Mathematical Gettier Cases and Their Implications

Neil Barton*

18 May 2023†

Abstract

Let *mathematical justification* be the kind of justification obtained when a mathematician takes themselves to have *proved* a theorem. Are Gettier cases possible for this kind of justification? At first sight we might think not: The standard for mathematical justification is proof and, since proof is bound at the hip with truth, there is no possibility of having an epistemically lucky justification of a true mathematical proposition. In this paper, I challenge this idea by arguing that there is conception of mathematical justification which is *fallibilist* (in addition to *infallibilist* accounts). I argue that for the fallibilist conception, nontrivial Gettier cases are possible (and indeed actual). I indicate some upshots for mathematical practice, in particular regarding folklore theorems and pluralism.

Keywords: Gettier case, mathematical justification, proof, folklore theorem, pluralism

*IFIKK, University of Oslo. E-mail: n.a.barton@ifikk.uio.no.

†I am very grateful to Carolin Antos, Zoe Ashton, Kieran Britt, John Burgess, Kenny Easwaran, Eva-Maria Engelen, Ben Fairbairn, Carrie Jenkins, Deborah Kant, Konstantinos Konstantinou, Lukas Koschat, Daniel Kuby, Beau Madison-Mount, Tim Maudlin, Colin McLarty, Thomas Müller, Vadim Kulikov, Jillian Roberts, Chris Scambler, Dima Sinapova, Martin Steenhagen, Colin Rittberg, Fenner Tanswell, Dan Waxman, Keith Weber, and Jack Woods, as well as audiences in Cambridge, Konstanz, New York, and Princeton for helpful discussion. Special mention must be made of Silvia De Toffoli—several long discussions about the paper helped improve it considerably. I would also like to thank the FWF (Austrian Science Fund) through Project P 28420 (*The Hyperuniverse Programme*), the VolkswagenStiftung through the project *Forcing: Conceptual Change in the Foundations of Mathematics*, and the Research Council of Norway through project *Infinity and Intensionality: Towards A New Synthesis* for their support.

Introduction

Let *mathematical justification* be the kind of justification obtained when we take ourselves to have *proved* a result (so the sort of thing that appears after the term "*Proof.*" in a mathematics journal or textbook). This paper concerns the (epistemologically familiar) phenomenon of Gettier-cases, how these might relate to mathematical justification, and what the upshots for mathematical practice might be.

A standard Gettier case is a situation in which an agent has justified true belief in a proposition, but that justification is *lucky*. One might think that it such Gettiering is impossible for mathematical justification. If one thinks that mathematical justification is obtained by knowing a proof from axioms, epistemic luck is *impossible* since the axioms are *true* and the rules of proof *preserve truth*.

However there *do* seem to be examples of Gettier cases in mathematics. In [Voevodsky, 2014a], Vladimir Voevodsky (a 2002 winner of the Field's Medal) relates the following episode from the development of his work on motivic cohomology:[1]

> The approach to motivic cohomology that I developed with Andrei Suslin and Eric Friedlander [relied][2] on my paper 'Cohomological Theory of Presheaves with Transfers', which was written when I was a Member at the Institute in 1992–93. In 1999–2000, again at the IAS, I was giving a series of lectures, and Pierre Deligne (Professor in the School of Mathematics) was taking notes and checking every step of my arguments. Only then did I discover that the proof of a key lemma in my paper contained a mistake and that the lemma, as stated, could not be salvaged. Fortunately, I was able to prove a weaker and more complicated lemma, which turned out to be sufficient for all applications. A corrected sequence of arguments was published in 2006.
>
> This story got me scared. Starting from 1993, multiple groups of mathematicians studied my paper at seminars and used it in their work and none of them noticed the mistake. [Voevodsky, 2014a, p. 8]

What we have here is, *prima facie*, an example of mathematical justification being *Gettiered*: The main results of Voevodsky's 'Cohomolog-

[1] I thank Lukas Koschat for pointing out that the talk [Voevodsky, 2014b] (a talk upon which [Voevodsky, 2014a] is based) contained particularly acute examples of the phenomenon I'll discuss throughout this paper.

[2] In this excerpt Voevodsky is talking about the avoidance of a different problematic 'lemma' (Bloch's lemma), and I have suppressed this detail for clarity.

ical Theory' (as I'll abbreviate the paper from hereon out) were correct, but the original 'proof' was flawed. Both Voevodsky and the scholars using his results in their proofs in the period when the error had not been diagnosed (1993–1999) had (or at least took themselves to have) *mathematically justified* beliefs—they were using (true) results that were widely accepted within the community they were a part of in proving more theorems. Intuitively speaking, however, they did *not* have knowledge since they were epistemically lucky that the relevant results (with flawed 'proofs') from 'Cohomological Theory' were, in fact, true.

There is a philosophical puzzle to be resolved here: How do we explain the apparent Gettier-phenomenon in mathematics whilst retaining the close link between mathematical justification and something like proof? Further, what might the upshots be of this phenomenon for how we do mathematics? I will argue for the following claims:

Main Claims.

1. Whilst one conception of mathematical justification is infallibilist, an important conception for understanding mathematics is *fallibilist* (i.e. not necessarily factive).
2. Such a fallibilist account indicates several *actual* instances of the Gettier-phenomenon concerning mathematical justification, and different kinds of *luck* at play.
3. These Gettier-cases indicate two dimensions in which mathematical justification can vary: (I) An *internalist* dimension (for the epistemic agent to have understood the conceptual dependencies of their justification) and (II) an *externalist* dimension (the parts of the justification to fit the mathematical facts).
4. In turn, these dimensions help to explain certain upshots for mathematical practice.

Here's how I'll argue for these claims: §1 outlines the different approaches one might take to *mathematical justification*. In particular, I'll point to the fact that there are both infallibilist and fallibilist conceptions. I'll press the point that the *fallibilist* conception is one that is operative in mathematical practice. We'll see that this kind of justification can be understood as underwritten by *simil-proofs*: Mathematical arguments that can be shared with and checked by the rest of the community. §2 provides some mathematical Gettier cases on the basis of this kind of justification via non-trivial errors or gaps (§2.1), difficulties with selection of the axioms (§2.2), and the use of lemmas in simil-proofs as 'black boxes' (§2.3), as well as isolating the different kinds of *epistemic luck* at play. I then argue (§3) that the manner in which these cases are generated suggests two dimensions in which an agent's understanding of a simil-proof can vary, an internalist dimension of understanding the conceptual dependencies and interrelationships of the resources involved in a simil-proof, and an externalist criterion of how well the steps taken in a simil-proof fit with the mathematical facts. These criteria, I argue, highlight some upshots for mathematical practice (§4), in particular concerning folklore theorems and pluralism in theory choice. I then conclude (§5) with a summary and some open questions for moving forward.

1 Mathematical Justification

No account of Gettier cases is complete without a thorough characterisation of the notion of *justification*. Before we get into the details of the cases we thus have to tackle the important and difficult question: *What is mathematical justification*?

This is a tricky problem: Say too much and one risks making one's arguments too narrow, too little and we run the risk of vicious imprecision. Rather than settling on one account, I'll examine two rough conceptions—fallibilist and infallibilist—and argue that each is important for understanding mathematical justification.

With this in mind, let's start by demarcating *mathematical justification* from *justification of a mathematical proposition*. The latter is an exceptionally broad notion that admits of easy Gettiering in a manner not tremendously interesting for mathematical practice. For instance, suppose that a close mathematician friend tells me that Goldbach's Conjecture is true. I trust them, and so I believe the conjecture. As it turns out, let's suppose that Goldbach's Conjecture is true but my friend was just playing a prank on me (and they had no idea whether Goldbach's Conjecture is true). Then I have justified true belief in a mathematical proposition, but intuitively speaking I don't have knowledge.[3]

Whilst much epistemology focusses on this kind of example, it is however *not* what I am interested in here. My focus is rather on the kind of epistemic justification conferred when a mathematician takes themselves to have *proved a theorem* and/or *possess* or *know* a proof of said theorem. My interest is thus in the kinds of justifications agents take themselves to have after working through arguments that occur after the term "*Proof.*" in mathematics texts like (reputable) journals

[3]See [Paseau, 2015] for some examples of justification of mathematical propositions that are not directly mathematical justification.

and textbooks. In particular, I want to examine if and how this kind of justification can be Gettiered, and, if so, whether there are upshots for the communal production and vetting of the relevant mathematical artefacts.

1.1 Warm up: Mathematical justification as possession of a proof

One historically prevalent conception of mathematical justification has been the following:

Assumption of Proof-Theoretic Justification. (APT) A subject S has mathematical justification for a believed proposition P just in case they have/know a proof of P from axioms for the relevant mathematical subject matter.

This assumption was widely taken to be the 'default' for much of the 20th century. For example, Giaquinto writes:

> It was simply assumed that mathematical knowledge would have to be a matter of proof, that is, deduction from the axioms... ([Giaquinto, 2007], p. 5)[4]

This assumption seems to vitiate the possibility of Gettier cases concerning mathematical justification. Since axioms are true, and since logical inferences preserve truth, it is hard to see how a mathematical justification could be lucky. By definition, it seems, the standards of mathematical justification prohibit the possibility of a Gettier case.

The problem is brought into sharper focus if we consider a standard template for generating a Gettier case. Linda Zagzebski showed that as long as there is a gap between justification and truth, Gettier cases are possible.[5] One simply takes a case in which a proposition is justified but false, and via epistemic luck modifies the case to make the proposition true. The example she gives is that of someone (Mary)

[4] Giaquinto continues:

> ...the only question, then, was how the axioms and inference rules of the relevant axiomatic systems could be justified. Thus, the epistemology of individual discovery simply dropped off the agenda. So did any concern with actual thinking in mathematics.

Some of these subtleties to do with justification of axioms and individual discovery will re-appear later.

[5] See [Zagzebski, 1994].

mis-perceiving her husband's brother (who looks somewhat like her husband) as her husband (whilst the brother sits in a chair). Taking her belief to be *My husband is sitting in the living room*, we can take her false-but-justified belief and 'make' it true by epistemic luck by having her husband sit out of eyeshot in the room. This makes it harder still to see how we might generate a mathematical Gettier case, since on the usual understanding of mathematical truth propositions are true or false by necessity. It is thus not possible to use the Zagzebski inescapability template to start with a *false* mathematical belief and then *modify the situation* to make it true by epistemic luck.

Emphatically *not* every accepted mathematical argument satisfies the APT, however. The example of Voevodsky from the introduction shows that there are often mathematical arguments that are accepted, and indeed become central in an area, that are not strictly proofs (in the sense of correct arguments from the accepted axioms). Work in mathematics journals is fundamentally *fallibilist* in nature: It is possible for accepted to exhibit minor errors (e.g. typos), significant gaps, and even major errors. With this in mind, let's turn to a fallibilist conception of mathematical justification.

1.2 Fallibilist mathematical justification and simil-proofs

The core idea at the centre of fallibilist conceptions of mathematical justification is that an agent can produce a convincing and subsequently accepted mathematical argument without it being a bona fide proof from true mathematical axioms (see [Dove, 2003], [Davis and Hersh, 1999], and [De Toffoli, 2021]). As we'll see, we *can* have mathematical Gettier cases for *fallibilist* accounts.[6]

Probably the most developed fallibilist account is Silva De Toffoli's [De Toffoli, 2021], and we'll use this to illustrate how fallibilist justification can lead to Gettiering. The key notion for us will be that of *simil-proof*, and the idea of *knowing* or *possessing* a simil-proof. Roughly speaking, a simil-proof is an argument that has been accepted by members of the community and meets certain minimal standards.

Before we explicitly define simil-proofs, we need the notion of *shareability* of a mathematical argument. De Toffoli characterises it as follows:

[6] Some readers may feel uneasy already and hold that such arguments do not constitute 'real' mathematical justifications. I will return to this point shortly (in §1.3).

> An argument is *shareable* if its content and supposed correctness could be grasped by relevantly trained human minds from a (possibly enthymematic) perceptible instance of a presentation of it. [De Toffoli, 2021, p. 830]

A simil-proof can then be defined as follows:

Simil-Proofs. An argument is a *Simil-Proof* (SP) when it is [(i)] shareable, and [(ii)] some agents who have judged all its parts to be correct as a result of checking accept it as a proof. Moreover, [(iii)] the argument broadly satisfies the standards of acceptability of the mathematical community to which it is addressed. [De Toffoli, 2021, p. 835, (i)–(iii) added]

We should immediately note that not every simil-proof is a proof. A mathematical argument can satisfy all of (i) to (iii), but ultimately be fallacious.

De Toffoli defines an agent *possessing a simil-proof* (or *having* a simil-proof) in the following manner:

Having-SP. S has an SP of C if and only if, when prompted to articulate a reason for her belief in C, in the appropriate context, S would (in good faith) share the SP. Moreover, S would be able to appropriately reply to challenges and hold related dispositions. For instance, if the validity of an inferential step of her SP is questioned, S would be disposed to abandon it if she cannot defend it. [De Toffoli, 2021, p. 839]

Often we talk of *knowing* a simil-proof in addition to *having* a simil-proof, and so I will use these terms (and their cognates) interchangeably. De Toffoli then defines mathematical justification as follows:

SP-Justification. A subject S's belief that mathematical claim C (in need of a proof) is mathematically justified if and only if S has an SP. [De Toffoli, 2021, p. 837]

We should note that SP-justification is compatible with many views on the nature of mathematical justification. [Rav, 1999] considers a view of mathematical practice that is irreducibly semantic, and the job of a mathematical justification is to indicate these semantic connections. Call this the **semantic account**. [Azzouni, 2004] considers a view on which proofs indicate the existence of derivations—the so-called **derivation indicator account** (see [Avigad, 2020] for a recent defence). Another view is the **recipe account** which holds that (at least some) simil-proofs are akin to recipes for reconstructing reasoning ([Tanswell, F] provides a recent proposal). We can also consider Catarina Dutilh Novaes' recent **dialogical account** of mathematical practice, on which mathematical justification is conceived of via particular dialogical games played between a prover and a sceptic (see [Novaes, 2020]). The account of mathematical justification in terms of knowing/having simil-proofs is compatible with each of these ways of spelling out mathematical justification. We can take simil-proofs to be presentations of semantic relations, indicating the existence of derivations, providing recipes for reconstructing reasoning, or particular prover-sceptic games.

Importantly: On each reconstruction of the role of simil-proofs we can have a *fallibilist* account. A simil-proof may fail to indicate semantic relations that really hold, it may fail to provide an appropriate indication of a derivation, the recipe may be fundamentally flawed, or the prover and sceptic may have failed in their duties.

Before we move on to consideration of an infallibilist way of construing mathematical justification, I want to forestall a couple of natural objections. One might feel that the definition of simil-proof is ambiguous in a couple of respects. First, it is unclear whether the 'agents' who are judging the parts of a simil-proof to be correct can be identical with the agent(s) who produced the proof. We might not want to rule out that a pioneer or isolated mathematician can be mathematically justified[7] or that justification is obtained at the point of discovery rather than acceptance of the mathematical argument. This point need not detain us—the cases we shall consider concern mainstream mathematical arguments that have been widely accepted.

Next, we might worry about the extent to which acceptability standards are addressed to specific communities. For instance, we might want communities to be able to criticise one another if they think that the standards are too lax (or perhaps too strict). In this case, we want to be able to say that one community *should* be accepting fewer (or more) mathematical arguments as simil-proofs. Making such criticisms seems to require acceptability standards that go beyond community-relative ones. This said, mathematics is quite a broad church, and acceptability standards may (as a point of fact) vary substantially from community to community. It may be that there is a 'core' of values shared across mathematics and hence an 'absolute' notion of what is to count has a simil-proof, or it may be that whether or not an argument is a simil-proof is an essentially

[7] De Toffoli in fact considers this notion (see, p.830 of her [De Toffoli, 2021]).

community-relative matter.[8] Again, whilst this is a subtle issue, it need not detain us further: Whether or not an agent is mathematically justified is 'community-relative' or 'absolute', we will still be able to generate certain kinds of Gettier-cases and analyse their upshots *relative* to a specific given standard.

Finally, we should identify an issue that will be important later (especially in §3 and §4): *being shareable* and *possessing a simil-proof* can be viewed as matters of *degree*.

Let's start with shareability. Arguments vary according to how easily they can be checked by suitably trained agents. Some simil-proofs are relatively hard to check or very gappy, others are clearer and easier to follow. Thus, whether a simil-proof is shareable can be viewed as a matter of degree—arguments can be more or less shareable. I do not think this is an objection against De Toffoli's account—it is reasonable to suppose that there is a minimal bar (possibly with fuzzy boundaries) that arguments must clear in order to count as definitively shareable. Thus, whilst I will continue to talk of *degrees of shareability*, I will also talk of arguments being *shareable* (simpliciter). However, it's important to note that this is a dimension along which a simil-proof can vary, and this will be important when we come think of what the upshots of the Gettier phenomenon might be.

Let's also note at this point that having or knowing a simil-proof can also be spoken of as a matter of degree—more or less effort might be required to defend particular steps. A familiar feeling to many mathematicians is that of simil-proofs beginning to fade in memory if they have not been checked or prepared recently. In this sense, we might say that we know or possess a simil-proof to different degrees dependent upon how easily we can defend particular steps. Again, I don't think that this is an objection—there is a (possibly fuzzy) bar that we can take to be cleared when we want to say that an agent has a simil-proof *simpliciter*. However, it is important to note that the modal form of simil-proof possession in De Toffoli's characterisation is somewhat tricky to articulate, in particular regarding the level of idealisation permitted. I do *not*, for instance, have or know a simil-proof of the Poincaré Conjecture simply by carrying around a copy of Perelman's proof with me wherever I go, and being prepared to defend the

[8] We should note that in really egregious cases, there definitely could be arguments published in mathematics journals that do not count as simil-proofs on any reasonable standard. For example, a corrupt editor who accepts a mathematician's request to have their (wholly error strewn) paper sent to a friend can result in the publication of a non-simil-proof mathematical argument, since the purported simil-proof might fail to meet the standards acceptable to any reasonable mathematical community.

relevant steps if prompted (perhaps after several years or lifetimes of study). I do think that there is a reasonable sense in which there are simil-proofs for which I could articulate particular steps given enough time (e.g. I would probably have to revisit a textbook to refresh my memory for the proofs of the incompleteness theorems) and others which I could not (e.g. Perelman's proof is currently out of reach for me). It's important to note though that even in cases where it's very plausible that I know a simil-proof of a proposition (e.g. the incompleteness theorems) this might not be backed up by perfect simil-proof possession, and there are others where my understanding of the relevant simil-proof is middling at best. We will return to this issue in §3 when we isolate different dimensions in which simil-proofs and possession can vary further.

At this point, some readers will feel that fallibilist justification strays too far from the classic APT-based account of justification. In the rest of this section, I want to return to infallibilist accounts and suggest that both infallibilist and fallibilist conceptions are interesting for mathematical practice.

1.3 Infallibilist mathematical justification: redux

Of course some authors will reject the claim that possession of a flawed simil-proof provides mathematical justification. There are various *in*fallibilist accounts. This includes infallibilist accounts of *mathematical* justification (e.g. the earlier discussed APT) but also infallibilist accounts of justification more widely (e.g. Littlejohn's account in [Littlejohn, 2012]). For such accounts, SP-justification does not match up with mathematical justification, since the former is fallible and the latter is not.

There are some attempts to support such an account of mathematical justification. In [Gödel, 1953], Kurt Gödel describes a proof as "a sequence of thoughts convincing a sound mind" [Gödel, 1953, p. 341]. This suggests a conception of mathematical justification on which it is not enough that we merely have a simil-proof, but rather that the agent has *understood* the relevant steps in the argument. Developing this idea, Leitgeb writes:

> the mathematical community's sense of proving a statement from other statements involves connecting the latter statements to the former by intermediate steps (i) which preserve truth and (ii) which make it evident why truth is preserved from one step to the next. [Leitgeb, 2009, p. 270]

This conception of mathematical justification is related to the APT—it provides a sharpening of what it is to have (provided) a mathematical justification. For Leitgeb, a mathematical justification is possessing an argument where it is *evident why* each step preserves truth. Clearly such an account of mathematical justification is infallibilist (presuming that we start from true principles).

Whilst adjudicating the debate between fallibilist and infallibilist conceptions of mathematical justification won't be possible here (really I think this will need multiple papers), some remarks on the role of each will help to elucidate the place of Gettier cases within the discussion. Really, I think that there are two legitimately interesting conceptions of justification in play here, and each deserves attention. Part of what I will do by specifying Gettier cases is to try and draw some relationships between the two. Let's start by noting that SP-justification (and fallibilist conceptions more widely) are clearly interesting for understanding mathematical practice. This is because the kinds of artefacts that we are *actually* confronted with as epistemic agents practising mathematics are fallibilist (as the example of Voevodsky neatly illustrates). Indeed, it may well be that infallibilist mathematical justification is rather inscrutable. But obviously the infallibilist conception is important too; indeed some take this kind of self-evidence of properly understood arguments to be a *hallmark* of mathematical practice. When relating the two, we would like as many cases of fallibilist justification to be instances of the infallibilist conception. And as I'll show, by examining the Gettiering of the fallibilist conception of mathematical justification, we can learn about possible desiderata for increasing the liklihood that we have an instance of the *infallibilist* conception too. With this in mind, let's move on to the cases.

2 Some mathematical Gettier cases

We now have the the distinction between infallibilist and fallibilist conceptions of mathematical justification. In this section, I'll explain how fallibilist accounts of mathematical justification lead to actual kinds of Gettier case. This will divide into three main kinds: (§2.1) significant gaps/errors, (§2.2) the selection of axioms, and (§2.3) the use of lemmas as black boxes. On the way, I'll make explicit the different kinds of *luck* at play in each case. First though, we need to make a couple of preliminary comments dealing with motivating the consideration of *Gettier* cases (as opposed to merely mathematically justified *false* beliefs).

Of course simil-proofs can lead to falsehoods, and indeed this phenomenon is interesting from a philosophical perspective. However, such falsehoods are often diagnosed *relatively* quickly. For example, in set theory [Džamonja and Shelah, 1999] claimed to have shown that there are models of set theory in which both ♣ (a statement about the combinatorics of sets) is true but there are no Suslin trees. Their results, it was subsequently discovered, contradicted a well-established theorem (namely Miyamoto's Theorem) and so the simil-proof was recognised to be flawed. However, [Džamonja and Shelah, 1999] contains much useful material, even if one result fails to go through. Moreover, the diagnosis of the error proved to be mathematically useful in articulating some relevant details (as explained in [Brendle, 2006]).[9] [Rav, 1999] provides a host of other examples.

The Gettier phenomenon is importantly different, in that the error might be harder to diagnose in virtue of the truth of the conclusion. Since the proposition in question is *true*, contradictions will not show up in the rest of our mathematical reasoning (putting aside dialethic conceptions of mathematics). The Gettier phenomenon is thus especially interesting as it identifies cases where our mathematical justification is not (intuitively) leading to knowledge, but also where this might be hard to diagnose—it is far harder to spot an error in a simil-proof (by and large) than to simply realise *that* a simil-proof must be flawed (without knowing where) because it leads to a falsehood.[10]

We should start by setting aside some easy and trivial Gettier cases. For example, consider a case where computer assistance is being used to prove a theorem. This now occurs often, as can be seen with the proofs of the Four Colour Theorem or Kepler Conjecture, and the pervasive use of GAP in classification problems in group theory.[11] However, we don't need anything so complicated, one can see such uses as roughly analogous to the use of an electronic calculator in computing steps in a standard high-school paper (where of course calculator assistance is allowed as part of the set-up!). Let's suppose that the relevant computational device is malfunctioning in some way, but complementary malfunctions just happen to cancel each other out or be tailored to the problem at hand. Here we might say that intuitively speaking we do not have knowledge despite having relevant mathematical justification (i.e. possession of a simil-proof). It is not that there is

[9] See [Brendle, 2006], p. 45, footnote 1 for some discussion and further references. I thank Daniel Soukup for bringing this example to my attention.

[10] Indeed this is the structure of some of the dialectic in [Voevodsky, 2014a], a counterexample to a simil-proof was found, but nowhere was the specific error identified.

[11] GAP is a system for computational discrete algebra, and in particular is used to computer check properties of finite groups. Thanks to Ben Fairbairn for helpful discussion and bringing this example to my attention.

anything wrong with our mathematical justification *per se*, but rather that something has gone wrong with extraneous empirical facts upon which that justification depends for verification. In this way, the situation is somewhat analogous to the Russelian 'stopped watch' case in which at (say) 7am an agent checks a (normally well-functioning) watch that has stopped indicating the correct hour (in this case with hands pointing at 7 and 12). Here, the relevant epistemic agent has done nothing wrong in checking a piece of external apparatus, and the relevant malfunction has just 'happened' to correspond neatly onto the way the world is at that point of time. The upshots of this kind of error are relatively obvious (e.g. where possible run software verification on multiple different pieces of hardware) and so I won't consider these kinds of case any further.[12]

2.1 Non-trivial gaps and hard-to-detect errors

More interesting are the following kinds of cases: There are situations in which an agent has a simil-proof of a true proposition, but where the simil-proof is defective in some way. It might, for example, contain a significant gap that (unaware to the agent) requires significant patching, or perhaps even a non-trivial flat-out error. One such example of this kind was mentioned in the introduction concerning some of Voevodsky's results in his 'Cohomological Theory' paper: Voevodsky himself had a simil-proof, but not *knowledge* since it turned out to be *lucky* (in virtue of the error) that many of the propositions contained therein were in fact true. Examples can easily be multiplied, especially by looking at long or complicated proofs. In the enormous literature proving the theorem classification of finite simple groups into various kinds (cyclic, alternating Lie, or one of the sporadic exceptions) several substantial gaps were found across the history of specific results involved in establishing the theorem (see [Solomon, 1995] for discussion). These gaps were significant and required fixing—in contrast to the present day where many agents are regarded as having mathematical justification of some of these results, even if small and trivial errors in the simil-proofs remain.[13,14]

[12] These kinds of examples, and a modification of the safety criterion to deal with them, are considered by [Pritchard, 2012].

[13] We should note that whilst there is a simil-proof of the classification of finite simple groups, no one individual knows the proof—it's just too large. This will become relevant in §2.3.

[14] Another example identified by Voevodsky:

> The groundbreaking 1986 paper "Algebraic Cycles and Higher K-

However, it need not be the case that SP-justification be flawed as a result of extravagant complexity. Another recent example concerns the use of the theory "ZFC without Powerset". This is often used with ultrapower constructions (for example in the theory of iterated ultrapowers). However, many authors (including some in textbooks) describe this theory as "ZFC with the Powerset axiom removed/deleted" (or similar). However for many simil-proofs using this theory there is an inferential gap—the sequence of propositions (so written) does not constitute a proof of the relevant proposition (this is [Fallis, 2003]'s definition of an inferential gap). In particular, this theory (simply removing the Powerset Axiom from the usual formulation of ZFC) does not suffice for many applications, notably there are models of ZFC with Powerset just deleted in which the Łoś Theorem fails (an essential theorem for the ultrapower construction). One instead requires—in addition to the Powerset Axiom being removed—that the Replacement Scheme be substituted by the schemes of Collection and Separation, and the usual formulation of the Axiom of Choice be replaced with the principle that every set can be well-ordered (these various formulations are equivalent in the presence of Powerset). Thankfully, many of the usual models one wants to perform iterated ultrapowers with *do* satisfy this stronger theory. In this way, many simil-proofs constructing ultrapowers without the awareness of these subtleties can be construed as Gettier-cases—the relevant ultrapowers can be constructed, but only because the usual models *happen* to satisfy the stronger theory and *not* because they satisfy ZFC with the Powerset Axiom simply deleted.[15] But here the case is relatively simple (certainly in contrast to motivic cohomology or the classification of finite simple groups) and arises through a lapse due to over-familiarity with the ZFC-context, rather than extreme complexity encoded within the simil-proof.

We can use these cases to identify different kinds of *epistemic luck* at play. We first isolate:

> theory" by Spencer Bloch was soon after publication found by Andrei Suslin to contain a mistake in the proof of Lemma 1.1. The proof could not be fixed, and almost all of the claims of the paper were left unsubstantiated.
>
> A new proof, which replaced one paragraph from the original paper by thirty pages of complex arguments, was not made public until 1993, and it took many more years for it to be accepted as correct. [Voevodsky, 2014a]

[15]See [Gitman et al., 2016] for a description of the situation with ZFC-Powerset, including some identifications of where the incorrect theory is stated. I thank Jonas Reitz for pointing out this example to me.

Logical luck. We say that agent S's mathematical justification of a true proposition ϕ exhibits a higher degree of *logical luck* iff more of the important steps in the simil-proof do not logically follow from the previous steps and/or these mistakes are less easily fixed.

The example of ZFC-Powerset is a clear case of logical luck. Whilst for the many of the relevant structures of interest the relevant propositions do follow from the premises, it is not the case that the conclusions (e.g. the Łoś Theorem) follow from the premises (ZFC with the Powerset Axiom simply deleted). So it is *lucky* that the desired theorems actually do hold in the relevant contexts.

It may be, however, that whilst the various steps do follow logically from one another there is still a kind of luck at play, as when a difficult-to-fill non-trivial gap is left unintentionally. This is indicative of:

Enthymematic luck. We say that agent S's mathematical justification of a true proposition ϕ exhibits a higher degree of *enthymematic luck* iff unbeknownst to S, important steps are missed (even if all the steps do follow logically from one another) and it is harder to fill in the details of these steps.

Not every instance of enthymematic luck is an instance of logical luck. A simil-proof that leaves a gap where there should be further mathematical justification of a key lemma (say because it is assumed that the lemma is obvious when it needs proof) might exhibit enthymematic luck but no logical luck.

We will discuss some upshots for these cases and some dimensions of mathematical justification later (§§3, 4). For now, we should note that (in contrast to the cases I'll examine below) the authors (and referees) of a flawed simil-proof are at least somewhat *epistemically blameable* for publishing a flawed argument. The fact that the simil-proof produced is not a proof is underwritten by a mistake that they themselves have made. Of course, mistakes can happen, and publishing a flawed simil-proof does not necessarily make an author morally or otherwise culpable. This point bears emphasising—even if an author is epistemically culpable for a defective simil-proof, the work often contains many valuable insights and isn't *bad* or *shoddy* (indeed the examples given in this section all represent significant contributions). They are, however, cases where our conception of fallibilist mathematical justification has come apart from the infallibilist one—it is certainly not *evident why* truth is preserved through the steps of the simil-proof.

As we will shortly see, however, there are cases where an agent is in a Gettier-type situation, but has not done anything epistemically blameworthy (or at least the level of epistemic blame is substantially diminished). The two kinds we shall see concern *axiom selection* and the use of lemmas as *black boxes*.

2.2 Axiom selection

Let's consider how axiom selection might play into a Gettier case. The central point here is that while we may have a perfectly good simil-proof from some set of assumptions, we may be lucky in the choice of these assumptions.[16]

The situation is especially acute in *set theory*. There, the discovery of the independence phenomenon has precipitated programmes suggesting several new axiom candidates. We will consider just two here. We won't go into too much technical depth, but rather will simply discuss some philosophically relevant properties.

One option is to use *forcing axioms*. These state that the universe has been saturated under certain kinds of sets (namely generic filters for certain partial orders and families of dense sets), and in this sense seek to maximise the subsets available.[17] A strong axiom of this kind is the *proper forcing axiom* (or PFA).

A different option is *Ultimate-L*. Under this axiom (V=Ultimate-L) the universe contains many ordinals with certain strong properties (so called 'large cardinals'). The axiom tries to capture the notion of the universe exhibiting pattern (as opposed to being chaotic) whilst containing the widest variety of structures possible.[18]

Some observations about these axioms are important: (1.) They agree on *Projective Determinacy* (PD), a statement about certain 'orderly' sets of real numbers (both imply a positive answer to PD), and (2.) they differ on the truth of the *Continuum Hypothesis* (CH); PFA implies that CH is false where V=Ultimate-L implies that CH is true. (3.) Proponents of the different axioms believe they have good justifications for why their axioms are true.

We can now consider the following (albeit fantastical) case:

The PFA-Lykovs. Suppose a family lives in complete isolation from the rest of the mathematical world. In our example they will be keen and talented set theorists who learnt the basics of ZFC and the iterative

[16] This problem, though not the kind of Gettier situation it has elicits, is pointed to by [Fallis, 2003] (see p. 45).

[17] See, for example, [Bagaria, 2006] for a discussion of this idea, and [Bagaria, 2005] for some philosophical remarks.

[18] It would take too long to explain these notions of 'pattern' and 'chaos' here. The details are available in [Bagaria et al., 2019].

conception of set before their isolation. They develop forcing axioms, and after lengthy discussion decide that they are the right axiomatisation of the set concept. They adopt PFA (believing themselves to have chosen the right axioms on the basis of a strong justificatory story), prove that PFA implies PD, and thereby come to be mathematically justified in believing PD.[19]

Suppose then however that V=Ultimate-L is in fact correct, and that both the Continuum Hypothesis and Projective Determinacy are true. What should we say about the PFA-Lykovs' belief in PD? On the one hand it is *correct*, and it seems justified in virtue of their having a satisfactory proof. However, one might dispute whether they have *knowledge*; their axioms get the nature of set-theoretic reality fundamentally *wrong*.

The example of the PFA-Lykovs, whilst somewhat far-fetched, plausibly has similarities with actual epistemic situations. Though mathematicians might not be literally physically isolated from their peers, culture and research specialisation can keep them apart. Consider Thurston's remarks concerning the state of mathematics more widely:

> Much of the difficulty has to do with the language and culture of mathematics, which is divided into subfields. Basic concepts used every day within one subfield are often foreign to another subfield. Mathematicians give up on trying to understand the basic concepts even from neighboring subfields, unless they were clued in as graduate students. [Thurston, 2006, pp. 42–43]

Thurston's point is that mathematicians are becoming increasingly specialised and unable to engage with the work in fields other than their own. One can then easily imagine a mathematician becoming 'siloed' within a community, with insufficient communication between researchers from different traditions needed in order to appreciate that a given different approach is a live axiomatic possibility. In this context, it's hard to sustain the claim that a mathematician who proves a true proposition using what they take to be well-justified axioms (to which they might think there are no serious alternatives) has done anything epistemically blameworthy simply by proving a theorem from their culturally inculcated axiom system.

[19] The Lykovs were a family of Russian Orthodox Christians who fled religious persecution and lived in almost total isolation between 1936 and 1978.

There are several immediate objections to the setup of the example that help to articulate what is required for the case to work. One objection starts by noting that I have presupposed a very strong form of set-theoretic realism: I assumed that every set-theoretic question (or at least PD and CH) have determinate answers. One might dispute this claim by arguing that the subject matter of set theory is rather composed of a plurality of different universes satisfying different sentences.[20] On this view, there might be no fact of the matter about PFA or V=Ultimate-L (or possibly even PD and/or CH), dissolving the problem. I accept that under this conception of the semantics of set-theoretic language, the problem dissolves. However, this is somewhat orthogonal to my point: It is enough to be able to conceive of situations in which we have competing siloed communities who disagree on what axioms are true of that subject matter and do not even view the other as offering a live justificatory possibility.

A second more subtle response is to argue that we should revise how we interpret agents holding false but consistent axioms as concerned with the *models* of those axioms rather than the *whole* universe (possibly because of charity considerations). In this context, we reinterpret the PFA-Lykovs as talking about *models* in which PFA holds (instead of the *whole* universe). Their simil-proof would then justify their belief that PD is *true in those models* and it wouldn't be lucky in any sense. This kind of response is connected with the view of mathematics in [Balaguer, 1998], where we view any consistent set of axioms as being *true* of some *part* of the mathematical realm.

I think that this response essentially gives in to the idea of the multiversist account of truth. If all we have to do in order to be speaking truthfully with our axioms is be *consistent*, then essentially we have said that how our thought and language relates to mathematical reality is given by consistency, even if there is some universe of sets beyond. For the purposes of mathematical practice, it would make the distinction between *true* and *consistent* axioms immaterial.

Moreover, we can *still* generate Gettier-style cases by pulling back to consideration of *consistency* questions.[21] For, so long as it is possible to have a justified but inconsistent axiom, one can come to have a mathematically justified belief in a true proposition whilst being lucky. For example, Frege's mathematical justification of Hume's Principle (and hence the axioms of second-order arithmetic) obtained by prov-

[20]This claim is made, for example, by [Hamkins, 2012].

[21]I thank Konstantinos Konstantinou for suggesting that cases of the kind I'm discussing transfer to this context, and John Burgess and Keith Weber for some further discussion of this point.

ing it from his axioms (and before the discovery of Russell's Paradox) is just such an example. Frege thought (pre-Russell's Paradox) that Basic Law V was a *logical truth*, and certainly felt that he had strong mathematical justification from true axioms for Hume's Principle and theorems of Peano Arithmetic. This kind of example is fairly common, and can occur in cases where we are able to prove a proposition from a set of axioms with strictly fewer consequences than the ones we in fact use. For example, many results concerning braids and left-distributive algebras are provable on the basis of ZFC, but the original proofs proceeded via (often very strong) large cardinals (e.g. measurable cardinals) before the reduction in strength was discovered. Dehornoy writes:

> It seems to us that the role of set theory in such cases is quite similar to the role of physics when the latter gives heuristic evidence for some statements that mathematicians are to prove subsequently. In both cases, the statements are first established rapidly but at the expense of admitting some additional hypotheses or approximative proof methods—observe that adding a set theoretical axiom is nothing but adding a new proof method—and the subsequent task is to give a proof that does not use the additional hypotheses any longer. [Dehornoy, 2000, p. 600]

On the assumption that these large cardinals were well-justified, any subsequent inconsistency discovered in them would result in a Gettier case.[22] So even allowing for every consistent theory correctly describing part of the mathematical realm, it's still possible to get a Gettier case from faulty axioms.[23]

One might feel that these cases are not ones of real mathematical justification since any real mathematical justification should begin with a consistent set of premises. So long as we think that we have a determinate grasp on arithmetic, I contend that these issues of axiom selection can be made quite common, even when we restrict to consistent theories. For, whilst the case of set theory is especially interesting for consideration of competing axioms and foundational questions, the problem can be made relatively concrete. All we need is to have a situation in which we prove a true proposition on the basis of

[22]See [Barton et al., 2020] for some discussion of the example of braids and [Dehornoy, 2000] for the original mathematics.

[23]Moreover, any subsequent discovery of reasons other inconsistency to reject the requisite large cardinals after propositions have been proved from them will also yield Gettier cases.

some false 'axiom' or other, and this isn't limited to set theory. Let's suppose that one is convinced by the robust evidence in support of the claim that not every problem solvable by a non-deterministic Turing machine in polynomial time is solvable by a deterministic Turing machine in polynomial time (i.e. one believes the statement known as $P \neq NP$ holds on the basis of the strong evidence supporting it).[24] Let's suppose that with the passing years and growing frustration with the difficulty of proving $P \neq NP$, it comes to be adopted as an *axiom* in many areas of computer science.[25] Now suppose we are given some problem Pr that is known to be NP-complete, and ask if a particular deterministic Turing machine $\mathbb{M}$ can solve Pr in polynomial time. Obviously at this point we can provide a very quick simil-proof to the contrary, Pr is NP-complete, and if we could solve Pr using $\mathbb{M}$ then $P = NP$, but this contradicts the $P \neq NP$ axiom. Let's suppose though that in this hypothetical scenario, that $P = NP$ is true on the standard model of the natural numbers, but is in fact independent of the axioms of our base theory (let it be ZFC for the sake of argument).[26] In other words, there are models of ZFC on which $P = NP$ holds and (non-standard) models of ZFC on which $P \neq NP$ holds. But lets suppose that in fact $\mathbb{M}$ *doesn't* solve Pr anyway, since the complex and hard-to-follow program doesn't do the job for Pr even if there *is* another Turing machine $\mathbb{M}'$ that *does*. Here we have a Gettier case—the grounds for the agent's assertion that $\mathbb{M}$ doesn't solve Pr are based on a false principle ($P \neq NP$), and the only way to have their simil-proof be connected to truth is to interpret them as talking about non-standard models (which, patently, they are not trying to do, unless one has a *very* strong kind of relativism in play on which we are not even capable of singling out the standard natural numbers).

If desired, one can come up with a more general statement of the problem for models of arithmetic. The $P = NP$ question can be formulated as a Π^0_2-statement. If we move to consideration of Π^0_1-statement of arithmetic ϕ (the Goldbach conjecture or the consistency statement for your favourite recursive theory extending the base theory will do),

[24] See [Aaronson, 2016], §3 for a survey of such evidence.

[25] It is at least plausible that we're are almost there in certain areas. In particular, many cryptographic protocols (e.g. RSA-cryptography) depend on the hardness (i.e. **PTIME** unsolvability) of problems in NP. This, however, is a somewhat delicate issue (e.g. a non-constructive proof of $P = NP$ wouldn't pose any immediate threat to cryptographic systems, even if it might shake confidence) and so I'll set the details aside.

[26] [Aaronson, 2016] regards the formal independence of $P = NP$ from suitable set theories as somewhat unlikely, but it remains an epistemic possibility which is good enough for current purposes.

then any situation in which we come to have good evidence for $\neg\phi$, prove some statement ψ on that basis, but then ϕ turns out to be independent from one's base theory and ψ is true will have the same flavour: ψ is true but $\neg\phi$ can only be true on non-standard models.[27] I won't go into this further—the point has already been made—but it bears mentioning that this kind of Gettier-case is not necessarily an isolated phenomenon in set theory, but has the potential to be quite widespread given the desire to expand our axiom systems in the face of ignorance.

For the case of axiom selection, no errors need be made in the proof to get the Gettier case. In this way, it highlights that the infallibilist conception of mathematical justification, in order to be truly infallibilist, must throw in the truth of the assumptions we start with (and such a move is non-trivial). We therefore need to identify another kind of luck with respect to mathematical justification:

Luck in origin. We say that agent S's mathematical justification of a true proposition ϕ exhibits a higher-degree of *luck in origin* iff more of the 'axioms' used in (more) important steps in the simil-proof are false.

The cases we have considered here are mostly ones solely of luck in origin—we are *lucky* that our false axioms yield true statements. Still, luck in origin can be combined with logical and enthymematic luck, we might make mistakes in the *steps* of our reasoning *in addition* to the *axioms* chosen.

Note that in examples of luck in origin we have a falsehood at the root of the problem (a 'false lemma' in Harman's sense of the term) namely (depending on the example) PFA, $P \neq NP$, or the negation of our Π^0_1-statement ϕ. We might then ask if it is possible to get rid of this false lemma and have a situation in which the agent is neither clearly epistemically blameworthy nor uses a false mathematical proposition. As I'll now argue, this is possible (and indeed actual).

2.3 Black box lemmas

The kind of example we will consider concerns the use of lemmas as *black boxes*. What does it mean to say that we are using a *black box lemma*

[27] Note that any Π^0_1-statement of arithmetic independent from the axioms of ZFC is true on the standard model. Intuitively speaking, when ϕ is independent, then there are models of ϕ and models of $\neg\phi$, but since every model of arithmetic includes the standard model as a submodel, and since ϕ consists in universal quantification over some quantifier-free χ (i.e. $\phi = \forall x_0, ..., \forall x_n \chi$, where χ is quantifier free), ϕ has to be true on the standard model.

or a result *as a black box*? In the course of constructing a simil-proof, we may use other results. Some of these results are well-understood by the prover, others less so. More generally, we can define:

Black box lemma. We say that a proposition ϕ is being used as a *black box lemma* by an agent S who knows a simil-proof SP, iff:

(i) ϕ is used in the simil-proof SP.

(ii) ϕ is accepted by the mathematical community as a theorem.

(iii) S does not have a simil-proof of ϕ (in the sense that S would not be able to defend the steps of ϕ's simil-proof themselves).

The metaphor of a 'black box' reflects this state of affairs. ϕ may be very useful to S in proving a theorem, but the simil-proof of ϕ might not be well-known to S (say because ϕ's simil-proof is very complex, or uses resources from a field different from S's expertise). The use of black box lemmas is *fruitful* and very important for the development of mathematics as a *community*: It allows mathematicians with diverse expertise to use each other's results. Moreover, it is *unreasonable* to expect mathematicians to understand all dependencies in their knowledge, as this may be too time costly.

With increases in the complexity of mathematics and the interlinking of various established results, it might be the case that the use of black box lemmas is *essential*. Consider the theorem on the classification of simple finite groups. This theorem is (a) accepted by the mathematical community, and (b) no one single person really *knows* its simil-proof—the simil-proof has been established by a huge network of researchers proving their own small part of it and the combined work totals many thousands of pages. Now, any use of this theorem in a simil-proof (say by showing that a group is both simple and finite and therefore must belong to one of the relevant classes) is *essentially* using the theorem as a black box, one *cannot* have a simil-proof of it—the argument is just too large. Proving results using the classification of finite simple groups as a black box lemma is now so common in group theory that it has its own acronym (CFSG).[28] Whilst the example of the CFSG is especially vivid—a human agent *cannot* possess the currently

[28] I thank Ben Fairbairn for discussion of the use of the CFSG in group theory. A survey of some problems solvable with the CFSG is available in [Cameron, 1981], some of the problems therein (e.g. Schreier's Conjecture that the outer automorphism group of every finite simple group is solvable) are still open without the CFSG. It should be mentioned that some group theorists (including Cameron) express a degree of unease about using a theorem with a simil-proof that cannot be known by any (human) mathematician.

accepted simil-proof—it can also be made more mundane, this kind of phenomenon will occur wherever the difficulty of all lemmas relied on in the proof exceeds the mathematician's ability to know those proofs.

The same goes for cases outside one's field of expertise. Often it is simply too labour intensive or inefficient to learn an entire area that one is not familiar with in order to understand a complicated lemma that is useful for a given proof. This attitude is backed up (to a degree) by some mathematicians' opinions on the matter. In a recent interview study [Andersen et al., 2020] present the following observations concerning a question on MathOverflow about when it is acceptable to use a lemma as a black box:

> [Sauvaget, 2010] raises the question of when one should check the results of others before using them in one's own proofs. And similar questions have been discussed elsewhere on MathOverflow... In his response to Sauvaget's question, Fields medalist Timothy Gowers suggests that, "If a result is sufficiently accepted by experts you have good reason to trust, then the result can be trusted." Matthew Emerton writes that, "If a result is generally certified by experts, is well-established, and widely used and understood (even if not by you personally), then there is surely no problem in quoting it, applying it, and relying on it." Another Fields medalist, Terence Tao, gives a similar comment to a blog post [Kowalski, 2009]. Tao writes that, "If [the result] is prominent enough, and one trusts the practitioners of that field, then presumably it has been checked and understood by the experts, and it would be safe to cite." [Andersen et al., 2020, p. 3]

Further, the interview partners they consider generally regard the use of lemmas as black boxes as acceptable (though there are some exceptions). Kowalski (in a blog post) points to this labour-saving feature (though he also acknowledges that he has some reservations about the practice):

> To give a concrete example, I have no doubt that the Riemann Hypothesis over finite fields is true, but although I have really done a lot of reading about it, and can claim to have gone in great detail in the first proof of Deligne, I can not yet claim to have mastered the second—which I've used much more often in my work (and even with the first, I have certainly not a full mastery of the total amount

> of background material, such as the complete proof of the Grothendieck-Lefschetz trace formula). This example is not academic at all: many analytic number theory results depend on estimates for exponential sums which are not accessible at all without Deligne's work and its extensions, but very few are the analytic number theorists who understand the full proof. [Kowalski, 2009]

What we have here is a feature of mathematical justification analogous to the use of testimony in epistemology more widely (indeed, this is how [Andersen et al., 2020] cash out the phenomenon). This then leads to the familiar kinds of Gettier-case that we see in the wider literature on epistemology where testimony from a reliable source results in Gettiering. Suppose that I prove some theorem ϕ, on the basis of a lemma ψ that I am treating as a black box (where the simil-proof of ψ is widely accepted). As it turns out though, the simil-proof of ψ is flawed, but ψ is true. So my mathematical justification in ϕ is somewhat lucky—it depends on using a true proposition that is widely accepted, but for which the accepted simil-proof is flawed.

This luck made clearer when we consider the modal behaviour of my doxastic and epistemic states. In the world at which the flaw is spotted early, and the publication of the lemma delayed until after I am dead, I don't come up with my simil-proof, and hence do not take myself to have a mathematically justified belief in ϕ. Presumably, it is also plausible that for many cases there will be proofs of $\neg\psi$ using similar flaws, and in such cases I might end up going on to 'prove' $\neg\phi$ instead (say when my proof-strategy is to show that ϕ and ψ are equivalent). In this way, it seems that my beliefs and justification are not *modally robust*; there are relevant ways in which the world could be lightly perturbed and in which the same (or a similar) justification would either (a) not be regarded as a proper justification, or (b) lead me to a false 'theorem'. In the case where my simil-proof depends upon a legitimate black box lemma, it's plausible that stronger perturbations in modal space will be required to dislodge my mathematical justification.

Despite this problem with my mathematical justification, I have not done anything obviously epistemically *culpable* just by using a lemma as a black box and indeed *my* simil-proof is in perfectly good working order. At least, we may want to say that my level of epistemic culpability is *substantially dimished* compared to the situation in which the error originates in my own work. However, it does not seem right to say that I *know* since my belief in ϕ is based upon ψ which has a flawed simil-proof. Again, here we have a case where the fallibilist notion of justification comes apart from the infallibilist one—it is not evident how the steps in the simil-proof of the black box communicate truth to their descendants in my proof.

Importantly, there are *actual* examples of the flavour just described. Voevodsky's example from the introduction is an interesting case. In particular, some of the results from 'Cohomological Theory' were widely used before the error was found and diagnosed.[29] The use of the flawed results by authors other than Voevodsky constitutes a Gettier-case; they were *lucky* that the parts of the paper relevant for applications could, in fact, be salvaged.

Examples are easily multiplied. Dehn's Lemma (a topological theorem about the mappings of a disk) was thought proved by Dehn in 1910 (in [Dehn, 1910]) a flaw was found in 1929 by Kneser (see [Kneser, 1929]) and it was finally proved only by Papakyriakopoulos in 1957 (in [Papakyriakopoulos, 1957]). The 'result' in 1910 'resolved' an important problem in topology at the time, and likely would have been used in the 'unknown' period 1910–1929.

The Four Colour Theorem (that any map can be coloured with four colours) was thought proved 1879–1891, but it wasn't until Appel and Haken's computer-assisted proof in 1974 that it was finally proved. Any use of the Four Colour Theorem (or indeed the Five Colour Theorem—the proof of which can be salvaged from Kempe's proof) between 1879 and 1891 can then be regarded as a black box Gettier-case.[30]

Moreover, this can happen with unsurveyable proofs. The case of the classification of finite simple groups will again serve as our example. As Solomon relates:

> The literature of the Classification was always challenging, coming in massive 200-page papers. Nevertheless, there were always individuals and seminar groups that made serious efforts to read and digest most of the papers which appeared during the years 1960–1975. At least 3,000 pages

[29] For ease, I repeat part of an earlier quotation from [Voevodsky, 2014a]: "Starting from 1993, multiple groups of mathematicians studied my paper at seminars and used it in their work and none of them noticed the mistake." Voevodsky is especially clear about the usefulness of the work in his lecture [Voevodsky, 2014b] (upon which [Voevodsky, 2014a] is based), and emphasises the usefulness of the results in 'Cohomological Theory' between 14:20 and 17:00 of the lecture (see especially the remarks occurring at around 15:30).

[30] See [Sipka, 2002] for a discussion of the history of the Four Colour Theorem. I thank Ben Fairbairn for directing me to the Four Colour Theorem example, and Vadim Kulikov for pointing out Dehn's Lemma. Both Dehn's Lemma and the Four Colour Theorem are also discussed by [De Toffoli, 2021].

> of mathematically dense preprints appeared in the years 1976–1980 and simply overwhelmed the digestive system of the group theory community. Mason's 800-page quasithin typescript has achieved some notoriety, inasmuch as it has never been published. More accurately, it is an extreme point on the spectrum of incompletely assimilated manuscripts from the latter years of the Classification. Indeed, it was not until 1989 that it was noticed that certain small subcases of the problem remained untreated in Mason's typescript, a gap which Aschbacher filled in a typescript distributed in 1992. [Solomon, 1995, p. 236]

These errors (which lay undiagnosed for a time) might have been fatal to the proof, for all we knew. And in this case it is *clearly* unreasonable to expect the relevant agent to follow up all dependencies of the relevant proof—that is to ask for an impossibility. These kinds of cases are indicative of the following kind of epistemic luck:

Luck in dependence. We say that agent S's mathematical justification of a true mathematical proposition ϕ exhibits a higher-degree of *luck in dependence* iff more of the mathematical justifications of accepted propositions on which S's simil-proof of ϕ depends (which may in turn be justified by a simil-proof held by someone other than S) exhibit either luck in origin or luck in reasoning.

For many cases of luck in dependence, in particular where the only kind of luck at play in the flawed simil-proof on which S's simil-proof depends is enthymematic luck (i.e. there is no logical luck or luck in origin), it is hard to say that (a) S is epistemically blameworthy (after all, the error does not lie with *them*, and since there is no logical luck, the error may be very hard to diagnose), and (b) there are no obvious false mathematical propositions at play in S's mathematical justification.[31]

Apart from the differences in the kind of luck at play, there is a question in all this as to how *different* the case of black box lemmas and axiom selection are. For, one might think, in the case of a black box lemma the agent is (roughly speaking) treating the black box lemma as a new axiom within their system. So, for example, when a group theorist uses the CFSG in proving a theorem, they are actually working in their usual base theory augmented with the CFSG as an axiom. For

[31] Of course one thing we might say is that S has the false belief that their black box lemma has a legitimate proof. This seems, however, to be a proposition of a more *sociological* flavour compared to a *mathematical* proposition.

now let's note a couple of disanalogies between the two cases. First, as noted above, many of the black box lemmas we're considering are *true*, in contrast to false 'axioms'. Second, in the case of a false 'axiom', the agent has a flawed belief in an area that they understand *well* whereas in the case of a black box lemma it is the agents *lack* of understanding that necessitates the use of a black box.

Despite these complications, I take it as clear that the fallibilist account of mathematical justification in terms of possession of a simil-proof produces Gettier-style cases for mathematical knowledge—one can be mathematically justified in a true proposition ϕ but intuitively speaking not know ϕ.

3 Externalist and internalist dimensions of simil-proof-possession

What do these cases tell us about mathematical practice and the generation of mathematical justification? In this section, I'll argue that there are two main dimensions in which an agent's possession of a simil-proof can vary, namely externalist fit and internalist understanding. As we'll see, when a simil-proof has a high degree of externalist fit and is possessed by an agent with a high degree of internalist understanding, the Gettier phenomenon is more easily avoided. Moreover, I'll suggest that these notions provide a link between the fallibilist and infallibilist conceptions of mathematical justification.

We can start by considering what we want out of simil-proofs. I see at least two roles (aside from the enjoyment of mathematical activity):

(1.) We want proofs to tell us that some proposition(s) is (are) *true*.

(2.) We want proofs to deliver *mathematical understanding*, we want them to show us how our mathematical concepts relate to one another.

The idea that simil-proofs fulfil something like these roles appears variously throughout the literature. I take it that (1.) is beyond reproach. An awareness of (2.) is also prevalent, however. It perhaps the central claim of [Thurston, 2006] that mathematical understanding is what is really desired, rather than merely discovering the turth of propositions. [Rav, 1999] is explicit about the value of proofs beyond the verification of truth. [Gowers, 2000] speaks of "two cultures" of mathematics, and in particular that for some mathematicians the point of solving problems is to understand mathematics better and for others the point of understanding mathematics is to solve problems better.

Each facet of mathematical activity motivates a different facet of mathematical justification that can be fulfilled to a greater or lesser degree. I'll deal with each in turn. Regarding *truth*, we want our simil-proofs to map on to the mathematical facts appropriately. We therefore define the following:

Externalist Fit. A simil-proof (held by a subject S) exhibits a *higher (lower) degree of externalist fit* iff more (less) of the steps that S takes in the simil-proof fit the mathematical facts (i.e. each step is *true* and follows logically from the previous steps).[32]

A few clarifications are in order here: (1.) By 'mathematical facts' I do not mean to commit to any platonistic or correspondence theory of truth. Most accounts of mathematical ontology have some account of what mathematical truth should come down to, even if they are non-platonistic in nature. This might require a paraphrase, a fictionalist (for example) can still talk about degrees of externalist fit even though they think that (strictly speaking) mathematical claims are false, for them it is just that mathematical truth comes down to truth within the relevant fiction. (2.) A requirement of *fitting the facts* does not necessitate simil-proofs being composed of sequences of propositions. S's

[32]One related externalist idea (that served as the starting point for this condition) is through considering [Jenkins, 2008]'s account of mathematical knowledge. Jenkins suggests that S knows that P iff:

- S believes P.
- P is true.
- P is a good explanation for S believing that P, for someone not acquainted with the particular details of S's situation (an 'outsider').

Where an 'outsider' O is defined as follows:

(1.) O is rational, and can understand the content of S's belief that P (i.e. is capable of entertaining the proposition P).

(2.) O is aware of commonplace facts about people and their mental lives, i.e. facts about what it is like, in general terms, to be a rational thinking person.

(3.) O is not aware of any special facts about A or A's situation. O is aware that A is a person and that A believes that P, but that is all.

For Jenkins then, to be a knower is to have ones beliefs be explained by the mathematical facts, and this fact to count as a good explanation to an observer. Jenkins' account of knowledge is itself subject to particular kinds of Gettier case (as [Tennant, 2010] points out), however it presents one way of thinking about an externalist desideratum on mathematical justification and is suggestive of the aspect of externalist fit.

production and presentation of a picture-proof, for example, can perfectly well correspond better or worse to the mathematical facts, even if one takes a view of mathematical justification on which such diagrammatic reasoning does not constitute a sequence of propositions. (3.) The dimension of externalist fit attaches to the *simil-proof* (in contrast to the agent's possession thereof). (4.) There is no obvious way to *guarantee* that our simil-proofs do track the facts (although as we'll discuss later there are things we can do to increase our *confidence* that they do).

A high degree of externalist fit is clearly desirable, but does not tell the whole story. In particular, it does not guarantee that an agent has a high-level of *mathematical understanding* in knowing a particular simil-proof. For example, blindly following a known formula may exhibit an exceptionally high degree of externalist fit, but tell us little beyond the fact that a certain proposition is *true*. (Many of us will be personally familiar with the example of blindly following the quadratic formula $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ at school without a clue as to why it works.) We therefore want to examine the following desideratum on the possession of simil-proofs:

Internalist Understanding. S's possession of a simil-proof exhibits a *high degree of internalist understanding* iff S understands how the pieces of their simil-proof fit into a wider framework of knowledge, and understands the conceptual dependencies of the resources employed in their simil-proof. We define a *low degree of internalist understanding* in the obvious way.

Of course this condition is itself up for philosophical interpretation, what constitutes *mathematical understanding* is no easy matter. I am, however, happy to work with the notion on an intuitive level for the purposes of this paper (though I remain open to debates concerning how it should be sharpened). It may be, of course, that mathematical understanding is *itself* a multi-faceted notion. Note that internalist understanding is a dimension in which a *subject's possession* of a simil-proof can vary, rather than any property inherent in the simil-proof. Whatever one's characterisation of mathematical understanding, it seems to be a pretty clear desideratum on possession of a simil-proof that it be accompanied by understanding.

The dimensions of externalist fit and internalist understanding are naturally related, but can be independent. Sadly of course, S's possession of a simil-proof may have neither—it may be both founded on mathematical sand and S may have a poor understanding of the concepts. When S has knowledge of a very gappy simil-proof that uses many black box lemmas, we may exhibit nigh-perfect externalist fit but a low degree of internalist understanding—S does not understand how the relevant steps of the simil-proof are conceptually related to other areas and does not understand the conceptual dependencies of the black boxes. On the other hand, an agent S may have an *excellent* understanding of an area, but produce a simil-proof that has a reduced degree of externalist fit, as when a strong established researcher simply makes an error and produces a flawed simil-proof (e.g. the earlier discussed examples of ♣, 'Cohomological Theory', and ZFC-Powerset). In what one might think is the ideal case, we have both—a perfect externalist fit between the steps in the simil-proof and mathematical reality, and a clear understanding of how the simil-proof fits into our wider mathematical apparatus. Indeed, one would hope that the dimensions *are* related—as we obtain better internalist understanding of a simil-proof, we become more likely to externally fit the mathematical facts, and a high-degree of externalist fit offers more opportunity for internalist understanding. And these are closely (if not exactly) related to what the advocate of the infallibilist conception wants; they want (i) to start from true axioms (externalist fit), (ii) for each step to guarantee truth preservation (externalist fit), and (iii) for it to be evident why this is so (internalist understanding).

The fact that possession of a simil-proof can vary in degree with respect to both externalist fit and internalist understanding suggests an interesting phenomenon: There is a sense in which the Gettier-phenomenon *itself* is a matter of degree. It is *very* rare (except perhaps in trivial and easy cases) that we have a *perfect* correspondence between the steps of the simil-proof and reality as well as understand *all* the dependencies of our proof, *especially* where difficult research-level mathematics is concerned. In this sense, our mathematical justification via possession of simil-proofs is *almost always Gettiered to a degree* and imperfect.[33] This presses the following problem: Given that a degree of imperfection is *ubiquitous* in mathematics, and given that it may be *very hard to diagnose*, what should we do about the matter?

4 Upshots for mathematical practice

We now have a robust sense of the Gettier phenomenon regarding mathematical justification, and two important dimensions in which possession of a simil-proof can vary (namely externalist fit and internalist understanding). Thes highlight several epistemic upshots for

[33] I thank Deniz Sarikaya for this suggestion.

mathematical practice.

Some 'upshots' are trivial or clearly an established part of mathematical practice, and so I relegate them to merely being mentioned in passing. Obviously enormous non-trivial gaps in proofs are to be avoided and researchers should endeavour to work with as much care as possible. Furthermore, it's clearly to be regarded in a result's favour if (i) its simil-proof has been checked by multiple agents (increasing our confidence in externalist fit) with multiple different backgrounds (increasing our confidence that there's good internalist understanding to be had), and (ii) been integrated in other areas (again increasing our level of internalist understanding).[34] However there are some further respects in which we might regard our two dimensions as yielding upshots for mathematical practice.

Folklore theorems. A phenomenon which is pervasive in mathematics is the existence of theorems that are 'folklore'. These can often take the form of unpublished notes (that have subsequently been lost) or proofs that are regarded as easy and known, but have not been written down. The present discussion shows that the acceptance of many folklore results is fundamentally *bad* practice. Such theorems fall into the following categories:

(1.) A result where there is unclear attribution (e.g. because the result is very old and/or was proved outside of the published literature by multiple people independently) but *has* appeared in textbooks/graduate theses etc. (with attribution to the 'folklore').

(2.) Results that are genuinely trivial.

(3.) Results that are merely believed to be easy or known.

Cases (1.) and (2.) are benign and so I set them aside. However case (3.) is problematic, we lose shareability of the proofs and the ability to scrutinise both externalist fit and internalist understanding, and may invite in logical, enthymematic, and dependence luck. A recent example, discussed in [Rittberg et al., 2020], concerns the attempted publication of a paper containing a result in topos theory that was rejected on the grounds of the result already being part of the folklore. On examination, however, the 'standard' folkloric proof was found to be flawed (although the theorem was true). As such, use of the folklore theorem constituted a Gettier-case before discovery of the flaw. Discussing the issue André Joyal (a prominent category theorist) writes:

[34]This checking from multiple perspectives is considered by [Andersen et al., 2020].

> Although considered "folkloric" by some experts, the result does not appear in the literature. I had believed that one could directly deduce it from the theory of classifying toposes of Makkai and Reyes. It is only recently, in the context of a discussion with Caramello, Johnstone and Lafforgue, that the latter attracted my attention to an aspect of Caramello's proof which I had missed... Surprised by this observation, I tried to exhibit the "folkloric" proof that I thought I had of this theorem. With my great astonishment, it took me a night of work to construct a proof based on my knowledge of the subject, and the proof *depended only partially on Makkai-Reyes' theory*! [Joyal, 2015, italics mine]

Key to note here is that not only were we in a Gettier situation with this theorem, but we see clearly problems along the dimension of internalist understanding. In addition to the mere *fact* that the simil-proof was flawed, it was flawed in such a way that the conceptual dependencies were *not well understood* by experts who felt they had a simil-proof (in the end, the resulting proof depended only partly on resources it was thought to hinge upon).[35]

There is thus value in writing folklore theorems down and making them shareable. If they are really trivial, the proofs can be given to students as exercises or included in textbooks. If they are merely thought to be 'easy', they can either be given to graduate students for presentation in theses[36] or should be published (either in a relevant journal or on a public pre-print archives like the arXiv).

Methodological pluralism in mathematics. The previous example speaks to a practical change we can make in avoiding one source of logical, enthymematic, and dependence luck. There are also practical steps we can take to assuage worries of luck in *origin*. The example of

[35] It may also be that the original folklore simil-proof contained inferences that did not logically follow from one another, and so exhibited poor externalist fit. To show this conclusively, however, we would have to exhibit models witnessing the failure of these logical implications, and this seems to be in itself a significant mathematical question.

[36] Three nice recent examples from logic are (i) Regula Krapf's PhD thesis [Krapf, 2017] contains simil-proofs of some folklore results concerning global choice and bi-interpretations between second-order arithmetic and set theories in which every set is countable, (ii) Francesco Parente's clarification of some folklore proofs relating to forcing axioms (in [Parente, 2012]) and (iii) Jeroen Hekking's presentation of Zermelo's Quasi-Categoricity Theorem in contemporary notation (in [Hekking, 2015]). Further examples of such results being clarified are now easy to find in mathematics, and often PhD theses begin by setting up some known folklore proofs to feed into the main results later.

the PFA-Lykovs (and related cases closer to home) showed that we can be mathematically justified in a proposition, have done *little wrong* in our proof, but still fail to have good externalist fit. Whilst a degree of focus can be mathematically beneficial—great strides are made when intelligent people focus solely on a specific range of problems from within a specific axiom system—it comes at epistemic cost: Mathematical monism results in the possibility of luck in origin and our beliefs being explained by falsehoods. This suggests that a *methodological* pluralism in foundations is advisable—we should encourage the study of multiple different axiom systems and cross fertilisation between these different systems. This is so even if one is *ontologically* a believer that there is a *true* axiom system, the possibility that one is wrong should motivate one to accept the study of competing frameworks, in order that we miss as few live possibilities for well-motivated axioms as possible.[37]

Within this context, the desire for externalist fit indicates further epistemic value in mathematical practice to what [Koellner, 2009] calls 'overlapping consensus'. There are certain statements (e.g. PD) that are agreed on by multiple strong theories. The development and study of multiple axiom systems, and finding what lies in the intersection of all their consequences, increases the chance that at least *one* of the systems is the correct explanation for believing particular statements, decreasing the risk of luck in origin. Oddly, a strong belief that there are final answers to independent questions motivates consideration of a pluralism concerning the study of different theories—if only to ensure that we have the correct system within our purview.

5 Conclusions and open questions

In this paper I've argued that there are two important conceptions of justification in mathematics; one fallibilist and one infallibilist. Under fallibilist accounts of mathematical justification, mathematical Gettier cases are not just possible, but in many cases are in fact actual. I've also argued that this highlights some important upshots for mathematical practice, in particular concrete steps that can be taken to make mathematical claims more epistemically tractable with respect to internalist understanding and externalist fit. In doing so, we can try to ensure that a higher number of our fallibilist mathematical justifications fit the infallibilist conception too.

[37]Similar arguments are made with respect to *ignorance* in mathematics in [Barton, 2017].

I want to close by highlighting a couple of questions. First: We have seen a some upshots that we might take from the Gettier phenomenon for the dimensions of externalist fit and internalist understanding. However, there may be many more. I therefore ask:

Question. What further upshots of the Gettier-phenomenon, externalist fit, and internalist understanding are there for mathematical justification and the philosophy of mathematical practice?

A second and broad-ranging question concerns the agents involved in mathematical justification. Throughout, I have been concerned with the kind of epistemic status conferred when a *single* person has a similproof. However, we might think that the appropriate agent is actually the *community as a whole*. Indeed both the example of black box lemmas and axiom selection depended on looking at the epistemic states of an isolated individual embedded within a community. This suggests that the following question is of key importance:

Question. How are (i) the status of the Gettier-phenomenon, and (ii) the relevant upshots for mathematical practice, affected by a move to *communal* rather than *individual* epistemic agents?

References

[Aaronson, 2016] Aaronson, S. (2016). $P \stackrel{?}{=} NP$. In Nash, J. F. and Rassias, M. T., editors, *Open Problems in Mathematics*. Springer.

[Andersen et al., 2020] Andersen, L. E., Andersen, H., and Sørensen, H. K. (2020). The role of testimony in mathematics. *Synthese*.

[Avigad, 2020] Avigad, J. (2020). Reliability of mathematical inference. *Synthese*.

[Azzouni, 2004] Azzouni, J. (2004). The Derivation-Indicator View of Mathematical Practice. *Philosophia Mathematica*, 12(2):81–106.

[Bagaria, 2005] Bagaria, J. (2005). Natural axioms of set theory and the continuum problem. In *Proceedings of the 12th International Congress of Logic, Methodology, and Philosophy of Science*, pages 43–64. King's College London Publications.

[Bagaria, 2006] Bagaria, J. (2006). *Axioms of generic absoluteness*, page 28–47. Lecture Notes in Logic. Cambridge University Press.

[Bagaria et al., 2019] Bagaria, J., Koellner, P., and Woodin, W. H. (2019). Large cardinals beyond choice. *The Bulletin of Symbolic Logic*, 25(3):283–318.

[Balaguer, 1998] Balaguer, M. (1998). *Platonism and Anti-Platonism in Mathematics*. Oxford University Press.

[Barton, 2017] Barton, N. (2017). Independence and ignorance: How agnotology informs set-theoretic pluralism. *Journal of Indian Council of Philosophical Research*, 34(2):399–413.

[Barton et al., 2020] Barton, N., Ternullo, C., and Venturi, G. (2020). On forms of justification in set theory. *The Australasian Journal of Logic*, 17(4):158–200.

[Brendle, 2006] Brendle, J. (2006). Cardinal invariants of the continuum and combinatorics on uncountable cardinals. *Annals of Pure and Applied Logic*, 144(1):43 – 72.

[Cameron, 1981] Cameron, P. J. (1981). Finite permutation groups and finite simple groups. *Bulletin of the London Mathematical Society*, 13(1):1–22.

[Davis and Hersh, 1999] Davis, P. J. and Hersh, R. (1999). *The Mathematical Experience: Illustrated Edition*. Mariner Books.

[De Toffoli, 2021] De Toffoli, S. (2021). Groundwork for a fallibilist account of mathematics. *Philosophical Quarterly*, 7(4):823–844.

[Dehn, 1910] Dehn, M. (1910). Über die Topologie des dreidimensionalen Raumes. *Mathematische Annalen*, 69(1):137–168.

[Dehornoy, 2000] Dehornoy, P. (2000). *Braids and Self-Distributivity*. Springer.

[Dove, 2003] Dove, I. (2003). *Certainty and Error in Mathematics: Deductivism and the Claims of Mathematical Fallibilism*. Rice University.

[Džamonja and Shelah, 1999] Džamonja, M. and Shelah, S. (1999). $\clubsuit$ does not imply the existence of a Suslin tree. *Israel Journal of Mathematics*, 113(1):163–204.

[Fallis, 2003] Fallis, D. (2003). Intentional gaps in mathematical proofs. *Synthese*, 134(1-2):45–69.

[Giaquinto, 2007] Giaquinto, M. (2007). *Visual Thinking in Mathematics: An Epistemological Study*. Oxford University Press.

[Gitman et al., 2016] Gitman, V., Hamkins, J. D., and Johnstone, T. A. (2016). What is the theory ZFC without power set? *Mathematical Logic Quarterly*, 62(4-5):391–406.

[Gödel, 1953] Gödel, K. (1953). Is mathematics syntax of language? In Feferman, S., Dawson, J. W., Goldfarb, W., Parsons, C., and Solovay, R. M., editors, *Kurt Gödel Collected Works, Vol. III*, pages 334–364. Oxford University Press.

[Gowers, 2000] Gowers, T. (2000). The two cultures of mathematics. In Arnold, V., Atiyah, M., Lax, P., and Mazur, B., editors, *Mathematics: frontiers and perspectives*, pages 65–78. American Mathematical Society.

[Hamkins, 2012] Hamkins, J. D. (2012). The set-theoretic multiverse. *The Review of Symbolic Logic*, 5(3):416–449.

[Hekking, 2015] Hekking, J. (2015). Natural models, second-order logic & categoricity in set theory. Bachelor's Thesis, Leiden University.

[Jenkins, 2008] Jenkins, C. S. (2008). *Grounding Concepts: An Empirical Basis for Arithmetical Knowledge*. Oxford University Press.

[Joyal, 2015] Joyal, A. (2015). Public letter to Olivia Caramello. Available from: `http://www.oliviacaramello.com/Unification/JoyalLetter.html`. Accessed 25 October 2021.

[Kneser, 1929] Kneser, H. (1929). Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten. *Jber. Deutsch. Math. Verein*, 38:248–260.

[Koellner, 2009] Koellner, P. (2009). Truth in mathematics: The question of pluralism. In *New Waves in Philosophy of Mathematics*, pages 80–116. Palgrave Macmillan.

[Kowalski, 2009] Kowalski, E. (2009). Quoting the great unknown. Blog post: `https://blogs.ethz.ch/kowalski/2009/02/14/quoting-the-great-unknown/`, retrieved 25 October 2021.

[Krapf, 2017] Krapf, R. (2017). *Class forcing and second-order arithmetic*. PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn.

[Leitgeb, 2009] Leitgeb, H. (2009). On formal and informal provability. In Bueno, O. L. O., editor, *New Waves in Philosophy of Mathematics*, pages 263–299. Palgrave.

[Littlejohn, 2012] Littlejohn, C. (2012). *Justification and the Truth-Connection*. Cambridge University Press.

[Novaes, 2020] Novaes, C. D. (2020). *The Dialogical Roots of Deduction: Historical, Cognitive, and Philosophical Perspectives on Reasoning*. New York, NY, USA: Cambridge University Press.

[Papakyriakopoulos, 1957] Papakyriakopoulos, C. D. (1957). On dehn's lemma and the asphericity of knots. *Proceedings of the National Academy of Sciences*, 43(1):169–172.

[Parente, 2012] Parente, F. (2012). Boolean valued models, saturation, forcing axioms. Master's thesis, Università di Pisa. Master's Thesis available from: `https://core.ac.uk/download/pdf/79618812.pdf`.

[Paseau, 2015] Paseau, A. (2015). Knowledge of mathematics without proof. *The British Journal for the Philosophy of Science*, 66(4):775–799.

[Pritchard, 2012] Pritchard, D. (2012). Anti-luck virtue epistemology. *The Journal of Philosophy*, 109(3):247–279.

[Rav, 1999] Rav, Y. (1999). Why Do We Prove Theorems? *Philosophia Mathematica*, 7(1):5–41.

[Rittberg et al., 2020] Rittberg, C. J., Tanswell, F. S., and Van Bendegem, J. P. (2020). Epistemic injustice in mathematics. *Synthese*, 197(9):3875–3904.

[Sauvaget, 2010] Sauvaget, T. (2010). Published results: when to take them for granted? MathOverflow. URL:https://mathoverflow.net/q/23758 (version: 2015-03-18), retrieved 25 October 2021.

[Sipka, 2002] Sipka, T. (2002). Alfred Bray Kempe's "proof" of the Four-Color Theorem. *Math Horizons*, 10(2):21–26.

[Solomon, 1995] Solomon, R. (1995). On finite simple groups and their classification. *Notices of the AMS*, 42(2):231–239.

[Tanswell, F] Tanswell, F. (F). Go forth and multiply: On actions, instructions and imperatives in mathematical proofs. Forthcoming in Joshua Brown and Otávio Bueno (eds.) *Essays on the Philosophy of Jody Azzouni* Cham: Springer.

[Tennant, 2010] Tennant, N. (2010). Review of C. S. Jenkins, *Grounding Concepts: An Empirical Basis for Arithmetical Knowledge*. *Philosophia Mathematica*, 18(3):360–367.

[Thurston, 2006] Thurston, W. P. (2006). *On Proof and Progress in Mathematics*, pages 37–55. Springer New York, New York, NY.

[Voevodsky, 2014a] Voevodsky, V. (2014a). The origins and motivations of univalent foundations. *The Institute Letter Summer 2014*, available from `https://www.ias.edu/ideas/2014/voevodsky-origins`.

[Voevodsky, 2014b] Voevodsky, V. (2014b). Univalent foundations: New foundations of mathematics. Conference presentation at the *Institute for Advanced Study*, March 26, 2014. Available from `https://www.youtube.com/watch?v=E9RiR9AcXeE&t=25s`.

[Zagzebski, 1994] Zagzebski, L. (1994). The inescapability of Gettier problems. *The Philosophical Quarterly (1950-)*, 44(174):65–73.