Proofs, Reliable Processes, and Justification in Mathematics

Yacin Hamami*

To appear in the *British Journal for the Philosophy of Science*

Abstract

Although there exists today a variety of non-deductive reliable processes able to determine the truth of certain mathematical propositions, proof remains the only form of justification accepted in mathematical practice. Some philosophers and mathematicians have contested this commonly accepted epistemic superiority of proof on the ground that mathematicians are fallible: when the deductive method is carried out by a fallible agent, then it comes with its own level of reliability, and so might happen to be equally or even less reliable than existing non-deductive reliable processes—I will refer to this as the *reliability argument*. The aim of this paper is to examine whether the reliability argument forces us to reconsider the commonly accepted epistemic superiority of the deductive method over non-deductive reliable processes. I will argue that the reliability argument is fundamentally correct, but that there is another epistemic property differentiating the deductive method from non-deductive reliable processes. This property is based on the observation that, although mathematicians are fallible agents, they are also self-correcting agents. This means that when a proof is produced which only contains repairable mistakes, given enough time and energy, a mathematician or a group thereof should be able to converge towards a correct proof through a finite number of verification and correction rounds, thus providing a guarantee that the considered proposition is true, something that non-deductive reliable processes will never be able to produce. From this perspective, the standard of justification adopted in mathematical practice should be read in a diachronic way: the demand is not that any proof that is ever produced be correct—which would amount to require that mathematicians are infallible—but rather that, over time, proofs that contain repairable mistakes be corrected, and proofs that cannot be repaired be rejected.

*Centre for Logic and Philosophy of Science, Vrije Universiteit Brussel, Brussels, Belgium. Email: yacin.hamami@gmail.com. Website: www.yacinhamami.com.

Reliable processes capable of determining the truth of mathematical propositions come in different kinds. The most common one is the *deductive method* which consists in establishing the truth of a mathematical proposition by producing and verifying a proof of it. But there exist also a variety of *non-deductive* reliable processes able to determine the truth of particular classes of mathematical propositions. Typical examples are probabilistic methods such as the Miller-Rabin primality test (Rabin, 1980) for determining that a given natural number n is prime,[1] or computational methods such as the PSLQ algorithm (Ferguson et al., 1999) for determining integer relations between the elements of real or complex vectors.[2] Although non-deductive reliable processes are commonly used in mathematical practice for various purposes, they are not considered to be *acceptable* or *legitimate* ways of establishing mathematical truths, however reliable they may be—one cannot claim to have established the truth of a mathematical proposition unless one has produced and verified a mathematical proof of it. This means that mathematical practice, as an epistemic practice, enforces a standard or norm of justification according to which proof is the only accepted or legitimate form of justification for mathematical propositions.[3] The epistemic ground for this normative choice is, presumably, that the existence of a correct proof for a mathematical proposition Φ is a *guarantee* that Φ is true,[4]

[1]The Miller-Rabin primality test is a probabilistic algorithm (Motwani and Raghavan, 1995) whose design is based on the existence of a certain property identified by Miller (1976)—the property for a number b to be a *witness to the compositeness* of n—and on the following two associated mathematical facts which hold for any natural number $n > 4$: if n is prime, then there are no witnesses to the compositeness of n; if n is composite, then more than $3/4$ of the numbers $1 \leq b < n$ are witnesses to the compositeness of n (Rabin, 1980, p. 130). The algorithm proceeds by picking randomly and independently k natural numbers strictly smaller than n and by checking whether one of them is a witness to the compositeness of n. If such a witness is found, then the algorithm outputs that n is composite, otherwise the algorithm outputs that n is prime. Given the first mathematical fact just mentioned, if the algorithm says that n is composite, then it is always correct. But given the second mathematical fact, when the algorithm says that n is prime, there is always a possibility that it is mistaken, for it might happen that the algorithm has picked only nonwitnesses to the compositeness of n while n was indeed composite—the chances that this happens, and so that the algorithm is mistaken when it judges n to be prime, is then smaller than $1/4^k$. For epistemological discussions of the Miller-Rabin primality test, see Fallis (1997, 2000, 2002, 2011), Peressini (2003, p. 221), Easwaran (2009), Jackson (2009), Paseau (2014), and Smith (2016, pp. 48–50). Motwani and Raghavan (1995) provide an overview of probabilistic methods based on probabilistic algorithms in different areas of mathematics such as computational geometry, graph theory, number theory, and algebra.

[2]Let $x = (x_1, x_2, \ldots, x_n)$ be a vector of real or complex numbers. The vector x is said to possess an *integer relation* if there exist n integers $(a_i)_{i \in [\![1,n]\!]}$, not all zero, such that $a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$. The PSLQ algorithm is a so-called *integer relation detection algorithm*, that is, an algorithm which takes as input a vector $x = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$ or $\mathbb{C}^n$, and either (1) identifies n integers $(a_i)_{i \in [\![1,n]\!]}$, not all zero, such that the sum $a_1x_1 + a_2x_2 + \cdots + a_nx_n$ is very close to zero, or (2) establishes that no such integer relation holds for vectors of integers $(a_i)_{i \in [\![1,n]\!]}$ lying within a certain distance to the origin. The PSLQ algorithm is typically applied to vectors of numerical values consisting of high precision approximations of mathematical constants, and so most often requires high-precision arithmetic. In this case, when the PSLQ algorithm identifies an integer relation for a vector x, there is always a possibility that it is mistaken, for the relation might hold due to insufficient numerical precision. For mathematical descriptions of the PSLQ algorithm, see Bailey and Broadhurst (2001) and Borwein and Bailey (2004, pp. 230–232). For epistemological discussions of the PSLQ algorithm, see Corfield (2003, pp. 64–66) and Sørensen (2010). For an appraisal of the philosophical significance of the increasing use of computers in mathematics, see Avigad (2008) and Borwein (2009).

[3]This is not to say that one could not be justified in believing a mathematical proposition on the basis of the outcome of a non-deductive reliable process. Of course, one would be justified in believing, for instance, that a number n is prime on the basis of a run of the Miller-Rabin primality test. But from the perspective of mathematical practice, this would not count as a *proper* justification insofar as it does not meet the justificatory standards of this epistemic practice.

[4]Provided that the propositions that do not appear as conclusions of inferences in the proof are themselves true. To ease the presentation, I will always assume in the following that this is the case.

while there is no such guarantee in the case of non-deductive reliable processes. This is usually taken as a reason to consider the deductive method *epistemically superior* to non-deductive reliable processes in determining the truth of mathematical propositions.[5]

As we shall see below, several philosophers and mathematicians have contested this epistemic superiority of the deductive method over non-deductive reliable processes on the ground that mathematicians are *fallible*. The key observation is that when a fallible agent has produced and verified a mathematical proof for a mathematical proposition Φ, there is always a possibility that she has made a mistake in the process, in which case the proof produced might not be correct and thus might not constitute a guarantee that Φ is true. This means that when the deductive method is carried out by a fallible agent, the process possesses its own level of reliability, and so appears to be on a par with non-deductive reliable processes as far as reliability is concerned.[6] Furthermore, some non-deductive reliable processes might even happen to be *more reliable* than the deductive method in establishing the truth of certain mathematical propositions. One may then conclude from these considerations that the deductive method is not always epistemically superior, and in some cases might even be inferior, to non-deductive reliable processes. Call this the *reliability argument*.

My aim in this paper is to examine whether the reliability argument forces us to reconsider the commonly accepted epistemic superiority of the deductive method over non-deductive reliable processes. If this were the case, then this would call into question the main epistemic reason behind the standard of justification adopted in mathematical practice and thus undermine the predominant role that proof has in the justification of mathematical knowledge. This question is not only important for the epistemology of mathematics, it is also key to understand the epistemic status of non-deductive methods[7] in mathematics such as probabilistic proofs[8] and those used in experimental mathematics[9], and to compare them with the deductive method. I shall examine below the reliability argument and argue that it leads to the conclusion that, as far as reliability is concerned, the deductive method and non-deductive reliable processes are indeed on an equal footing. But for this to constitute a reason to reconsider the epistemic superiority of the deductive method, reliability would have to be the *only* epistemic property with respect to which the deductive method and non-deductive reliable processes could

[5] In comparing the deductive method with non-deductive reliable processes, I will be exclusively concerned in this paper with the capacity of these reliable processes to provide justification for mathematical propositions. There are many other reasons why mathematicians may attribute a privileged status to proof in mathematical practice. Maybe only proof can provide the type of mathematical understanding or explanation that mathematicians are aiming at. As a reviewer pointed out, maybe proof is the most suitable way to establish mathematical propositions that are necessary and general. There may also be sociological and pedagogical reasons as to why proof is preferred in practice, as recognized by Fallis (1997, p. 166). Finally, it may simply be in the essence of the mathematical activity to establish mathematical propositions through proofs, as noted by Paseau (2014, p. 795).

[6] Given that non-deductive reliable processes are designed and implemented by human agents, the issue of fallibility also applies to them. However, it is negligible with respect to the issues addressed in this paper, for the effect of fallibility mainly consists in impacting the reliability of the considered non-deductive reliable processes, and so could easily be taken into account by applying a decreasing factor to the value of their reliability computed in the infallible case.

[7] For an overview of the philosophical discussions surrounding non-deductive methods in mathematics, see Baker (2020).

[8] For the philosophical debate on the epistemic status of probabilistic proofs, see Fallis (1997, 2000, 2002, 2011), Easwaran (2009), Jackson (2009), and Paseau (2014).

[9] The nature of experimental mathematics has attracted some attention in the philosophy of mathematics. I refer the interested reader to Van Bendegem (1996, 1998), Baker (2008), Sørensen (2010), and McEvoy (2013).

be compared. I shall argue, however, that there is another epistemic property that the deductive method possesses and that non-deductive reliable processes lack which constitutes an epistemic reason to consider the deductive method epistemically superior to non-deductive reliable processes. This property is based on the observation that, although mathematicians are fallible agents, they are also *self-correcting* agents—both as individuals and as collectives—which means that, although they can make mistakes in the production and verification of mathematical proofs, they also have the capacity to correct these mistakes *over time*.

Before entering into this discussion, it is important to put the reliability argument in context and to recall the purpose to which it has been used by the authors who have advanced it. To my knowledge, the reliability argument has been used mainly in order to defend the epistemic status of (certain) non-deductive reliable processes against the predominant role attributed to proof in mathematical practice. For instance, the mathematicians Borwein and Bailey (2004) have appealed to the reliability argument in order to defend the epistemic status of computational methods used in the field of experimental mathematics:

> [O]ne can argue that many computational results are as reliable, if not more so, than a highly complicated piece of human mathematics. For example, perhaps only 50 or 100 people alive can, given enough time, digest *all* of Andrew Wiles' extraordinarily sophisticated proof of Fermat's Last Theorem. If there is even a one percent chance that each has overlooked the same subtle error (and they may be psychologically predisposed to do so, given the numerous earlier results that Wiles' result relies on), then we must conclude that computational results are in many cases actually *more* secure than the proof of Fermat's Last Theorem. [...]
>
> Many may still insist that mathematics is all about formal proof, and from their viewpoint, computations have no place in mathematics. But in our view, mathematics is not ultimately about formal proof; it is instead about secure mathematical knowledge. (Borwein and Bailey, 2004, pp. 9–10)

Paseau (2014) has relied on the reliability argument as part of his more general argumentation aiming to defend the epistemic status of inductive methods in mathematics, his main thesis being that, in some cases, one can have knowledge of a mathematical proposition on the basis of inductive evidence alone:

> Although deduction is by and large more reliable than induction as a general method, in mathematics and elsewhere, and although the best deductive arguments may be more reliable than the best inductive ones, it is not true that knowledge-generating deductive arguments are always more reliable than inductive ones. The Rabin test exemplifies just how reliable inductive evidence can be, namely, extraordinarily reliable even by the high standards of the exact natural sciences. More generally, although most mathematicians quite reasonably regard deductive evidence as generally speaking more secure than inductive evidence, they would concede that a deductive argument is not always a better guarantee of its conclusion's truth than an inductive one. Indeed, keen 'experimentalists'—mathematicians who develop and make much use of such methods—point out that strong inductive arguments are more convincing than the specious deductive arguments one regularly encounters in mathematical journals. (Paseau, 2014, p. 785)

Finally, Fallis (2002) has appealed to the reliability argument within his general defense of the epistemic status of probabilistic proofs:

> [M]athematicians are much more (epistemically) risk averse in their inquiries than other scientists.
>
> Unfortunately, their higher degree of caution is not sufficient to explain their rejection of probabilistic proofs (and only probabilistic proofs). As I will argue below, some probabilistic proofs of primality (e.g., the Rabin test) are more reliable than some deductive proofs of primality (e.g., the trial division test). Thus, even if mathematicians are extremely risk averse, they should still prefer using certain probabilistic proofs to using certain deductive proofs. (Fallis, 2002, p. 379)

The epistemological importance of the reliability argument lies in the fact that it opens the door to an eventual revision of the standard of justification adopted in mathematical practice. After all, if non-deductive reliable processes can be as reliable as—and in some cases even more reliable than—the deductive method, why not accept them as legitimate ways of establishing mathematical truths? Such a revisionist position seems to be suggested by Borwein and Bailey in the above quote.[10] Fallis (2002) and Paseau (2014) have also noticed that the various arguments they advance in defense of inductive evidence in mathematics raises the question of an eventual revision of the standard of justification commonly accepted in mathematical practice, although neither of them endorse such a revisionist position. But whether or not one wishes to go as far as to defend a revisionist position, it remains an important epistemological issue to decide whether the deductive method is epistemically superior, on a par, or inferior, to non-deductive reliable processes.

We shall begin with a closer examination of the reliability argument. A key observation here is that the reliability of the deductive method as well as that of non-deductive reliable processes, such as the Miller-Rabin primality test and the PSLQ algorithm, is always dependent on some *reliability parameters*. The reliability of the deductive method depends on the time and effort invested into the proof verification process, on the dedication and performance of the agent(s) involved in it, and on the length and difficulty of the proof under consideration.[11] The reliability of the Miller-Rabin primality test depends on how many numbers are randomly drawn by the probabilistic algorithm, while the reliability of the PSLQ algorithm depends on the numerical precision of the vector given as input as well as on the detection threshold set in the algorithm.[12] It follows from this that it is meaningless to compare the deductive method with non-deductive reliable processes in an *absolute* way—such a comparison only makes sense *relatively* to given and fixed reliability parameters associated to the processes to be compared. So when Borwein and Bailey compare in the above quote the status of Wiles' proof of Fermat's

[10] Indeed, Borwein explicitly endorses a revisionist position in his philosophical essay on the philosophical implications of experimental mathematics: "In my view it is now both necessary and possible to admit quasi-empirical inductive methods fully into mathematical argument" (Borwein, 2009, p. 34). Borwein writes, furthermore, that: "Today, while I appreciate fine proofs and aim to produce them when possible, I no longer view proof as the royal road to secure mathematical knowledge" (Borwein, 2009, p. 35).

[11] Of course, it will be very hard to calculate or estimate precisely the reliability of the deductive method from given specifications of its reliability parameters. However, nothing in the present paper rests on the assumption that this reliability can be calculated or estimated precisely. All that is assumed is that this reliability exists and that it can be increased or decreased in certain ways.

[12] The detection threshold is the value given to the algorithm for which the sum $a_1x_1+a_2x_2+\cdots+a_nx_n$ should be considered very close to zero.

Last Theorem—which is highly complicated and has only been verified by a handful of mathematicians—with computational methods of a very high level of reliability, they are taking the extrema of the spectrum where the reliability of the deductive method is at its lowest and the reliability of computational methods at its highest. But if we now compare a proof that is more simple and that has received a lots of scrutiny—e.g., Euclid's proof of the infinitude of primes—with a computational method of low or moderate reliability, then we will be led to draw the opposite conclusion. Indeed, one shall observe that, in principle, there is no theoretical upper bound to the reliability of the deductive method as well as to the reliability of the non-deductive reliable processes considered here.[13] This means that, given a mathematical proposition Φ and a non-deductive reliable process NDed capable of determining the truth of Φ, if δ is a set of reliability parameters for the deductive method[14]—hereafter denoted by Ded for short—one can always find a set of reliability parameters μ for NDed such that $r_{\mathsf{NDed}}(\mu) > r_{\mathsf{Ded}}(\delta)$, and conversely, if μ is a set of reliability parameters for NDed, one can always find a set of reliability parameters δ for Ded such that $r_{\mathsf{Ded}}(\delta) > r_{\mathsf{NDed}}(\mu)$—here $r_{\mathsf{Ded}}(\delta)$ and $r_{\mathsf{NDed}}(\mu)$ refer respectively to the reliability of Ded with the set of reliability parameters δ and to the reliability of NDed with the set of reliability parameters μ. Thus, the right conclusion of the reliability argument is that, as far as reliability is concerned, the deductive method and non-deductive reliable processes are on an equal footing.[15]

As we have just seen, comparing the deductive method and non-deductive reliable processes in terms of their reliability can only make sense with respect to *fixed* reliability parameters. Such a comparison is therefore inherently *static*, and thus disregards

[13]Is the assumption that there is no theoretical upper bound to the reliability of the deductive method warranted? We are considering *theoretical* upper bounds, which means that we are assuming that we have an unlimited, though always finite, storage of resources to allocate to the process—in this case, an unlimited number of agents to verify the proof and an unlimited amount of time and energy to dedicate to the verification process—in the same way as we are assuming that we have unlimited computing resources to allocate to non-deductive reliable processes such as the Miller-Rabin primality test and the PSLQ algorithm. With such unlimited access to resources, it is hard to see what could be at the origin of such a theoretical upper bound. Although it may be argued that there is such a theoretical upper bound, it is unclear on which basis such an argument will proceed. By contrast, consider the simplest way to model the reliability of the deductive method. We can consider a verification system composed of one or several agents carrying out the verification process. When the system has verified the proof under consideration and said that it is correct, there is a probability p that it has made a mistake in the verification process. By repeating the verification process an arbitrary number of times, we can then bring the probability p that the system has made a mistake as close as we want to 0, and the reliability as close as we want to 1. Of course, this model is overly simple. Yet, it is hard to see how, by refining it, one would introduce a theoretical upper bound to the reliability of the process.

[14]We assume here that the truth of Φ can be determined by the deductive method. This means that the deductive method has the capacity to produce and verify a proof of Φ. In some cases, the deductive method may even be able to produce different proofs for the same mathematical proposition. In these situations, we will assume that the deductive method has produced one of these proofs. Which proof is being produced does not matter for the present reconstruction of the reliability argument.

[15]I have assumed that there was no theoretical upper bound to the reliability of the reliable processes under discussion, but of course there may be *practical* upper bounds relative to their concrete implementations in the real world. For instance, at any given point in time, there will always be a finite number of agents who have the right abilities to verify a given proof, and there will always be a finite amount of computing resources that can be recruited to implement non-deductive reliable processes such as the Miller-Rabin primality test or the PSLQ algorithm. One could argue for the superiority of non-deductive reliable processes over the deductive method by identifying such practical upper bounds. The problem, however, is that (i) it is not clear how to establish such practical upper bounds, and (ii) these practical upper bounds are changing over time, for instance, due to improvements in computing technologies. So, although plausible, it would seem that this strategy is unlikely to succeed.

the *dynamic* behavior of these reliable processes *over time*. This dynamic behavior is, however, central to the way these reliable processes are used in practice. When a proof has been produced and verified by a mathematician or a group thereof, it will then be verified by other agents, if only through the reviewing process, thus increasing the time and effort invested into the proof verification process and eventually the dedication and performance of the agents involved into it. If one has used the Miller-Rabin primality test to determine that n is prime for a given $n \in \mathbb{N}$, one can always improve the situation by running the test again, eventually by drawing randomly more numbers during the process. And if one has used the PSLQ algorithm to determine that an integer relation holds between the elements of a real or complex vectors, one can always increase the numerical precision of the vector given as input as well as the detection threshold of the algorithm. In all these cases, one has determined at t_0 that a mathematical proposition Φ is true using a reliable process whose reliability parameters and reliability are those associated to the process at t_0. But all of these processes can always be *continued* in such a way as to *improve* their reliability parameters, and so to *increase* their reliability. From such a temporal perspective, a reliable process shall be conceived as running from t_0 to the current time t. The reliability parameters, and therefore the reliability of the process at time t, must then be associated to the whole run of the process starting at t_0 and finishing at t.

I shall now argue that there is a fundamental epistemic difference between the deductive method and non-deductive reliable processes with respect to their behavior *over time*. Let Φ be a true mathematical proposition and let NDed be a non-deductive reliable process capable of determining the truth of Φ. For the comparison to be fair, we shall also assume that Ded has the capacity to determine the truth of Φ—i.e., Ded has the capacity to produce a proof of Φ—at the instant we are considering. We now consider the situation at t_0 in which both Ded and NDed have determined that Φ is true, that is, the processes $\mathsf{Ded}_{\delta(t_0)}(\Phi)$ and $\mathsf{NDed}_{\mu(t_0)}(\Phi)$ have both yielded that Φ is true, where $\delta(t_0)$ and $\mu(t_0)$ are respectively the sets of reliability parameters of the processes Ded and NDed at t_0. As we saw earlier, Ded and NDed can be compared at t_0 with respect to their reliability associated to the sets of reliability parameters $\delta(t_0)$ and $\mu(t_0)$. But as we mentioned in the previous paragraph, these two processes can also be compared with respect to their dynamic behavior between t_0 and any later time $t > t_0$. In this respect, I claim that there is a fundamental epistemic difference between Ded and NDed which can be stated as follows: although Ded might not provide at t_0 a guarantee that Φ is true, it will provide one after a *finite* amount of time—I will refer to this as the *finite convergence property*; by contrast, NDed will *never* produce a guarantee that Φ is true after a finite amount of time without turning itself into a deductive process.

To see that Ded indeed has the finite convergence property, let P_0 be the proof that has been produced and verified by Ded at t_0. If P_0 is a correct proof of Φ, then Ded has indeed produced a guarantee that Φ is true after a finite amount of time. Suppose that P_0 is not a correct proof of Φ. There are two cases: either P_0 contains mistakes that cannot be repaired to turn P_0 into a proof of Φ and shall then be abandoned as a failed proof attempt, or P_0 contains mistakes that can be repaired by any mathematical agent with the proper abilities. Since we assumed earlier that Ded has the capacity to determine the truth of Φ, although Ded might indeed initially produce a proof that cannot be repaired and that should be abandoned, after a finite amount of time it should be able to reach a proof that can be repaired, and so we can focus on the second case. In mathematical practice, when a proof has been produced and verified by a mathematical agent or a group thereof, it will then undergo a certain number of verification-correction rounds by members of the mathematical community, that is, by agents who have the proper abilities to verify and eventually correct the proof. During each verification-correction round, one or more mistakes may be detected in the verification stage which are then corrected in the correction stage. Since P_0 contains mistakes that can be repaired, after a certain finite number n of rounds, the proof P_n must indeed be a correct proof of Φ. For if this was not the case, this would mean that there are one or more mistakes in the proof that would *never* be detected by the relevant mathematical agents. But this would simply mean that there is something *wrong* with the agents under consideration: if there is a mistake in the proof, an agent with the right abilities must be able to find it given a finite amount of time and energy. Otherwise this would simply mean that this agent is *necessarily* and *systematically* making a mistake in her verification of the proof, that is, the agent is more than fallible, she is *defective*—a fallible agent is one that *can* make mistakes, not one that *systematically* make mistakes. So the mistakes in P_0 should be detected and repaired after a finite amount of verification-correction rounds by mathematical agents with the right abilities. This means that the process should converge towards a correct proof of Φ after a finite amount of time, that is, Ded indeed possesses the finite convergence property stated above.

In the above argument that Ded has the finite convergence property, I have only considered the most common situation in mathematical practice where only mistakes may be detected as such and where the agents are indeed able to correct the mistakes identified. However, a fallible agent may detect a mistake where there is none and, in trying to correct it, may introduce one or more *new* mistakes. Furthermore, when a fallible agent has correctly identified a mistake, she may as well replace it by one or more new mistakes. This may lead to a stagnation, and even an increase, of the number of mistakes in the proof. Couldn't this prevent Ded to converge towards a correct proof of Φ? This issue needs to be addressed if we are to establish that Ded will indeed converge towards a correct proof of Φ in *all* cases. To do so, it is necessary to be more explicit on the characteristics we attribute to mathematical agents. I have assumed that mathematical agents are non-defective agents with the right abilities to detect and correct the mistakes in the proof under consideration. This follows from the implicit assumption that mathematical agents are *competent* in performing verification-correction rounds on the considered proof. What does it mean to say that an agent is competent in performing a task $\mathcal{T}$? One necessary condition seems to be that, after a finite number of trials, a competent agent should be able to perform $\mathcal{T}$ *correctly*. For instance, a heart surgeon who repeatedly make mistakes in performing heart operations, or a musician who repeatedly fails to perform a given piece correctly, would hardly qualify as a competent agent in the relevant task.[16] In our case, a competent mathematical agent is one that can perform a verification-correction round correctly after a finite number of trials, that is, that can identify all the mistakes in the proof under consideration at this stage and correct them successfully so that the resulting proof is indeed correct. Thus, starting with a proof P_0 of Φ which contains a finite number of mistakes, a competent agent should be able to perform a verification-correction round correctly after a finite number of trials, that is, to produce a correct proof P_n of Φ after a finite number n of verification-correction rounds. This means that Ded will indeed converge on a correct proof of Φ after a finite amount of time. Now, we can weaken this hypothesis and assume not that there is an agent

[16]These examples suggest that a competent agent is one that can perform a task $\mathcal{T}$ correctly with a certain frequency, and maybe even in most trials. For the present argument, it is enough to consider the weaker condition that a competent agent is one that can perform $\mathcal{T}$ correctly after a finite number of trials.

in the community that is competent in verifying and correcting the repairable proof P_0 of Φ, but rather that the mathematical community as a *collective agent*[17] is competent in doing so. This would better reflect the distributed nature of the verification process that is typical for long and difficult proofs which often require a team of agents with different expertise to verify the different parts of the proof. Under the assumption that the mathematical community is competent as a collective agent to verify and correct P_0, we obtain again that a correct proof of Φ will be reached after a finite number of verification-correction rounds by the mathematical community or a subgroup thereof. In sum, the above argument that Ded has the finite convergence property rests on the assumption that the mathematical community is composed of *fallible* and *competent* agents: although they can make mistakes in verifying and correcting proofs, they can also *get it right* after a finite number of trials.[18]

One may undermine the above argument that Ded has the finite convergence property by arguing against the assumption that mathematical agents are competent. It is not immediately clear how one could argue that mathematicians are not competent in the sense specified above. By contrast, there is concrete evidence that mathematicians are competent in verifying and correcting proofs. First, we can simply observe that there is a certain stability over time of the proofs that have been accepted as correct by the mathematical community, at least for the proofs of major results that have received significant scrutiny. In particular, it rarely happens that a proof that was accepted as correct by the mathematical community turned out later on to be flawed, and that the associated theorem returned to the status of conjecture—one of the few noticeable exceptions is Alfred B. Kempe's incorrect proof of the four-color theorem (I will come back to this case below). This stability is confirmed by the very low retraction rate in mathematics journals, at least for the major ones, although I recognize that some retractions may sometimes be disguised as "Errata" and "Corrigenda", or in a "Comments on" and "Reply to" exchange, which may significantly alter the original proofs and/or results as pointed out by Grcar (2013), but even those are relatively rare. Second, a significant portion of the mathematical literature has now been formally verified using proof assistants, including long and complex proofs of contemporary mathematical results such as the four-color theorem, the odd order theorem, and the Kepler conjecture.[19] In the vast majority of cases, it was possible to formally verify the proof one was starting with from the literature, thereby providing strong evidence that the initial humanly verified proof was indeed correct. Taken together, these observations provide evidence

[17] I have in mind a conception of collective agents such as the one proposed by Bird (2014) in what he calls the 'distributed model' (Bird, 2014, pp. 44–46) which is inspired by Hutchins (1995). In the distributed model, a collective agent is composed of different members, who have specific roles, and who will accomplish specific sub-tasks in order to accomplish an overall task. Through specific mechanisms of coordination, the collective agent will be able to perform the overall task in a distributed manner, where each member will contribute its part. The distributed model seems particularly adapted to account for the way several mathematical agents may collaborate to verify and/or correct a proof in mathematical practice.

[18] In saying this, I am by no means implying that getting to the point where a proof has been entirely verified and all the mistakes have been corrected is an easy matter. For instance, it took 12 referees for 4 years to verify the 300-pages of the Kepler conjecture that Thomas Hales submitted to the *Annals of Mathematics* (Hales, 2005). Another extreme example is Grigori Perelman's proof of Thurston's geometrization conjecture, which in turn implies the Poincaré conjecture. Perelman's proof appeared as a series of preprints posted on ArXiv in 2002 and 2003, but it took until 2006 for several independent groups of mathematicians to verify the proof, and thus for the proof to be accepted by the mathematical community (The Clay Mathematics Institute, 2010).

[19] For recent overviews of the field of formal verification, see Avigad and Harrison (2014) and Avigad (2018).

that mathematical agents, or at least the mathematical community taken as a collective agent, are competent in verifying and correcting proofs, and that when a proof has been accepted as correct by the mathematical community, it is most often correct.

There is thus a fundamental epistemic difference between the deductive method and non-deductive reliable processes in that the latter will never produce a guarantee that a given mathematical proposition is true in a finite amount of time without turning itself into a deductive process. This is due to the fact that, for non-deductive reliable processes, the uncertainty is *inherent* to the process and cannot be eliminated. For instance, and as noted earlier, the reliability of the Miller-Rabin primality test or the PSLQ algorithm can theoretically be increased as much as one wants, and so the uncertainty be reduced accordingly, but still this uncertainty can never be entirely eliminated, unless the considered process turns itself into a deductive process—this will happen in cases where the Miller-Rabin primality test would have picked *all* numbers strictly smaller than the number n to be tested or where the PSLQ algorithm would have carried out an *exact* computation.

It could be useful to illustrate this epistemic difference with a concrete example. A mathematical proposition whose truth has been determined both by the deductive method and by a non-deductive reliable process is the so-called Bailey-Borwein-Plouffe or BBP formula for π (Bailey et al., 1997b):[20]

$$\pi = \sum_{k=0}^{\infty} \frac{1}{16^k} \left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6} \right).$$

As reported in Bailey et al. (1997a, 2013), this formula was first discovered using the PSLQ algorithm. The search proceeded by identifying a list of potential mathematical constants $(\alpha_i)_{i \in [\![1,n]\!]}$ and by running the PSLQ algorithm to determine whether a linear relation of the form

$$a_0\pi + a_1\alpha_1 + a_2\alpha_2 + \cdots + a_n\alpha_n = 0,$$

may hold for some integers $(a_i)_{i \in [\![0,n]\!]}$. After a while, the search yielded the following formula:

$$\pi = 4 \ \cdot {}_2F_1 \left(\begin{matrix} 1, \frac{1}{4} \\ \frac{5}{4} \end{matrix} \middle| -\frac{1}{4} \right) + 2 \arctan \left(\frac{1}{2} \right) - \log 5, ^{21}$$

which can be rewritten to yield the above BBP formula for π. As discussed previously, when the PSLQ algorithm identifies such an integer relation among a list of mathematical constants, there is always the possibility that it is mistaken due to insufficient numerical precision in the approximations of the mathematical constants involved. The level of uncertainty in the final result—i.e., the chances that the PSLQ algorithm is mistaken in saying that the relation holds—can be reduced as much as one wants by increasing the numerical precision of the approximations of the mathematical constants, but it can never be entirely eliminated. On the other hand, the BBP formula can also be established through an ordinary proof. The following elementary proof was found shortly after the discovery of the BBP formula, reproduced here verbatim from Bailey et al. (2013, p. 847):

[20]The interest of this formula lies in the fact that it allows the n^{th} hexadecimal or binary digit of π to be computed directly without computing the digits that precede it.

[21]The first term corresponds to the evaluation of a Gauss hypergeometric function.

Proof. First note that for any $k < 8$,

$$\int_0^{1/\sqrt{2}} \frac{x^{k-1}}{1-x^8} dx = \int_0^{1/\sqrt{2}} \sum_{i=0}^{\infty} x^{k-1+8i} dx$$
$$= \frac{1}{2^{k/2}} \sum_{i=0}^{\infty} \frac{1}{16^i(8i+k)}.$$

Thus one can write

$$\sum_{i=0}^{\infty} \frac{1}{16^i} \left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right)$$
$$= \int_0^{1/\sqrt{2}} \frac{4\sqrt{2} - 8x^3 - 4\sqrt{2}x^4 - 8x^5}{1-x^8} dx,$$

which on substituting $y := \sqrt{2}x$ becomes

$$\int_0^1 \frac{16y-16}{y^4-2y^3+4y-4} dy = \int_0^1 \frac{4y}{y^2-2} dy - \int_0^1 \frac{4y-8}{y^2-2y+2} dy = \pi,$$

reflecting a partial fraction decomposition of the integral on the left-hand side. □

Now, there is always a possibility that this proof contains one or more mistakes. But as we discussed previously, such mistakes should be identified and corrected after a finite number of verification-correction rounds by agents with the appropriate abilities (in this case, the verification of the proof only requires an undergraduate training in analysis). This means that the deductive method will be able to provide a guarantee that the BBP formula is true after a finite amount of time, while the PSLQ algorithm will never produce such a guarantee, however high the level of precision of the approximations of the considered mathematical constants might be.

Mathematicians are fallible agents but they are also *self-correcting* agents, that is, they are able to recognize and correct their mistakes. This process of self-correction takes place at the individual level—when an agent corrects her own mistakes—but also at the social level—when agents correct the mistakes of others. Needless to say, this process is an essential element to the functioning of mathematical practice which is present in many forms of interaction, from office and email discussions, to conference presentations and the reviewing process. As the mathematician William Thurston put it: "People are usually not very good in checking *formal correctness* of proofs, but they are quite good at detecting potential weaknesses or flaws in proofs" (Thurston, 1994, p. 169). This means that the practice instantiates numerous mechanisms which have the capacity to ensure that when a proof of a mathematical proposition Φ has been produced by one or more agents, although the proof may initially contain a number of mistakes, if these mistakes can be corrected, then the verification and correction process will eventually converge towards a correct proof of Φ, and so will eventually produce a guarantee that Φ is true, provided that enough time and energy have been invested into the verification and correction process.[22] This is how the standard of justification adopted in mathematical practice should be read: the demand is not that any proof that is ever produced be

[22] Avigad (forthcoming) identifies several epistemic mechanisms present in mathematical practice that contribute to the robustness and reliability of the deductive method when carried out by fallible agents.

correct—which would amount to require that mathematicians are infallible—but rather that, over time, proofs that contain repairable mistakes be corrected, and proofs that cannot be repaired be rejected. From this diachronic perspective, the demand for a correct proof of a mathematical proposition Φ, and so for a guarantee that Φ is true, shall then be conceived as an *epistemic goal* to be reached *over time*. Insofar as (1) the practice has the means to ensure that this epistemic goal is reachable through self-correction processes, and (2) a guarantee that a mathematical proposition is true is always to be preferred to other forms of justification, there is thus an epistemic ground for the normative choice adopted in mathematical practice to only accept proof as a legitimate form of justification for mathematical propositions.

One may object that I have not provided an estimation of the time it may take for Ded to converge towards a correct proof of a mathematical proposition Φ from the moment Ded has reached a repairable proof P_0 of Φ. For all we know, this could take an amount of time that would completely exceed what may be acceptable in mathematical practice. If this were the case, it would strongly undermine the capacity of the finite convergence property to serve as an epistemic reason for accepting proof as the only legitimate form of mathematical justification in practice. I do recognize that it is particularly difficult to estimate exactly the amount of time it would take for Ded to detect and repair all the mistakes in P_0, as this would depend on many factors such as the length and difficulty of P_0, the number of remaining mistakes, and the performance level of the agents carrying out the verification-correction rounds. However, there is evidence that, in most cases, such an amount of time would be considered acceptable in practice.

First, it is often considered that mathematicians are very good at detecting mistakes in proofs, as witnessed by Thurston's quote above. Two examples of recent mathematical history may illustrate this. The first is the famous detection of a gap in Andrew Wiles' initial proof of Fermat's Last Theorem. The gap was identified by Nick Katz, a colleague of Wiles at Princeton, who was assigned as one of the reviewers of the manuscript Wiles initially submitted to *Inventiones Mathematicae*.[23] Katz discovered the gap while working through the part of the proof that was assigned to him. Wiles first thought that this was only a minor error that could easily be fixed, which led him to send an email to Katz explaining how to correct it. Katz was not satisfied with Wiles' answers and kept pressuring Wiles for more details and explanations. It is through this dialogue that Wiles realized that the error was not innocuous, and that there was in fact a serious flaw in this part of the proof. Another example is the identification of a fatal error in a proof advanced by Edward Nelson in 2011 that Peano arithmetic is inconsistent. Nelson released an outline of the proof which he made available on his website.[24] Shortly after, and only based on an outline of the proof, Terence Tao and Daniel Tausk both identified a major flaw in the proof which quickly led Nelson to withdraw his claim. To my knowledge, cases of mistakes in proofs that would have remained undetected despite intensive scrutiny by the mathematical community are rare. The only example that comes to mind is Alfred B. Kempe's "proof" of the four-color theorem that he published in 1879 (Kempe, 1879). There was an important mistake in Kempe's proof that remained undetected until 1890 when it was uncovered by Percy J. Heawood (1890). This mistake was particularly hard to identify because it concerned one subtle subcase that was omitted by Kempe and for which the general method he developed did not work (Sipka, 2002). The difficulty of identifying it was enhanced by Kempe's general expository style which was more verbose than formal and by the fact that his decomposition of cases was based

[23]For the story behind Wiles' proof of Fermat's Last Theorem, see Singh (1997).

[24]For an updated version of this document, see Nelson (2015).

on intuitive reasoning making it hard to ensure that all cases have been covered. If mathematicians were not efficient in detecting mistakes in proofs, we would regularly witness cases like Kempe's proof where a previously accepted proof is later on recognized as flawed and the theorem proved is returned to the state of conjecture, but this is not the case.

Second, in the majority of cases, mathematicians have been able to reach a consensus as to when enough time and energy have been invested in verifying and correcting a proof, and it seems that these amounts of time have been considered acceptable by the practitioners. For instance, it is considered today by the mathematical community that enough time and energy has been invested to verify, say, Wiles' proof of Fermat's Last Theorem or Perelman's proof of Poincaré conjecture. By contrast, it is considered that not enough time and energy has been invested in verifying Shinichi Mochizuki's proof of the *abc* conjecture so that the mathematical community could consider the proof to be correct. This means that the mathematical community takes itself to be able to estimate the moment when Ded has converged on a correct proof. These estimations are particularly important as they are used to decide when a mathematical result can safely be added to the body of mathematical knowledge on which other works can be built. Of course, the mathematical community may be mistaken in these estimations, and may entirely underestimate the amount of time and energy to be invested in verifying and correcting a proof in order to converge towards a correct proof. Here again there is evidence that the mathematical community may not be off track with these estimations, and that when it considers that enough time and energy has been invested in verifying and correcting a proof, that is, when it considers that a correct proof has been attained, Ded has indeed converged on a correct proof. The evidence here is the same as mentioned earlier, namely (1) there is a certain stability over time of the proofs that have been accepted as correct by the mathematical community, (2) when we try to formally verify the proofs accepted in the mathematical literature, we realize that, in the vast majority of cases, it is possible to turn them into formal proofs, thereby providing strong evidence that the initial proofs were indeed correct.

Two important remarks are in order here. First, it should be noted that estimating the convergence time of Ded in different situations is an empirical question about the mathematical community, its processes, and the agents that compose it. This is why my reply to the above objection appeals to some sort of empirical evidence from the ordinary functioning of mathematical practice. Second, it should be noted that, for the finite convergence property to constitute an epistemic ground for the norm of justification adopted in mathematical practice, it is not necessary to possess an *exact* estimation of the convergence times of Ded. All that is needed is that estimations of these convergence times are considered acceptable in practice, that obtaining a correct proof is within practical reach, so that the adopted norm of justification gives us "a game we can play". What is considered "acceptable" in this case should of course be decided by the practitioners. As we have just seen, the mathematical community is able to estimate the amounts of time and energy needed to verify and correct proofs of various lengths and difficulties, and seems to consider those to be acceptable.

A further objection is that, although Ded may converge towards a correct proof in an acceptable amount of time, the agents engaged in the verification-correction rounds may never know for sure when all the mistakes have been corrected. Unfortunately, this is inevitable for fallible agents. Assuming the opposite would be to assume that mathematical agents are *infallible* in their meta-evaluation of the verification process, that is, in their capacity to tell whether they have carried out the verification process correctly. Here again, this does not undermine the finite convergence property to support the norm of justification adopted in mathematical practice. As we have just discussed, mathematical agents are able to estimate when they may have reached a correct proof. And as we have seen, there is evidence that these estimations are not off track. All that is required to support the norm of justification adopted in practice is that the practitioners have evidence that all the mistakes in a repairable proof can be detected and corrected in an acceptable amount of time.

How does the present discussion relate to the arguments advanced by Fallis (1997, 2000, 2002, 2011) and Paseau (2014)? Fallis has argued that mathematicians do not have good grounds for rejecting probabilistic proofs as a legitimate way of gaining mathematical knowledge, while Paseau has argued that one can know a mathematical proposition on the basis of inductive evidence alone, that is, in the absence of proof. They both adopted an argumentative strategy which consists in identifying several epistemic properties that could distinguish between probabilistic proofs and deductive proofs and arguing, for each of them, that they do not succeed.[25] Fallis and Paseau both acknowledge the possibility that such an epistemic property may exist. They offer it as a challenge to those who want to argue for an epistemic superiority of deductive proofs over probabilistic proofs to exhibit such an epistemic property. As we will discuss shortly, this is exactly what Easwaran (2009) has proposed by arguing that deductive proofs possess a property called 'transferability' that probabilistic proofs lack. This paper proposes another epistemic property distinguishing the deductive method from non-deductive reliable processes which concerns their convergence behavior over time. As I argued, this provides an epistemic reason to consider the deductive method epistemically superior to non-deductive reliable processes. It thus constitutes a direct challenge to Fallis' and Paseau's theses. The specificity of the approach proposed here is that it gives central stage to the dynamic dimension of the deductive method and non-deductive reliable processes, while Fallis and Paseau have only attended to their static aspects. As we have just seen, this dynamic dimension is essential to the way these different reliable processes function in practice.

As we just mentioned, Easwaran (2009) has proposed an epistemic property to distinguish between deductive and probabilistic proofs which he called 'transferability'. An argument for a mathematical proposition is 'transferable' whenever it can be verified by an agent with the right expertise without having to resort to external information or to the testimonies of other agents. Deductive proofs are clearly transferable since it suffices for the agent to evaluate the validity of each step in the proof (assuming that the proof is written with the appropriate level of detail so that agents in the intended audience—i.e., with the appropriate expertise—will be able to verify it). Probabilistic proofs are not transferable because their evaluation requires some external information on the processes involved, e.g., in the case of the Miller-Rabin primality test, that the numbers selected by the algorithm were indeed picked randomly and independently. Although Easwaran's proposal and the one developed above are different, there are interesting connections between the two. First, the above argument that the deductive method possesses the finite convergence property presupposes that proofs are transferable. More specifically, I have construed the deductive method as involving verification-correction rounds by members

[25] Paseau (2014) has also advanced further arguments in defense of his thesis which appeal to the general nature of knowledge, to the potentially inductive basis of our knowledge of mathematical axioms such as those of ZFC set theory, and to the fact that we can know certain mathematical propositions by deriving them from physical propositions. These further arguments are, however, not directly concerned with comparing the deductive method and non-deductive reliable processes in their capacity to provide justification for mathematical propositions.

of the mathematical community. This requires that a purported proof proposed by a given agent could be verified by any other agent of the community with the appropriate level of expertise. Second, as emphasized by Easwaran, transferability is inherently a social property, one that is tied to the social dimension of mathematical practice. The finite convergence property, as described above, also has a strong social dimension insofar as the process of verifying and correcting a proof over time most often requires the involvement of several agents in addition to the author(s). These additional agents have a strong responsibility with respect to the mathematical community since they will be the ones certifying that the proof is correct. As Easwaran pointed out: "if non-transferable proofs were accepted, then the community could not engage in this constant self-monitoring—some argumentative steps would be hidden behind appeals to authority, or particular historical verifications" (Easwaran, 2009, p. 356). There are thus some intimate connections between the finite convergence property of the deductive method and the transferability of deductive proofs.

So is there any epistemic reason to consider the deductive method *epistemically superior* to non-deductive reliable processes in determining the truth of mathematical propositions? The reliability argument suggests that there is none, and this has been exploited by some authors to defend the epistemic value or status of certain non-deductive reliable processes against the predominant role of proof in mathematical practice. As we have seen, the reliability argument is correct, but it only instantiates a *static* comparison between deductive and non-deductive reliable processes, disregarding their *dynamic* behavior *over time*. There is, however, a fundamental epistemic difference between the two types of processes at this latter level which lies in the capacity of the deductive method to produce a guarantee for a mathematical proposition within a finite amount of time—what I have called the finite convergence property—while a non-deductive reliable process will never produce such a guarantee without turning itself into a deductive process. Mathematical practice instantiates many self-correction mechanisms, both at the individual and the social level, to ensure that when a proof is being proposed which only contains mistakes that can be repaired, the verification-correction process will eventually converge towards a correct proof, provided that enough time and energy have been invested into the process. This means that, from a dynamic perspective, the deductive method is indeed epistemically superior to non-deductive reliable processes, and so that there is an epistemic ground for the standard of justification adopted in mathematical practice. For fallible but self-correcting agents, the deductive method should then be conceived as a *dynamic process* whose purpose is to converge towards correct proofs through successive rounds of verification and correction, i.e., to provide *over time* guarantees for the truth of the mathematical propositions to be established, something that non-deductive reliable processes will never be able to produce, however reliable they may be.

References

Jeremy Avigad. Computers in mathematical inquiry. In Paolo Mancosu, editor, *The Philosophy of Mathematical Practice*, pages 302–316. Oxford University Press, Oxford, 2008.

Jeremy Avigad. The mechanization of mathematics. *Notices of the American Mathematical Society*, 65(6), 2018.

Jeremy Avigad. Reliability of mathematical inference. *Synthese*, pages 1–23, forthcoming.

Jeremy Avigad and John Harrison. Formally verified mathematics. *Communications of the ACM*, 57(4):66–75, 2014.

David H. Bailey and David J. Broadhurst. Parallel integer relation detection: Techniques and applications. *Mathematics of Computation*, 70(236):1719–1736, 2001.

David H. Bailey, Jonathan M. Borwein, Peter B. Borwein, and Simon M. Plouffe. The quest for Pi. *The Mathematical Intelligencer*, 19(1):50–56, 1997a.

David H. Bailey, Peter B. Borwein, and Simon M. Plouffe. On the rapid computation of various polylogarithmic constants. *Mathematics of Computation*, 66(218):903–913, 1997b.

David H. Bailey, Jonathan M. Borwein, Andrew Mattingly, and Glenn Wightwick. The computation of previously inaccessible digits of π^2 and Catalan's constant. *Notices of the American Mathematical Society*, 60(7):844–854, 2013.

Alan Baker. Experimental mathematics. *Erkenntnis*, 68(3):331–344, 2008.

Alan Baker. Non-deductive methods in mathematics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)*. URL = <https://plato.stanford.edu/archives/sum2020/entries/mathematics-nondeductive/>, 2020.

Alexander Bird. When is there a group that knows? Distributed cognition, scientific knowledge, and the social epistemic subject. In Jennifer Lackey, editor, *Essays in Collective Epistemology*, pages 42–64. Oxford University Press, Oxford, 2014.

Jonathan M. Borwein. Implications of experimental mathematics for the philosophy of mathematics. In Bonnie Gold and Roger A. Simons, editors, *Proof and Other Dilemmas: Mathematics and Philosophy*, pages 33–59. The Mathematical Association of America, Washington, DC, 2009.

Jonathan M. Borwein and David H. Bailey. *Mathematics by Experiment: Plausible Reasoning in the 21st Century*. A K Peters Ltd., Natick, MA, 2004.

David Corfield. *Towards a Philosophy of Real Mathematics*. Cambridge University Press, Cambridge, 2003.

Kenny Easwaran. Probabilistic proofs and transferability. *Philosophia Mathematica*, 17 (3):341–362, 2009.

Don Fallis. The epistemic status of probabilistic proof. *The Journal of Philosophy*, 94 (4):165–186, 1997.

Don Fallis. The reliability of randomized algorithms. *The British Journal for the Philosophy of Science*, 51(2):255–271, 2000.

Don Fallis. What do mathematicians want? Probabilistic proofs and the epistemic goals of mathematicians. *Logique & Analyse*, 179–180:373–388, 2002.

Don Fallis. Probabilistic proofs and the collective epistemic goals of mathematicians. In Hans Bernard Schmid, Marcel Weber, and Daniel Sirtes, editors, *Collective Epistemology*, pages 157–175. Ontos Verlag, Heusenstamm, 2011.

Helaman R.P. Ferguson, David H. Bailey, and Steve Arno. Analysis of PSLQ, an integer relation finding algorithm. *Mathematics of Computation*, 68(225):351–369, 1999.

Joseph F. Grcar. Errors and corrections in mathematics literature. *Notices of the American Mathematical Society*, 60(4):418–425, 2013.

Thomas C. Hales. A proof of the Kepler conjecture. *Annals of Mathematics*, 162(3): 1065–1185, 2005.

Percy J. Heawood. Map colour theorem. *Quarterly Journal of Pure and Applied Mathematics*, 24:332–338, 1890.

Edwin Hutchins. *Cognition in the Wild*. Number 1995. MIT Press, Cambridge, MA, 1995.

Jeffrey C. Jackson. Randomized arguments are transferable. *Philosophia Mathematica*, 17(3):363–368, 2009.

Alfred B. Kempe. On the geographical problem of the four colours. *American Journal of Mathematics*, 2(3):193–200, 1879.

Mark McEvoy. Experimental mathematics, computers and the a priori. *Synthese*, 190 (3):397–412, 2013.

Gary L. Miller. Riemann's hypothesis and tests for primality. *Journal of Computer and System Sciences*, 13(3):300–317, 1976.

Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, Cambridge, 1995.

Edward Nelson. Inconsistency of primitive recursive arithmetic. *arXiv preprint arXiv:1509.09209*, 2015.

Alexander Paseau. Knowledge of mathematics without proof. *The British Journal for the Philosophy of Science*, 66(4):775–799, 2014.

Anthony Peressini. Proof, reliability, and mathematical knowledge. *Theoria*, 69(3):211–232, 2003.

Michael O. Rabin. Probabilistic algorithm for testing primality. *Journal of Number Theory*, 12(1):128–138, 1980.

Simon Singh. *Fermat's Last Theorem: The Story of a Riddle that Confounded the World's Greatest Minds for 358 Years*. Fourth Estate, London, 1997.

Timothy Sipka. Alfred Bray Kempe's "proof" of the four-color theorem. *Math Horizons*, 10(2):21–23, 2002.

Martin Smith. *Between Probability and Certainty: What Justifies Belief*. Oxford University Press, Oxford, 2016.

Henrik Kragh Sørensen. Exploratory experimentation in experimental mathematics: A glimpse at the PSLQ algorithm. In Benedikt Löwe and Thomas Müller, editors, *PhiMSAMP. Philosophy of Mathematics: Sociological Aspects and Mathematical Practice*, pages 341–360. College Publications, London, 2010.

The Clay Mathematics Institute. Prize for resolution of the Poincaré conjecture awarded to Dr. Grigoriy Perelman [Press release]. Retrieved from `http://www.claymath.org/sites/default/files/millenniumprizefull.pdf`, 2010.

William P. Thurston. On proof and progress in mathematics. *Bulletin of the American Mathematical Society*, 30(2):161–177, 1994.

Jean Paul Van Bendegem. Mathematical experiments and mathematical pictures. In Igor Douven and Leon Horsten, editors, *Realism in the Sciences: Proceedings of the Ernan McMullin Symposium Leuven 1995*, pages 203–216. Leuven University Press, Leuven, 1996.

Jean Paul Van Bendegem. What, if anything, is an experiment in mathematics? In Dionysios Anapolitanos, Aristides Baltas, and Stavroula Tsinorema, editors, *Philosophy and the Many Faces of Science*, pages 172–182. Rowman & Littlefield, Lanham, 1998.