could be precisely formulated in linguistic terms. Roughly, there are restrictions on how quantifiers can be used.

The outcome is that assertions with suitably simple formulations are independent of the existence of infinitesimals. If they can be proved with infinitesimals then they must be correct even in models without them. For a while it seemed that this might justify re-introduction of infinitesimals in education, but it turned out that

¹⁶There were also philosophical concerns about what it means to know something. Today these seem quaint and irrelevant but they were troubling at the time.

¹⁷By Abraham Robinson, see *Nonstandard Analysis* in [6].

the necessary linguistic subtleties are worse than the old-fashioned $\epsilon - \delta$ fix. It does provide a useful tool for research, however.

The final example involves transfinite methods, the Continuum Hypothesis and the Axiom of Choice in particular. The Axiom of Choice was first identified as a hidden assumption in vigorous set-theoretic constructions and had to be made explicit to keep the methods error-displaying. It has a great many useful equivalent forms and—though a certain unease persisted—over time it became widely used. The unease was resolved when it was shown to be independent of the basic axioms of set theory¹⁸. In other words this axiom can introduce an error only if there is already an error in other more basic axioms. With this reassurance it is now accepted as a basic tool.

The Continuum Hypothesis is also known to be independent of the other axioms so is also available for use, or denial if that seems more profitable.

2.4.4. *Summary.* The current core methods of mathematics are essentially Zermelo-Frankel set theory with the Axiom of Choice. There are minor variations and packagings that seem not to effect the mathematics built on them, see §10.1.

This core is effective and extremely well-tested and unlikely to change in any substantial way. Qualitatively new methods are likely to be of two types:

- Subtle elaborations of what is already there, e.g. model theory or computability and complexity theory; or
- Methods that can be shown to be compatible with (or independent of) the core, e.g. the Continuum Hypothesis.

2.5. **Computers in Mathematics.** So far no mention has been made of computers because from this perspective computation is not qualitatively different from core methods. Moreover, potential problems associated with computation are not qualitatively different from other problems faced by mathematics. [[[unfinished]]]

3. MATHEMATICAL SCIENCE

[[[unfinished]]]

4. PROOF AND DISCOVERY

Proofs are the operational form of the description of mathematics as error-displaying: arguments that actually would display errors. The formulation reflects my conviction that proofs are first and foremost an enabling technology for users. The more usual question “what is a proof?” leads to a passive, static picture, see “spectator proofs” in §4.3.

User features are clarified by describing proofs in two stages. First, “potential proofs” use formats and methods that should display errors but they are not required to be correct. Second, a prospective proof becomes a proof when it is checked and found to be free of errors. The main complications concern who should be able to find errors, and who might be qualified to declare something error-free. This is followed by a discussion of the processes used to develop proofs, and the role of proof in mathematical discovery.

¹⁸We refer here to Zermelo-Frankel (ZF) set theory, see [28]. There are other approaches to set theory, most notably Quine’s [17], but ZF is relatively clear and by far the best tested.

Use of proofs (broadly interpreted) as an everyday methodology in mathematics education is discussed in §8 and in more detail in [63]. Slightly more elaborate examples are given in a proofs-for-teachers context in [62].

4.1. Components of Proof. The validation criterion used in mathematics is: *if* error-displaying methods are used, and *if* the argument is given in enough detail to display any errors, and *if* checking does not reveal errors, *then* the conclusion is reliable. “Proof” is the name given to such an argument.

Making this precise, eg. how much detail is “enough”, can be complicated and is discussed in §4.2 below. The point here is that the format and methodological requirements are as important as correctness, and in some senses more important. To clarify this I introduce another term:

4.1.1. *Potential Proof.* An argument is said to be a *potential proof* if it:

- uses only methods tested and accepted as having the error-displaying property, and
- provides a record or transcript sufficiently detailed to allow checking for errors.

In other words, a potential proof satisfies all the methodological and format requirements of a proof, but is not known to be error-free. Potential proofs are fundamental to the proof development process because they can be developed without being too anxious about whether they will hold up under checking (i.e. at this stage “potential” is more important than “proof”). Checking often does reveal errors, but then corrections are developed and incorporated to get a better potential proof. The process is then repeated, and with skill and luck may eventually lead to a real proof. Focusing too much on correctness inhibits this process.

This description makes it clear that potential proofs should be sufficiently detailed that the *person developing the proof* can genuinely check them. Checking by others is essential in the discovery process and important in everyday use, and imposes further requirements that are discussed below, but the developer has primary responsibility for checking, and feedback from checking is an essential part of the process.

4.1.2. *Actual Proof.* The criterion at the beginning of the section can now be reformulated as: a *proof* is a potential proof that has been checked and found to be error-free, and consequently can be expected to have reliable conclusions.

Most accounts of proof emphasize correctness over methodology: an argument can be considered correct even if the methodology is defective, and this correctness can somehow make up for methodological weakness. In our terms, a proof need not be a potential proof.

Overemphasis on correctness often leads to misuse of checkability criteria. Teachers, for instance, often accept “proofs” with big gaps if the outcome is right, there are no outright errors, and the argument more-or-less conforms to a proof known to the teacher. But the gaps may hide unreliable extrapolation from other problems or various leaps of faith. This is “teacher checkability” that does not reveal problems to the student.

At the other end of the spectrum, an argument whose checking takes twenty years of serious research and a major contribution by a Fields Medalist cannot be

considered a potential proof, and therefore should not qualify as a proof even if in the end it has no outright errors, see §10.6.

Another problem involves unjustified intuitive leaps or “physical reasoning”. These may be appropriate in work that does not have to be fully mathematical, see §3, and are always appropriate in exploratory development, but they are not error–displaying. They cannot be relied on to *always* produce correct conclusions even if there are instances where they do. Again too much focus on correctness obscures this.

Finally there is a confusion involving truth and knowledge. It is common to think of arguments as either correct or wrong: checking may reveal which, but checking does not change whether or not the argument is correct. In particular an error–free argument is considered to be a proof whether or not the error–freeness has actually been verified. But this viewpoint is misleading. Something whose status is unknown cannot be relied on no matter how it eventually turns out. Checking may not change truth, but it certainly changes our knowledge and this is what matters in practice. The picture here reflects this: a potential proof does not become a proof until it has been appropriately checked and found to be error–free.

4.2. Verifier–Specific Proofs. The objective here is to make the phrase “sufficiently detailed to allow checking for errors” more precise.

A *potential proof for a person X* is a potential proof (i.e. record of reasoning, using tested and accepted mathematical methods) that *X* can check for errors. Similarly a *proof for a person X* is a potential proof that *X* has checked and believes to be free of errors.

This idea can be used to formulate expectations. In education, for instance, the ideal goal is for students to learn to find and fix their own errors. But for this to work they must be able to diagnose their arguments, or in other words they should be expected to give “potential proofs for themselves”. This depends on the individual. If a student can reliably do several arithmetic operations without writing intermediate steps, then omitting these steps from the record still gives a potential proof for that student. Another student may make mistakes that he does not catch when he attempts similar omissions. Potential proofs for this student require the intermediate steps. Further, what qualifies as a potential proof for a given student will change over time.

Verifier–specific potential proofs can also be used to make sense of behaviors seen in the research literature. There is quite a bit of variation¹⁹ but for the most part mathematicians write papers with a particular audience in mind. In other words, the goal is potential proofs for this group. Most mathematical specialties are sparsely populated and papers are relatively detailed to make them accessible to nearby areas and later times. In densely populated or rapidly evolving areas the target verifiers are much more expert and papers correspondingly less detailed.

4.2.1. Verifier–Independent Proofs. This verifier–specific version is useful and logically unproblematic but rarely used. The general feeling is that once a potential proof has been carefully checked and found to be error–free then anyone else who checks it should also find it to error–free, and consequently there is no need to list specific verifiers.

¹⁹Particularly among elite mathematicians, see §2.3.7.

Logically this just postpones the problem to describing what “carefully checked” should mean, and who is entitled to certify something as carefully checked. In practice there are many conventions that work pretty well, but not so well as to render the verifier-specific version unnecessary.

- *General expectation*: “appropriately qualified” people should be able to check it. This lack logical force and has been used to dodge criticism.
- *For publication*: an argument is a proof for (i.e. has been checked by) an anonymous referee.
- *Extreme high-visibility cases*: usually the community develops agreement on specific experts whose conclusions will be accepted as definitive.
- *Secondary literature*: proofs in advanced texts and monographs have usually been checked and refined by many people and errors are almost unknown. Proofs in lower-level texts are frequently unreliable.
- *In Education*: There are two issues:
 - (1) Spectator proofs (see below) that students are expected to read but not emulate. Standards are low, generally aiming for “understanding” rather than precision, and these frequently do not qualify as genuine proofs.
 - (2) Student proofs. Expectations are established by examples and are frequently unclear, partly because the goals of proof are poorly understood.
- *Machine proofs*: These have yet to stabilize. Computer calculations in human-designed proofs are still being treated on a case-by-case basis. Computer generated proofs are still a cottage industry that has not produced an example significant enough to force the issue, see §2.5.

Finally, experience shows that at the research level a *proof for the prover* (i.e. known only to, or checked only by the author) does not provide reliability and therefore should *not* be accepted as a proof. This is sometimes disputed by elite mathematicians whose intuitions are good enough to keep them out of trouble most of the time, and who don’t want to be slowed down by careful writing, see [3], §§7, 2.3.6, and 11. The overwhelming majority in the community do not accept this practice.

4.3. Incidental Benefits and Spectator Proofs. It turns out that proofs often have benefits beyond providing reliable conclusions. This plays out somewhat differently at the research frontier and in education.

4.3.1. *Discovery.*

- The process of correcting and refining an argument to be a genuine proof almost always leads to deeper and sharper understanding.
- This process also leads to proofs of complexity well beyond direct human capability.
- Understanding and checking a proof frequently illuminates the significance and core meaning of an assertion or the underlying definitions.

The first two points relate to an important feature of the proof process. The immediate objective is to give an argument that uses mathematical reasoning and can be checked for errors, but without any great hope that it is actually error-free. In other words a potential proof. Next the errors are found. The process then proceeds by iteration: fixes for the errors are developed and incorporated to give a hopefully better potential proof. Then errors are found in this, Along the

way ideas get sharpened, definitions often improved, and dealing with subtle errors steers thinking in new directions.

The final point is that deeper understanding is often a byproduct of the disciplined analysis needed to produce error-free arguments. It is the discipline, however, that ensures that the insights are correct and useful. There have been many casual arguments that were compellingly insightful but wrong.

4.3.2. *Spectator Proofs in Education.* Most educators think of proofs as opportunities for learning or “bearers of mathematical knowledge” (Hanna-Barbeau, [22], and [70]). This is much more passive than the view here of proofs as an enabling technology for users. In particular the focus is on what appear here as incidental benefits (understanding, insight etc.) rather than on the methodology and discipline that make them beneficial.

A name helps with exploring the difference. An argument designed to communicate understanding or evoke insight, but not intended to be emulated or deeply analyzed, is called a *spectator proof*. These suppress (or lie about) details to keep ideas clear, so cannot be checked for errors and are not potential proofs.

Spectator proofs are certainly valuable. Everyone—from elementary students to cutting-edge researchers—needs to learn, and fully rigorous proofs are a poor way to do this²⁰. Complex research papers often begin with a spectator proof (a sketch) that helps readers orient themselves, and this is often sufficient for many readers.

It is important, however, to remember that a spectator proof is not the real thing. Ideally they should be condensed from real proofs to ensure that the “insights” are accurate. They should retain as much as possible of the logical precision and basic structure, and should provide a skeleton that could be expansion to a real proof.

Spectator proofs in education rarely have these virtues. Focus on “understanding” etc. to the exclusion of precision often leads to vague, unfocused, and heuristic discussions. In some approaches brevity and precision seem to be considered bad things, and the goal of “understanding” seems to be to avoid the need for discipline and precision. This leads to weak meanings for “understanding” and similar terms, see [60]. The point, again, is that understanding and insight are *incidental* benefits of proofs, not part of the definition. They are important and precious benefits but they get their value from the connection to the more disciplined aspects of proof. Attempts to disconnect understanding and discipline weaken the benefits and undercuts further learning.

4.4. Non-Proofs. ‘Proof’ has a specific technical meaning that gives special significance to statements that have been proved. The terms

- physical proof
- visual or dynamic proof
- heuristic proof
- proof by example

²⁰The “Moore Method”, see the Wikipedia entry http://en.wikipedia.org/wiki/Moore_method and the legacy project at <http://www.discovery.utexas.edu/rlm/method.html>, does introduce mathematics through rigorous proof. A single such course at the undergraduate or beginning graduate level can be valuable: I had one and certainly benefited from it. But this is unsatisfactory as a general approach.

indicate methods that do not do the same job. There are also terms that do not contain the word ‘proof’ but are sometimes proposed as having similar status. These include

- understanding
- physical insight
- common knowledge or experience
- intuitively clear

Less-rigorous explanation, reasoning and argumentation are essential but there is a clear priority: proof establishes goals for relaxed reasoning and relaxed reasoning is evaluated by reference to proof. In more detail,

- Explanations of things already proved, are good or bad depending on their success in promoting understanding of things revealed by proof. A proof without an explanation is a sad thing but this does not effect its status as a proof. In contrast, an explanation that leads to understanding inconsistent with outcomes of proof is a bad explanation.
- A heuristic argument for something not yet proved can be impressive and compelling but turn out to be bad. The claim may be wrong, or it may be right for different reasons and the heuristic argument impedes rather than gets one closer to a precise understanding. The real value of heuristic reasoning cannot even be evaluated until a proof is available.

Mathematicians need and use heuristic reasoning and relaxed explanation. They celebrate insightful speculation and in fact the most influential contributions to mathematics have been conjectures and problems, not theorems. However speculation can provoke development only if it is presented as *needing* development, i.e. as a conjecture.

To put it another way, presenting heuristic reasoning as a finished product robs it of much of its value, and makes it hard to celebrate no matter how insightful it is. The claim of completeness must be withdrawn or widely rejected before development can proceed. Making such a claim or refusing to withdraw it is often seen as a parasitic attempt to cash in on the hard-earned reputation of real proof without doing the hard work. See [27] for details and examples, and §10.5 for discussion.

5. MATHEMATICAL OBJECTS

Technically, mathematical objects are ones that can be studied with error-displaying methods. The discussion begins on this level with generalities about structure of definitions, especially the use of axioms.

Deeper levels concern the way humans perceive these objects, and strategies used to develop these perceptions. It seems to me that the human aspects of objects are more basic and perhaps more complicated than the corresponding story for proofs. They are certainly more widely misunderstood. Discussion of human aspects continues in §6 on Mathematical Developments.

5.1. Level 1: Definitions. Mathematical objects are described in terms of axioms, and other objects that are in turn defined with axioms. This section provides an overview of this practice primarily as a context for later cognitive discussions.

5.1.1. *Axiomatic Format.* Just as with rules of argument §2.2.6, the axiomatic format of mathematical definitions is forced by the requirements of error-displaying methods: any property that is not made explicit cannot be used without appearing to be an error.

The axiomatic format has an unfortunate public image. The American Heritage Dictionary listing is

Axiom: . . . *Mathematics & Logic.* An undemonstrated proposition concerning an undefined set of elements, properties, functions, and relationships.

In other words, accepted without question or justification. Some otherwise very sophisticated people (including some philosophers) think of axioms this way, and believe axioms are a matter of taste and could be adjusted to get any desired outcome.

Another common belief is that axiomatic formulations rigidify a field and stifle further development, c.f. Devlin [14] and Lakatos [34], discussed in §10.3, §10.5 respectively.

Axioms would be necessary even if these unflattering beliefs were true because, again, error-displaying methods require everything to be made completely explicit and precise. But in fact axiomatic definitions are one of the most powerful and effective tools of modern mathematics.

5.1.2. *Development of Objects.* Standard basic definitions were developed and refined over long periods and with great effort. They were *not* simply a codification of an intuitive understanding.

In nearly every case, early attempts to codify intuitive ideas were unsuccessful in some way. Understanding the problems with an attempted definition usually leads to a sharper and more sophisticated intuitive idea. The new idea is codified and used to form new intuitions, problems with it are discovered, and the cycle repeats. Eventually a definition is found that is “right” in the sense that it is functional and captures the target idea. Quite often these right definitions have unanticipated further benefits. “Right” definitions are often reorganized as some features come to be seen as part of a wider context and are separated from features special to the object, but substantial changes are rare. This development process is frequently a community effort. The vision of initial discoverers may be clouded by the untidy complexity of the discovery process and leftover fragments of early versions. Clarity often comes through challenges by unencumbered minds trying to learn it and make it their own.

Finally, the primary criterion for correct definitions is functionality. Curiously this means they may not be recognized immediately. In fact, technically correct definitions of many major objects (e.g. limits, differentiability, groups, manifolds, categories, . . .) appeared years before they were accepted because there was insufficient technique to demonstrate their functionality.

5.1.3. *Benefits and Usage.* Perspective gained during the discovery process is often packed into the final formulation for the benefit of later users. The ideal definition has a very concise formal statement so it can be essentially memorized and quickly reviewed. It is often accompanied by suggestions on how to think about it or what it is supposed to do. And frequently minor propositions are provided to develop basic features and let the user see how the definition functions. The objective is

to develop an accurate internal representation (intuition) as quickly as possible so the definition becomes a tool and an anchor point rather than an obstacle; see the discussion below.

Expert mathematicians know this and when trying to understand a new topic usually start with the definitions, see Wilkerson-Jerde, Wilensky [80].

The deliberate identification and definition–exercise packaging of cognitive units is a relatively new development. It is probably the primary factor in the explosive expansion of complexity in the last century.

5.1.4. *Non–Mathematical Objects.* Things that *cannot* be completely and explicitly described with a definition *cannot* be studied with error–displaying methods. They are in an essential sense non–mathematical.

At present all physical phenomena are excluded. Most professionals acknowledge this by distinguishing carefully between phenomena and models. Mathematical conclusions about models are reliable but say nothing about the fit between models and nature. Failing to make this distinction negates the special virtue of mathematical conclusions. See §8.6 for discussion in an educational context.

5.1.5. *Summary.* In brief,

- Random definitions, or mutations of standard definitions, are useless and have no consequences.
- Poorly formulated definitions are painful to work with even when they do describe natural objects.
- Well–developed definitions efficiently encapsulate properties of apparently natural objects in insightful and powerful ways; and
- they serve as solid anchors for both learning and research.

Good definitions are a precious part of our mathematical heritage and should be exploited and celebrated, not avoided or degraded!

5.2. **Level 2: Perception.** This section concerns the way humans perceive things, and some implications for teaching, learning, and understanding mathematics.

5.2.1. *Perceptual Mechanisms.* People live in the world, through their senses. People deprived of sensory input for any length of time tend to become mentally unbalanced. It is common to conclude from this that things that cannot be sensed (mathematical objects in particular) are humanly inaccessible, meaningless, or even non–existent. But this is based on a misunderstanding of human perception; a more nuanced view shows that our limitations put constraints on how things should be organized or presented in order to be comprehensible, but not on the things themselves.

The first point is that people do not see what they look at, nor hear what they listen to. We live primarily in internal worlds of pre–packaged objects that are evoked by actual perception rather than determined by them. For the most part this is a good thing: it means we perceive what we *know* about things, not just what we could deduce from their appearance, and what we know usually far exceeds what we could perceive directly.

There is, for example, a thriving area in psychology that explores curious mismatches between reality and knowledge–influenced perception. Some of these mismatches are even different at different levels of awareness. There are magic tricks, for instance, that deceive the conscious mind but, judging from eye movements, do

not deceive whatever it is that controls eye movement. Presumably this means that the knowledge component is integrated into perception at a fairly high level in the brain.

The next point is that our internal models encode patterns, relationships, behaviors and facts, but with no particular restrictions on what it is that has the patterns, behaviors, etc. Through study and experience we can develop an enormous variety of effective models that allow us to “perceive” a great deal from very little, and very odd, input. Examples are radiologists who look at X-rays and see diseases; trackers who can follow nearly-invisible trails; stock traders who seem to see the economy in stock prices; scientists who see deep and intimate features of nature in their data streams; and art experts who see volumes in a brushstroke.

A final point is that highly-developed and heavily-used models become transparent and automatic. I cannot see printed English words without reading them; I cannot hear spoken English without understanding it. As a child I spent a lot of time in cornfields looking for arrow points. Now I cannot walk along a dirt path without scanning for points, and I automatically recognize characteristic features even when very little is visible. And, for better or worse, I cannot look at student work in mathematics without automatically diagnosing errors. All this happens so transparently that it feels like primitive direct perception.

By contrast I recognize very few characters from other alphabets and I am not sure I can claim to really “see” them. In other spoken languages I cannot even distinguish individual words. I cannot see pottery fragments where I see arrow points. And it is an effort to make sense of student work in most other subjects. In other words my genuinely direct perceptual abilities are primitive and unsatisfactory, and seeing *anything* clearly is dependent on well-developed internal models.

5.2.2. *Mathematical Perception and Reality.* Mathematical objects are toward the abstract end of the spectrum of things that people model internally, but are evidently well within reach of a system that can learn to read X-rays or diagnose electric circuits with a voltmeter. In fact elementary mathematical objects are quite accessible by comparison with these. The point, however, is that mathematics uses standard perceptual and cognitive facilities. It is not a separate ability or qualitatively different activity. This has quite a few consequences, including:

Mathematical objects are as real as objects in the physical world.

This is a cognitive rather than a philosophical assertion²¹:

- Because we use cognitive and perceptual abilities that evolved to help us make sense of the physical world (nothing else is available, after all), mathematical objects *appear* just as real to us as the physical, psychological, and other things we model with these abilities.
- Mathematical perception is not qualitatively different from what we think of as physical-world perceptions because almost *everything* we perceive is heavily filtered through highly-developed internal models.

In other words mathematics is as real to us as anything else. Philosophical anxieties about whether it “actually exists” are irrelevant and worrying about it interferes with the doing of mathematics.

²¹In particular this is not support for the Platonic view of mathematics. That approach begins with vague but pre-conceived notions of “real” and claims that mathematical objects qualify. Here “real” is used to describe the way people relate to things, not some abstract quality.

The cognitive form of the observation in §2.2.8 is that we should think of the mathematical world as independent of the physical world, as well as real. It is a commonplace observation that most things are better understood on their own terms than by forcing them into some other mold (trying to put square pegs in round holes). This is particularly true of mathematics because it has fundamentally different rules.

5.2.3. *Historical Note.* This is a recent point of view. Until the end of the nineteenth century it was an article of faith that mathematics is an abstraction of the physical world. Separating the two was a major part of the transformation in the early twentieth century, §9, and no doubt a major reason why so many people found it disturbing. Traditionalists like Poincaré and Felix Klein, for instance, objected because they wanted to use physically or experientially based methodology. Some philosophers objected (and are still objecting) because they believe mathematics only has “meaning” if it manifests in the physical world. Educators objected (and are still objecting) because they believe their students can only perceive and understand things that manifest in the physical world.

Physicists tend to feel that “important” mathematics is an abstraction of physics. It is therefore ironic that the needs of early twentieth-century physics (relativity and quantum mechanics) were major drivers of the conceptual separation: the mathematics needed was too complex and abstract to be accessible to the physically-oriented methodology.

Finally I note that while essentially all contemporary core mathematicians *act* as though mathematics is a real, separate world, this is rarely articulated. In fact it is not uncommon to profess to believe in a dependence on the physical world. There are two points about this:

- This is understood as a philosophical issue, and the physical connection is still philosophical orthodoxy. In particular, lack of philosophically defensible alternatives means that if a mathematician addresses the issue at all he is pretty much limited to this version.
- In §2.3.3 I observe that believing something false does not cause problems unless the belief is used in a way that essentially uses its falseness. In particular this belief is now relatively harmless because the methodology no longer uses it.

Believing in the linkage to the physical world is *not* harmless for elementary mathematics education because the methodology does still use it.

5.3. Level 3: Concept Formation. The point at level 1 is that the structure of mathematics requires the use of precise, axiomatic definitions. The point at level 2 is that humans have to relate to such things the same way they relate to features of the physical world. There is an obvious tension between these two. Educators usually try to resolve this tension by trying to soften and humanize mathematics, but the results are unsatisfactory. The mathematical approach is to develop human-adapted strategies to deal with mathematics as it really is. This has been successful but is still largely limited to the professional community.

In this section I discuss some of the professional strategies developed to help people understand mathematics. The first topic is the bottleneck: human cognitive facilities.

5.3.1. *A Model for Cognition.* For simplicity we consider human thinking as taking place on two levels: conscious and subconscious.

- The conscious level is where critical thinking, judgement, and effective error correction take place. The drawback of conscious thought is that it is quite limited in how many things and how much explicit complexity can be dealt with at one time. The saving grace is that these things can have great *implicit* complexity, managed by the subconscious.
- The subconscious—properly trained—can accurately represent or model very complicated constructs. These are perceived by the conscious level as single objects, with properties or behaviors that are more-or-less transparently managed by the subconscious.
- The limitation of the subconscious is that it seems to be mostly incapable of critical thinking and judgment. It can find possible connections, generate flashes of insight, or bring concerns to the attention of the conscious level, but for the most part conscious attention is required to evaluate and solidify them.

Effective thinking requires a distribution of effort so strengths of each level can compensate for weakness in the other. In mathematics the general plan is:

- (1) Organize as much as possible of the static material into conceptual units (objects, structures, definitions, standard methods, etc.).
- (2) Internalize (learn, understand, conceptualize, etc.) these units, or in other words, train the subconscious to accurately model their behavior.
- (3) Consciously manipulate these units.

The following sections expand on this.

5.3.2. *Accurate Internalization.* The biggest challenge in concept formation is accuracy: successful work obviously requires accurate subconscious models, and bad models are the main source of error. The reason this is a challenge is that errors in subconscious models *are almost never corrected by the subconscious*.

- Conceptual errors usually have to be consciously diagnosed, either by the individual or a mentor.
- The concept must then be reopened to conscious processing, and changes have to be imprinted over the faulty version.

Elementary concept errors can sometimes be corrected in the subconscious but the process is long, uncertain, and painful because it is driven by trying to work with a defective version. Diagnosis and conscious correction is faster and more effective, but in any case corrections are considerably harder than initial internalizations. Therefore there is a substantial advantage to getting concepts right the first time.

Experienced mathematicians take this seriously: before releasing a concept from conscious development they exercise it and are alert to inconsistencies and other indications of error. Well-written mathematical material is designed to support this: good conceptual units are identified and packaged in ways that facilitate internalization, and are accompanied by propositions and exercises designed to expose likely errors, see §5.1.3. Serious conscious work with conceptual units is usually attempted only after the internalization process is reasonably complete and reliable.

Experienced mathematicians can often quickly absorb new material. I believe the main factor is careful attention to accuracy in concept formation.

5.3.3. *Development of Clarity.* Development of clarity—finding better ways to organize and package the material for effective human use—is a more extensive part of mathematical research than is generally acknowledged. The neglect is not just cultural bias, however:

- The search for a better understanding of known material is usually driven by a desire to extend it. The better understanding then accompanies the advance that it enables, and it is the new result rather than the new clarity that is celebrated. In other words, conceptual improvements are often found embedded in papers with big new theorems.
- “Better” is something that has to be demonstrated by improving functionality of the concept. In research this usually means enabling an advance. It can also be demonstrated by making known material more accessible, but the bar for this is now pretty high: unless the improvement is dramatic it is usually better to wait for an advance-enabling clarification.
- Another way to put this is that concept development is best done with the guidance of applications. Applications, so to speak, “force us to get it right”²².
- “Better” also depends on context: different views of a concept might be better for different uses.

An important point here is that ambitious use forces the development of both technically effective concepts and packaging that enables accurate internalization. This suggests that educators might start with mathematical formulations and see what has to be done to adapt the material to their own context. The closer they can stay to the professional version the more effective and functional it will be, and the better it will be as a foundation for later learning. The current practice is to rely on conceptual formulations that are centuries if not millennia old: modern conceptual clarity is an untouched resource.

5.3.4. *Objects vs. Structure.* An important organizational strategy is to distinguish between “objects” and “structures” and think of them in different ways. The distinction is cognitive rather than mathematical so is imprecise and usually kept vague.

- Formats and development are essentially the same. In particular “structures” are objects in the general sense used in this chapter.
- Structures at one level of sophistication may be objects at another; and
- Many topics can be approached either way.

The conceptual distinctions are:

- Structures are properties, relationships or patterns of behavior. Generally they need to be internalized (“automatic facility” acquired) but the pattern and its consequences are more important than the “meaning” in any particular instance. Thinking in terms of structure is powerful because we can transfer internalizations and work habits from simple examples.

²²For example the Bourbaki project in France between 1935 and the mid-1980s tried to systematically organize and clarify core areas of mathematics. In the early period other materials were of low conceptual clarity (by current standards) and their work did in many cases significantly improve access. However some of their more ambitious organizational efforts did not mesh too well with later needs.

- Objects have, are defined by, or are related by, structures. “Understanding” an object frequently means recognizing and learning use its structures.

For example “addition” is a structure: explicitly, an associative, commutative binary operation with a unit and inverses. There is a rich collection of methods and consequences, and these apply to anything that has the structure. The corresponding “objects” are things that have the structure. These include the usual number systems; polynomials and power series; functions; rings and modules; many categories, etc.

The cognitive point is that we can internalize the properties of addition by working in the integers, and as long as it is learned *as a structure* this learning will serve us well in dealing with other additive systems. This is illustrated in [62] where the ring structure (additive and multiplicative structures together) of the integers is used as a model for the structure of general commutative rings.

Incidentally, individual numbers do not qualify as either objects or structures in this sense. They have properties and significance mainly as members of a larger ensemble, and may have different properties in different settings. The proper cognitive view, in other words, is that numbers are elements of a number system, *not* that a number system is composed of numbers.

The structural aspect of arithmetic is almost absent from elementary education. The connections between integer arithmetic and counting, for example, is specific to the example, not part of the structure, and does not transfer. Focusing on the “meaning” of integer arithmetic therefore interferes with internalization of structure. It is my experience (see §8.7.1 and [61]) that current use of calculators also interferes with internalization of structure by hiding the structure of numbers and making numbers seem completely different from algebraic expressions. A structural approach to arithmetic in elementary education may be impossible, but surely it should be possible to develop approaches that are not counterproductive.

5.3.5. *Variations on Primary Concepts.* Mathematicians deal with a complex landscape of concepts by collecting them in families around primary exemplars. The main function seems to be to help transfer internalized structure, so the development of these families seems to throw light on the scope and limitations of the transfer process. I illustrate this with some examples from algebra but these do not begin to represent the full picture seen deeper in algebra and in topology and geometry. I suspect that there is much more commonality and more opportunities for transfer in some applied areas but do not know enough to trace them out.

Understanding how this works might help organize educational approaches to number systems, for example.

The first example is the group family. a *group* (set with associative binary operation, unit and inverses) is considered a primary concept and variations include:

- *abelian* group (the operation is commutative);
- *finite* group (finite underlying set)
- *semigroup* (may not have a unit or inverses)
- *groupoid* (operation not always defined)
- special classes such as arithmetic groups, permutation groups, and linear groups.

Note that the primary concept is not the most general: two of the related concepts are obtained by *omitting* properties rather than specializing. This may seem strange

logically, but the objective is to support human understanding and it seems to work. Also, the deep theories for different variations are often very different, but again the objective seems to be to provide useful perspective and reference points, and particularly for beginning work, not complete analogs.

The boundaries between families are also determined by human usefulness. A *category*, for instance, could be described as a “semigroupoid” but familiarity with groups does not seem to transfer in any useful way to categories. Conversely a group could be described as a category with one element and invertible morphisms, but internalization of category structure does not provide a useful basis for understanding groups. Consequently categories are in another family, and in fact are the primary concept in the family.

Rings provide another example. A *ring* has two associative binary operations: one (thought of as addition) is commutative and has inverses; the other (thought of as multiplication) distributes over addition and is usually required to have a unit. The integers are the prime example. Variations include:

- *commutative* rings (multiplication commutes)
- *semirings* (addition not required to have inverses)
- special classes such as Noetherian, Artinian, nilpotent, semisimple, Euclidean, etc.

Note that the *non*-commutative version is considered primary even though people worked in commutative rings for millennia before these made an appearance. Also rings have underlying group structures, but the distributive relationship between the two operations is so powerfully influential that the group perspective is not particularly useful.

The final example is fields. A *field* is a ring as above, but additionally the multiplication commutes, and every non-zero element is required to have a multiplicative inverse. The prime examples are the rational and real numbers. Variations include

- finite fields and fields with nonzero characteristic;
- algebraically closed fields;
- *skew* fields, in which the commutativity condition on multiplication is dropped;
- special classes such as number fields, function fields, Galois extensions, etc.

This family is interesting because even though it is a subset of rings and the structures are very similar, it is considered a separate family. Invertibility leads to a theory different enough that ring theory is not a useful starting point for field theory, and vice versa. It is also interesting that the primary structure in fields requires commutative multiplication, while the primary structure in rings does not. Apparently non-commutativity changes the development too much to really exemplify field theory. Skew fields are often called “division rings” and located in the ring family.

It is important to realize that all of these variations were developed in response to applications, and are not the result of random exploration. Skew fields, for example, arise in a powerful structural result for rings: roughly, a semisimple ring is a product of matrix algebras over skew fields.

Other curiosities:

- The primary exemplar in a family is often the oldest or best-developed variation, but this probably should be seen more as winning a Darwinian

fitness struggle than simple historical accident. Wussing’s account [83] of the development of the group concept describes some rather heated competition among preliminary versions.

- Different communities may have different needs and different exemplars. Semigroups, for instance, are important as time-independent dynamical systems and in theoretical computer science, and these communities have their own terminology and conceptual organization.
- Corfield [10] and §10.4 describes a community that feels groupoids should be considered more central. This community, however, is not large enough to develop its own conceptual tradition and their technical developments do not rival the group concept.

5.3.6. *Summary.* The third-level discussion of mathematical objects is concerned with concept formation. The most important point is the need for accuracy, and some of the strategies for avoiding, finding, and correcting errors are described. The discussion then touches on:

- Thinking of some patterns of behavior as “structures” rather than objects, to aid in transferring internalizations between domains.
- Grouping structural concepts in families around a primary exemplar. The rough idea is to develop an understanding of structure in the exemplar and then, as the need arises, transfer to other objects in the family.
- The evolutionary process that identifies structures, families, exemplars, efficient approaches to internalization, etc.

There are related themes in the next chapter, including dealing with difficult concepts, and fitting concepts together in extended developments.

It should be kept in mind that all of this is a first approximation. I hope it will be useful for orientation and general perspective but I am sure that better clarity and functionality will evolve over time, just as in mathematics itself.

6. MATHEMATICAL DEVELOPMENTS

Most mathematical developments require multiple stages of concept formation. In this section I discuss consequences of this, including relationships with reliability and tradeoffs between difficulty of individual stages and complexity associated with multiple stages.

6.1. **Cognitive Difficulty.** [[unfinished]] Because it is a cognitive feature it depends on the individual and is not well-defined.

6.2. **Cognitive Complexity.** The *cognitive complexity* of a mathematical development is the number of distinct stages of concept formation it requires.

[[unfinished; refer to lecture notes.]]

7. MATHEMATICIANS

This section explores the ways in which mathematicians adapt themselves to the field, and some of the consequences of this adaptation.

[[unfinished]]

7.1. Adapted to Mathematics. High-achievers in any area mold themselves to the subject and a large part of the process is development of an accurate subconscious model as described in §??.

[coordinate with §5.3]

M. Gladwell in *Outliers* [18] proposes the “10,000 hour rule”: This seems to be how much fully focused professional activity is needed to become completely expert in (fully adapted to) almost any area. Gladwell’s examples are software pioneers, musicians and chess masters; mathematicians certainly fit the pattern.

The *way* people become adapted can be described in terms of the “dual process” model of thinking (c.f. Stanovich-West [75]). There is a fast, unconscious and heuristic mode, and a slower, deliberative and conscious mode.

- For most people, reliable work in mathematics requires inhibition of heuristic thinking and careful use of the slow, deliberate mode (Simpson [73]).
- Fully adapted mathematicians have trained their fast heuristic facilities to use mathematical methods. Their fast unconscious thinking is as good as most people’s careful conscious logic. This may well reflect a physical reorganization of the brain.
- Frequently expert thought processes become so transparent that they are unaware of what they are doing. Some have claimed that they no longer need careful logic or formal proof (c.f. responses [3] to Jaffe-Quinn [27]). This is silly: logic has been internalized, not bypassed.

Simultaneous chess exhibitions dramatically illustrate this sort of adaptation. There are, roughly speaking, two kinds of chess games: long games with lots of conscious deliberation, and speed chess which relies on fast heuristic thinking. Heuristic thinking is a disaster for non-experts because errors are easy to make and usually fatal. In a simultaneous exhibition one master player takes on a large number of non-experts; often twenty or thirty. The non-experts are playing carefully and deliberately and the expert is playing purely heuristically, but the expert nearly always wins. Again, their fast unconscious thinking is as good as most people’s careful conscious logic.

Fully-adapted mathematicians are like chess masters or concert violinists in that the technical aspect of their art has become transparent and their conscious focus is on on creativity and expression. They often have trouble giving coherent explanations of what they do, and these explanations are rarely good guides for beginners. Chess masters and violinists, however, have no hesitation in identifying practice—and lots of it—as the key to proficiency, and proficiency as a necessary precursor to real creativity and expression.

8. EDUCATION

In this section I explore educational implications of the perspective described in this essay.

The first topic is why conventional educational approaches seem unable to deal with the problems observed. In a nutshell, mathematics education is based on nineteenth century mathematics, and contemporary problems in mathematics education are closely related to difficulties that led mathematics itself to change in the early twentieth (see §9). Mathematics education may not be really effective until it also goes through this transformation.

After the general introduction I offer explicit suggestions how contemporary approaches might improve student learning. Strictly speaking many of these are not “implications” in the sense that they are neither deduced from nor validated by the perspective. In fact the reverse is true: these are ideas developed through extensive direct work with students, and an important objective in writing this essay was to make sense of them.

The education topics are student use of proofs, both formal and informal; the use of abstract structure to guide teaching; the role of diagnosis; and an approach to word problems. §§? and [63] expand on the idea that more care with organization, and more focus on diagnosis, could bring proof-like benefits to elementary mathematics even if the word ‘proof’ is never mentioned. Another opportunity for change concerns definitions and the development of mathematical concepts.

8.1. Educational Philosophy. The basic point is that current mathematics education is based on nineteenth-century mathematics. Rather than attempt a general discussion I focus on the role of Felix Klein. Good sources are Gray [21], especially §4.2.1.1 (pp. 197–199) and 4.8.3 (p. 277), and the Klein Project of the ICMI, [33].

Klein was very influential in educational reform around 1905 through his course and curriculum designs, his book *Elementary Mathematics from an Advanced Standpoint*, [31], and his philosophical perspective. His work and point of view are still central to the current educational system.

Klein was also universally recognized as a preeminent mathematician, but by 1905 the nature of his influence in mathematics was shifting. The modernist transformation (§9) was well under way and Klein, a product of the nineteenth century, was a strong advocate for the old ways. His work was still important and admired but for the new generation he was no longer a model to be emulated.

I describe some of the issues in the transformation that remain important in education.

8.1.1. *Meaning.* Up through the nineteenth century the purpose of mathematics and meaning of mathematical ideas were philosophical rather than mathematical issues. Leading mathematicians were, perforce, involved in philosophy, and many philosophers had mathematical interests. The main theme held that mathematics is an idealization of the physical world and receives meaning from it. Mathematical things unrelated to the physical world are meaningless, if not wrong.

This is the point of view so compellingly developed by Kant, or at least the way people seem to have understood it. This is also why non-Euclidean geometries were so troubling at an earlier time. Not because the old master was wrong—many of his methods were already being discarded—but because they drove an undeniable wedge between mathematics and the physical world. Klein was not at all bothered by non-Euclidean geometry. However this may have been because he saw a place for them in a deeper understanding of the physical world; he remained committed to the physical-realist philosophy.

The new view is that mathematics is a world of its own, with wonderful things not dreamt of in their philosophy. Much of it doesn’t relate to the physical world (yet anyway), and there is nothing wrong with that. The philosophical framework with its prescriptions about “meaning” inhibited rather than aided mathematical development.

8.1.2. *Intuition.* The role of intuition, particularly intuition derived from our experience in the physical world, is closely related to the philosophical connection described above. If mathematics is a faithful mirror of the world then our physical intuitions should be valid in mathematics. This point of view became increasingly problematic in the late nineteenth century and was abandoned as untenable early in the twentieth²³.

Klein was a strong proponent of intuition. In lectures in the 1890s Klein agreed that the “arithmetization” of geometry and analysis had brought new precision and advances. He nonetheless opposed it because it came at the expense of space intuitions, understanding, perhaps even meaning. In one lecture he seemed to offer a compromise: “naive” intuition was unreliable and had to be “refined” by careful work with axioms. He still insisted on the importance of psychological and experiential aspects of intuition. His discomfort, it seems, was with the increasing tendency of people to develop “mathematical” intuition directly from their work with axioms, and to give this artificial intuition precedence over that coming from the physical world (see also §9.4.2).

8.1.3. *Modern Difficulties.* Early in the twentieth century Klein was no longer a leader in mathematics because no one would follow him: talented young people saw giving up physical intuition in exchange for technical power as a good deal and voted with their feet. It was then that he turned his focus to education, where he was more successful. To this day, and in no small part due to his efforts, mathematics education follows nineteenth century mathematical ideals.

School mathematics rarely even contains eighteenth-century material so perhaps, at least for elementary education, this is not a bad model. In fact education had many problems but the basic model seems not to have been one of them until the 1990s when three things changed:

- Technology, particularly calculators, enabled elimination of rote and skill-oriented work that was never part of the model but could not be avoided as long as students had to work by hand. It turns out that students were learning quite a lot that wasn’t being deliberately taught and doesn’t fit the model.
- A vigorous reform movement, particularly the National Council of Teachers of Mathematics [42] in the US, put strong emphasis on pretty much exactly the things in the nineteenth-century model that were abandoned as problematic in the twentieth.
- An educational-research community developed, both deeply committed to nineteenth-century ideals and more-or-less designed to validate them (see [57]).

To be more explicit, and repeat some of the points above, the following commonalities are revealed by comparison of the NCTM Standards and the writings of Klein and other nineteenth-century proponents:

- An emphasis on intuition and experience from the physical world as a basis for mathematical learning;
- a conviction that the meaning and value of mathematics derives from the physical world (ie. from applications);

²³This refers to the use of intuition in official justifications (proofs). Intuition as a source of insights is unproblematic and thriving.

- an emphasis on heuristic understanding and intuition over rote or formal work with soulless axioms; and
- a preference for discussion of goals and methods on a high (philosophical) level rather than in terms of skills and mathematical structure.

The last point helps explain the persistent focus on Standards Documents despite plentiful evidence that they are ineffective, [58]. In any case these are all convictions that had to be abandoned or substantially revised before twentieth-century mathematics was possible, and may also be blocking progress in education.

These observations resolve a puzzle: current educational philosophy is coherent, attractive, and well articulated, even if deeply flawed. This is a significant accomplishment and I used to admire the NCTM writing teams for it. But it is now clear that they cribbed—knowingly or not—from the masters of nineteenth-century mathematics.

8.1.4. *Is Change Possible?* This discussion reveals some substantial barriers to change:

- (1) The coherence and attractive articulation inherited from nineteenth-century mathematics prevents piecemeal change. The whole model has to be replaced, just as it was in mathematics, before significant progress is possible.
- (2) Mathematicians have been unable to articulate convincing reasons for their objections: they typically follow current mathematical norms in focusing on technical issues and this simply does not make contact with discourse in the education community. The differences therefore appear to be a matter of opinion, and since they conflict with a coherent and well-established philosophy they are rejected.
- (3) The natural-selection pressures that drove the mathematical transformation do not apply. In mathematics the new methodology offered substantially more power, and this was clear enough to the next generation that they abandoned the old. Modern mathematical methodologies do not offer immediate advantages to the next generation of educators.
- (4) Mathematics education has become ideological and politicized, particularly in the US. Change is now a matter of winning or losing rather than improving, and the educational establishment is far more powerful politically than the mathematical community.

A principal objective of this essay is to develop a conceptual basis for educational change. In particular, my description of the nature of contemporary mathematics in earlier sections is an effort to address (1) and (2).

Items (3) and (4) show that change cannot happen through high-level discourse so a conceptual description alone is not sufficient. Concrete and detailed suggestions might avoid political problems and give selective advantage some purchase. The remainder of this section, [63], and a number of other essays on [64] are attempts in this direction.

Finally, (4) means that some changes are so firmly believed to be wrong that even concrete suggestions will be rejected. These include descriptions of ways technology could—or will not—support learning, and the use of definitions and ideas from this essay in bits of elementary education. It seems these must be approached on a very primitive level: products have to be developed and offered for use without

explanations that would trigger resistance. This is the motivation for my problem-list project [67] and the EduT_EX educational software project [68].

8.2. Informal Student Proofs. According to §4 a *potential proof* is an argument that uses accepted mathematical methods and is presented in sufficient detail to be checked for errors. This becomes a proof once it *has been* appropriately checked and found to be free of errors. Remarkably, proofs and potential proofs in this sense are very common at all levels: a teacher’s instruction to “show work”, for instance, is very close to “give a proof”. I expand on this below.

This should mean that being more conscious and systematic about showing work should give students at all levels access to the procedural and cognitive benefits of proof. Good practice should include:

- write intermediate expressions correctly and carefully enough that someone else can read it;
- write steps in order or clearly indicate the order, for instance by numbering;
- explain notation used to formulate a problem (especially word problems).

The goal is work that should be reasonably easy for someone to follow and check for errors.

It is important to distinguish between checkable and “beautiful”. Crossed-out mistakes are fine as long as it is clear what is in and what is out. Students should not worry about beauty while they are trying to do math, and should not be expected to copy work over unless it is genuinely hard to decipher.

8.3. Focus on Diagnosis. The thesis of this article is that the reliability associated with mathematics is obtained by making mathematical arguments that can be checked for errors, *checking them*, and correcting any errors found. The previous section describes how checkable arguments could become a routine part of mathematics education. However they won’t produce benefits unless *checking* also becomes a routine part. To be explicit: diagnosis and error correction should be key focuses in mathematics education.

- Answers are important mainly as proxies for the work done. Incorrect answers indicate a need for diagnosis and correction and ideally *every* problem with a wrong answer really ought to be diagnosed and corrected.
- Mathematics uniquely enables quality so the emphasis should be on quality, not quantity. In other words, doing fewer problems to enable spending more time on errors is a good tradeoff.
- An important objective is to teach students to routinely diagnose their own work. The fact that diagnosis is possible and effective is the essence of mathematics, so teaching self-diagnosis is mathematics education in the purest sense.

Ideally teachers would regularly go through students’ work with them so they can see the checking process in action. Students should be required to redo problems when the work is hard to check, not just when the answer is wrong. I note that the goal is to establish work habits that will benefit students, but students respond to feedback from teachers, not long-term goals.

8.4. Complicated Problems, Formal Proofs. Formal proofs differ from informal versions in requiring explicit mention of main steps and perhaps use of a structured format [36, 45]. The purpose is still to make it possible to find errors: formality is required by complicated problems and sneaky errors, not a new purpose.

Formal proofs as an organizational tool could be introduced when first needed and then used again from time-to-time as appropriate. For example integer fractions are probably the first significant conceptual structure students have to deal with explicitly. A version of formal proof might be used until they develop enough facility to work accurately without aid. Formal structures could be used again later when a more abstract and symbolic treatment of fractions is given, see §??.

As a general point, many conceptual structures and subtle procedures must be mastered before mathematics becomes useful in a serious sense, and each of these has a learning curve. If we provide good tools—such as formal proof—for developing conceptual mastery, students will go further, more easily.

8.5. Teacher Preparation. I have two suggestions for preparation of teachers of elementary mathematics. First, ensure they have a mathematically correct understanding of the material, and second, develop an understanding of the way mathematics works, but *not* through lots of mathematics.

8.5.1. *Elementary Mathematics from an Advanced Viewpoint.* The section title comes from the important and influential book by Felix Klein, c.f. [33] though I may interpret it in a slightly different way.

Learning mathematics divides roughly into two parts: developing intuitive (subconscious) models of mathematical objects, and learning to consciously work with them, see §5.3.1. In practice models developed by elementary students are frequently inadequate for work at their current level and almost never good enough to use at the next level, so in practice there is a third part: identifying problems with internal models and retraining the subconscious to fix them. This third part is intellectually and psychologically difficult and probably accounts for much of the difficulty students have with mathematics.

I believe the greatest help teachers can give their students is by carefully—often subliminally—guiding them toward effective internal models, and by diagnosing and correcting problems before they get too deeply embedded. However for this to be possible the teacher must have very accurate and well-established models, and these should be informed by an Advanced Viewpoint.

It would also help if teachers used contemporary approaches to concept formation, §5.1.3, ??.

The essay *Proof Projects for Teachers* [62] was written to try out these ideas. It gives perspectives on fractions that will probably be new and useful to even the most seasoned teacher.

8.5.2. *Understanding How Mathematics Works.* To preface the suggestion, the almost universal feeling in the mathematical community is that teachers must know considerably more mathematics before good outcomes can be expected. The real issue is probably not content, but the habits and mindsets mathematicians develop while learning content. Things like understanding the need for skepticism and rigor, or appreciation of the balanced beauty and rich content of a good definition, are acquired through intense study as mathematicians adapt themselves to the subject, see §7.1.

Since these things are not learned explicitly they tend to be visceral rather than conscious. To determine if something is “mathematical” mathematicians consult their gut feelings: if the gut feelings are good the math is good, otherwise something is wrong.

I suspect that when mathematicians say “teachers need to know more math” they probably mean “my guts are unhappy about what they are doing. If they had studied enough mathematics to have effectively internalized its nature then their gut feelings would keep them from doing that.” They identify insufficient content as the problem because they themselves acquired the mindset through extensive work with content.

It is unrealistic and inappropriate to expect teachers to go through professional mathematical preparation. The mindset, however, might be developed using less-ambitious work together with explicit descriptions of how it illustrates general features. The essay [62] is an attempt to flesh out this idea. It is a draft of a course for teachers, showing modern methodology (formal definitions, abstractions and proofs) at work in the contexts of fractions and area of polygons.

8.6. Applications and Word Problems. Word problems have two components: a physical-world formulation and a mathematical model. According to the analysis presented here the physical formulation is *not* part of the mathematics.

Material that is not strictly mathematical has an important role to play in mathematics courses. However it should not be presented as mathematics and student should be taught to distinguish between the two. In particular students should be taught to separate translation of a word problem to a mathematical model from analysis of the model. One reason is intellectual honesty: these are different activities with different rules. However I find the cognitive reasons just as compelling, and see the abstract considerations almost as guidance on where to draw the line.

- Translating a word problem to a form suitable for mathematical processing is a different cognitive process than doing the processing. Separating the two reduces cognitive interference and errors.
- Models are abstractions that frequently apply to many different physical situations. Cutting a string and sharing out a pitcher of juice can both be modeled by the same division operation. Connecting mathematics too closely to physical illustrations obscures this and can impede later learning.
- In genuine real life there are always mismatches between models and the situations they are supposed to model. This is a modeling problem, not a mathematics problem.

Word problems can engage students. Word problems with a reality/model disconnect offer further opportunities for engagement even as they promote valuable long-term perspective and work habits.

8.6.1. Sample Problem. Bubba has a still that produces 700 gallons of alcohol per week. If the tax on alcohol is \$1.50 per gallon, how much tax will Bubba pay in a month? [Set up and analyze a model, then discuss applicability of the model.]

I have given an example with obvious cultural bias because I am not sure I could successfully avoid it. At any rate students in my area in rural Virginia would think this problem is hilarious. We have a long tradition of illegal distilleries and they would know that Bubba has no intention of ever paying any tax.

8.7. Counterproductive Practices. My experiences have convinced me that some educational practices are genuinely counterproductive. The perspective in this article can help identify reasons they are counterproductive and suggest ways to avoid some problems. However these were identified as problems through direct experience with students, not by abstract deduction from the perspective.

I focus on two examples in which disconnects from the nature of mathematics are easy to see.

8.7.1. *Calculators.* I am convinced that a computational environment for mathematics education is urgently needed [65]. Unfortunately, calculators as they are currently used are not satisfactory.

- Calculators do not record steps so calculator work cannot be diagnosed for errors, see §??.
- Calculators seem not to make contact with important modes of learning [61]. This contributes to deficiencies in symbolic skills and internalization of qualitative geometric structure that are serious disadvantages when students reach the university level.
- Calculators encourage thinking in terms of algorithmic calculation rather than logical structure and algebraic expressions [66].
- Calculators encourage focus on numerical work, contributing to deficiencies in symbolic skills.
- Calculators even reduce number sense, as I explain below.

Numbers have structure, and the ability to recognize and exploit this structure is the number sense needed at the university level. A basic structure encoded in our notation is that multiplication by powers of ten corresponds to shifting the location of the decimal point. Traditional students are well aware of this because it is a big time-saver, and use it automatically. However I often see students using calculators to multiply by powers of ten. Colleagues tell me they have seen students use a calculator to multiply by one.

To calculator students, all numbers are the same and their automatic response to a multiplication problem is to start pressing keys. Scanning for special cases requires a different mode of thinking, more attention, and increases exposure to errors. Even decimal-point shifting does not save time. It might save time when it works, but scanning for the opportunity to use it is not automatic for these students and the additional time and attention required for routine scanning would cause a net time loss.

Another example lies on the boundary between number sense and symbolic skills. When presented with $(56 \times 7 \times 233)/7$ traditional students will generally see the opportunity to cancel the two sevens. Many scan for things like this automatically because it is a big time-saver. Calculator students generally do all the arithmetic. Scanning for cancellations requires more attention and saves little if any time.

Canceling rarely makes much difference when dealing with numbers but becomes important with expressions such as $(56 \times a \times 233)/a$. Many traditional students are already scanning for cancellations and in this mode are dealing with numbers as symbols anyway so quickly adapt to literal symbols. Calculator students aren't scanning, and their number skills—mostly concerned with keystrokes—don't transfer to symbols.

To repeat, a good computational environment for students is necessary, even urgent. However current calculators do not provide a good environment and improvement seems unlikely until their drawbacks are acknowledged and addressed. An attempt to imagine a good environment, shaped mostly by the need to avoid calculator problems, is described in [65].

8.7.2. *Euclidean Geometry.* I firmly believe that geometry is an essential part of mathematics education. However traditional Euclidean geometry is not satisfactory, and in particular it is unsatisfactory as the principal setting for proofs.

Section 2.2.10 describes ways in which Euclidean methods fail to be error-displaying and are correspondingly nonmathematical. The fact that it is essentially proof-by-example is particularly problematic. Section ?? explains how the conceptual organization is out of step with current practice and is close to dysfunctional. On top of all this there are educational shortcomings:

- In later work students will need coordinate geometry with parametric lines, function graphs, etc. The Euclidean approach of excluding coordinates and taking points and lines as primitive objects connects poorly with this.
- Students certainly need exposure to proofs. However they will need these more-careful ways of thinking in *all* mathematical work, not just a limited context. Euclidean methods do not provide good models for work in other courses even if they are upgraded to actually be proofs.
- Coordinate plane geometry is pretty obviously a baby version of three-dimensional geometry and most of the methods extend directly to higher dimensions. Euclidean plane geometry is very tied to the plane, and these days that is an uncomfortably small box.

In sum, the primitive Euclidean approach to geometry has become a barrier that hinders access to other dimensions and other mathematical worlds. These worlds are too rich and too easily accessible to modern approaches for this to be acceptable.

9. HISTORY: THE MODERNIST TRANSFORMATION AND LATER

Modern mathematical practice dates from the early twentieth century, and differs in substantial ways from earlier practice. In this section I describe the transition, contrast practice before and after, and explain what drove the change. Jeremy Gray has a better view than most, and the discussion here owes much to his analysis, particularly his book *Plato's Ghost* [21]. The term “modernist transformation” is taken from this book, and the book is discussed briefly in §9.4.

I also describe how mathematics was influenced by events later in the twentieth century, and speculate on some upcoming challenges.

It is well-known that something was going on in mathematics in this period (roughly 1890–1930). The general belief seems to be that in the end nothing significant happened, and the mathematics that emerged was essentially the same as it always was. This is a major puzzle to me: the mathematics I know and describe in this essay seems nearly a different discipline. Therefore an objective in this section and the next is to look for explanations for this lack of understanding.

9.1. **Sketch of the Transformation.** Up through the early nineteenth century mathematics was thought of as an idealization of the physical world. There were no “mathematical objects” per se, just abstractions or idealizations of features of physical experience. Such objects could be specified by example or by reference to

the physical world, and what we now think of as mathematically-precise definitions were not considered necessary. Similarly, it was believed that reasoning about such objects could safely depend on physical intuition.

All this changed. More precisely, mathematical practice changed so this description is no longer valid even though it remains the dominant view outside the profession. I describe the change in terms of stages; the reality was obviously not so tidy.

9.1.1. *Stage One: Objects.* The first stage in the transformation was the development of definitions in the modern sense: concise and *definitive* formulations rather than examples or explanations intended to bring to mind or refine some feature of physical experience. This seems to have taken place mainly in Germany in the mid nineteenth century, led by Dirichlet, Dedekind, and Riemann in Göttingen.

Development of precise definitions seems to have been driven by the need to investigate objects that were so subtle and far removed from previous experience that they could only be understood this way. A good example is provided by the work of Kummer on Fermat’s “Last Theorem”, see [2] volume III p. 345. In the middle of the nineteenth century Kummer found a proof, assuming uniqueness of prime factorization in the cyclotomic integers. Uniqueness is a key tool in the ordinary integers, and consequently deeply embedded in intuition developed there. Unfortunately it fails in cyclotomic integers. Moreover, to investigate factorizations in this context one must work with prime *ideals*, not just primes in the ring. This is a complicated and subtle business, and the chances of getting it right without precise definitions are very near zero. Using these tools, however, Kummer, Dedekind and others went on to do deep work in the subject.

Precise definitions also enable identification and rapid development of new mathematical concepts, so a consequence of the new methodology was a profusion of new—or newly useable—concepts. Put another way, imprecise definitions put strong limits on complexity (in the sense of §6.2) of mathematical development. The development of precise definitions made cognitively-complex subjects more accessible, and the power of the innovation caused it to spread relatively quickly to parts of the community interested in these subjects.

Precise definitions were not universally accepted. They gave little advantage in areas of low cognitive complexity²⁴, and practitioners in these areas tended to denigrate the movement. Another barrier was the apparent lack of connection to physical reality. Philosophers and many traditional and applied mathematicians saw the new material as abstract and frivolous, done apparently “for its own sake”. However definitions alone were not seriously problematic: people who needed them used them, and people who didn’t need them could ignore them.

This sketch depends on hindsight and inference. Tracing out the actual development in the historical record would give a much richer and possibly different story. I think this is an important issue: I have come to see modern approaches to concept formation as just as significant, particularly for education, as modern approaches to proof.

9.1.2. *Stage Two: Rigorous Proofs.* The second stage in the transformation centered on more-precise ways of working, particularly proofs as described in §4.

²⁴Recall that cognitive complexity is not the same as difficulty, §6.1. There are very hard subjects with low complexity, see §10.2.

Modern proofs both require precise definitions, and effectively exploit their precision. For this reason proofs came most easily to the new areas—most notably “abstract algebra”—made accessible by the use of precise definitions.

The situation in analysis was more complicated and better illustrates the general situation. Modern rigor is usually considered to have its roots in the work of Cauchy, Bolzano, Gauss and others sorting out problems with convergence. They did make progress, but it was limited for quite some time by the lack of a precise definition of convergence and, more fundamentally, a precise description of the real numbers. This latter problem was recognized by Bolzano by 1817 and he and others began an effort to describe the real numbers in terms of the integers. The effort was brought to fruition by Weierstrauss, Dedekind, Cantor and several others around 1872 (c.f. Boyer, [8], §25). This development put full rigor within reach in analysis, and in the twentieth century it would be considered an essential basic part of mathematics, but it was not universally welcomed.

The first problem was that this effort revealed that the integers themselves were not completely precisely understood. Cantor, Frege and others set out to fix this with logic and set theory, and their efforts were plagued by even deeper problems. One might have thought Peano’s axiomatic description of the natural numbers in 1889 would have brought some clarity, but it was one of the factors that launched the main 1890–1930 transformation: it was very effective technically, and very unsatisfying philosophically (see the discussion in the next section). The outcome was that the description of the reals using the integers did in fact provide a solid base for analysis, but it took nearly sixty years for the deeper foundational crisis to work itself out so people could feel completely comfortable with it.

The second problem was that some leading analysts and geometers were hostile to this development. Felix Klein referred to it derisively as the “arithmetization” of analysis and continued to oppose it well into the twentieth century. Poincaré described it as a disease that needed a cure. Hermann Weyl described it in 1918 as “in an essential part built on sand.” He evidently felt that the foundational problems made “arithmetization” less reliable than his own intuition.

The consequence was that analysis, the ostensible nursery of modern rigor, was actually a bit tardy in becoming fully modern.

9.1.3. *Stage Three: The Transformation.* According to Gray, who mapped it out in some detail in [21], the transformation became a major issue around 1890 and was pretty much over by 1930. Göttingen was again at the center of the development, this time led by Hilbert. This section gives an outline from the perspective of this essay; §9.4 describes the connection to Gray’s work in more detail.

That this was a time of turmoil is well-known even to those who do not see that anything came of it. It seems to me that the turmoil had three main factors:

- The need for uniformity of practice, and particularly the rejection of the use of intuitive conclusions in proofs;
- The rejection of philosophy and physical reality as the main sources of meaning; and
- The foundational crisis.

First, uniformity of practice. Precise definitions and modern proofs can bring substantial benefits to a mathematical community, but only if *everyone uses them*.

Work that is not scrupulously careful is not fully reliable and potentially dangerous. Consequently, once a nucleus of fully-modern material was developed the practitioners began to exclude old-fashioned work, describing it as incomplete or preliminary. The elite-practitioner syndrome described in §2.3.6 and 7, where leaders with excellent intuitions want to rely on these intuitions and not bother with fiddly details, came into play with full force. In some areas, e.g. analysis and geometry with Klein, Poincaré and Brower, completing the transformation was literally a matter of waiting for the older generation to fade away.

Second, the distancing from philosophy and physical reality. Self-contained definitions and high-precision methods do not depend on either philosophical or physical guidance. Students trained in the new methods found meaning, insight, and intuition directly in the mathematics, and the mathematical meaning and intuition frequently proved to be deeper and more powerful.

I illustrate this with the Peano axioms for the natural numbers. In this formulation the natural numbers are a set with a “successor” function that moves everything up one place. “1” is the unique element that is not the successor of anything. “2” is the successor of 1, and so on. Philosophical problems with this include:

- (1) This describes *models* for the natural numbers, not *the* natural numbers. Any set of the appropriate size has a successor function so gives such a model. In particular every model has a different “1” and “2” so this does not provide a “coordinate-free” definition of numbers.
- (2) The connection to counting real things is indirect and derived rather than primary.

The modern attitude toward (1) is that it doesn’t matter if each mathematician has his own private model: they are all canonically equivalent so one as good as any other. And no, you should not try to get a universal model by taking equivalence classes because this leads to ugly and irrelevant set-theoretical issues²⁵. The point in (2) is that counting is an *application*, not a definition. Trying to turn counting into a definition leads to a dysfunctional mess (compare Peano with Russell-Whitehead).

Returning to the items on the list, “separation from reality” means, as in (2) just above, taking the mathematical world on its own terms as primary. Clues, goals and inspiration still frequently come from physical reality, but these lead us to mathematics rather than define it. This attitude is often described as “formal and meaningless”, “divorced from reality”, etc. These can indeed be problems, but the truth is that taking mathematics as primary gives power and simplicity, and eventually leads to deeper insights and better applications in the physical world.

Finally, the foundational crisis. This was resolved in a technically effective but philosophically unsatisfactory way. First it developed that “higher mathematics” is robust in the sense that it depends on the existence of *some* set theory, but not on troublesome details of any particular set theory. This is nicely explained by Mazur in [37] using a categorical point of view; again see §10.1. For example Frege’s attempt to derive the integers from set theory was generally thought (e.g. by Frege) to have been severely damaged by Russell’s set-formation paradox. In fact this is

²⁵See Mazur [37] for a similar discussion, and §10.1 for more about [37].

a “troublesome detail” that can be resolved in a number of ways, and none of them effect the final result²⁶.

The second peculiarity of the resolution was the ‘experimental’ basis described in §2.3.2 and 2.3.4. The philosophical goal of *proving* that mathematics is consistent was never realized. Instead a somewhat threadbare set theory (Zermelo–Fraenkel) was developed that was good enough to support mathematics and has (so far) resisted all attacks. This was philosophically unsatisfying but this and robustness were good enough for mathematicians.

The outcome of all this was that mathematics emerged from the transformation powerful and ready to undertake the vast developments of the twentieth century. The connection with philosophy had been lost, however, and the relationship with the physical world had been weakened and became in some ways problematic.

9.2. Confusion, Obscurity, and Philosophy. Implicit in this story is the fact that philosophers went from being the primary guides to methodology and meaning in mathematics, to objects of ridicule, in less than forty years. How did this happen, and does it matter?

9.2.1. *Left Behind.* The mathematical side of the story is largely told in the previous section. In brief, the high–precision methodologies developed in the mid to late nineteenth century worked very well—better, in fact—without philosophical input. Philosophers seemed not to notice this. They were preoccupied with new insights into antiquity, grappling with the significance of non–Euclidean geometry, or the beginning of the foundational crisis. Furthermore the new development took place in technically–complex areas largely inaccessible to philosophers.

During the 1890–1930 transformation philosophers concerned with current mathematics were largely focused on the foundational crisis. By their standards the efforts to resolve it were failures. For example Frege’s program for arithmetic and Hilbert’s algorithmic approach to foundations are both thought of as failures: Frege brought down by Russell, and Hilbert by Gödel. Further, much is made of the disagreements between the two (see Blanchette, [45]). It seems to me, however, that the “fatal flaw” description had more to do with philosophical interpretation of the significance of the difficulty than with technical issues. To me the problems seem to be the kind one expects in any ambitious program: requiring adjustment of goals or techniques but not requiring starting over.

In fact these “failed” programs were both instrumental in developing contemporary methodology, and Hilbert’s was particularly successful. It seems to me that the major implicit goal of his program was a mathematical meaning for the word “true”. Philosophers had controlled this word for millennia and burdened it with connections to meaning and knowledge that encouraged word games and led to paralysis.

For example if an axiom system has nothing to do with “reality” then more-or-less by definition (according to philosophers) we cannot really “know” anything about it, it cannot have “meaning”, and “true” or “false” don’t even apply. This is extremely confining. I certainly “know” and find “meaning” in lots of things that as far as I can tell have no connection to *physical* reality. I see them as part of a larger reality, possibly because my brain was rewired by a hundred thousand

²⁶See Burgess [9] and Heck [24] for a resolution that would seem to have been easily accessible to Frege. The puzzle is why he did not use it.

hours spent peering into the void. However I would *not* want to argue that because they seem real *to me* they must have some validity. This is playing the game by philosophical rules and leads to a worthless user-dependent meaning for “true”.

Hilbert’s proposal was to define “mathematically true” as “algorithmically decidable from the axioms”, and essentially refuse to address the “meaningfulness” of the axioms. Gödel proved, with fully-modern rigor, that in usefully complex axiom systems there had to be statements that are “true” but that cannot be proved in this sense. This refuted the technical statement but it seems to me that it accomplished the larger goal: it established Gödel’s technical interpretation of “true” as legitimate for mathematical use. Philosophers accepted this as part of the price of having Hilbert refuted, and did not realize that it cut them out of the game²⁷.

The outcome was that mathematics emerged from the transformation free from philosophy, and apparently much better off without it.

To paraphrase an old joke, philosophers declared the operation a failure but the patient emerged much healthier. Some philosophers have spent the last eighty years or so trying to convince the patient he is still sick (e.g. with strange and scary variants on logic or set theory), only to be ignored or met with derision. Others still go through the motions of describing mathematics but as far as I can tell very few have been able to (or even feel the need to) penetrate the technical thickets of modern practice far enough to see what it is really about, or even that it changed in the last century. They seem unaware that the patient left and isn’t coming back.

9.2.2. *It Does Matter.* The separation of mathematics and philosophy has unfortunate aspects. We don’t need someone to tell us what we *should* do, but we could use help in understanding and publicizing what we *are* doing.

There were vigorous debates during the transformation and—unsurprisingly—the philosophical arguments were far better articulated. In particular the charges that a separation from physical reality deprived mathematics of its meaning, and that reining in the use of intuition reduced it to sterile symbol manipulation, were quite powerful. These arguments remain unrefuted on a conceptual level and the real nature of contemporary mathematics remains obscure. This causes problems, and will continue to so until an articulate and convincing account is given.

The educational community, for instance, still finds the arguments against contemporary mathematics compelling, and still follows the nineteenth-century model, see §8.

Mathematicians have little objective awareness of how they work, for good reason: it interferes with the work itself (see §7). Some still subscribe to the old philosophical description even though it is almost unrelated to how they actually work, see e.g. §5.2.3. Nearly all emphasize intuition and physical connections in non-technical descriptions of their work, and this perpetuates the confusion (see §10.3). In short it seems unlikely that mathematicians themselves will sort this out.

What we need is a “science of mathematics” community that could objectively study the field, as it really is, and explain it both to us and to the world. David Corfield makes a related observation in [11]:

²⁷It seems clear that philosophers (and Gödel himself, see [21], §7.3.8) intended this interpretation of “true” to apply only in a limited technical context, and on higher levels intended truth to remain connected to meaning. However, as one quickly discovers when learning proofs, intentions have no force in mathematics. Once a technical interpretation was on the table the clarity and power it offered made it irresistible.

... debates about the proper ordering of the goods of mathematics should be organised, not just allowed to happen in unsatisfying ways, as outbursts and releases of tension. The Jaffe-Quinn debate was the most public of these outbursts in recent years (Jaffe & Quinn 1983). There was no role for a philosopher in this debate, which, I like to think, was the poorer for it.

As one of the principals in this debate I would like to go further: this was a job for a philosopher, not a mathematician. But there was no role for a *contemporary* philosopher because they have no understanding of *contemporary* mathematical issues. Their attempts to contribute just displayed the inadequacy of their understanding; see §10.5, 10.4 for discussion. Philosophers would have to abandon most of their rich heritage as irrelevant before progress would be possible, so help from that quarter is unlikely.

9.3. Later Developments. The internal discord that marked the transformation mostly ended about 1930, but later events effected the structure of the community just as profoundly and must be understood to fully understand the current situation. In this section I offer a very broad outline. Significant details are missing and the picture may change as these are filled in.

9.3.1. *The Second World War.* An early casualty, even before shots were fired, was the German mathematical community. Göttingen, the cradle of modern methodology, was gutted. Mathematics essentially came to a halt in the rest of continental Europe when the war started, and many of the most talented left, often for the United States. The story is well-known and does not need to be repeated here.

Long-term consequences included:

- The United States became the primary center for mathematics for the rest of the century.
- The mathematical community was homogenized and became genuinely international (mostly; see below). The days of parallel English, French, and German development with national groups either ignorant of or deliberately ignoring the others were over (mostly), and norms reflecting the new standards of precision became universal (mostly).
- English became the standard language for mathematical communication, and eventually the language of record for publication. At first this was a consequence of war damage, but the advantages were so great that the practice persisted even after other national communities were reconstituted.
- Applied mathematics became a serious subject of major importance.

The French community retained enough coherence to attempt linguistic independence, and enough idiosyncrasy to develop the Bourbaki movement, but even they were largely internationalized by the end of the century.

Japan developed a substantial mathematical community in the last half of the century. It was well-connected to the international English-based community, but the mechanism may not be the obvious one. After the war ambitious and talented students from all over the world came to the U.S., and later Europe, for training. Most of them stayed. It may be that changes during the U.S. occupation made returning to Japan more attractive, and significant numbers of returning mathematicians well-versed in contemporary methods set standards and direction for the community.

9.3.2. *The Soviet Union.* The only major mathematical community that did not participate in the postwar homogenization was in the Soviet Union. This community was largely isolated from developments in Europe in the early twentieth century as well, so was not fully converted to the new methodologies.

In the description of §9.1 they had gone through Stage One (precise definitions, §9.1.1) but not Stage Two (fully rigorous proofs, §9.1.2). Great emphasis was put on intuition and “big ideas”, and intuitive leaps were accepted in published arguments. One result was an uncomfortably unreliable literature that the international community protected itself from by largely ignoring much of it.

[[this section to be expanded]]

The existence of a large and powerful but untransformed mathematical community concerned with core issues was a potential threat to the stability of the modern international community. This was unexpectedly resolved in the early 1990s by the dissolution of the Soviet Union and the consequent collapse and dispersal of the community. There are residual groups large enough to maintain some of the customs (Moscow, Saint Petersburg, and perhaps IHES in Bures-sur-Yvette, France), but they are reduced enough that they should eventually be assimilated into the international community.

9.3.3. *Summary.* In this essay I have repeatedly emphasized that real success with contemporary mathematical methods depends heavily on a reliable literature and high and reasonably uniform standards of rigor. In §9.3.1 above I suggested that homogenization and internationalization after the second world war accelerated the spread and acceptance of these standards in the major communities outside the Soviet Union. In §9.3.2 above I suggested that problems with the Soviet community were defused by its collapse following the dissolution of the Soviet Union.

Putting these points together suggests that the spectacular growth of mathematics in the second half of the twentieth century, as well as its current state of health, owe much to political events that we all hope will not be repeated. The discipline will face other challenges and it would be better to do so consciously and thoughtfully than rely on cataclysms to fix them for us.

One such challenge is the relationship with applied mathematics discussed in §??. Another is the development of new strong and relatively isolated national communities, in China in particular. It is almost inevitable that the Chinese–language literature will be problematic enough that the international community will have to ignore most of it, much like the twentieth–century Russian–language literature. However the Chinese are publishing a great deal more in English than the Russians did, and in my area of geometric topology there have already been problems, [41]. If they are unable to maintain good standards then the international community may, as a matter of self–preservation, discount their English–language literature as well.

9.4. **Gray’s History.** Gray [21] describes the 1890–1930 transformation in great detail, including the issues at stake; arguments, as they developed over time; the convictions and contributions of dozens of participants; social context; and many other factors. He demonstrates convincingly that something happened, but does not have a good explanation *why* it happened.

This essay provides an explanatory framework for Gray’s observations. Roughly speaking the transition was a Darwinian struggle between two ways of working: one

depending more on philosophy and intuition, and the other on technical definitions and rigorous formal argument. The latter provides substantially more power in complicated or delicate situations, so increased the “fitness” of its users. Natural selection determined the outcome. This explanation is implicit in the sketch in 9.1.3 that describes the transformation as the third stage of a larger development.

This section provides more detail about how this essay and Gray’s work fit together, and will be mainly useful to those who actually read [21].

9.4.1. *Influences.* Gray describes a number of factors that influenced the transformation but rejects them as causes. The “Darwinian struggle” viewpoint provides a context in which to understand these influences. It does not reduce their validity or significance. Including them in the picture enriches it and deepens understanding, but here I can only list the issues.

Gray’s observed influences include:

- The old sociological standby, a power struggle between groups. Certainly there were elements of this, and at the end there were winners and losers, but it doesn’t explain change of this magnitude in something as well-developed as mathematics. How could Hilbert overcome Poincaré, Brouwer, Klein, and other giants of the late nineteenth and early twentieth centuries? Particularly if Hilbert’s explicitly articulated program “collapsed”?
- The mysterious force that also drove the modernist movements sweeping through art, music and literature at roughly the same time. There are structural similarities that aid analysis: his somewhat unfocused working definition of “modernist” ([21], §1.1.1) reflects this, and the term “modernist” comes from this. However he rejects it as an explanation, and eventually seems dissatisfied with the definition ([21], §7.4).
- The political turmoil of the time. Gray credits Herbert Mehtens [40] with the first focused discussion of the transition. Mehtens was mainly concerned with developments in Germany, where in fact most of the action took place. Unfortunately that picture is clouded by the political destruction of the German mathematical community shortly after the transition. This leads Mehtens to suggest a connection between politics and the technicalities of mathematics, but Gray does not find support for this in the larger picture.
- The growing importance of applications, and the professionalization of the community. Again these play roles but do not explain the changes. Why, for instance, would interest in applications favor formal argument over intuition? A professional community would stabilize the results of useful change, but why were these useful?

9.4.2. *Anxiety About Error.* Gray makes a number of observations for which he has no explanation but that make sense in our framework. I give one example; others await those who read the book.

Gray observes in [21] §4.8 (“Anxiety”):

What has not been sufficiently discussed by historians of mathematics is the note, hesitant at first but growing to a crescendo around 1900–1914, of anxiety. The mathematics of the nineteenth century is marked by a growing appreciation of error. For although

mathematicians—with some notable exceptions—have traditionally had a low tolerance of errors, during the nineteenth century the awareness of errors grew and became a source of anxiety. [...] Indeed, once the safe havens of traditional mathematical assumptions were found to be inadequate, mathematicians began a journey that was not to end in security, but in exhaustion, and a new prudence about what mathematics is and can provide.

The perspective of §2 and particularly §2.2.8 (Mathematics vs. Science) sharpens this: error is part of science, and nineteenth century mathematics was a science. The position taken by Poincaré, Klein and others can be paraphrased as: it is the role of intuition and the connection to physical reality that makes mathematics a science, and by extension meaningful. This brings with it exposure to error, but that is simply a cost of being a science and meaningful.

The real oddity is not that there were errors but that some mathematicians began to believe they could be avoided. They saw that precise definitions and rigorous proofs reduced the error rate, and more precision meant fewer errors. Increased anxiety about the remaining errors was a consequence of the growing possibility of, and need for, error-free conclusions. In the terminology of §2.1, the possibility of complete reliability. In other words, anxiety was a symptom of the evolution of mathematics from a science to something completely different: something with a strange and physically unreal domain, but with unprecedented power within that domain.

The objections of Poincaré and Klein to the new methodology were essentially that they amounted to the end of mathematics as a science. They were right, but wrong about the consequences.

9.4.3. *Conclusion.* Gray's account weakens, as historical accounts tend to, when it turns from description to analysis. In §7 he provides a clear and insightful description of some of the unclear and inconclusive philosophical discussion that followed the transformation. He does not see much support for any of these themes in his own work, but rather than conclude—as I do—that none of them had much relation to actual practice, he is respectfully neutral and the situation remains murky.

In §7.4, “Did Modernism ‘Win’?”, Gray tries to identify “modernists” in the post-transition period, with very mixed results. The working definition of §1.1.1 that served well enough as a vague organizing principle is not up to this challenge, and in Britain for instance it singles out the modernizers rather than those who were modernized. He recognizes the insufficiency and tries to use the philosophical discussion to refine the definition. He does not, however, achieve a clear formulation, and does not show that the refinement is more effective at meeting the various challenges.

Gray's final section (§7.5 “The Work is Done”) concerns the philosophical summary more than the rich historical account, so is essentially inconclusive. The Yeats poem from which the title of the book is drawn ends with:

“The work is done”, grown old he thought,
 “According to my boyish plan;
 Let the fools rage, I swerved in naught,
 Something to perfection brought”;

But louder sang that ghost, [Plato] “What then?”

The very title of the book (“Plato’s Ghost”) suggests dissatisfaction with the outcome.

A clear central principle strongly related to the historical story would have permitted the triumphant finale I feel the work deserves. It seems to me that the “quest for complete reliability” theme developed here provides such a principle.

My respect for Gray’s achievement is increased rather than diminished by the fact that he did not have a good explanation. A reasonably good conceptual framework makes identifying and describing something like this transformation straightforward. Doing it without the guidance of a framework requires scholarship and wisdom.

10. DESCRIPTIONS: OTHER ACCOUNTS OF MATHEMATICS

In this section I briefly discuss a few other accounts of mathematics. One objective is to try to understand why such accounts are so diverse and mostly—it seems to me—irrelevant when they all ostensibly concern the same thing. The mainstream philosophy of mathematics literature seems particularly irrelevant, and the reasons shallow and uninteresting, so only two are considered here. Essays by people with significant mathematical background often have useful insights, and when they seem off-base to me the reasons are revealing. The essay by Mazur is not off-base.

10.1. Barry Mazur, ‘When is One Thing Equal to Some Other Thing?’

This essay begins:

One can’t do mathematics for more than ten minutes without grappling, in some way or other, with the slippery notion of *equality*.

This leads to a beautiful riff on equality, equivalence, isomorphism, canonical isomorphism, and the ways mathematicians have found to organize it all. The language is casual, clear, and—thankfully—non-philosophical. There are quotations from Plato and Aristotle to suggest that some of the ideas have roots in their work. This is a bit like finding significance in horoscopes, but is too charming to be objectionable.

The heart of the article is a clear description of how category theory explains why so much of “higher mathematics” is robustly independent of the troublesome details of set theory. Category theory requires an underlying set theory, but it is sensitive only to certain basic features. Apparently almost any set theory will do, a situation Mazur describes as “bring-your-own”. The basic reason is that most mathematics depends on some form of equivalence rather than actual equality; categories provide a systematic framework for such equivalences; and equivalence is less sensitive than literal equality to set-theory issues.

It is important to remember that “insensitive” is not the same as “unimportant”. For instance construction of functors (Kan extensions, limits, etc.) requires careful attention to actual identity even though the end result may only be defined up to equivalence. This is still a simplification made to keep the picture clear. Here I want to re-introduce some complication to suggest a different moral: not “set theory doesn’t matter”, but “set theory does not really come into play until one goes deeply into certain subjects”.

10.1.1. *The Axiom of Choice.* The first example is the axiom of choice. There are models of set theory in which it fails in general, and category theory does not provide a substitute. There are also many mathematical constructions that depend on it. Genuinely set-theory independent statements of these results read something like “For all spaces small enough that the axiom of choice applies, ...”, rather than “For all spaces, ...”. This quickly becomes very annoying. There are also no known benefits, so in standard practice ‘choice’ is considered part of basic set theory and used without comment.

10.1.2. *Deep Structure of Spectra.* This example relates to a point made in §2.3.3 (Short-Term Success of Erroneous Formulations): it is harmless to believe or ignore something false if you never use it in a way that depends on its falsity. This is used in §2.3.3 to explain why in history we often see things going well for a while but then run into problems.

The point for the present discussion is that set-theoretic details probably seem irrelevant because we rarely do things bold enough to depend on them. I’ll describe a question I’ve wondered about for nearly forty years, related to Mazur’s discussion of representable functors. It is rather long and technical (we have to go through four or five levels of generality before set theory matters), and not particularly relevant to the rest of the essay so can be skipped without loss.

The question concerns applications of “the set of all things”, illustrated here with vector bundles. There is a classifying space B_O and a bundle $E_O \rightarrow B_O$ over it that is universal in the following sense: given *any* vector bundle over *any* space, say $V \rightarrow X$, there is a map $X \rightarrow B_O$ and a bundle isomorphism of V with the pullback of the universal bundle. Moreover this map and isomorphism are well-defined up to equivalence involving homotopy of maps and bundle isomorphisms.

In many ways this is very effective. It gives an equivalence between isomorphism classes of vector bundles over X , usually denoted $K_O(X)$, and homotopy classes of maps $[X, B_O]$. Further, the well-defined aspect of the universal property specifies what the equivalence relations on bundles and maps should be to make this true. In the terminology of Mazur §7.5, $K_O(X)$ is a representable functor of X , and B_O is a representing object.

We now get more ambitious. Given vector bundles $V \rightarrow X$ and $W \rightarrow Y$ we can form the (external) direct sum to get a bundle $V \oplus W \rightarrow X \times Y$. Apply this to two copies of the universal bundle to get $E_O \oplus E_O \rightarrow B_O \times B_O$. Applying the universal property to this bundle gives a map $B_O \times B_O \rightarrow B_O$. We would like to think of this as a group structure on B_O , and in fact an abelian group structure. The reason is that direct sum gives $K_O(X)$ an abelian group structure; a group structure on B_O would give the set of homotopy classes an abelian group structure; and the representing isomorphism $K_O(X) \simeq [X, B_O]$ would then be an isomorphism of abelian groups.

The problem is that the universal property only defines a map $B_O \times B_O \rightarrow B_O$ up to homotopy. There is no chance that it will satisfy the exact identities required for a group structure. At this level this is not a big problem: the group identities are satisfied *up to homotopy*, and this is enough to give a genuine group structure on the set of homotopy classes $[X, B_O]$.

The problem grows at the next level of ambition. We want B_O to be an infinite loop space, not just a group-up-to-homotopy. For this we need something like exact group identities, not just up to homotopy.

To get exact identities we need a construction that gives specific classifying maps, not just up to equivalence. Choose a big set U and consider the vector bundles $V \rightarrow X$ with $V \subset U$. Each fiber over X is a vector space, and a subspace of U . We can get a specific model for B_O by taking “the space of all vector spaces that are subsets of U ”. Then, to $V \subset U$ we get a specific map $X \rightarrow B_O$ by taking a point x to the fiber of V lying over x . Moreover there is a universal bundle $E_O \rightarrow B_O$ whose fiber over a point $[W] \in B_O$ is W itself, and the pullback of this over classifying map $X \rightarrow B_O$ gives the original bundle V *identically*, not just up to isomorphism.

We would like for the identity map of B_O to be the classifying map of the universal bundle. This doesn’t work as it stands because the total space is not a subset of U . We can get around this by relaxing $V \subset U$ to: there is a map $V \rightarrow U$ that is an injection on each fiber.

Note that we have already had to avoid a set–theoretic problem by restricting to subsets of a set U , and this has forced minor modification of the definition, but we do now have specific, well–defined classifying maps. Moreover we are dealing with actual equality of bundles, without the protection of a categorical–equivalence cushion. We are outside the safety zone described by Mazur.

The next step is to form the direct sum of the universal bundle with itself, $E_O \oplus E_O$, and consider the classifying map of that. This sum is a structure on $E_O \times E_O$ so it comes with a map $U \times U$, not a map to U , and does not immediately have a classifying map. We might try to fix this by assuming that U is an abelian group and using the composite $E_O \times E_O \rightarrow U \times U \rightarrow U$. But this won’t always be injective: over the point (x, x) , for instance, we have the fiber $W \times W$ and the product will divide out by the involution that interchanges the factors.

Another approach to the sum is to suppose U is an infinite–dimensional vector space and use “the set of all (finite dimensional) vector subspaces of U ” as our model for B_O . We could then define a sum operation that stays in U by taking $V + W$ to be the linear span of V and W . Again there is a disjointness problem: if these spaces have positive–dimensional intersection then the sum is not a direct sum, with the consequence that adding two bundles may not give a bundle. This is not just a technical detail. “Linear span” is an associative and commutative operation and ignoring disjointness problems seems to give a topological abelian monoid structure on B_O . However it is a theorem of John Moore that a topological abelian monoid has trivial k –invariants. This is false for B_O so being too casual with identity issues would actually lead to false numerical calculations.

I was trying to do constructions like this for my thesis and at this point gave up and used other approaches. There are elaborate homotopy structures that do this particular job (with B_O) with difficulty but without set–theoretic problems.

There are ambitions beyond this, however, that are yet harder. We can take tensor products of vector bundles as well as direct sums, and this should give B_O something like a commutative ring structure. The advantage a commutative ring has over homotopy versions is that one could define tensor products of modules. In the early 1970s I tried to use the approach above and set–theory tricks to construct some useful approximation to such a structure. Again I failed, but the attempt motivated Peter May’s development of E_∞ ring spectra, [38] (see the last paragraph on p. 3). This was satisfactory for some purposes but not others, and extremely complex.

Twenty-five years later *two* useful approximations to commutative ring structures were finally obtained, [15] and [26], via commutative smash products for spectra. These are effective but again technically complex.

We finally get to the question: can the set-theoretic approach be used to get simpler and easier to use constructions of some sort of commutative ring spectra? If so it would probably involve more-sophisticated handling of actual-identity issues, but in standard set theory. I am sure there are deeper and more-ambitious levels beyond this, and it seems possible that a set theory with “designer” features will eventually be useful.

The purpose of this digression was to give a specific example of a significant mathematical area that genuinely involves identity and set-theory issues. Set theory may matter to higher mathematics when goals are sufficiently ambitious.

10.2. Jonathan Borwein, ‘Implications of Experimental Mathematics for the Philosophy of Mathematics’. [[incomplete]]

10.3. Keith Devlin, ‘What Will Count as Mathematics in 2100?’ Keith Devlin is a masterful popularizer of mathematics: I particularly enjoyed his *Mathematics: The New Golden Age*. [13]. He has a sweeping view of mathematical history in terms of topics and conclusions, but seems to be relatively insensitive to methodology and sociological issues. I’ll describe the problem and suggest an explanation.

In [14] he describes the “Göttingen revolution” in the mid nineteenth century as the last major change in mathematics. The shift is

[from] performing a calculation or computing an answer [to] formulating and understanding abstract concepts and relationships. This represented a shift in emphasis from *doing* to *understanding*. [...] Mathematical objects, which had been thought of as given primarily by formulas, came to be viewed rather as carriers of conceptual properties.

In §9.1.1 (stage one of the transformation) I identify this proliferation of conceptual material as a consequence of the development of precise, self-contained definitions. Devlin missed the role of the enabling technology.

Furthermore it is not true that the Göttingen group lost interest in computation. Rather, they discovered that increasingly subtle and difficult computations depended on increasingly subtle use of abstract structure. The ancient practice of imprecise definition and dependence on physical-world intuition was not successful in dealing with these abstractions, but one could be successful if these objects are carefully defined and studied more-or-less on their own terms before attempting concrete applications. Finally, it was exactly studying them on their own terms (rather than reflections of physical reality) that made them “carriers of conceptual properties”.

Next, in describing this as the *last* major shift he misses the significance of the 1890–1930 transformation. He certainly knows about the sociological turmoil in the community at the time, and recognizes the later emphasis on precision and distancing from physical reality. However he sees this as a “passing fad” (see below) and feels that nothing really essential has changed. I see profound methodological changes, why does he not?

Later in [14] he extrapolates into the future:

Whereas the Göttingen revolution changed the *nature* of mathematics, but left it *looking* on the surface much as it always had, I believe the next revolution will leave the fundamental *nature* of mathematics unchanged but will lead to something that *looks* very different on the surface.

The change he expects is a much greater incorporation of conceptual and heuristic material directly in mathematics. He has learned much of what he knows through heuristic explanations that hide a great deal of technical detail. He knows details are missing, but sees them as technical intermediaries for the concepts rather than as the actual mathematics. Eventually, he predicts, we will be able to dispense with the intermediaries and work more directly with concepts and heuristics, and thereby greatly increase the scope of mathematics. He provides a good number of interesting opportunities for such a development.

On the surface his vision is completely at odds with the picture presented here. It is a reasonably good description of actual practice in the nineteenth century, but this was already unsatisfactory for twentieth-century mathematics, and it certainly won't work in the twenty-first. On the other hand the vision is quite similar to my description of "mathematical science" in §3. In other words a complement to, rather than a replacement for, contemporary core mathematics. Moreover his vision may be a good prediction for *explanations* of even core mathematics in the twenty-first century. I am sure it will seem increasingly magical as we have to gloss over increasingly complex details and downplay the contortions needed to connect the domain of mathematics to the physical world.

Finally:

Contemporary mathematics may have declared its goal to be the formulation of precise definitions and axioms and the subsequent deduction of theorems, but that is a fairly recent phenomenon and likely just a passing fad. It is also, I suggest, a foolish one that serves no one particularly well. In the last hundred years or so, mathematics has parted company with (and even distanced itself from) theoretical physics, statistics, and computer science, and even split internally into "pure" and "applied" mathematics only to find that some of the most exciting and productive developments within core mathematics (i.e., those parts that have not been cast out) have come from other disciplines.

We both see the consequences of the transformation, but where he sees a foolish passing fad I see a difficult adaptation to deep and unexpected features of mathematics itself. He finds problematic the fact that mathematics no longer officially includes the sources of some of its inspiration. I find the same fact exciting: the domain in which mathematics is powerful is very limited, and it is wonderful that it can be enriched by clues from other areas. Again, why the difference?

I believe our different viewpoints come from our different backgrounds. I spent thirty years dealing—successfully—with complex and subtle technical details before it even occurred to me to think about mathematics as a discipline. He came to mathematics essentially from the outside, as a sophisticated connoisseur and popularizer of mathematics rather than as a mathematician. He has learned what he knows—a great deal indeed—through high-level conversation that emphasized concepts and heuristics rather than by slogging through details. Twentieth-century

high-level conversations, concepts and heuristics are much the same as those of the nineteenth century and before. The difference is that in the twentieth century these are ways to *explain* mathematics and make it accessible by hiding details, while in the nineteenth they came close to actually *being* the mathematics. The difference is in the technical work needed to get things done: impossible to miss when you are slogging through details, nearly invisible in high-level explanations.

10.4. David Corfield, ‘Towards a Philosophy of Real Mathematics’. David Corfield is one of the very few philosophers who has made a serious attempt to understand modern mathematical practice, and his book [10] has useful insights and addresses real issues. However, because he is ahead of the field his ideas have not been digested and refined by his community; there are missteps due to lack of professional mathematical experience and perspective; and he has relied too uncritically on the views of participants, a practice well-known to Historians to be fraught with danger. In fact I found only a few things I felt were exactly right, and more that a few that seemed wrong. It still seems to me to be an important step in the right direction and worthy of analysis, if only to reveal pitfalls to be avoided in the next such attempt.

10.4.1. *Analogy, Axioms and Belief.* Mathematical practice is a compromise between human abilities and the demands of the subject, and confusion about what is human and what is mathematics is a perennial source of error. Corfield does better with this than most but still makes missteps.

Analogy, for instance, is a human cognitive device. Sometimes it enables us to see similarities between different areas long before we see the structural reasons for these similarities, and indeed this is often how we are led to deep structure. Nonetheless analogy is neither a feature of mathematics nor a mathematical method. In Chapter 4, Corfield confuses unexpected connections between fields (something about mathematics) with our *noticing* unexpected connections (human cognition, often using analogy) and describes these as “a riddle begging philosophical treatment” (p. 81).

At the end of the chapter (p. 98) he writes:

... we have seen mathematicians operating with ‘bite-size’ chunks of mathematical concepts. . . and it seems to suit them because they need only employ, in a recursive fashion, a manageable number of them.

This is a very good observation about human cognition (c.f. §??) with deep ramifications for mathematical practice. However Corfield uses it only to describe the gulf between human practice and automated (computer) proving.

In his discussion of Lakatos’s philosophy of mathematics, Corfield criticizes the idea that axiomatization has stifled development (p. 152):

I shall suggest instead that the appropriate use of rigorous definition and axiomatization has not acted as a hobble on the creativity of mathematicians, but rather an invaluable tool in the forging of new mathematical theories and the extension of old ones.

The success of twentieth-century mathematics should make this a no-brainer, but the evils of axiomatization are still an article of faith among most philosophers (and educators) so in context this is actually a big step. However he does not take it further, to understand *why* it has not acted as a hobble. It seems to me that

Lakatos is a hobble (see §10.5 for a brief discussion) and though he has reservations, Corfield has not broken free.

Chapter five concerns Bayesianism (subjective probabilistic methods, i.e. belief) in mathematics. He correctly cites Peano arithmetic as something believed but not proved, but does not realize that mathematicians worked quite hard to eliminate such things, with the result that this (in the guise of consistency of set theory) is essentially the *only* such example, see §2.3.2 and 2.4.1. He discusses the role of belief in the development and evaluation of new material but these are human aspects of the process, not part of the end product. He has good company in seeing Bayesian methods as part of mathematics (see §10.3 on Devlin), but not good reasons. The discussion in Chapter 6, where belief concerns modeling more than the mathematical analysis of models, is much more useful.

10.4.2. *Conceptualization.* In Chapter 9 Corfield discusses apparent cultural resistance to conceptualization in mathematics. However, as mathematicians know and philosophers endlessly point out, there are infinitely many possible mathematical ideas and most of them are worthless. There *should* be cultural resistance to unimportant ideas. A claim of *unwarranted* resistance needs evidence that the concept deserves more attention, and not just complaints from partisans. Another problem is that ‘conceptual’ is used in several different senses and careful interpretation is necessary. I discuss two of Corfield’s points from these perspectives.

On page 206 he quotes Gromov as attributing a lack of appreciation for his book on the h -principle in partial differential equations to resistance to conceptual work. I suspect Gromov is using ‘conceptual’ here as ‘lacking technical detail’: he is a product of the Russian school that puts more emphasis on ideas than precision, see §9.3.2, and the h -principle book (developed from his thesis) reflects this. Gromov’s complaint may be that the concept/precision ratio is considered too high for direct modern use rather than that the ideas themselves have been neglected.

I turn to the concepts involved. Gromov generously shares his ideas and a community has developed devoted to exploring and extending them. The ‘conceptual’ nature of the book should therefore not be much of a barrier if the ideas warrant attention. However this is doubtful:

- The basic principle was already reasonably well understood and more general. It applies to smooth and PL structures on topological manifolds, for instance.
- Most of the main applications for the differential version had already been developed, and new applications were not compelling enough to justify the somewhat unwieldy systematic treatment.
- The information obtained is “flabby” and tends to clarify where the real problems are rather than solve them.

To illustrate the last point, Gromov’s theorem shows that symplectic structures on non-compact manifolds are classified (in a weak sense) by homotopy information. Symplectic structures on compact manifolds are much richer and more rigid, or to put it another way, much of the richness of compact symplectic manifolds unravels if one removes a single point. Gromov has contributed powerfully to this area, and said of his earlier homotopy-oriented work “I did not understand symplectic topology so I was trying to destroy it”.

A similar argument shows that smooth structures on noncompact topological manifolds are homotopic structures, but doesn't give any information about the classifying space. The much more sophisticated work of Kirby and Siebenmann in dimensions ≥ 5 both evaluated the homotopy groups and showed the classification theorem extends to compact manifolds. I extended the calculation to dimension 4, [49], and this together with the topological h -principle showed that every noncompact (connected) 4-manifold has a smooth structure. But this is flabby in the same way as the symplectic result: Donaldson showed that most compact 4-manifolds are not smoothable and the theory is very far from being homotopic in nature. Again removing a single point causes the subtle complications to unravel; see [16] for more discussion, and a proof that avoids the heavy h -principle machinery.

Finally, 'conceptual' here does not mean 'non-axiomatic'. The basic approach is that a derivative map is defined from a space of geometric structures to a space of corresponding tangent structures. The tangent structures are essentially homotopic in nature so are characterized by a few basic properties (axioms). When one can show that the space of geometric structures also satisfies the axioms then it follows that the derivative map must be an equivalence. Gromov's h -principle gives a criterion for the axioms to hold. The use of axioms here is essential, and this particular use is so well established that it is now considered conceptual.

The next point is Corfield's use of groupoids as an example of cultural resistance to new concepts (§9.2, page 208). The implicit claim, again, is a value judgement: 'groupoids' are so important that their lack of recognition must be due to *unwarranted* cultural resistance.

An idea has to prove itself to be recognized as important, either by doing an important job particularly well, or through a rich and useful structure theory, or both. Two examples:

- Groups have a beautiful structure theory and do many jobs well. It took a long time, though, for these to emerge, and even quite a while to discover the exact formulation of the concept (definition) that has these virtues.
- Categories have a great deal of structure (adjoint functors etc.) and do many organizational jobs very well. Again it took a long time for this to emerge and be accepted. When I was a student categories were still generally seen as linguistic tool for organizing complicated naturality properties. Now they are promoted as a universal setting for mathematics (see §10.1 and the next section).

Groupoids are one of the many things that fit between groups and categories. They would have to do something quite unique and special to qualify as an important distinct concept, rather than a variation on these powerful neighbors; see §5.3.5 for a discussion. Corfield found some enthusiastic partisans but I did not see evidence that the idea is undervalued. I'll also give an example from my own work where groupoids seemed promising but were inadequate.

Many subtle invariants in geometric topology follow this pattern: start with a space X ; choose a basepoint x_0 and form the group ring of the fundamental group at this basepoint, $Z[\pi_1(X, x_0)]$. Then apply the algebraic K -theory functor to this ring. The choice of basepoint (to get a group) is artificial and interferes with naturality properties. For example suppose we have a map of spaces $Y \rightarrow Z$ and want to apply the construction coherently or continuously to the family of point inverses; see [48] and [59] for sophisticated applications of this. A coherent choice

of basepoint corresponds to a section of the map and these usually don't exist, so the fundamental-group approach is inadequate. This seems to be an opportunity for groupoids since they incorporate a multiplicity of basepoints.

Perhaps we could use the groupoid obtained by taking all possible basepoints? But then the whole space is included in the data and the groupoid structure is redundant and unnecessary. If the spaces are simplicial then perhaps we could use the groupoid with objects the vertices. But then we get unpleasantly involved with lax maps of these objects. The moral of the story seems to be that the 'group' part of groupoids comes from taking equivalence classes (here homotopy classes of paths) and in these applications this is just as unnatural as choosing a basepoint. My experience is that when groups don't work neither will groupoids, and I ended up working with the category of paths in the space.

10.4.3. *Higher-Dimensional Algebra.* The final chapter in [10] is concerned with higher-dimensional algebra, by which Corfield means n -categories. One of his goals is to give a glimpse of current research "at the *coal-face*", something he notes is almost entirely missing from contemporary philosophy. He also cautions that this is "insufficiently philosophically 'processed' ", and that he is a mathematical neophyte and learned the material mostly from a single source. Nonetheless he speculates that this subject will become the foundation of twenty-first century mathematics.

His point that philosophy desperately needs contact with actual mathematics is certainly right and his attempt to do this is admirable, but the account is problematic in several ways. I'll discuss some of the problems, and prospects for n -categories.

The greatest problems with this account have to do with limitations of the primary source, John Baez, and others in the immediate area. Baez's main background is in physics and he has a physicist's penchant for sweeping extrapolation from his own experience. However he shows little awareness of large areas of mathematics relevant to his concerns and his extrapolations are correspondingly problematic. A related problem is that the n -category development is too internal. These are extremely complicated axiom systems with many "degrees of freedom" in their formulation, and developments like this tend to be sterile unless guided and challenged by applications. There is nothing particularly wrong with any of this, but it is not a good example on which to base a philosophy.

One important area of application for fancy category theory is in homotopy theory, because structures on a category lead to structures on classifying spaces. Output structures include iterated loop structures, operad actions, various kinds of spectrum structures, and symmetric (or "new age") ring structures. Starting with one of these outputs and carefully investigating what makes it work has led to a number of special categorical structures, and identifying these has in turn led to important naturality results for the output structures. The emerging picture is of elaborate category theory as a toolkit: there is no single answer, and fine-tuning is often necessary in difficult cases. The structures useful in homotopy theory, however, do not seem to fit particularly well in the n -category context described by Corfield.

Another area of possible application is in geometry and topology, and through them, physics. Again, though, we should start with questions rather than answers. The general structure of manifolds and related objects of dimension five and above,

and topological four-manifolds [16], have all been deeply and effectively explored with classical techniques. This is not to say everything is known, but there are no easy opportunities. This area is my primary specialty and the only real possibilities I see involve the homotopy theory of classifying spaces as described above.

Topics that resist classical techniques and so may offer opportunities are:

- Finite 2-dimensional complexes (the Andrews–Curtis and Whitehead conjectures);
- 3-dimensional manifolds; and
- Smooth 4-dimensional manifolds.

Special structures such as symplectic manifolds above dimension four also resist classical techniques but the theory that is beginning to take shape doesn't strike me as particularly field-theoretic or categorical.

The main geometric motivation, and the one that informs Baez and colleagues, comes from 'topological quantum field theories' on 3-manifolds, with variations such as knots and links. Precursors were the Jones polynomial for knots and the Casson invariant for homology 3-spheres. The theory hit the big time with Witten's heuristic construction [81] of field theories with 'integrals' used in physics. Reshetikhin and Turaev gave a completely different construction with the properties predicted by Witten. It quickly became clear that the input for this and analogous constructions is a category with extra structure, and development of such things was a major industry during the 1990s. It never had major impact, however, and lags behind many other technologies including the Ricci flow, essential surfaces and laminations, exploitation of conformal structures on surfaces, and Heegaard–Floer homology.

I spent several years working in this area in the early 1990s, c.f. [51], and felt that there were indications that dimension three is not a good place for the method. Its greatest value to philosophy might be as an instance where received wisdom from physics impeded mathematical progress.

I feel that 2-complexes and smooth 4-manifolds are the most promising areas for topological field theories, see [51] for an explanation. There are analytic theories on 4-manifolds (Donaldson, Seiberg–Witten, and Osvath–Szabo) with field-like behavior. I spent ten years constructing and doing numerical computations in field theories on 2-complexes, intended as a warm-up for constructions on 4-manifolds. However:

- The analytic 4-dimensional theories do not have a full composition structure, so do not come from categories.
- The known categorical field theories on 2-complexes do not do anything interesting.

I suspect that the categorical approach, and particularly the n -category approach, is not likely to have much success. I'll give two reasons beyond the failure to date.

Topological field theories on 2-manifolds can be characterized in terms of Frobenius algebras. The modular ones (roughly the ones coming from 2-categories) correspond to semisimple Frobenius algebras. Semisimple algebras are 'measure zero' in unrestricted algebras and have much simpler structure. This indicates that requiring higher-order decomposition properties corresponding to higher categories enormously constricts the field theories. To get more power we apparently need to *reduce* the categorical order rather than increase it.

A heuristic explanation for the weakness of categorical field theories is that the composition property is a form of locality. Roughly, if two points are separated by a wall then any connection between two has to pass through the wall. More precisely, the relationship between what happens at the two points can be reconstructed from their individual relationships to what happens on the wall. One might think this would be seen as a description of ‘classical’ field theories rather than ‘quantum’ ones. In topology this corresponds to a form of excision. Excision is essentially the defining property of homology, so again is a measure of how ‘classical’ a theory is. In any case the power of the analytic 4–dimensional theories seems to be related to the *failure* of the decomposition property, i.e. the extent to which they are genuinely ‘nonclassical’, and consequently non–categorical.

10.4.4. *The Jaffe–Quinn Debate.* In section 7.6, pages 171–173, Corfield briefly discusses an exchange between myself and Arthur Jaffe, and William Thurston. The issue essentially concerns where the line is located between the human and mathematical aspects of the enterprise, and is discussed in more detail in the section on Thurston’s paper §10.6.

Corfield seems not to have fully come to grips with the issue in the book section. A later blog post²⁸ shows more clarity, unfortunately based on a misunderstanding:

... I came to realise what was really at stake in Thurston’s contribution to the debate. In essence, he and a few others were being accused by Jaffe and Quinn of acting very irresponsibly. By tossing out a conjecture and sketching how a proof might go, they were not acting for the good of mathematics. Big names shouldn’t behave like this, as it causes confusion, misleads the young, and discourages people from sorting out the field. Thurston’s response was to agree that responsibility is precisely what is at stake, but he goes on to say that responsibility involves so much more than maintaining standards of rigour. ...

In fact I believe that “tossing out a conjecture and sketching how a proof might go” is a wonderful thing, a great contribution to the field, and not at all irresponsible. My objection is that Thurston insisted on calling his conjecture a theorem, and claimed the full credit associated with a full proof. This made a huge difference to his students and followers who were trying to get jobs and promotions: one could become famous for proving a *conjecture* of Thurston, but filling in details of a *proof* of Thurston would be taken as a routine activity indicating a lack of real talent. It is nice that Thurston and philosophers can afford to think on a higher level, but arrogant and irresponsible for them not to acknowledge that it makes a real difference to less–generously endowed mathematicians.

10.5. **Michael Stöltzner; ‘Theoretical Mathematics: On the Philosophical Significance of the Jaffe–Quinn Debate’.** Stöltzner’s paper [76] concerns [27], a precursor of this article, so I begin with a brief discussion. First, however, I want to apologize for the term ‘Theoretical Mathematics’.

Our point was that misrepresenting conjectural or heuristic material as ‘proved’ is bad practice and damaging in several ways. On the other hand we value conjecture very highly when it is presented as such, and indeed think mathematics needs more

²⁸ http://golem.ph.utexas.edu/category/2006/10/wittgenstein_and_thurston_on_u.html, Blog post October 11, 2006.

of it rather than less. I had the bright idea that a dignified *name* for the activity might allow a more positive and constructive approach: instead of ‘conjecture is *not* mathematics’, we could say ‘conjecture *is* [something else]’. Unfortunately ‘theoretical mathematics’ seemed to irritate people who understood the point and confuse those who didn’t²⁹.

10.5.1. *Background.* Jaffe and I were concerned that a century of difficult adaptation to precision might be eroded by a new wave of heuristic work. Witten in particular was selling his wares to the mathematical community, but insisted on playing by physics rules. Further, he seemed to be supported in this by figures like Atiyah, who had been the very model of clarity and precision in his youth but was so enthusiastic about Witten that at one point he attributed Morse theory to him.

In sounding a warning we felt we had to cite specific contemporary examples, to try to make the issues clear and show that this is a real and ongoing problem. The people we mentioned have made profound and powerful contributions to mathematics and we had no intention of denying this. Our point, rather, was that they might have been *even more* influential if incompleteness of some of their work had been acknowledged, and further, it was their power that enabled them to behave this way: lesser talents would have been virtually banished from the community.

The editor (Richard Palais) solicited responses from the persons named, and others likely to have different viewpoints. Most of the responses in [3] were in general agreement, but felt that we undervalued the importance of intuition and conjecture. Thurston [77] disagreed strongly; see §10.6 for a discussion. Outsiders who read the article and the responses got the impression that our stand was controversial. It certainly did strike a nerve among theoretical physicists. Among mathematicians in general however, and somewhat to my surprise, our view was seen as so obviously correct that there was no interest in discussing it.

10.5.2. *Our Article.* Our article was written for mathematicians and depended on shared experience for understanding. The points can now be made more explicit.

In this essay I have emphasized that contemporary methodology is an enabling technology for users, from elementary school to the research frontier. In particular it enables rank-and-file mathematicians to make significant contributions. The cumulative nature of science is often suggested by the phrase “we see further because we stand on the shoulders of giants³⁰”. In mathematics it would be more accurate to say “we see further because we stand on the shoulders of men and women of average height”. The shoulders of giants are higher but frequently not strong enough to stand on. It is true that the ‘little people’ are often working out the ideas of giants, but this is no excuse for overlooking the great value they add.

The point is that mathematical progress is more dependent on proper function of the wider community than may be the case in other sciences. In any science it would be mean-spirited to make exaggerated claims that denigrates the work of students and followers trying to get jobs and promotions. In mathematics it also damages the field.

²⁹In the present development, conjectural work falls into what I call ‘mathematical science’ §3. However, having learned my lesson, I am certainly not proposing this as a name for the activity.

³⁰Attributed to Newton. Newton, however, said it of someone else, presumably thinking of himself as the giant.

This difference in mathematics derives ultimately from the validation criterion. Other sciences have external criteria for correctness, with the result that wisdom, insight, experience, intuition etc. play important roles in validating ideas. In mathematics, deep intuition etc. may be necessary to figure out what to *try* to prove, but the validation criterion is internal (error-free proof) and quite accessible even to those without profound intuitions. Further, *effective* mathematical intuition derives largely from the structure and details of proof. Speculation is rarely sharp enough to support further speculation, and this is the sense in which the ‘shoulders’ of speculative giants are too weak to stand on.

10.5.3. *Stöltzner’s Article.* Our article does not connect with conventional wisdom in philosophy. The usual approach is to find an interpretation that is easily dismissed, dismiss it, and when this is found to be unsatisfying find another such interpretation, dismiss that one, and so on. Stöltzner made a genuine effort to understand the issues from perspectives available to him. In the end he is not successful because the perspectives are inadequate, but it is a worthy effort.

The main problem is that Stöltzner relies on Lakatos’s *Proof and Refutations* [34] as a model for mathematical practice. As a baby model for the discovery process it is not too bad, but for the same reason that mathematics comes out looking like other sciences: this is the human aspect of the endeavor. Mathematicians have the same limited perceptual and cognitive facilities as other scientists. Their strategies for using these facilities to investigate nature are therefore much the same.

Lakatos’ story is a caricature of an antique that does not go past the early nineteenth century. The subject-adapted features of twentieth century practice are completely missing:

- The end result of the argument and refutation dialogue should be a completely precise, technically functional definition of the objects identified, *and* a logically complete proof that the conclusion is guaranteed by properties following from the definition.
- In particular the endpoint is *not* simply a matter of running out of refutations, with the possibility that the issue might be reopened at a later date.
- There is then the vital step of internalizing the definition and the functionality revealed by the proof. This replaces the pathetically limited cases that can be visualized directly and opens up new worlds.
- Finally, ‘island mathematics’ is sterile. It is interconnections, challenges, and applications that keep the internal verification criteria from leading to self-satisfied stasis.

Had Lakatos followed his story further he could have seen polyhedra of arbitrary dimension; cell complexes; the homological reasons an alternating sum works so well; apparently abstract but accessible examples such as projective spaces and Grassmann manifolds; apparently concrete but hard-to-access examples such as zeros of multivariable polynomials; the attempt to count cells in a specific dimension (Betti numbers, circa 1877) rather than mix them all together; the Noetherian revolution that brought clarity and power by replacing numerical invariants of matrices with abstract algebra (groups, rings, modules); the resulting view of Euler characteristics

as elements of some sort of representation ring rather than mere numbers; applications of this using Burnside rings and algebraic K -theory (e.g. the Wall finiteness obstruction as an Euler characteristic); etc., etc.

The unfortunate consequence is that Stöltzner is quite wrong when he writes (p. 206):

Jaffe–Quinn and Lakatos share a common starting point: Proof stands for a thought-experiment or quasi-experiment which suggests a decomposition of the original conjecture into subconjectures or lemmas, thus embedding it in a possibly quite distant body of knowledge. (Lakatos, 1976, p. 9)

This now-obsolete meaning for ‘proof’ has many well-known drawbacks, including cultural relativism. And again, it has more to do with the human aspects of research and discovery than the constraints imposed by the subject matter. Stöltzner makes a valiant and interesting attempt to make sense of our article as an effort to deal with these drawbacks. But mathematics itself dealt with these drawbacks by evolving a new meaning for ‘proof’ in the early twentieth century, and our paper concerns new problems connected to this new meaning.

10.6. William Thurston; ‘On proof and progress in mathematics’. Thurston’s paper [77] has been influential in education and philosophy, though not in mathematics itself.

There is no question that Thurston revolutionized three-dimensional topology. Some in the older generation—Armand Borel for instance—were unhappy with his oracle-like certifications that things were true without offering proof. But most of what he claimed was true, and his outlines provided material for many dissertations and some careers. There was a price: his followers were often seen as scribes recording the thoughts of the master, rather than independent researchers, because he had already claimed a proof. Ambitious students had to distance themselves from his program rather than work on it.

The high point was his geometrization conjecture and his claim to have proved it for Haken manifolds. Over the next ten years much of his sketch was filled in, but the non-fibered case got stuck on a key fixed-point theorem on Teichmüller space. There was widespread skepticism among experts that Thurston had fully addressed the problem. R. Kirby put it back on his list of open problems, [30]. In 1989 Curt McMullen [39] proved the necessary fixed-point theorem, but even so complete proofs were not available until the late 1990s, twenty years after Thurston’s announcement, cf. [44].

[[incomplete]]

10.7. Constructivism. Constructivism begins with the observation that mathematics is a human activity and a great deal of it is influenced as much by the human aspect as anything else. This is certainly true, and is an important consideration in any account of mathematics. Constructivism goes further, to assert that the human aspect is dominant, or even the whole story.

It is useful to distinguish two levels of constructivism:

- *Strong*, or explicit constructivism: mathematics is a social construct, and different people or different societies may very well have different versions. No version should be given a privileged role.

- *Weak*, or implicit constructivism: there is indeed something special going on in mathematics. However mathematicians themselves have many different views of it (pure, applied, experimental, etc.), and *these* differences are social constructs. What constitutes mathematics could be decided by a vote of mathematicians, and in particular contemporary pure mathematics should not have a privileged role.

The strong version is relatively rare. Supporters have to work pretty hard to align facts with their convictions, so this version is discussed in the section on religion.

Implicit constructivism is very common. Many of the accounts described in detail here (Gray, Borwein, Devlin, and Thurston) support defining mathematics by vote, and some push it strongly on the grounds that taking rigorous core mathematics as a model has inhibited development. Devlin writes in [13]:

However, the case I am trying to make in this essay is based not so much on historical precedent, but rather on a highly pragmatic view of the discipline as a human activity carried out by a human community. [...] As I see it, the question is, who gets to say what is and what is not to be called “mathematics”?

I observe in §3 that there is substantially more biomass in non-core mathematics than in the core even if one restricts to people with published work that qualifies for review in Math Reviews. A vote would indeed change the definition, and the change would be even more dramatic if the “human community” is taken to be all those who identify themselves as mathematicians.

The problem is that the subject is defined by the subject, not people. People can decide which methodologies to use but they cannot decide how effective these methodologies are. The choices in mathematics itself are:

- Contemporary core methods, highly disciplined and not particularly people-friendly but very effective; or
- Older methods, more relaxed and comfortable but significantly less well-adapted to the subject.

The conclusion is that the *solution* proposed by constructivists would undo the gains of the last century and severely damage mathematical practice. It seems to me, however, that constructivists are motivated by real problems, and rejecting the proposed solution does not make the problems go away. The motivating problems seem to be:

- There is a significant discontinuity between contemporary core methods and those used up through the nineteenth century; in education through the first few years of college; and in applied and computational areas.
- Many people are uncomfortable with these methods because they are highly disciplined; not human-friendly; and disconnected from physical reality and “meaning”.

These cannot be fixed by a vote, but there are other possibilities:

- People could realize that core mathematicians do this because they have to, not because they want to. Differences could be accepted with tolerance if not sympathy.
- People who dislike the methods could identify themselves as mathematical scientists, §3, rather than as mathematicians. This is a larger and more

diverse community and has goals that do not require such unfriendly methods.

- The gap could be reduced by bringing education up to date. A goal here and in related articles ([63], [?], [?]) is to suggest that a more contemporary approach to elementary mathematics education would not be difficult and might substantially improve outcomes.

It would also help if people who write about mathematics would understand it a bit more carefully. For instance the use of precise definitions and axioms is often described as unnecessary (“a foolish [passing fad] that serves no one particularly well”; Devlin, quoted above) and as making mathematics random and meaningless. But the methodology requires this approach, see §2.2.6, and this is not a deep observation. It is true that the axiomatic approach invites useless exploration, and entire fields (e.g. point-set topology) have drifted into irrelevance for this reason. However, ambitious mathematicians are keenly aware of the danger and have strategies for avoiding it. Further, most major definitions evolved over long periods exactly because early versions were still a bit on the random side and did not quite describe the objects of study. Definitions become standard—and highly prized—exactly because they are not random and not meaningless. Again this is not a deep observation.

10.8. **Religion.** [[unfinished]]

11. SOCIOLOGY: CULTURE IN MATHEMATICS AND PHYSICS

This is an extract from an older essay, [56], that uses a comparison with physics to illustrate the influence of reliability on mathematical culture. It duplicates material elsewhere in the essay, though I hope the different viewpoint will be useful. Finally, it is a travelogue rather than an attempt at scholarship.

By “physics” I will mean theoretical high-energy and quantum physics: areas where sophisticated mathematical apparatus is needed even to organize or interpret real data. By “mathematics” I mean strict areas in which abstract mathematical apparatus is developed. Experimental physics and applied mathematics are excluded to provide a clearer picture.

For most of history physics and mathematics were inseparable. Even after differences in subjects led to divergence there was a great deal of interaction between them. Physicists rely on mathematical work, and many mathematical structures are inspired by physics. The shared use of mathematical apparatus and long history of interaction has led to a great deal of shared language and strong superficial similarities.

Topics include reliability and its cultural consequences §11.1; efficiency as a selection force on norms §11.3; and construction of the literature §11.4.

11.1. **Reliability.** The key difference is reliability. Mathematics routinely achieves essentially complete reliability and this is a defining characteristic, see §2.1. Physical conclusions may be excellent but are—so far—never perfect.

The significance of reliability is illustrated by attitudes toward elaborate logical arguments. Such arguments magnify errors. Mathematicians exploit this in proofs by contradiction: an elaborate argument is built on a doubtful hypothesis in hopes that flaws in the hypothesis will be magnified to become obvious. However an error at the end of a long argument only indicates an error *somewhere* and one can

conclude that the initial doubtful hypothesis must be wrong only if there are *no* other sources of error. Further, since the objective is to magnify errors the other ingredients must be genuinely error-free, not just well-established.

Physics does not work this way. Even if nature is “error-free” physical descriptions have limits of applicability and at least minor discrepancies, so error-magnifying methods can be counted on to produce errors. Short insightful arguments tend to be more robust than logical proofs even when they are wrong in detail.

11.1.1. *Cultural Consequences.* Because mathematical methods are sensitive to errors, customs have developed to carefully distinguish between perfect and imperfect information. The conclusion of a mathematical plausibility argument is traditionally called a “conjecture”, while the result of a rigorous argument is called a “theorem.” Theorems can be used without fear in a contradiction argument; conjectures are a possible source of error.

The theorem/conjecture distinction is not useful in physical arguments because the methods cannot take advantage of the difference. Physicists frequently do not understand the significance in mathematics either, partly because it is not part of their own culture but there are other reasons:

- When mathematics works for physicists it becomes transparent precisely because it is reliable; they don’t think much about it because they have more urgent concerns.
- Well-tested mathematical conjectures often work as well as theorems because physical methods are too robust to run afoul of subtle flaws.
- Theorems often come with a lot of fine print that sometimes includes vital restrictions. Physicists, accustomed to robust statements and arguments, tend to ignore fine print and are sometimes misled as a result. For them there is little useful difference between a theorem whose fine print is missing (a conjecture) and one whose fine print is incomprehensible.

These differences lead to culture shock when the areas interact. In physics conclusions from intuition and plausible argument have first-class status. Mathematicians tend to describe these conclusions as “conjectures” still needing proofs, or dismiss them as hopelessly imprecise. Physicists resent this. Conversely physicists tend to be disdainful of mathematical rigor as being excessively compulsive about detail, and mathematicians resent this. But there are good reasons for the values held in both disciplines. The problems come from customs adapted to the subject, not xenophobia or a power contest.

11.1.2. *Ethical Behavior.* The effects of these mutual ill-adaptations are not symmetric. Mathematical practices used in physics are inefficient or irrelevant, but not harmful. Physical practices (no distinction between conjecture and theorem) used in mathematics can cause harm: it jeopardizes standard techniques such as proof by contradiction. There is widespread feeling among mathematicians that violating these ‘truth in advertising’ customs should be considered misconduct³¹.

Presenting heuristic conclusions as finished products is productive in one field and verges on misconduct in another, not because of contingent historical development of “standard practice”, but because the subjects are different.

³¹Not all mathematicians agree, see §7.

11.1.3. *The Interface.* The interface between mathematics and physics is populated roughly by three groups: mathematicians, physicists, and some straddlers.

People working in mathematical physics are typically in mathematics departments and identify themselves as mathematicians. They are usually very careful about following mathematical norms as a matter of survival. Physics can be fast and flashy where mathematics proceeds at a glacial pace. The fact that physics has a tendency to evaporate while mathematics doesn't is highly valued by mathematicians, less so by physicists (see §11.3). Consequently if mathematical physicists carefully follow mathematics norms they can be well-respected mathematicians. If they don't they would be at best second-rate physicists.

From time to time physicists (as opposed to mathematicians inspired by physics) have direct impact in mathematics. This was rare in the mid-twentieth century but in the 1980s and 90s several lines of thought in theoretical quantum physics led to truly inspirational ideas. In fact these ideas were more celebrated in mathematics than in physics: string theory for example was ridiculed by much of the physics community as being fatally out of touch with reality, c.f. Woit [82]. The result was a number of powerful and insightful physicists trying to sell their products to the mathematical community but wanting to play by physics rather than mathematical rules.

The physicists' suggestion that more-relaxed physics rules might be a good idea even in mathematics had support from a number of eminent mathematicians [3]. Some of us were concerned that an erosion of standards might result [27] but in the end our concern was not needed. The idea was rejected without interest or comment by the overwhelming majority of the mathematical community, and students in the area realized that if they wanted jobs they would have to play by mathematical rules.

Finally, straddlers are people who identify themselves as mathematicians when speaking to physicists, and as physicists when speaking to mathematicians, thereby dodging scrutiny by either community. I have met a number of them but don't know how common they are. They seem to have little effect on either community.

11.2. **Process versus Outcome.** The decisive criterion for correctness in theoretical physics is agreement with experimental observation. This focuses attention on outcomes, not process. This point of view permeates even internal efforts in theoretical development that do not make direct contact with experiment. When a model is developed it is checked against others believed to be relevant: special cases, the "classical limit", etc.

Mathematicians, as explained in §§2.2.7, 2.2.8, are unscientific in their use of an internal criterion for correctness.

There are many ramifications of this difference in focus. For example mathematicians are more tolerant of apparently pointless exploration, as long as it conforms to internal standards of rigor and quality. This tolerance is amply justified by the many times that exploration turned out not to be pointless at all, but the point was beyond human ken.

Physicists tend to be more relaxed about precision and more judgmental about significance, as illustrated by Wolfgang Pauli's dismissive description of a paper as "not even wrong". In physics this attitude is justified: without an internal criterion to keep a project "correct" while waiting for a mission to appear, the chances of anything worthwhile coming out are near zero.

11.3. Efficiency. Customs well-adapted to the subject should maximize return on resource investment. How this works is not always obvious, especially when adaptations are not consciously understood.

11.3.1. *Slow but Sure.* As an illustration we consider the different paces in mathematics and physics. Years often pass between an understanding satisfactory to physicists and a rigorous mathematical demonstration. Is insistence on rigor self-indulgence enabled by being sheltered from the demands of the real world? Or is it more efficient in some way?

Explaining the efficiency of rigor begins with another fundamental aspect of mathematics: since it is (usually) right the first time, it is not discarded. Over time it may become uninteresting or insignificant, but it does not become incorrect. As a result mathematics is an accretive activity.

In physics (and other sciences) material must be checked and refined rather than simply accreting, and customs have evolved to support this. Duplication is tolerated or even encouraged as “replication”. The primary literature has to be sorted and refined; a great deal of material is discarded; and there is a strong secondary literature to record the outcome of the process. These activities use resources. Mathematics lacks many of these mechanisms: the payoff for working slowly and getting it right the first time is savings in the refinement process³². In principle the same payoff is available to physics: if complete reliability were possible then the most efficient approach would be to seek it even at great sacrifice of “local” speed. But this is not possible, and an attempt to import this attitude into physics would be ill-adapted and counterproductive.

This adaptation has produced a vulnerability in mathematics. A group or individual can disregard the customary standards and seem to make rapid progress by working on a more intuitive level. But the output is unreliable. Mathematics largely lacks the mechanisms needed to deal with such material so this causes problems ranging from areas frozen up for decades, to unemployable students, to the outright collapse of entire schools of study (see [27]).

11.3.2. *Moveable Feasts.* As a second illustration of efficiency we consider the “fad” phenomenon in (theoretical) physics. It sometimes happens that an area becomes fashionable. There is a flurry of publication, with a lot of duplication. Then most of the participants drop it and go off to the next hot area.

Physics has been criticized for this short attention span. But this behavior is probably adapted to the subject matter. First, the goal is development of intuition and understanding, and this is an effective group activity. Duplication in publication is like replication of experiments: several intuitions leading to the same conclusion increase the likelihood that the conclusion is correct. And after a period the useful limits of speculation are reached, and it is a better use of resources to move on than to try to squeeze out a bit more.

If all activity ceased after a fad then eventually all of theoretical physics would become unsuitable for further development. Different activities continue: experimentalists test the testable parts. Mathematicians clean up the logical parts. A few physicists remain to distill the material into review and survey articles. And

³²Note this is a long-term payoff for the community, not individuals. This illustrates the point in §2.3.5 that selection acts primarily on communities and norms rather than individuals.

after a period of solidification the area is ready for another round of theoretical development.

Mathematics has occasional fads, but for the most part it is a long-term solitary activity. *Mathematical Reviews* and *Zentralblatt* divide mathematics and related areas into roughly 5,000 subtopics. Some of these are outside mathematics but there are still a great many areas for the size of the community and most are sparsely populated.

Mathematicians tend to be less mobile between specialties for many reasons: a greater technical investment is needed for progress; big groups are seldom more efficient; and duplication is unnecessary and usually discouraged. These factors tend to drive mathematicians apart.

In consequence the community lacks the customs evolved in physics to deal with the aftermaths of fads. If mathematicians desert an area no one comes in afterwards to clean up. There is less tradition of review articles: since the material is already right there is less sifting to do, and less compression is possible. Shifts of fashion may be an efficient behavior in physics, but they are not a good model for mathematics.

11.3.3. *Funding.* These considerations suggest ways funding programs might be fine-tuned to mesh with cultural nuances.

- Collaborative activity is often not efficient in mathematics and it is counterproductive to try to force it.
- Physics group activity is often focused at conferences, while mathematical interactions need to be alternated with solitary digestion and exploration. This suggests that mathematical conferences should (on average) be shorter, smaller, and more focused.
- Lack of large-scale cleanup mechanisms makes mathematical areas vulnerable to quality control problems. There are a number of once-hot areas that did not get cleaned up and will be hard to unravel when the developers are not available. Funding agencies might watch for this and sponsor physics-style review and consolidation activity when it happens.

11.4. **Publication.** Papers in pure mathematics and theoretical physics often look similar, treat the same subjects, and often even reside in the same library. However publication customs and uses of the literature are quite different.

11.4.1. *Purpose of Publication.* Physicists tend not to use the published primary literature. They work from current information (preprints, personal contacts), and the secondary literature (review articles, textbooks). The citation half-life of physics papers is short, and there are jokes about “write-only” journals that no one reads. Duplication and rediscovery of previously published material are common. In contrast many mathematicians make extensive use of the literature, and in classical areas it is common to find citations of very old papers.

There are differences in the construction of the literature as well. In mathematics the refereeing process is usually taken seriously. Errors tend to get caught, and detailed comments often lead to helpful revision of the paper. In physics the peer-review process has low credibility. Reviewers are uninterested, and their reports do not carry much weight with either authors or editors. Published papers are almost always identical to the preprint version.

One view of these differences is that the mathematical primary literature is user-oriented: genuinely useful to readers. In physics it is more author-oriented, serving largely to record the accomplishments of writers.

These differences again reflect differences in the subject matter. The theoretical physics primary literature is not reliable enough to make searching it very fruitful. It records the knowledge development process rather than the end result. If material is incorporated into the secondary literature or some shared tradition then it is reasonably accessible, but it is often more efficient to rediscover something than to sift the primary literature. A consequence is that there is not much benefit in careful editing or refereeing. This leads to journals that are, in the words of one mathematician, “like a blackboard that must periodically be erased.” In contrast, the mathematical literature is reliable enough to be a valuable asset to users.

11.4.2. *Social Consequences of Quality Control.* The differences in literatures have led to differences in social structure. As noted above, mathematics has many sparsely populated specialties. More accurately these could be described as larger communities distributed in time and communicating through the literature. This works even though the communication is one-way, because the material is reliable: when puzzled one can find clarification in the paper itself even if the author is long gone.

Less-reliable material requires two-way give and take. As a consequence working groups in physics are more constricted in time and appear larger because they are all visible at once.

This also works the other way: a large working group with a lot of real-time interaction weakens the benefits of reliability, and in fact larger groups in mathematics often do become more casual about quality control. This in turn leads to a curious problem in the mathematical infrastructure. The leadership in the professional societies and top journals tends to come from larger and more active areas. As a result they tend to underestimate the importance of quality control to the community as a whole.

Twenty years ago I was concerned that electronic preprints and publication might undermine quality control mechanisms [52]. The context was that in the previous twenty-five years or so it had become much easier to circulate paper preprints that were considered “off the record” and sometimes very casual about quality. It may be that quite a few people had been depending on referees to enforce discipline in their work, and this was an unwelcome peek behind the scenes. At any rate a flood of electronic preprints with similar low standards could erode the painfully-acquired benefits of a reliable literature.

My concerns were unneeded for an interesting reason related to community norms. Paper preprints are temporary in the sense that they have limited circulation and after a time can be hard to find. Electronic preprints are available everywhere, can last forever, and are impossible to kill if they turn out to be an embarrassment. Functionally they are *not* off the record. The sheer level of exposure led most authors to use much higher standards than for paper preprints and this has, in effect, been incorporated into norms. Formal refereeing is still vital for maintaining quality but it no longer seems to be the first line of defense it once was.

REFERENCES

- [1] Bibliography unfinished!
- [2] Aleksandrov, A. D.; Kolmogorov, A. N.; Lavrent'ev, M. A. (editors); *Mathematics: its Content, Methods, and Meaning* M.I.T. Press, second edition 1965.
- [3] Atiyah, M. et al; *Responses to Theoretical Mathematics : Towards a Cultural Synthesis of Mathematics and Theoretical Physics*, Bulletin of the American Mathematical Society, **30**(2) (1994) pp. 178-207.
- [4] van Atten, Mark; *Luitzen Egbertus Jan Brouwer*, *Stanford Encyclopedia of Philosophy*, 2008.
- [5] Balaguer, Mark; *Platonism in Metaphysics*, *Stanford Encyclopedia of Philosophy*, 2009.
- [6] Bell, John L.; *Continuity and Infinitesimals*, *Stanford Encyclopedia of Philosophy*, 2005.
- [7] Blanchette, Patricia; *The Frege–Hilbert Controversy*, *Stanford Encyclopedia of Philosophy*, 2007.
- [8] Boyer, Carl; *A History of Mathematics*, Wiley 1968.
- [9] Burgess, John *Fixing Frege*, Princeton University Press, 2005.
- [10] Corfield, David; *Towards a Philosophy of Real Mathematics*, Cambridge University Press 2003.
- [11] _____; *How Mathematicians May Fail To Be Fully Rational* (version 21 November 2005)
- [12] Devlin, Keith; *Mathematics: The Science of Patterns* (1994).
- [13] _____; *Mathematics: The New Golden Age* (Second Edition 1999).
- [14] _____; *What will count as mathematics in 2100?*, pp. 291–311 in [19].
- [15] Elmendorf, A. D.; Kriz, I. ; Mandell, M. A.; and May, J. P.; *Rings, modules, and algebras in stable homotopy theory*, Mathematical Surveys and Monographs, vol. 47, American Mathematical Society, Providence, RI, 1997.
- [16] Freedman, Michael; and Quinn, Frank; *Topology of 4-manifolds*, Princeton University Press, 1990.
- [17] Forster, Thomas; *Quine's New Foundations*, *Stanford Encyclopedia of Philosophy*, 2006.
- [18] Gladwell, Malcolm; *Outliers*, Little, Brown and Co., 2008.
- [19] Gold, Bonnie and Simons, Roger (editors); *Proof and Other Dilemmas: Mathematics and Philosophy*, Math. Assoc. Amer. 2008.
- [20] Goldman, Alvin; *Reliabilism*, *Stanford Encyclopedia of Philosophy*, 2008.
- [21] Gray, Jeremy, *Plato's Ghost: The Modernist Transformation of Mathematics*, Princeton University Press (2008).
- [22] Hanna, G.; Barbeau, E. *Proofs as bearers of mathematical knowledge*, *Mathematics Education* **40** (2008) pp. 345–353.
- [23] Hardy, Godfrey H., *A Mathematician's Apology*, Cambridge University Press, London (1941).
- [24] Heck, R. *The development of arithmetic in Frege's Grundgesetze der Arithmetik* *The Journal of Symbolic Logic*, **58** (1993) 579–601.
- [25] Hondé - Tzourio-Meyer, *Neural foundations of logical and mathematical cognition* (2003).
- [26] Hovey, Mark; Shipley, Brooke; and Smith, Jeff; *Symmetric Spectra* *J. Amer. Math. Soc.* **13** (2000), no. 1,
- [27] Jaffe, A.; Quinn, F. *Theoretical Mathematics: Towards a synthesis of mathematics and theoretical physics*, Bulletin of the American Mathematical Society, (1993)
- [28] Jech, Thomas; *Set Theory*, *Stanford Encyclopedia of Philosophy*, 2002.
- [29] JSTOR Archive of scholarly publications, <http://www.jstor.org/>.
- [30] Kirby, Robion, *Problems in Low-Dimensional Topology*, retrieved from <http://math.berkeley.edu/~kirby/>.
- [31] Klein, Felix, *Elementary Mathematics from an Advanced Standpoint*
- [32] Klein, Morris, *Mathematical Thought from Ancient to Modern Times*. Oxford University Press 1972.
- [33] The Klein Project of the International Commission on Mathematical Instruction, <http://www.mathunion.org/index.php?id=805>.
- [34] Lakatos, Imre, *Proofs and Refutations: The Logic of Mathematical Discovery*, edited by John Worrall and Elie Zahar, Cambridge University Press, Cambridge 1976.
- [35] Lin, Chun-Chi *How can the game of hex be used to inspire students in learning mathematical reasoning?*, Proceedings of ICMI Study 19 (2009).

- [36] Manilla, Linda; Wallin, Solveig, *Promoting students' justification skills using structured derivations*, Proceedings of ICMI Study 19 (2009).
- [37] Mazur, Barry, *When is One Thing Equal to Some Other Thing?*, essay in [19].
- [38] May, J. Peter; *E_∞ Ring Spaces and E_∞ Ring Spectra* (with contributions by Frank Quinn, Nigel Ray, and Jørgen Tornehave) Springer Lecture Notes in Math. **577** (1977).
- [39] McMullen, C., *Iteration on Teichmüller space*. Invent. Math. **99** (1990) pp. 425–454.
- [40] Mehrrens, Herbert; *Moderne Sprache Mathematik: Eine Geschichte des Streits um die Grundlagen der Disziplin und des Subjekts formaler Systeme*, Suhrkamp, Frankfurt 1990.
- [41] Nasar, Sylvia and Gruber, David *Manifold Destiny: A legendary problem and the battle over who solved it*, New Yorker magazine, August 28, 2006.
- [42] National Council of Teachers of Mathematics, see <http://www.nctm.org/>.
- [43] *The right organism for the job*, a collection in J. History of Biology 26 (1993) 233–368.
- [44] Otal, Jean-Pierre, *Thurston's hyperbolization of Haken manifolds*. Surveys in differential geometry, Vol. III pp. 77–194, Int. Press, Boston, MA, 1998.
- [45] Peltomaki, Mia; Back, Ralph-Johan, *An Empirical Evaluation of Structured derivations in high school mathematics*, Proceedings of ICMI Study 19 (2009).
- [46] von Plato, Jan; *The Development of Proof Theory*, [Stanford Encyclopedia of Philosophy](#), 2008.
- [47] Priest, Graham; Tanaka, Koji; *Paraconsistent Logic*, [Stanford Encyclopedia of Philosophy](#), 2009.
- [48] Quinn, Frank, *Ends of maps, II*, Invent. Math. 68 (1982) 353–424.
- [49] _____ *Ends of maps III, dimensions 4 and 5*, J. Diff. Geom. 17 (1982) 503–521.
- [50] _____ *A role for libraries in electronic publication*, EJournal vol 4 no. 2 (1994), reprinted in Serial Review 21 (1995) 27–30.
- [51] _____ *Lectures on axiomatic topological quantum field theory* in the IAS/Park City mathematics series vol. 1, Freed and Uhlenbeck ed., American Math Soc. 1995 pp. 323–459.
- [52] _____ *Roadkill on the electronic highway? The threat to the mathematical literature* Notices of the American Mathematical Society (1995); expanded version in Publishing Research Quarterly **11** (Summer 1995) pp. 20–28; [Link to text file](#).
- [53] _____ *A digital archive for mathematics*, unpublished 1996, retrieve from <http://www.math.vt.edu/people/quinn/epub/archive.html>.
- [54] _____, *Consequences of electronic publication in theoretical physics*, in *Scholarly Publication at the Crossroads: A Subversive Proposal for Electronic Publishing*, A. Okerson and J. O'Donnell ed. Association of Research Libraries (1995) pp. 170–174.
- [55] _____ and McMillan, Gail *Library copublication of electronic journals*, Serial Review **21** (1995) pp. 80–83.
- [56] _____ *Cultural adaptation in mathematics and physics* (1996, unpublished) [HTML format](#), [PDF format](#).
- [57] _____ *Evaluation of Methods in Mathematics Education*, March 2006 unpublished, retrieve from [64].
- [58] _____ *Dysfunctional Standards Documents in Mathematics Education* December 2004 unpublished, retrieve from [64].
- [59] _____ *Controlled K-theory I: basic theory* Preprint Feb 2004 <http://arXiv.org/abs/math.KT/0402396>, accepted for publication 2009.
- [60] _____ *Communication between the mathematical and math-education communities* October 2008; retrieve from [64].
- [61] _____ *K-12 Calculator Woes* Opinion Column, Notices of the AMS, May 2009.
- [62] _____ *Proof Projects for Teachers*, June 2009 unpublished; retrieve from [64].
- [63] _____ *Contemporary Proofs for Mathematics Education*, January 2010; retrieve from [64].
- [64] _____ *Education web page* <http://www.math.vt.edu/people/quinn/education/>
- [65] _____ *Student computing: interface design*, unpublished; retrieve from [64].
- [66] _____ *Beneficial High-Stakes Tests: An Example*, unpublished; retrieve from [64].
- [67] _____ *AMS Working Group on Preparation for Technical Careers* web site, <http://amstechnicalcareers.wikidot.com>
- [68] _____ *The EduT_EX project*, Wiki at <http://www.edutex.tug.org>.
- [69] _____ *Map of the historical development of manifolds*, with incomplete notes. Retrieve from <http://www.math.vt.edu/people/quinn/philosophy/index.html>.
- [70] Rav, Y. *Why do we prove theorems?*, Philosophia Mathematica **7**(3), (1999) pp. 5–41.

- [71] Reshetikhin, N.; Turaev, V. G. *Invariants of 3-manifolds via link polynomials and quantum groups* Invent. Math. 103 (1991), no. 3, 547–597.
- [72] Scholz, Erhard, *Geschichte des Mannigfaltigkeitsbegriffs von Riemann bis Poincaré* Birkhäuser, 1980.
- [73] Simpson, Adrian, *Mathematics: proof, logic and inhibitions* Proceedings of ICMI Study 19 (2009).
- [74] *Stanford Encyclopedia of Philosophy*, Edward N. Zalta et. al. editors, URL <http://plato.stanford.edu/>
- [75] Stanovich-West, *Individual differences in reasoning: implications for the rationality debate* (2000).
- [76] Stöltzner, Michael; *Theoretical Mathematics: On the Philosophical Significance of the Jaffe-Quinn Debate*, pp. 197–222 in *The Role of Mathematics in Physical Sciences: Interdisciplinary and Philosophical Aspects*, Boniolo, Budinich and Trobok eds. Springer 2005.
- [77] Thurston, W. *On proof and progress in mathematics* Bull. Amer. Math. Soc. **30**, (1994) pp. 161–177.
- [78] Wanko, Jeffrey J. *talking points: experiencing deductive reasoning through puzzle discussions*, Proceedings of ICMI Study 19 (2009).
- [79] Weinburger, Shmuel, *Computers, rigidity, and moduli. The large-scale fractal geometry of Riemannian moduli space*, M. B. Porter Lectures. Princeton University Press, Princeton, NJ, 2005.
- [80] Wilkerson-Jerde, Michelle H.; Wilensky, Uri *Understanding Proof: Tracking Experts Developing Understanding of an Unfamiliar Proof* Proceedings of ICMI Study 19 (2009).
- [81] Witten, Edward, *Quantum field theory and the Jones polynomial*, Comm. Math. Phys. **121** (1989), no. 3,
- [82] Woit, Peter, *Not Even Wrong* Basic Books (2006).
- [83] Wussing, Hans, *The genesis of the abstract group concept* 1969, English translation by Abe Shenitzer MIT press 1984.

INDEX

- Atiyah, Michael, 64
- Borwein, Jonathan, 56
- Bourbaki project, 31
- Calculators
 - problems with, 42
- Cognition
 - model of, 29
- Cognitive
 - complexity, 34
 - difficulty, 34
- Cognitive Structure
 - internalization, 29
 - objects vs. structure, 31
- Constructivism
 - strong, 66
 - weak, 66
- Corfield, David, 3, 34, 48, 58
- Criteria for Correctness
 - external (Science), 11
 - internal (mathematics), 8, 70
- Culture
 - comparison of mathematics and physics, 68
- Culture Shock
 - math/physics, 69
- Definitions
 - mathematical, 25
- Devlin, Keith, 56
- Education
 - philosophy of, 36
- Efficiency
 - of adapted methodologies, 70
- Error-Displaying Methods
 - definition, 7
 - legal analogy, 10
 - methodological requirements, 10
 - problems with, 9
- Ethics
 - in mathematics and physics, 69
- Euclidean Geometry
 - educational problems, 42
 - mathematical defects, 12
- Evolution
 - of methods, 14
- Fads in physics, 71
- Funding
 - math-adapted, 72
- Göttingen
 - innovation at, 44
- Gladwell, Malcolm, 34
- Gray, Jeremy, 3, 43
- Insights
 - from editorial work, 17
 - from education, 9
 - from history, 7
 - from publication, 6
 - from sociology, 6
- Jaffe, Arthur, 4, 17, 63
- Jaffe-Quinn Debate, 48, 63, 66
- Kant, Emmanuel, 36
- Khayyam, Omar, 12
- Klein, Felix, 36
- Kummer, Ernst, 44
- Lakatos, Imre, 65
- Manifolds
 - history of, 7
- Mathematical
 - methods, evolution of, 14
 - science, 20
- Mathematical Objects
 - axiomatic definition of, 25
 - development of, 26, 44
- Mathematical objects
 - functional reality of, 28
- Mathematicians, 34
 - adaptation of, 34
 - problems with elite, 16
- Mathematics
 - compared to science, 11
 - contemporary, 1
 - core, 1
 - definition of, 11
 - experimental basis of, 15
 - limits of, 27
 - norms, 15
 - publication practices, 6
 - sociology of, 6
 - success of wrong, 14
- Mazur, Barry, 53
- Mehrtens, Herbert, 51
- Methodological Norms
 - beneficiaries, 16
 - description, 15
- Methods
 - development of, 17
 - testing of, 18
- Modernist Transformation
 - leadup to, 44
 - subsequent development, 49
 - the event, 45
- Norms
 - beneficiaries of, 16
 - methodological, 15

- Palais, Richard, 64
- Philosophy
 - irrelevance of, 2, 47, 65
 - need for, 48
- Proof, 20
 - benefits of, 23
 - by contradiction, 8
 - conventions for checking, 22
 - for an individual, 22
 - for the prover, 23
 - formal student, 39
 - informal student, 38
 - non-examples, 24
 - potential, 21
 - spectator, 24
- Publication
 - math/physics differences, 72
- Reliability
 - characteristic feature of mathematics, 5
 - methods for achieving, 7
- Riemann Hypothesis, 18
- Robinson, Abraham, 19
- Scholz, Erhard, 7
- Second World War
 - consequences of, 49
- Soviet Union
 - mathematics in, 49
- Stöltzner, Michael, 63
- Struik, Derik, 12
- Thurston, William, 16, 66
- Unfinished
 - Bibliography, 74
 - Borwein, 56
 - Cognitive Complexity, 34
 - Cognitive Difficulty, 34
 - Computers in Mathematics, 20
 - Mathematical Science, 20
 - Mathematicians, 34
 - Physics and Computing, 9
 - Religion, 68
 - Soviet Union, 50
 - Thurston, 66
- Vulnerability, of mathematics, 15, 69, 71
- Witten, Edward, 64
- Word problems
 - components of, 41
 - reality/model disconnects, 41
- Wussing, Hans, 33
- Zermelo-Frankel set theory, 20