

Rigour and Thought Experiments: Burgess and Norton

James Robert Brown
Philosophy, University of Toronto

Contents

1	Introduction	2
2	Burgess on Mathematical Rigour	3
3	Proofs and Formal Proofs	5
4	Norton on Thought Experiments	7
5	Discovery vs (Two-stage) Justification	11
6	Problems with Burgess	15
7	Problems with Norton	23
8	Concluding Remarks	30

Abstract

This article discusses the important and influential views of John Burgess on the nature of mathematical rigour and John Norton on the nature of thought experiments. Their accounts turn out to be surprisingly similar in spite of different subject matters. Among other things both require a reconstruction of the initial proof or thought experiment in order to officially evaluate them, even though we

almost never do this in practice. The views of each are plausible and seem to solve interesting problems. However, both have problems and would seem not able to do justice to some interesting examples. They fail in similar ways. More pluralistic accounts of proof and of thought experiment could embrace aspects of each, while rejecting their claimed universality. An ideal account (not provided here) would contribute to *explanation* and *understanding*. These are important topics for future work.

1 Introduction

Many academic disciplines could be called spectator sports. We can read about them, watch them, enjoy them. Sports events are, of course, spectator sports, but so are musical concerts and theatre performances. One does not have to work through them, as it were, to enjoy them. We simply listen to a piece of music or watch a football game and are consequently excited and entertained. Intellectual life is often similar. We can be told about the expanding universe or how we evolved from earlier life forms. As exciting as it is, this sort of learning is still a passive activity.

Mathematics and thought experiments (TEs) are not spectator sports. To take in a piece of mathematics one has to work through it. It is an active, not a passive process. TEs are similar in this regard. One cannot simply read a TE to take it in. One has to actually work through it to understand the claim and how the reasoning goes. Like mathematics this is an active exercise, not a passive one. Active evaluation is an intrinsic part of it, the way it is not an intrinsic part of watching a tennis match or listening to jazz. We can read about an actual experiment in a scientific journal and we take what the experimenter is telling us as a report of the observational facts.¹ Proofs and TEs are quite different from this. We don't simply

¹Of course, this is oversimplified. Experiments often include some interpretive theorizing, with which we might quibble. But raw data is usually presented as a *fait accompli* and passively accepted as such by the reader. The distinction between the active and the passive might be fuzzy, but there is a distinction none the less.

read the conclusion; we actually have to work through the proof or the TE. We are given the evidence and we must check for ourselves. It should come then as no surprise that mathematics and TEs might have something important in common.

John Burgess and John Norton have proposed strikingly similar views within their respective fields of interest. Burgess has given an account of how mathematical rigour is achieved and Norton an account of how thought experiments work. Their virtues are similar and significant, and they rightly impress us by taking some of the mystery out of reasoning in mathematics and in thought experiments, respectively. They also run into similar problems, which I will try to make clear.

2 Burgess on Mathematical Rigour

Burgess has recently produced a thorough and insightful study of the troublesome notion of rigour, *Rigor and Structure* (2015). Many would concur when he says, “Rigor requires that new mathematical material be logically derived from old, and ultimately from first principles.” (2015, 51) Actual published proofs are seldom explicit derivations, so how can we reconcile the ideal with the actual? After surveying a long history of candidates, Burgess adopts and clarifies the definition of Hyman Bass: “A proof is what convinces the mathematician that a formal proof exists.” Burgess takes “a formal proof here to mean something like a deduction in which every inferential step is a fairly simple one recognized as valid by essentially all logic texts...” (2015, 96) Crucially, he adds “To be a genuine proof, a supporting argument must not only produce conviction that a formal proof exists, but it must do so in the right way... [Moreover] ‘the right way’ to produce conviction that a formal proof exists is primarily by supplying enough steps of the purported formal proof.” (2015, 97)

In sum, Burgess’s account of rigour includes four key features. (1) A rigorous (informal) proof is one that convinces mathematicians that a formal proof exists. (2) A formal proof is a derivation from first principles or from earlier established

theorems. (3) A rigorous (informal) proof has a structure that indicates the structure of the formal proof by containing key steps, at least in outline. (4) Evidence for a theorem ultimately rests on the existence of a correct formal proof.

Others have offered versions of proof in the same neighbourhood. Saunders Mac Lane, a prominent mathematician and one of the fathers of category theory, expresses the standard view as follows:

A mathematical proof is rigorous when it is (or could be) written out in the first order predicate language $L(\epsilon)$ as a sequence of inferences from the axioms ZFC, each inference made according to one of the stated rules.... To be sure, practically no one actually bothers to write out such formal proofs. In practice, a proof is a sketch, in sufficient detail to make possible a routine translation of the sketch into a formal proof.... What is at hand is not the *practice* of absolute rigor, but a *standard* of absolute rigor. (Mac Lane 1986, 377)

There is a huge range of alternative views about proofs, which I will not pursue. A small selection of views include: Azzouni (2013a, 2013b), Lakatos (1976), Horgan (1993), Rav (1999), Avigad (2020), Detlefsen (2008)²

²I will, however, mention a couple of things in this footnote. Michael Detlefsen is skeptical of the common view, but puts it forward this way: “(i) proper proofs are proofs that either are or can be readily be made rigorous; (ii) proofs that are or can be readily be made rigorous are formalizable; therefore (iii) all proper proofs are formalizable.” (Detlefsen 2008, 17) Jeremy Avigad is skeptical of the standard view, which he describes as follows: “Modern logic offers idealized accounts of what it means to be correct, based on the notion of a formal derivation. On that account, the mathematical objects we talk about are defined in a suitable formal foundation, our theorems are expressed as formal statements in that system, and our proofs are required to conform to the axioms and rules.... Of course, ordinary mathematical texts are not nearly as regimented as formal proofs. But, on the logical story, they can be seen as an approximation to the ideal.” (Avigad 2020, 1) He adds two objections: “One is that a formal system cannot possibly be the operant standard because most mathematicians cannot even state the axioms of a particular formal system. The other is that there are various formal foundations on offer, and who is to say which is the right one?” (ibid., 1)

These are important remarks, causing us to clarify the Burgess position on his behalf. When Burgess says a rigorous informal proof convinces a mathematician of the existence of a formal system, we need not say she knows precisely which formal system or what the particular details of the formalization are. On a more liberal view all we need say is that there is some robust sense

The idea in general and Burgess's version in particular is a very clear and plausible characterization of mathematical rigour. It will appeal to many, perhaps to most mathematicians. A few things can be noted. Burgess's expression "produces conviction" probably means produces a sense of certainty. Some, still in the spirit of Burgess, might want to modify his characterization in favour of a fallibilist version. And "convinces the mathematician" could reasonably be understood as meaning convinces the kind of mathematician with the right training. We need not worry about the reaction of nonspecialists. Even among specialists, there are those who are convinced by a faulty proof and those who resist a good one. Burgess addresses these possibilities, but I will ignore them here.

I should also note that Burgess has interesting things to say about ideal elements,³ elementary vs non-elementary proofs,⁴ and other matters. These play a part in any complete account of rigour, but will be ignored here as not relevant to my chief interest, which is the four features of Burgess's account that listed that I listed above.

3 Proofs and Formal Proofs

Burgess's account fits a great deal of mathematical experience. I will illustrate with a non-trivial example taken almost at random from a standard text, Rudin, *Principles of Mathematical Analysis* (1976). Rudin follows the common definition-theorem-proof pattern; he begins with a definition of the Γ function,

of a formal proof that the working mathematician has in mind and that is sufficient for Burgess's purposes.

³Imaginary numbers and points at infinity are sometimes called "ideal elements." They simplify proofs but are considered by some to be only useful fictions. Hilbert took such a view.

⁴An elementary proof, say of a theorem in Peano arithmetic, is a derivation from the Peano axioms using only the concepts of that theory. It does not matter how long or complicated the proof is as long as it sticks to the elements of Peano arithmetic. A proof of the same theorem using, say, group theory or complex analysis is non-elementary, even if it is only an easy few lines long.

followed by a three-part theorem about the function, and then a proof of the theorem. I will comment briefly on the proof after reproducing the whole passage from Rudin (1976, 192).

Definition: For $0 < x < \infty$, $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$.

Theorem:

- (a) The functional equation $\Gamma(x + 1) = x\Gamma(x)$ holds, if $0 < x < \infty$.
- (b) $\Gamma(n + 1) = n!$, for $n = 1, 2, 3, \dots$
- (c) $\log \Gamma$ is convex on $(0, \infty)$

Proof: An integration by parts proves (a). Since $\Gamma(1) = 1$, (a) implies (b) by induction. If $1 < p < \infty$ and $(1/p) + (1/q) = 1$, apply Hölder's inequality [established earlier in the text] and the definition of Γ and obtain $\Gamma(x/p + y/q) \leq \Gamma(x)^{1/p}\Gamma(y)^{1/q}$. This is equivalent to (c).

This informal proof illustrates Burgess's account well. It tells us the main step in proving (a), assuming we already know how to integrate by parts. We can next prove (b) assuming we know how to prove things by mathematical induction. (c) is more tricky, since we might have to remind ourselves of the properties of convex sets. But in each of these three cases we can (perhaps with some effort) construct a formal proof, that is, a derivation, using the information given in the informal — but still rigorous — proof. Consequently, Rudin's informal proof counts as a rigorous proof by Burgess's lights. One need only look through standard textbooks or research journals to see similar examples in abundance.

Incidentally, the Rudin example also illustrates the worth of informal proofs. Rather than grinding through a formal proof, we often only need to be told that something follows, say, by induction. Frequently we can see how it might go right away, contributing to the often elusive mathematical *insight* or *understanding* that we all crave. Mathematics is always worthwhile, but it is the occasional *aha effect* that makes it fun.

Burgess rightly notes features such as this, including the fact that the informal proof is much more intelligible. He also notes that the informal proof might yield

greater certainty that the formal proof, paradoxically, because the latter is often so hard to follow that we might miss outright mistakes.

I will return to Burgess's account after describing John Norton's strikingly similar views on a different topic. But before leaving Burgess for Norton, I should say something about definitions. Much of what is said about proofs applies to mathematical definitions, as well. Burgess cites Weierstrass's definition of continuity and uniform continuity as an example. (2015, 30f) This was the beginning of ϵ - δ methods, which are now universal. Before Weierstrass there was much confusion about continuity and uniform continuity. This was one of those cases where a modicum of formalization clarified everything. In natural language we often use "for all there is" and "there is for all" interchangeably, usually without confusion, as in: "For everyone there is someone" and "There is someone for everyone." We know, in fact, that they do not mean quite the same as is illustrated with the pair: "For every number there is one greater" and "There is a number greater than every numbers." When it comes to continuity vs uniform continuity the difference between $\forall\exists$ and $\exists\forall$ makes all the difference but is easy to be confused about. Formalization sorts it out.⁵

4 Norton on Thought Experiments

Thought experiments present us with a wonderfully interesting problem: How is it possible, just by thinking, to learn something new about the world? There is a

⁵Here is the difference in detail. There are three distinct definitions involving continuity of a real valued function f defined on the interval $[a, b]$. f is *continuous at a point* c in $[a, b]$ if and only if $\forall\epsilon > 0 \exists\delta > 0 \forall x \in [a, b] (|c - x| < \delta \rightarrow |f(c) - f(x)| < \epsilon)$. If we wish to say that f is continuous at every point in $[a, b]$, then we add another quantifier at the beginning of the above. Thus, f is *continuous at every point in* $[a, b]$ if and only if $\forall c \forall\epsilon > 0 \exists\delta > 0 \forall x \in [a, b] (|c - x| < \delta \rightarrow |f(c) - f(x)| < \epsilon)$. For the stronger idea of uniform continuity we rearrange the quantifiers. Thus, f is *uniformly continuous on* $[a, b]$ if and only if $\forall\epsilon > 0 \exists\delta > 0 \forall x \in [a, b] \forall c \in [a, b] (|c - x| < \delta \rightarrow |f(c) - f(x)| < \epsilon)$. The order of quantifiers was successively: $\forall\exists\forall$ for continuity at a point, then $\forall\forall\exists\forall$ for continuity at every point, then $\forall\exists\forall\forall$ for uniform continuity. These examples show the enormous value of rigorous definitions. Without the benefit of explicit quantifier rules and quantifier symbols it would be very hard to keep track of the different types of continuity.

large spectrum of answers, from a form of rationalism to a staunch empiricism, with several plausible accounts in between.⁶ John Norton holds a view of thought experiments that is exceptionally clear and powerful, and has been highly influential. Norton's answer to the question is that we actually do not learn anything new, although we might draw out things that hadn't previously been recognized. He asserts that

Thought experiments are arguments which

- (i) posit hypothetical or counterfactual states of affairs and
- (ii) invoke particulars irrelevant to the generality of the conclusion.

The account is motivated by a staunch empiricism.

Thought experiments in physics provide or purport to provide us information about the physical world. Since they are thought experiments rather than physical experiments, this information does not come from the reporting of new empirical data. Thus there is only one non-controversial source from which this information can come: it is elicited from information we already have by an identifiable argument, although that argument might not be laid out in detail in the statement of the thought experiment. The alternative to this view is to suppose that thought experiments provide some new and even mysterious route to knowledge of the physical world. (Norton 1991, 129)

Galileo's wonderful argument in the *Discorsi* shows that all bodies, regardless of their weight, fall at the same speed (Galileo 1974, 66f.). It is easy to imagine, so no need for a diagram. It begins by noting Aristotle's view that heavier bodies (e.g., a cannon ball, H) will fall faster than light ones (e.g., a musket ball, L): ($H > L$). We are then asked to imagine that the heavy cannon ball is attached to the light musket ball, making a combined object ($H + L$). What would happen if it were

⁶My own view, which is on the rationalist side can be found in Brown (1991/2011, 2004). Norton holds down the empiricist pole in a number of articles including Norton (1991, 1996, 2004a, 2004b). Stuart et al (2017) is a useful source of a range of topics and issues.

released?

Reasoning in the Aristotelian manner leads to an absurd conclusion. First, the light ball will slow up the heavy one, acting as a kind of drag, so the speed of the combined system would be slower than the speed of the heavy ball falling alone ($H > H + L$). On the other hand, the combined system is heavier than the heavy ball alone, so it should fall faster ($H + L > H$). We now have the absurd consequence that the heavy ball is both faster and slower than the combined system. Thus, the Aristotelian theory of falling bodies is inconsistent; hence, destroyed.

But the question remains, Which falls faster? The right answer is now plain as day. The paradox is resolved by making them equal; they all fall at the same speed ($H = L = H + L$).

According to Norton, this thought experiment can be reconstructed as an explicit argument. Reconstructing is no easy matter and often involves considerable effort. Norton's reconstruction of the Galileo thought experiment in the form of an argument goes as follows:

1. Assumption for reductio proof: the speed of fall of bodies in a given medium is proportionate to their weights.
2. From 1: if a large stone falls with 8 degrees of speed, a smaller stone half its weight will fall with 4 degrees of speed.
3. Assumption: if a slower falling stone is connected to a faster falling stone, the slower will retard the faster and the faster speed the slower.
4. From 3: if the two stones of 2 are connected, their composite will fall slower than 8 degrees of speed.
5. Assumption: the composite of two weights has greater weight than the larger.
6. From 1 and 5: the composite will fall faster than 8 degrees.

7. Conclusions 4 and 6 contradict.
8. Therefore, we must reject assumption 1.
9. Therefore, all stones fall alike. (Norton 1996, 341-2)

I think this is adequate to get the idea behind Norton's proposal. However, he acknowledges that more detail might be necessary, but for our purposes, we need not include it here.⁷

After the reconstruction, the evaluation of a TE, according to Norton, takes the following process: Evaluate the premisses: Is there sufficient evidence of their truth? Typically, this is an empirical task. Second, evaluate the argument's

⁷For those wanting the full story, here are the extra details Norton provides.

This final step is actually quite tricky. Its analysis varies according to whether the thought experimenter makes a tacit assumption:

8a. Assumption: The speed of fall of bodies depends only on their weights.

However, in either case, with or without this assumption, it is hard to see any room for a Platonic leap in the final step. If assumption Ba is made, then the move from the refutation of Aristotle's theory (8) to Galileo's theory (9) is a simple inference. The most general viable theory is that the speed of fall of bodies is some arbitrary monotonic increasing function of their weights. (Any other choice would have to allow the function to decrease somewhere, which amounts to the possibility of a heavier body falling slower than a lighter one!) Galileo's own theory arises as the special case of a constant function. Now any admitted function other than Galileo's leads to a contradiction through a simple argument almost exactly the same as the argument steps 1-8 and immediately obvious to anyone who has followed Salviati's inferences. Consider a faster-falling, heavier body and a slower-falling, lighter body. If we connect them in their fall, the slower will retard the faster, so that the composite falls at an intermediate speed. This contradicts the consequence that the heavier composite must fall no slower than either of its lighter parts. Thus we can give the transition from 8 to 9 as:

8a. Assumption: The speed of fall of bodies depends only on their weights.

8b. Assumption: The speed of fall of bodies is some arbitrary monotonic increasing function of their weights.

8c. From 3, 5. If the function is anywhere strictly increasing, then we can find a composite body whose speed of fall is intermediate between the speed of fall of its lighter components.

8d. The consequent of 8c contradicts 8b.

9. From 8d. The function is constant. All stones fall alike. (Norton 1996, 342-3)

validity: Does the conclusion follow deductively or inductively⁸ from the premisses? If the answer to both questions is yes, then the TE is a good one. If it fails on either, then it is bad.

Norton is taken to be an empiricist, and this is largely correct, but it does need some qualification. He allows that there are TEs in mathematics and in ethics, which means that when reconstructed as arguments they will have non-empirical premisses. Even in some physics TEs verificationism might be a premiss, which, of course, is a normative principle, not an empirical fact. Norton's empiricism is secondary to his account. His claim that a TE can be reconstructed as an argument is his central point. It is also the feature that makes his view similar to Burgess's account, since the latter also stresses the importance of reconstructing a proof as a formal proof, an argument.

5 Discovery vs (Two-stage) Justification

How do Burgess and Norton compare? The main similarity is this: They rely, respectively, on a *reconstruction* of the initial proof or initial thought experiment in order to evaluate it. The reconstruction need not be manifest, but it must be thought to exist in principle. For Burgess an informal mathematical proof is rigorous provided there is a corresponding correct formal proof. For Norton, a thought experiment is a good one provided there is a corresponding sound argument. In both cases, the initial reasoning, that is, the intuitively natural line of thought produced by mathematicians or by scientists (in the form typically published in textbooks or journals), is of evidential value only in so far as it is

⁸Deductive logic is well understood, but not so inductive. Norton realizes this but is rightly not bothered in this context. He can leave that to others to sort out. It does, however, present a special problem for Norton, since he holds unusual views of about inductive logic. He claims that there are no general principles of induction and every case must pay attention to its own special features. He calls this "the material theory of induction." It's an interesting and important view. It does, however, get him in trouble with TEs which seem to require general principles of induction to play a role in the evaluation of the reconstructed argument. Stuart (2020) brings this problem to light. Fortunately, this plays no role in our concerns here.

linked to the reconstruction. It is the reconstruction that bears the ultimate evidential burden for both Burgess and Norton. To repeat, it seldom matters whether the reconstruction is made manifest; its acknowledged existence (in Plato's heaven?) is what matters. In both cases it must be obvious that the reconstruction can in principle be done.

The nature of the connection between the informal proof or TE and the reconstruction is one important place where they slightly differ. Burgess requires the right sort of connection between them for the informal proof to be rigorous. To repeat, "...the right way to produce conviction that a formal proof exists is primarily by supplying enough steps of the purported formal proof." (2015, 97) This suggests that an informal proof is an enthymeme, which means that we could construct the corresponding formal proof by filling in the missing parts. This is a rather strong requirement and might not be satisfiable in some cases, which I will explain momentarily. Norton, by contrast, has not specified any required connection between the initial TE and the reconstructed argument. All he demands for a good TE is a reconstruction that is a good argument. The connection might be as disparate as an apple falling on Newton's head and the theory of universal gravitation. He does, however, suggest that a typical TE will have an obvious argument form that can without much difficulty be turned into a specific argument. If pressed, Norton could easily add something similar to the Burgess requirement of making key argument steps part of being a good TE.

Given the similarity between the views of Burgess and Norton, we might think they ought to hold the same view on this point. But which way should they tilt? Given that published mathematical proofs are overwhelmingly informal, there really couldn't be a huge gap between what we read and find persuasive in mathematical texts and journals and a formal proof, which Burgess considers the ultimate evidence. An appeal to the usual justification-discovery distinction would no do. It would be like reading a chemistry book and when the structure of the benzene molecule is discussed, we are not told that it is a ring, only that Kekulé had a dream about a snake. Freud might find this interesting, but chemistry students

would be mistified. The connection between the informal proof or TE and the formal proof or reconstructed TE needs to be fairly close. The informal proof/TE, after all, is typically persuasive in its own right. The formal proof or reconstructed TE cannot consist in utterly different evidence. Burgess has this covered by insisting that to be rigorous the informal proof must contain the key steps of a formal proof (if only in a sketchy form).

Burgess acknowledges the traditional discovery-justification distinction but brushes it aside. “One of the first things to note about rigor is that it pertains only to what philosophers of science call the ‘context of justification,’ as opposed to the ‘context of discovery’.” (Burgess 2015, 8) This is important for a couple of reasons. Burgess is surely right to think of a rigorous proof as evidence, hence as justification. But it also implies that justification is a two-step process. First comes the informal rigorous proof, then the formal proof. Together they are the complete justification. The discovery stage was earlier, based on seeing a falling apple, having a dream about a snake, calculating lots of examples until a patterns seemed to arise, and so on. These are not evidence but they might suggest a possible theorem and even suggest a proof of that theorem.

Norton does something similar. That is, he notes the discovery-justification distinction but then puts everything important into justification — though he does not use the terminology in precisely this way. He also turns justification into a two stage process. I use “context of discovery/context of justification” differently than Norton does in the following passage, but that is a mere terminological matter.

Two forms of the thesis. The thesis that thought experiments are arguments requires some elucidation. Is the claim merely that thought experiments can do no more than argumentation when it comes to justifying claims? Or is it, in addition, that the actual execution of a thought experiment is just the execution of an argument? Following Norton (1996, p. 354), I intend the stronger version and urge both:

(1a) (*Context of justification*) All thought experiments can be reconstructed as arguments based on tacit or explicit assumptions. Belief in the outcome-conclusion of the thought experiment is justified only insofar as the reconstructed argument can justify the conclusion.

(1b) (*Context of discovery*) The actual conduct of a thought experiment consists of the execution of an argument, although this may not be obvious, since the argument may appear only in abbreviated form and with suppressed premisses. (Norton 2004b, 50)

To say, as Norton does, that the unreconstructed TE is a non-obvious argument, because it is abbreviated or has suppressed premisses, sounds a lot like Burgess's rigorous informal proof. At this stage we are already in the realm of justification. The discovery stage has long passed and now a Galileo or a Newton or an Einstein is presenting us with evidence for some theory. Norton's explicit arguments are like Burgess's formal proofs.

Norton seems to be saying implicitly what Burgess made explicit. A rigorous proof, according to Burgess, convinces mathematicians "in the right way" because it contains key steps in the corresponding formal proof. Norton is not so specific and allows that the argument character of a TE might not be obvious, but it nevertheless is an argument and might, for instance, merely need the suppressed premisses to be made explicit. At a minimum, the TE suggests a proper argument.

Discovery plays no role for either Norton or Burgess. Justification is everything. What is important about their common view is that justification is a two-stage entity, even if the second stage is normally not carried out. For Burgess a normal informal proof is rigorous because we are convinced that a formal proof exists. Only rarely do we bother to construct one. For Norton a TE, as we normally encounter them, is a good one, because there is a corresponding argument that is sound. In most cases we are convinced of this, but only rarely make the argument explicit.

Questions naturally arises. For Burgess: Are there any proofs that we could plausibly call rigorous that do not have corresponding formal proofs? For Norton: Are there any TEs that need not be reconstructed as arguments to be convincing in their own right? As we shall see, the answer is yes in both cases.

6 Problems with Burgess

There is one more aspect of Burgess's view that I want to consider before looking at specific problems. A great deal of mathematical activity does not involve proof, however we might conceive of proof. It is often called inductive or experimental mathematics, although those terms do not cover all. It is important for all sorts of reasons. For one thing, mathematicians need to have some idea of what problems might be worth tackling or suitable for assigning as a PhD topic. A young career could be ruined by stupid problems. Perhaps more interesting are kinds of evidence that differ from proofs. For instance, the evidence cited for speculations and conjectures made in conference talks or in blogs, which are an increasingly popular outlet for mathematical ideas in an early stage.⁹ Such conjectures are often supported by inductive evidence that would be considered stellar in the sciences, but fall short of proof, as normally conceived by the mathematical community.

Recall my summary of Burgess's account of rigour, which includes four aspects. (1) a rigorous (informal) proof is one that convinces mathematicians that a formal proof exists. (2) A formal proof is a derivation from first principles or from earlier established theorems. (3) A rigorous (informal) proof has a structure that indicates the structure of a formal proof by containing key steps, at least in outline. (4) evidence for a theorem ultimately rests on the existence of a formal proof.

My unease with the Burgess account of rigour turns on two aspects of his account.

⁹Here are two excellent ones by leading mathematicians. Both are rather advanced: Timothy Gowers's blog: <https://gowers.wordpress.com>, Terence Tao's blog: <https://terrytao.wordpress.com>. There are lots more on the Internet at various levels of sophistication.

(1) and (3) are the problem. (2) and (4) are either okay as they are or could be slightly modified into acceptable form. (For instance, I would reject (4) if it says that evidence rests *exclusively* on the existence of a formal proof.) Before proceeding, however, I stress that the Burgess account seems quite right for a great many cases, perhaps even most. I also think it would be accepted by the general mathematical community. What I am disputing is its claim to universality. The cases where it fails are important and shed light on the nature of mathematics.

Contra (1): (A rigorous informal proof is one that convinces mathematicians that a formal proof exists): First, on the Burgess account it is a mystery why we value multiple proofs of the same theorem. If the point of a proof is to convince us that a formal proof exists, then why multiple proofs when the first one did the job of convincing us? The usual answer to this question is that different proofs give us different kinds of insight into what is going on. To be fair, Burgess could rightly reply that he is only interested in what makes a proof rigorous, not what make it insightful. This is a perfectly acceptable reply, but we should always keep in mind the importance of explanation and understanding, even though they are very difficult to characterize.

Second, there are fields of mathematics where the first principles are not complete, thus, there will be propositions P such that neither P nor $\text{not-}P$ is derivable from those first principles. And yet, we might have an informal proof that P is true. Two examples come to mind. Gödel famously proved that any system strong enough to represent arithmetic would have true but unprovable sentences. He constructed a sentence G that could be understood to say of itself that it is unprovable, which indeed it was. So, it is plausible to claim (as did Gödel) that G is true and we have a kind of proof of it, but there is no formal proof.¹⁰

¹⁰Gödel's incompleteness theorem can be found in most advanced logic books. An interesting special case is described in Paris and Harrington (1977). They find an explicit example of a true but unprovable sentence of Peano Arithmetic.

A second example concerns the continuum hypothesis (CH) which is independent from the axioms of set theory, hence there is no formal proof of it or its negation. (This independence is from the rest of mathematics, if we think that it all reduces to set theory). I will briefly give an informal proof due to Freiling (1986) that CH is false. If the informal proof works, then it would be a case of a rigorous proof without a corresponding formal proof. I can here give little more than a hint of the actual argument, which is rather technical.¹¹

The CH is the claim that the continuum of real numbers \mathbf{R} has the cardinality \aleph_1 . Cantor's diagonal argument showed that it is larger than the natural numbers \mathbf{N} and the rational numbers \mathbf{Q} which are both \aleph_0 . CH is the claim that the continuum is the next infinite number. Since the cardinality of \mathbf{R} is 2^{\aleph_0} , CH is the claim that $\aleph_1 = 2^{\aleph_0}$. Thanks to the work of Gödel and Cohen we know that CH cannot be proved or refuted on the basis of existing set theory. That is, there exists no formal proof of CH or of not-CH. But, as any mathematical realist would say, CH is nevertheless true or false; we merely don't know which. Could there be some sort of argument for CH or against it?

Chris Freiling (1986) produced a remarkable argument that shows CH is false. Needless to say, it is somewhat controversial. However, it is sufficiently interesting to be worth considering and even if the proof fails to work, it can be highly instructive here. It provides an example of a kind of mathematical reasoning that is plausible and should be considered legitimate in principle.

We begin by assuming CH is true, aiming for an absurdity, then concluding it must be false. As background, we assume ZFC (Zermelo-Frankel set theory with the axiom of choice). This implies what is known as well ordering, which means we can line up the real number with the so-called ordinals, including the transfinite ordinals. It will be enough to restrict ourselves to the real numbers in the interval $(0, 1)$. A well ordering of $(0, 1)$ is not the usual ordering based on size, but something else. In fact, we have no idea of what a well ordering of the reals would

¹¹For a more detailed exposition, see Freiling (1986), Mumford (1999),

be, but ZFC implies that one exists. Our assumption that CH is true means that the size of $(0, 1)$ is \aleph_1 .

Next we imagine tossing two darts at the interval $(0, 1)$. Obviously, this is a thought experiment. The darts hit two real numbers p and q . They do this (1) at random, (2) independently from one another, and (3) symmetrically, in that the order of tosses is not relevant.

Each ordinal contains its predecessor. Thus, $3 = \{0, 1, 2\}$, and ω is defined as the set of all finite ordinals, $\omega = \{0, 1, 2, 3, \dots\}$. The ordinal number ω corresponds to the cardinal number \aleph_0 , which is the cardinality of a countable infinity. The set of all countable ordinals is Ω , which corresponds to the cardinal number \aleph_1 .¹² The ordinals, which are automatically well ordered, are:

$$0, 1, 2, 3, \dots, \omega, \omega + 1, \dots, \omega + \omega, \dots, \omega^\omega, \dots, \Omega, \dots$$

The well ordering theorem and the assumption that CH is true means we can line up $(0, 1)$ with these ordinals. They will jump around, not maintaining their normal order, but every number in $(0, 1)$, including the the ones we picked out with the darts, p, q , would correspond to some ordinal. Perhaps $p = \pi/4$ might correspond with $\omega + 284$ and $q = 0.9943$ might correspond with 17.

The person who tosses the dart that lands on p can argue as follows: No matter where p is in the well ordering, there will be a countable number earlier and an uncountable number later. Appealing to a standard result in measure theory, any countable set has measure zero. Interpreting probabilities as measures, it follows that the probability of the other dart being earlier in the well ordering is zero. (It could happen, since probability zero does not imply impossibility, but there is zero chance that the event will happen.)

The second dart thrower reasons the same way. No matter which number q I hit

¹²Different symbols are used. If we were going beyond the first uncountable ordinal, here called Ω , it would be best to use: $\omega_0, \omega_1, \omega_2, \dots$ which correspond to the cardinal numbers $\aleph_0, \aleph_1, \aleph_2, \dots$

with my dart, there will be a countable number earlier and an uncountable number later in the well ordering. So, p must be later. The probability of p being earlier is zero (though still a possibility).

We have reached an absurdity. In any pair of tosses, one must be earlier than the other in the well ordering, yet by the correct reasoning of each dart tosser the probability of either being earlier is zero. The assumption that CH is true led to this, so we conclude that CH is false. Of course, logically speaking, some other assumption might be to blame. Freiling (in correspondence) tells me that he has changed his mind and now blames the well ordering principle. That would require a significant overhaul of existing set theory. I will not pursue this issue here, nor will I discuss the critics of Freiling's dart TE, since my only purpose in using this remarkable example is to illustrate a type of reasoning that is plausibly rigorous and yet, *contra* Burgess, cannot be linked to a formal proof.

Contra (3): (A rigorous proof has a structure that indicates the structure of the formal proof by containing key steps, at least in outline): I doubt that all informal proofs (rigorous or not) have a structure that is captured by a formal proof, even if a formal proof exists. Picture proofs, for instance, would typically fail, though not always. Burgess could say that shows they are bad proofs, but he would be begging the question against those who claim pictures can be genuine, convincing, even rigorous evidence. A typical *modus ponens* vs *modus tollens* standoff. I will give examples, so that readers can judge for themselves.

Burgess acknowledges that diagrams would seem to violate the spirit of his characterization of rigour, but he gives the issue only a footnote.

One thing the proposed formulation does not seem to leave room for is a role for diagrams, such as one can find in almost any issue of almost any current journal in the library: Their inclusion may play a part in convincing readers that a formal proof exists, even though diagrams are not themselves deductive steps in a formal proofs.... But the issue is a large one, deserving a monograph of its own, and I will have to set it

aside. (Burgess, Ch 2, fn25)

He is right; it is a large issue. But it is central to my case, so I will not set it aside. Rather than rule diagrams out in consequence of not meeting Burgess's condition, we might let them in (at least some of them) and claim that they do indeed exhibit key features of a formal proof. A very nice example from Littlewood might illustrate. Burgess could accept it as an instance of a rigorous proof that fits his account.

First, Littlewood defends pictures in general: "A heavy warning used to be given that pictures are not rigorous; this has never had its bluff called and has permanently frightened its victims into playing for safety. Some pictures, of course, are not rigorous, but I should say most are (and I use them whenever possible myself)" (Littlewood 1953/1986: 54). He then goes on to illustrate with a nice example.

One of the best pictorial arguments is a proof of the 'fixed point theorem' in one dimension: Let $f(x)$ be continuous and increasing in $0 \leq x \leq 1$, with values satisfying $0 \leq f(x) \leq 1$, and let $f_2(x) = f\{f(x)\}$, $f_n(x) = f\{f_{n-1}(x)\}$. Then under iteration of f every point is either a fixed point, or else converges to a fixed point. For the professional the only proof needed is Figure 1. (Littlewood 1953/1986: 55)

It would be plausible for Burgess to claim that this picture proof is indeed rigorous (as Littlewood claims), since it arguably would convince a mathematician that a formal proof exists. The crucial condition in the theorem is illustrated in the diagram, which could be seen as strongly suggesting what a formal proof would involve, namely, a convergent series of iterated terms.¹³

The great value of Littlewood's example is that it shows two important things. First, (some) diagrams can be rigorous.¹⁴ Second, (some) diagrams can satisfy the

¹³Burkill (1962, 36f) briefly remarks on the graphical representation of recurrence relations such as Littlewood's.

¹⁴There is no universal agreement on this. Giaquinto (2007) says "I would not call this a proof,

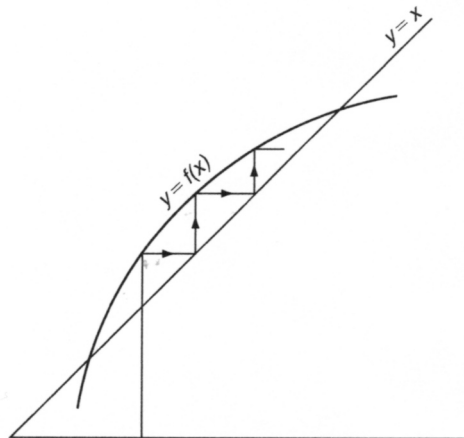


Figure 1: Littlewood's picture proof of the fixed point theorem.

Burgess condition of being convincing because they contain key steps in the formal proof. Chalk this one up for Burgess. And while we are at it, chalk up another: a simple diagrammatic proof that the number of real numbers in $(0, 1)$ is the same as in any other interval (a, b) and even the same as in \mathbf{R} . Here are diagrams that seem to convincingly show this. Figure 2 (left) shows the 1-1 correspondence between $(0, 1)$ and another finite interval (a, b) . What about the infinite interval $(-\infty, \infty)$, which is \mathbf{R} ? Figure 2 (right) makes it clear.

Of course, there are many analytic expressions that will also establish 1-1 equivalence of $(0, 1)$ and \mathbf{R} , for instance, $\tan((\pi/2)(2x - 1))$.

The question that naturally arises next is this: Are there rigorous picture proofs that clearly do not contain key ingredients of a formal proof? The answer is yes.

There is a theorem in number theory that is generally well known. The sum of the numbers up to n is equal to $(n^2 + n)/2$. Or, equivalently,

even for the professional... But it may be possible to substantiate a weaker claim than Littlewood's: seeing or visualizing the diagram can initiate a partly visual pass to discovery of the fixed point theorem." (2007, 177) He goes on to give a detailed analysis of how this might work (2007, 186-190). Notice Giaquinto's implicit appeal to the discovery-justification distinction.

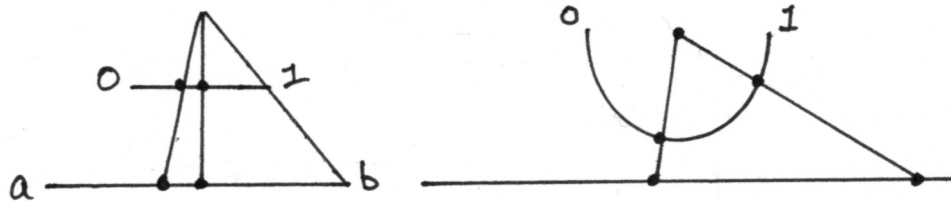


Figure 2: Equivalent intervals.

Theorem : $1 + 2 + 3 + \dots + n = n^2/2 + n/2$.

Here is the proof (Figure 3). Try to see how it works. I will explain below.

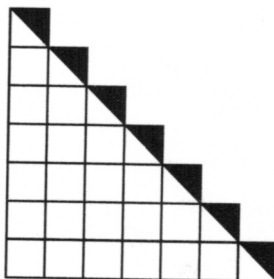


Figure 3: Picture proof of a number theory theorem.

The little squares represent numbers, so stating from the top and adding, we get: $1 + 2 + 3 + \dots + 7$. Next, consider the whole as if it were a 7×7 square cut in half by the diagonal. So the number of squares is $7^2/2$. The diagonal cut 7 squares in half; we want to add them back, $7/2$. Thus, the total number of squares is $7^2/2 + 7/2$. This proves the theorem for the special case of the number 7. But when we understand how this proof works, we are able to grasp all generality and see that $n^2/2 + n/2$ holds for any number n .

On any reasonable characterization of rigour, this picture proof is, I claim, perfectly

rigorous. But could it pass the Burgess test for an informal rigorous proof, that is, does it convince mathematicians of the existence of a formal proof by providing key steps in the formal proof? There is, of course, a formal proof. It is a proof by mathematical induction. Does the diagram suggest mathematical induction? No. The key step in induction is the passage $n \rightarrow n + 1$, whereas here the diagram could only suggest the next line, which is the passage $7 \rightarrow 8$. That is not enough. Grasping the picture proof is not an instance of some dim awareness of a formal proof by mathematical induction.¹⁵

The remarkable thing about some diagrams is that we can somehow grasp all generality. We literally see the special case of 7, but we intellectually grasp the proof for all n . This is a remarkable cognitive capacity that we have, a form of non-sensory intuition. Of course, it is fallible and sometime leads us astray, but then so do non-diagrammatic proofs.

The upshot of these examples is that the Burgess account of rigour cannot be completely correct. There are many examples where it does full justice. That is, there are proofs we can rightly call rigorous because they convince mathematicians of the existence of a formal proof, and the reason they are convincing is because they indicate key features of the formal proof. The number theory picture proof is not like that at all. It is clearly rigorous yet its correctness is not recognized by any connection to a formal proof. It stands alone.

The ability to stand alone as evidence is what links mathematical picture proofs with thought experiments. To pursue this I now turn to Norton's account.

7 Problems with Norton

Not all TEs lead to a conclusion. Some present us with a puzzle, not an answer. For instance, in the the TE known as the Ship of Theseus we are asked to imagine

¹⁵For more examples of picture proofs and for a discussion of various aspects of them, see Brown (1999/2008).

a ship that over time has every part replaced, so that eventually nothing remains of the original. We might further imagine the old parts recombined so as to form a ship. Which of these two is the true ship of Theseus? This is the fundamental metaphysical problem of identity. What makes a thing the thing that it is? The TE is a mere curiosity when it comes to ships but highly significant when it comes to people: What (if anything) makes a person the same person over time? No theory of identity is either supported or undermined by the Ship of Theseus TE. The point is to make the problem of identity manifest. Reconstructing the TE as an argument is not possible; there is no argument because there is no conclusion.

Some TEs are illustrations of a theory. They play a pedagogical rather than an evidential role. Newton's cannon TE, for instance, produces that wonderful sense of understanding how bodies such as the moon can stay in place by perpetually falling. We imagine a cannon firing cannon balls farther and farther, eventually circling the earth and returning to their starting point and, hence, repeating their orbital motion indefinitely. (See Figure 4.) Newton already calculated the result without a TE. He gave us the TE to help us understand. Again, an argument from premisses to conclusion is beside the point. Helping us to "see" is everything.

Other TEs try to establish how something is *possible*. Critics of Darwin, for instance, claim that so-called natural selection could not produce something like the eye or the opposable thumb. Darwin and biologists since him have been at pains to specify evolutionary pathways that would have these results. They are not always claiming that this is what happened. They only claim that known early conditions could be followed by a sequence of changes allowed by natural selection that resulted in the complex human eye. There is no deductive or inductive argument that begins with empirically justified premisses and concludes with the formation of the eye. It is only showing that the possibility of such a result is compatible with what we know. It is a "how possible," not a "how actual" demonstration. It is a refutation of the "not possible" claim.

I do not think cases such as these hurt Norton's main point, at least not much. They do, however, show that TEs, like real experiments, work in multiply different

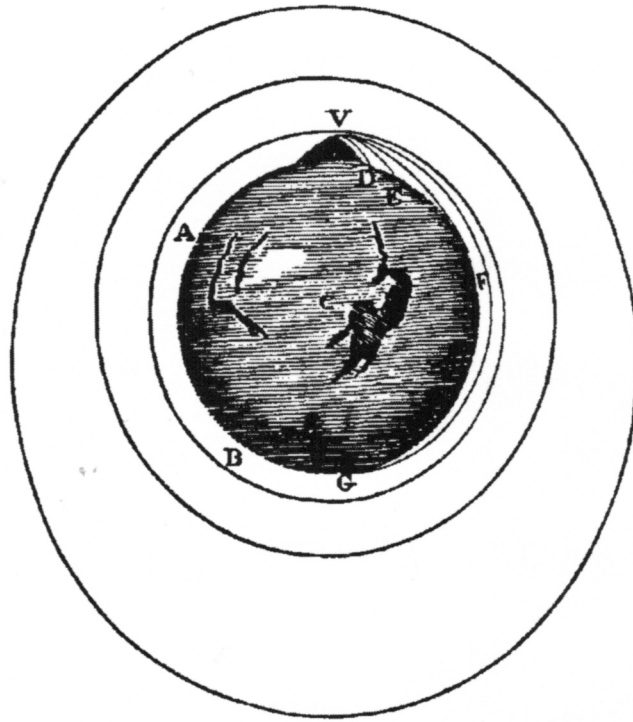


Figure 4: Newton's TE on projectile motion.

ways. We can ignore these diverse examples and restrict our discussion of Norton to TEs that play an evidential role of some sort in coming to a definite conclusion. Not all TEs can be arguments, but perhaps the main ones in physics can be. It would be a great achievement for Norton to account for the influential TEs of Galileo, Newton, Einstein, and others in terms of arguments. These are the ones we really care about.

Besides reconstructing specific examples of TEs as arguments, Norton offers another consideration for his argument view.

Reliability thesis. If thought experiments can be used reliably epistemically, then they must be arguments (construed very broadly) that justify their outcomes or are reconstructible as such arguments. (Norton 2004b, 52)

Understood strictly, Norton's point is that if a TE cannot be reconstructed as an argument then its conclusion would need to be justified some other way, such as *a priori* intuition. This is impossible on empiricist grounds. It is fair to say that the thesis assumes TEs can indeed be used reliably as evidence, and it is done on empiricist grounds. Moreover, TEs (most, if not all) can be used reliably, independently from their reconstruction as an argument. This reliability, independent from reconstruction, must be based on empirical grounds.

Let T be a specific TE and its reconstruction as an argument T' . According to Norton the thing to be evaluated ultimately is T' . As mentioned above, this is done by asking: Are the premisses true (or at least evidentially supported) and is the derivation of the conclusion deductively or inductively valid? If yes to both of these, then T' is a good TE.

However, to repeat, Norton rightly acknowledges that T can be reliably used in its own right independent from the reconstructed argument T' . "Thought experiments are commonly taken to be more than just generators of interesting hypotheses. Thought experimentation is also taken to be a reliable mode of inquiry." (2004b, 52) In other words, *TEs operate within the realm of justification, not discovery*. The reliability thesis is a welcome feature of Norton's account, since it embraces the obvious fact that normal TEs are in their own right taken to be evidence for their conclusion. Scientists, philosophers, and anyone else reading a standard TE will typically not need the reconstruction to be convinced.

Most TEs have not been reconstructed as arguments, just as most mathematical proofs have not been reconstructed as formal proofs. The views of Norton and Burgess turn out to be the same on this fundamental point. For both of them justification is a two-stage process, with the second not playing much of a role. The first stage typically convinces us by itself. But, according to Burgess and Norton, the first stage also convinces us of the existence of the second stage, the formal proof or the reconstructed argument. It is belief in the existence of a second stage that is so crucial for them.

The problem is this: if the first stage is rightly convincing by itself, why bother with the second, except, perhaps, in rare questionable cases? This is a problem for Norton and for Burgess, alike. I have already discussed the problem as it arises in Burgess. Now to the problem as it occurs in Norton.

Norton declares, “The reconstructions are generally rather straightforward and often differ little from the original narrative of the thought experiment.” (Norton 2004b, 50). Recall that Burgess says a rigorous proof typically contains the key steps in a formal proof. If Norton had made it a condition of a good TE that it outlined the structure a fully specified argument, then we would have exactly the same view. As it is, they come pretty close.

I will not attempt to provide a TE that cannot be reconstructed as an argument, but maybe we can come close.¹⁶ According to Norton, a good (but not yet reconstructed) TE is already an argument although only in rough form. Perhaps it is missing a premiss that we could fill in without too much difficulty. This is why it is both convincing and reliable. The relevant question is how close is the original TE to the reconstruction?

Before producing a TE that might pose a problem, we should consider some simple examples that strike us as non-arguments but nevertheless as plausible inferences. I am not asserting that any one of these is true, only that they illustrate candidates for an alternative to deductive/inductive inference as a way to extend out knowledge.

Descartes’s *cogito* strikes many as a good inference that is in no way an argument, either deductive or inductive. The proposition “*I think, therefore I am*” seems

¹⁶Permit me a biographical remark. From time to time I find a TE that I am confident cannot be reconstructed along Norton lines. I send it to John, who has been a friendly rival for many years, and gleefully wait for him to throw in the towel. Alas, within a day or two I am rebuffed with a plausible reconstruction, often cleverly done. I have now given up the hope of ever finding a clear counterexample. I am still not won over to his view, but I resist it for other kinds reasons. I am only concerned here with that aspect of Norton’s account of TEs that is similar to Burgess’s view of proofs. I leave other considerations aside. See Brendel (2017) for a survey of critical discussions of various aspects of Norton’s view.

correct to us as soon as we understand it. But there is no rule of inference taking us from the premiss to the conclusion.

Moore's paradox is another example where we grasp the falsity of a proposition without deductive or inductive rules playing a role. The paradoxical proposition is: "P is true, but I do not believe P." It is immediately recognized as somehow self-contradictory even though "P is true" does not contradict "I do not believe p." (Note that "P is true but I did not believe P" is unproblematic. It is only the present tense version that is paradoxical. We recognize a problem without any empirical input and without appealing to any rule of inference, deductive or inductive.

Colour exclusions are another candidate. "If X is red all over, then X is not blue all over." Neither deductive nor inductive logic will get us from the antecedent to the consequent. Nor is it an analytic truth, since the concept of *not blue* is not contained in the concept *red*. And yet it seems incontestably true.

These examples are controversial, but, to repeat, that is of little concern to us. I cite them only as possible candidates; they show the possibility of inferences that are not justified by logic. Thought experimental reasoning is sometimes more like these than like a deductive or inductive argument. They are examples of intuitions, a nonsensory form of cognition. Now to some TEs.

Galileo's TE on falling bodies can be reconstructed as an argument. The first part fits Norton's argument view very well; it is a *reductio ad absurdum* of Aristotle's account. The second part, however, involves some sort of insight and is not obviously an argument, but more like a direct perception. Of course, it could be both, an intuition and reconstructable as an argument.

Stevin's double inclined plane TE is also reconstructable as an argument, but in considering it, we come to have an epiphany. The background knowledge of the impossibility of perpetual motion machines and the intuition about symmetry do the trick. See Figure 5, top. Which way would the chain move on the frictionless double inclined plane? It is steeper on the right, but there is more weight on the

left. By adding the weights underneath, we easily solve the problem. The additional weights will pull each side symmetrically. If the whole starts to move (or, if it speeds up when already moving), then we would have the makings of a perpetual motion machine from which we could draw endless energy, which is impossible.

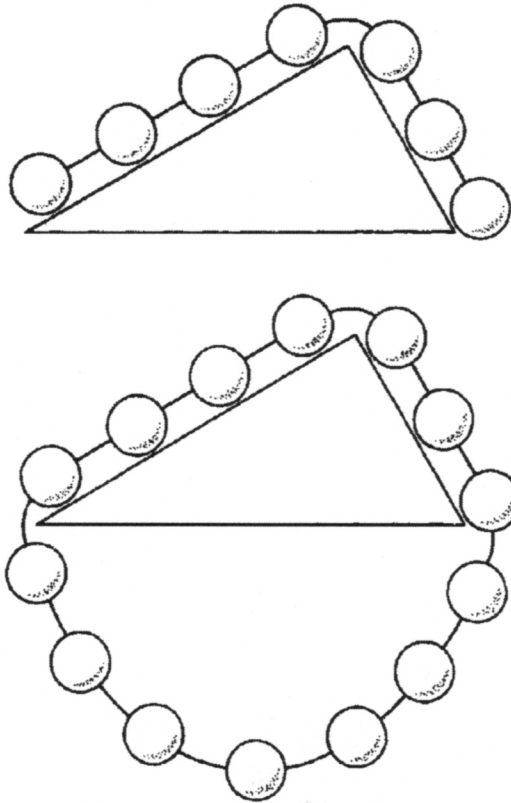


Figure 5: Stevin's double inclined plane TE.

Herman Weyl was a big fan of symmetry and had no qualms about *a priori* physics, which he saw as linked. "As far as I see all a priori statements in physics have their origin in symmetry." (Weyl 1952, 126)

Norton can reconstruct TEs as arguments with the same conclusion as Galileo or Stevin. However, I claim, the link between the initial TE and the argument

reconstruction is weak; it is not a mere filling in of missing but obvious premisses. Both of these TEs are correct; no one disputes them. And I will not dispute Norton's ability to provide a sound argument reconstruction for each. However, they also strike most people as obvious at first reading. This is the crucial point. No reconstruction is needed to appreciate their veracity. The reconstruction contributes nothing, just as the Burgess's formal proof contributes nothing to the evidence.

8 Concluding Remarks

Time to take stock. Burgess and Norton have put forward plausible, detailed accounts of mathematical rigour and thought experiments, respectively. In one important respect they are identical. Burgess justifies the standard evidence we find in mathematics, explaining that it is indeed rigorous because it convinces mathematicians of the existence of a formal proof, the gold standard. Moreover, the ordinary proof typically contains the key steps of the formal proof. This has the virtue of minimizing the epistemic gap between them. Norton similarly claims that any TE can be reconstructed as an argument. Moreover, he too, claims that typically we can easily recognize the argument structure of the initial TE, making the epistemic gap small, as well.

Against Burgess I presented examples, especially picture proofs, that would seem to be genuine trustworthy proofs that do not correspond to formal proofs. And against Norton I presented examples where the reconstruction would seem to be irrelevant, since the initial TEs are utterly convincing in their own right.

One particularly interesting feature of these two accounts is that they do not play on the traditional discovery-justification distinct. Whatever might go into the discovery of a proof or TE in the first place, is of no concern to the two-stage justification process. Burgess proofs and Norton TEs are already in the realm of justification. We might even call it early and late justification, corresponding to the initial proof or TE and then to the formal proof or reconstructed argument. What

I hope to have shown is that this way of looking at things is a mistake, at least if it is taken to be universal. The early stage is complete evidence for a theorem or a TE as it stands.

New questions arises: What is it about the initial proof or the initial TE that accounts for the recognition of its sufficiency as evidence? Burgess wants rigorous proofs to build on what went before. But how reliable is that? It could be genuine objective knowledge or it could be arbitrary stipulations. The examples and the history of mathematics suggest something closer to the former, while theorems about chess are, of course, more like the latter. I won't argue for it here, but I remain convinced of a broadly platonistic outlook, and think it intact after considering Burgess's account of rigour. I am inclined to say something similar in light of Norton's account of TEs. We seem to have some cognitive ability that allows us to grasp truths in the abstract realm, which is the source of our knowledge. I have argued for this elsewhere and will not repeat the case for it here.¹⁷ The initial proof or TE guides our intuitions (at least in some cases) of the abstract realm, which is why the proof or TE not only seems right but reliably is right.

The case against Burgess and Norton is not merely a refutation of their views, but also opens new possibilities for evidence. Mathematical proofs need not be tied to formal proofs, and TEs need not be tied to derivations from empirically justified premisses. New ways of learning are possible.

Let me end with a profession of hope. I think the Burgess account of mathematical rigour is plausible and probably right for most cases — but not for everything. And I think the Norton account of TEs is also right for many cases — but is not the whole story. The abstract realm, the home of proofs and TEs, is so rich and varied in its content that there is very likely not a single kind of evidence. A pluralistic outlook seems best. Wittgenstein remarked that mathematics is a “motley of techniques of proof.” (Wittgenstein 1956/78 46). He meant it somewhat

¹⁷See Brown (1991/2011 and 1999/2008.)

disparagingly — I do not. The same could be said about TEs. Yes, proofs are evidence, but so are diagrams, and intuitions, and statistical arguments, and much more. And they not only yield truths, but sometimes also explanations and understanding. (To be fair to both Burgess and Norton, explanation and insight are not intended to be part of their concerns.) We accept things because of the available evidence for them. But there is no reason to confine ourselves to formal proofs or to empirical observations and deductive or inductive logic as the only legitimate forms of evidence. Rational belief runs well beyond what these can justify by either Burgess or Norton means. In mathematics and in thought experiments — and everywhere else — we should be guided by these fundamental epistemic principles: (1) We have many cognitive capacities, including (fallible) intuition. (2) Truth and understanding are achievable, although certainty is not. (3) We are always in the middle working out in all directions, often using newly invented means as we go.

What do we want? Truth and explanation and understanding. We often get it with TEs; the initial version is usually sufficient. Reconstruction and rigour can't deliver much. They can give us truth and certainty only if we start with truth and certainty. How likely is that, given empirical fallibility or that mathematical axioms are fallibly self-evident or outright conjectures? Reconstruction and rigour do not even try for explanation and understanding. At best, they maintain them, if they are in the initial proof or TE. The problem with making much of rigour is that it leaves out a great deal of mathematical activity. Huge advances are made with definitions. Think of Dedekind's definition of real numbers. All sorts of important things rigorously follow from it; modern analysis largely rests on it. But no one has ever proved that it is correct. Instead, we see it as a brilliant solution to a problem of dealing with irrational numbers and we embraced it for that reason. It is outside the realm of rigour. But it is inside the realm of explanation and understanding. Is it certain? No, but we can reasonably hope it is true.

Acknowledgements

Thanks to Silvia De Toffoli who stimulated my interest in this topic. I have long been interested in TEs and in the philosophy of mathematics, but it was only after a talk by Silvia in the fall of 2020 that I recognized the interesting similarities between the views of Burgess and Norton. Of course, I am always indebted to John Norton for comments on anything connected to TEs, including valuable comments on this paper. We have been discussing the issue for longer than either cares to remember.

References

- Avigad, J. (2006) “Mathematical Method and Proof,” *Synthese* 153: 105159
- Avigad, J. (2008) “Understanding Proofs,” in Paolo Mancosu, editor, *The Philosophy of Mathematical Practice*, Oxford University, 317-353.
- Avigad, J. (2020) “Reliability of Mathematical Inference”, *Synthese*
<https://doi.org/10.1007/s11229-019-02524-y>
- Azzouni, J. (2013a) “The Relationship of Derivations in Artificial Languages to Ordinary Rigorous Mathematical Proof,” *Philosophia Mathematica* (III) 21 (2013), 247254
- Azzouni, J. (2013b) “That We See That Some Diagrammatic Proofs Are Perfectly Rigorous,” *Philosophia Mathematica* (III) 21 (2013), 323338
- Brendel, Elke (2017) “The Argument View: Are Thought Experiments Mere Picturesque Arguments?”, in Stuart et al (eds) (2017), 281-292
- Brown, James Robert (1991/2011) *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, London and New York: Routledge
- Brown, James Robert (1999/2008) *The Philosophy of Mathematics: An Introduction to the World of Proofs and Pictures*, London and New York:

Routledge

Brown, James Robert (2004) "Why Thought Experiments Transcend Empiricism," in Christopher Hitchcock (ed.) *Contemporary Debates in the Philosophy of Science*, Oxford: Blackwell

Burgess, John (2015) *Rigor and Structure*, Oxford: Oxford University Press

Burkill, J.C. (1962) *A First Course in Mathematical Analysis*, Cambridge: Cambridge University Press

De Regt, H., S. Leonelli, and K. Eigner, (2009) eds., *Scientific Understanding*, Pittsburgh: University of Pittsburgh Press

Detlefsen, M. (2008) "Proof: Its Nature and Significance," B. Gold and R Simons (eds) *Proof and Other Dilemmas*, Mathematical Association of America

Freiling, C. (1986) "Axioms of symmetry: throwing darts at the real number line," *The Journal of Symbolic Logic*, 51 (1): 190-200

Giaquinto, Marcus (2007) *Visual Thinking in Mathematics: An Epistemological Study*, Oxford: Oxford University Press

Horgan, J. (1993) "The Death of Proof," *Scientific American* 269, No. 4, 92-103.

Lakatos, I. (1976), *Proofs and Refutations*, Cambridge: Cambridge University Press

Mac Lane, S. (1986) *Mathematics: Form and Function*, New York: Springer-Verlag

Mumford, D (1999) "The dawning of the age of stochasticity," in *Mathematics: Frontiers and Perspectives 2000*, American Mathematical Society, 197-218.

Norton, John (1985) "What was Einstein's Principle of Equivalence?," *Studies in the History and Philosophy of Science*, vol. 16, 203-46.

- Norton, John (1991) "Thought Experiment's in Einsteins Work," in T. Horowitz and G. Massey (eds.) *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield.
- Norton, John (1996) Are Thought Experiments Just What You Thought? *Canadian Journal of Philosophy*, vol. 26, 333-66.
- Norton, John (2004a) On Thought Experiments: Is There More to the Argument? *Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association, Philosophy of Science*, vol. 71, 1139-51.
- Norton, John (2004b) "Why Thought Experiments Do Not Transcend Empiricism," in Christopher Hitchcock (ed.) *Contemporary Debates in the Philosophy of Science*, Oxford: Blackwell, pp. 44-66.
- Paris and Harrington (1977) "A Mathematical Incompleteness in Peano Arithmetic," in Jon Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland
- Rav, Y. (1999) "Why Do We Prove Theorems?" *Philosophia Mathematica*, (3) Vol. 7 (1999), 5-41
- Rota, Gian-Carlo (1997/2008) *Indiscrete Thoughts*, Birkhäuser: Boston
- Royden, H.L. (1988) *Real Analysis*, 3rd edition, New York: Macmillan
- Stuart, M., Y. Fehige, J.R. Brown (eds.) (2017) *The Routledge Companion to Thought Experiments*, London and New York: Routledge
- Stuart, M. (2017) "How Thought Experiments Increase Understanding," in Stuart et al eds. (2017)
- Stuart, M. (2020) "The material theory of induction and the epistemology of thought experiments," *Studies in History and Philosophy of Science* 83, 17-27
- Weyl, H. (1952) *Symmetry*, Princeton: Princeton University Press
- Wittgenstein, L. (1956/1978), *Remarks on the Foundations of Mathematics*,

Revised Edition, G.H. von Wright, R. Rhees and G.E.M. Anscombe (eds.);
translated by G.E.M Anscombe, Oxford: Basil Blackwell