Philosophia Mathematica (III) **17** (2009), 341–362.
doi:10.1093/philmat/nkn032 Advance Access publication November 6, 2008

Probabilistic Proofs and Transferability†

Kenny Easwaran*

In a series of papers, Don Fallis points out that although mathematicians are generally unwilling to accept merely probabilistic proofs, they do accept proofs that are incomplete, long and complicated, or partly carried out by computers. He argues that there are no epistemic grounds on which probabilistic proofs can be rejected while these other proofs are accepted. I defend the practice by presenting a property I call 'transferability', which probabilistic proofs lack and acceptable proofs have. I also consider what this says about the similarities between mathematics and, on the one hand natural sciences, and on the other hand philosophy.

One of the central questions in the philosophy of mathematics concerns the nature of mathematical knowledge. The version of this question familiar from [Benacerraf, 1973] asks how knowledge of *any* mathematical proposition could be consistent with any picture of the semantics of mathematical language (and in particular with the apparently abstract and acausal nature of mathematical objects). However, there is a further question: even granting existing knowledge of mathematical propositions, one may wonder what exactly it takes for a mathematician to come to know yet more propositions.

To begin to address this question, I note that there is some extremely close connection in mathematics between knowledge and proof. Mathematicians often say that a claim is not known until a proof has been given, and an account somewhat like this is presupposed in some naturalistic discussions of mathematical knowledge (see [Horsten, 2001, pp. 186–189], where he concedes that other means may provide knowledge of mathematical propositions, but suggests that proof must underlie a notion of 'mathematical knowledge').

However, for any account like this to work, it must be clear what counts as a proof. This question is in the end a normative one, about which proofs

† I would like to thank Branden Fitelson, John MacFarlane, and the audience at the Midwest Philosophy of Mathematics Workshop in 2006 for helpful comments on an earlier version of this paper. I would also like to thank two anonymous referees, and audiences at UNLV, Berkeley, Leeds, UT Austin, USC, UW Madison, NYU, University of Pittsburgh, Stanford, ANU, and Sydney University for helpful discussion on later versions. And I am especially grateful to Don Fallis for very helpful discussion throughout the project.

* Department of Philosophy, Research School of Social Sciences, Australian National University, Canberra, ACT 0200, Australia. easwaran@gmail.com

have a right to play a role in understanding mathematical knowledge. However, I will only start to address it, by instead focusing on the related question of which proofs mathematicians actually accept in mathematical practice.[1] This is a descriptive question, but it is still a question about norms—in particular, I claim that there is a property that some arguments have, of being mathematically acceptable. This property is similar to the notion of grammaticality, which some strings of words in a given language have, while most do not. Both properties are normative ones, so the mere fact that a competent speaker has uttered a string of words, or that a competent mathematician has published an argument, does not suffice to show that it is grammatical or acceptable. But these norms are both norms that are implicit in practices, rather than being objective in some further sense: there may be better ways to communicate meanings than by using the specific grammar of any natural language, and the acceptability of arguments to mathematicians may not be a guide to whether the arguments actually give knowledge of mathematical facts. An extreme version of the naturalism espoused in [Maddy, 1997] might deny this possibility—such a naturalist may believe that there is nothing to mathematical knowledge beyond the norms implicit in the practice. But even on a more moderate position, the success of mathematics suggests that the norm of acceptability that is implicit in the practice probably has some strong connection to the objective epistemic norms. Thus, the project of understanding this norm of acceptability (which this paper is a step towards) can itself be seen as part of understanding the epistemology of mathematics.

The proofs mathematicians accept are not complete formal proofs of the sort studied in proof theory, but are rather some sort of informal approximation to them. But as Don Fallis points out in [2003], in addition to being expressed in informal language (rather than formal symbolism), many steps are only gestured at, or even left out completely, perhaps with a note that 'It is easy to see that . . .'. In addition, at least since the 1976 publication of Haken and Appel's proof of the four-color theorem, it has been considered permissible to publish proofs, some of whose steps can only be carried out by computer calculation and will not fit into the published version. However, Fallis has gone on to argue in a series of papers [1997; 2000; 2002] that given the proofs that mathematicians *do* accept, there is no epistemic reason for them to reject 'probabilistic proofs' (which

[1] By 'mathematical practice' I here mean the official business of publishing arguments for conclusions. Mathematicians may accept many other sorts of arguments in non-official contexts. For instance, when deciding which conjectures to work on, and what methods to use, there are all sorts of heuristic arguments that will convince a mathematician that one is correct. They will also accept arguments based only on the testimony of experts, even though, as I will discuss later, these are not arguments that would be acceptable in official practice.

I will explain the details of later). That is, no epistemic purpose is served by the norm that tells mathematicians to reject probabilistic proofs while accepting the proofs that they do accept. He argues that the norms of practice and the epistemic norms come completely apart.

Like Fallis, I will look here at what epistemic purposes could guide mathematicians in their practices of acceptance and rejection of proofs for publication in journals. As in [Fallis, 2002, p. 2], 'the project of this paper is to identify epistemic objectives that will explain this rejection of probabilistic proofs.' However, though Fallis argues in his papers that none of the objectives he considers could possibly draw the line between probabilistic and non-probabilistic proofs, I will present one that does draw this line.

The criterion I propose is 'transferability'. I will discuss this in more detail later, but the basic idea is that a proof must be such that a relevant expert will become convinced of the truth of the conclusion of the proof just by consideration of each of the steps in the proof. With non-transferable proofs, something extra beyond just the steps in the proof is needed—in the case of probabilistic proofs, this extra component is a knowledge of the process by which the proof was generated, and in particular that the supposedly random steps really were random. I will present some reasons why transferability could be an epistemically useful criterion for proofs, but I think that further discussion is needed to decide whether this objective is really important enough that it ought to continue to shape mathematical practice in the way that I claim it actually does.

Interestingly, transferability is a *social* epistemic virtue, rather than an individual one. Thus, I think it sheds some light on the fact that mathematics is in fact a social practice, and not a solitary one—solitary mathematicians would have no need for transferability, and would thus also develop a very different practice than is actually observed. The epistemic defect of non-transferability is not relevant to a mathematician as an individual choosing to believe or disbelieve mathematical claims—it is only in her role as a member of the mathematical community that I claim that it is relevant.

Finally, I think that consideration of the value of transferability will help shed light on the different ways in which the social practices of the natural sciences and philosophy are each similar and dissimilar to the social practice of mathematics.

1. Probabilistic Proofs

1.1. *Probability in Mathematics*

In order to make sense of probabilistic proofs, we first have to have a notion of probability that makes sense in mathematics. In particular, I will be concerned with the Bayesian notion of probability as 'degree of belief',

which is a formalization of the ordinary notion of uncertainty. The basic tenets of Bayesianism are that belief comes in degrees (one can be more or less certain of a proposition), that these degrees of belief obey the standard axioms of probability, and that rational agents update their degrees of belief by taking their old conditional probabilities as their new unconditional ones. Bayesianism is so far the most successful and well-developed account of uncertainty, though what I say about probabilistic proof will most likely carry through on any of the alternatives (Dempster-Shafer theory, imprecise probability theory, *etc*.). But whichever framework is used, it is some account of degree of belief that I am interested in here, and not any notion of frequency or objective chance that one might associate with the word 'probability'.

There are some worries about applying this notion of probability in mathematics—in particular, traditional formulations of Bayesian probability theory require that a rational agent have a degree of belief in the conclusion of a valid argument that is at least as great as her degree of belief in the conjunction of the premises. (This is often known as the 'problem of logical omniscience'.) Since the fact of a number's being prime or not is always a consequence of the Peano axioms, and we can assume that most mathematicians are at least very highly confident of the Peano axioms, this would prevent them from being uncertain of whether a number is prime. Since this is clearly both descriptively and normatively false (mathematicians are in fact uncertain of whether or not various large numbers are prime, and they are not thereby irrational[2]), there is a challenge to the use of probability to measure uncertainty in mathematics.

However, there are ways to sidestep this problem. Garber [1983] recommends replacing all statements of interest in a particular application of probability by propositional atoms, and then applying probability theory to the resulting propositional language. In this way, the actual logical relations between statements are opaque to the system, so the requirement of 'logical omniscience' does not interfere with the possibility of uncertainty. This is an incomplete solution, because there is still logical omniscience in the resulting propositional language, even though agents can clearly be rationally uncertain of whether very long propositional statements are tautologies. Thus, this picture cannot be the right final story about uncertainty in mathematics, but it is a good start that lets us at least *represent* the relevant aspects of rational uncertainty in mathematics appropriately.

[2] Some might suggest that this is in fact a failure of rationality. But if this is right, then rationality is of no interest for mathematics—a rational agent in this sense could not do mathematics, because she would already know all the theorems of every system. Instead, whatever concept of rationality is of interest for mathematical practice must be such that being uncertain of the primality of a number does not itself guarantee irrationality.

Another approach is suggested by Gaifman [2004], where he advocates assigning probabilities only to a subset of the formulas of the language (perhaps the ones that are short enough for the mathematician to grasp), in accordance with a slightly modified set of probability axioms. In this new system, logical omniscience is only required for deductive inferences that can be derived only using sentences from the subset of the language grasped by the agent. This account may be more promising for our purposes—he specifically shows how this system deals with the probabilities involved in the probabilistic proofs of primality that I will discuss later. However, it is still not clear if this can be the right formal account of uncertainty—it seems plausible that a mathematician may be rationally uncertain about some result even if it has a proof all of whose statements the mathematician can easily grasp, if the proof is extremely long.

But whether or not either approach is the right solution to this problem, there are reasons to adopt something like a probabilistic framework for describing mathematical beliefs. Probability (or something like it) is widely seen as the right framework for understanding degrees of belief or uncertainty, and it is clear that some notion of this sort is required to understand mathematical epistemology adequately, or else we will ignore the role of conjecture, hypothesis, and failed proof in the development of mathematical knowledge.

1.2. *Probabilistic Proof Techniques*

The technique Fallis primarily concerns himself with in [1997] is 'probabilistic DNA proof'. This is a technique for showing that a particular directed graph has no Hamiltonian path.[3] For each vertex in the graph, the mathematician chooses a pair of distinct sequences of DNA bases. Each directed edge in the graph is then represented by a strand of DNA that links up in the appropriate ways to the strands representing its start and end vertices. The mathematician is then able to make many copies of the strands representing the edges and the vertices, and stir them in a test tube to ensure that many long strands are created. With only a little work, she can select out strands of exactly the right length, and use physical facts about the particular strands representing each vertex to keep all and only the strands containing the sequence representing that vertex. If any

[3] In this technical sense, a graph is a set of points, called 'vertices', together with a specification of which pairs of vertices count as adjacent. Such vertices are said to have an 'edge' between them. A path is a sequence of vertices, each of which is adjacent to the next one. In a directed graph, the edges may have a direction to them, specifying which direction they are allowed to be traversed in paths—they can go in one direction or the other, or in both directions. A Hamiltonian path is one that contains each vertex of the graph exactly once. Graphs and directed graphs are often used to model computer networks, social networks, highway systems, food chains, and many other things.

strands remain at the end, then there must be a Hamiltonian path, which can be read off from one of the strands. Conversely, if there is a Hamiltonian path, then (provided enough copies of the strands were created in the initial stages) it is extremely likely that some strands will remain at the end.

This procedure is not deterministic, but it yields a very high likelihood $P(strands|HP)$ (which can be made arbitrarily close to 1 by making enough copies of each DNA strand at the start) and a likelihood of 0 for $P(strands|\neg HP)$. Thus, it can be used to convince oneself whether a particular graph has a Hamiltonian path. It is clear that $P(HP|strands) = 1$. And by Bayes' Theorem, we can calculate:

$$P(HP|\neg strands) = P(\neg strands|HP)\frac{P(HP)}{P(\neg strands)}.$$

We want to make this value as small as possible, so that the lack of strands at the end gives us high posterior confidence that there is no Hamiltonian path. Since $\neg HP$ entails $\neg strands$, we see that $\frac{P(HP)}{P(\neg strands)} \leq \frac{P(HP)}{P(\neg HP)}$, which is the initial betting odds for the existence of a Hamiltonian path. Since this value is fixed (for a particular agent and graph), we can make the posterior confidence in the non-existence of such a path as high as we want, just by making $P(strands|HP)$ sufficiently high (that is, by making enough copies of all the strands at the beginning of the process). Thus, this technique can convince an agent to a very high degree of certainty of the non-existence of a Hamiltonian path. (It also generally finds a path if there is one, but for demonstrating a path, we can use completely non-probabilistic methods, even if the path were originally found in this way—non-existence is in general much harder to prove.)

The technique Fallis discusses in [2000] and [2002], to which I will pay greater attention, is the Miller-Rabin primality test. Standard ways to determine if a number is prime (divisible only by 1 and itself) are very slow when the number to be tested is large. In his [1976], Miller found a relation R such that if p is prime, then $R(n, p)$ never holds when $n < p$, and such that if p is not prime, then $R(n, p)$ holds most of the time. In his [1980], Rabin was able to prove that in the non-prime case, it holds for at least $3/4$ of the $n < p$, which enabled a probabilistic primality test. Testing whether $R(n, p)$ holds is very fast; so one can quickly generate many $n < p$ at random and check for each whether $R(n, p)$ holds. If it never does, then the number is declared prime; otherwise, it is declared composite, and the particular calculation of $R(n, p)$ gives a deductive proof that it is.

In this case, particular probability values are easy to come by. If we check k integers less than n, and we choose these numbers to check by some means independent of the process by which we chose n, then it seems

clear that we should have $P(yes|prime) = 1$ and $P(yes|\neg prime) \leq 1/4^k$. By a use of Bayes' Theorem similar to the previous case, we can see that $P(prime|no) = 0$ and $P(\neg prime|yes) \leq P(yes|\neg prime)\frac{P(\neg prime)}{P(prime)}$. Thus, if our threshold for belief is $1 - \epsilon$, then to convince ourselves that a number is prime, we just need to make sure that $P(yes|\neg prime) \leq \epsilon \frac{P(prime)}{P(\neg prime)}$.[4] Since $P(yes|\neg prime)$ goes down exponentially based on the number of trials, we see that we just need this number of trials to be proportionate to the logarithm of the prior betting odds against primality. Thus, no matter how unlikely the mathematician originally thought it was that the number was prime, she can use this test to convince herself fairly quickly that it is, or to find a witness to its compositeness otherwise.

1.3. *What Probabilistic Proofs Are Not*

There are other cases where mathematicians have 'probabilistic' evidence for the truth of a claim where no deductive proof is available. Goldbach's Conjecture says that every even number is the sum of two primes. Most mathematicians are quite convinced that this is true, because no counterexamples have been found among the first several million integers. One might be tempted to count this as a 'probabilistic proof' of the Goldbach Conjecture because it is the same sort of evidence we have for many pieces of non-mathematical knowledge. The Riemann Hypothesis is the claim that the non-trivial roots of the Riemann zeta function all have real part exactly equal to $1/2$. It entails many powerful and interesting results about the distribution of prime numbers, and is connected to many important ideas in seemingly disparate areas of mathematics. Most mathematicians are quite convinced that the Riemann Hypothesis is true, because (in addition to a lack of counterexamples) it has significant explanatory and unificatory power.

While there may be good reason to consider these sorts of arguments in a more general study of 'probabilistic proofs', I will not focus on them here. The methods I am focusing on involve testing a statement S with some test T, and giving precise values to $P(T|S)$ and $P(T|\neg S)$, so that a precise value can be arrived at for $P(S|T)$. By contrast, there is no clear consensus on the likelihood of finding certain types of counterexamples, or explanatory connections, given the truth or falsity of either the Goldbach Conjecture or the Riemann Hypothesis. I concede that the prior probabilities $P(S)$ for each of these statements may be equally vague,[5] but Miller and Rabin

[4] In fact, as I will suggest in the next section, these arguments do not just increase degree of belief beyond some threshold, but actually give an importantly better epistemic status because of their sensitivity to the truth or falsity of the claim that they justify.

[5] One might try to use the Prime Number Theorem to give a prior of $1/\log n$ to the claim that n is prime, but regardless of whether or not this is initially appropriate, just knowing

have established a clear upper bound of 1/4 for the probability that a given number below n will pass the test if n is prime, while there is no clear upper bound at all for the probability that a given number will satisfy the Goldbach Conjecture or Riemann Hypothesis if either statement is false. So unlike these other cases, we can at least get a clear sense of the strength of the evidence in the case of these probabilistic proofs, even if the initial degree of belief is no clearer. Thus, whatever epistemic problems there are for probabilistic proofs, these other arguments will have further problems in being accepted by the mathematical community. The cases I focus on are the best ones for Fallis's argument that mathematicians ought to accept at least some probabilistic proofs. If Fallis's claim fails here, then it will fail for these more general types of probabilistic proofs as well.

The feature of having a very high value for $P(T|S)$ and a very low value for $P(T|\neg S)$ also distinguishes these probabilistic proofs from other probabilistic procedures. These are stronger requirements than that $P(S|T)$ is high, which just says that a positive test result will rationally guarantee a high degree of belief. For instance, the case of probabilistic proof can be usefully contrasted with a standard lottery case—let S be the claim that the ticket I have in the lottery is a loser, and T be the background knowledge that it is a fair lottery with a million tickets and a single winner. Then $P(S|T)$ is very high, but if T is background already known by the agent, S is irrelevant to T. Therefore, $P(T|S)$ and $P(T|\neg S)$ are both high, and this test proposition T does not seem to do any work in tracking the truth of S. Although obeying these probabilistic tracking conditions may or may not be either necessary or sufficient for knowledge (see [Roush, 2005]), they certainly make the situation epistemically better than in a lottery case. The apparent value of these probabilistic proofs is not just that they give the mathematician a very high degree of belief in the conclusion, but also the way that this degree of belief depends on the truth or falsity of the claim in question.[6]

One might think that the right way to understand a probabilistic proof is as a *deductive* proof of a probability statement, rather than an *inductive* proof of a mathematical statement, as I have described them. But

some simple facts about n, such as its last digit for instance, will quickly move one away from this value. Thus, I do not want to assume that the prior is any more precise in the one case than in the others.

[6] Making sense of this probabilistic dependence on the truth or falsity of a mathematical claim is the reason why a solution to the problem of logical omniscience is so essential to this project. If S is a logical truth, then standard probability theories make $P(T|\neg S)$ undefined, and perhaps even senseless. However, I take it that before we actually perform the test T, it is clear that we have some degree of belief in T turning up positive conditional on S being true, and some conditional on S being false. Whatever this conditional degree of belief is, that is the essential point differentiating these probabilistic proofs from mere lottery arguments, and perhaps even making them potential sources of knowledge.

the proof that $P(S|T) = 1 - \epsilon$ is not the argument that is relevant to the agent. Rather, it is a proof for *us*, theorizing about the agent. What it means to say that $P(S|T) = 1 - \epsilon$ is that once the agent has rationally responded to the evidence that T is true, she will have a very high degree of belief in S. Thus, the agent just needs to acquire T as evidence, and she will thereby become highly confident of S, even though the connection between T and S is not deductive. The deductive argument about $P(S|T)$ itself has no evidential significance for S—it is useful to us as theorists only because it tells us facts about the agent's probability function, which would be true even if we did not make the deductive argument.[7]

Sometimes mathematicians make this mistake—they say that a certain number 'is a probable prime', so that they can try to retreat to having given a deductive proof of some probabilistic claim. However, there is no such mathematical property. In the Miller-Rabin case, the only property a number might have to be a 'probable prime' is the historical and relational property of having passed a Miller-Rabin test. But this is no more a mathematical property than the property of having been written in red ink. So the right way to phrase the result of a Miller-Rabin test is by saying that the number 'is probably a prime', rather than saying that it 'is a probable prime'.[8]

2. Fallis

Fallis's arguments focus on the fact that mathematicians accept proofs with gaps in them, very long and complicated deductive proofs, and computer proofs, while rejecting probabilistic proofs. His argument in each paper then basically proceeds by considering various criteria one might consider that could draw such a distinction, and then showing that none of them do. I shall not go through all the criteria here, but will outline his responses to the most important ones. Then I will go on to describe my own proposed criterion of 'transferability'.[9] Even if some of his particular claims can be contested, I will concede them all, because I think transferability still does

[7] An anonymous referee has pointed out that of course, the deductive proof Miller and Rabin gave for their theorem is important for the agent, and not just the theorist, to have. If the agent didn't know that theorem, then she wouldn't have the degrees of belief the theorist uses in the calculation.

[8] For a simpler primality test that is related to the Miller-Rabin test, it has been known for a while that there are numbers where the test will *always* go wrong, which are called 'Carmichael numbers'. However, there is no probability involved there.

[9] At the December 2005 meeting of the Association for Symbolic Logic, Michael Rabin suggested that probabilistic proofs are 'non-transferable', but he did not give a clear account of what this might mean. I will try to define more clearly a property (which, following Rabin,

a better job of isolating the distinction that mathematicians are actually after.

First of all, he is concerned only with proof 'as a means of establishing mathematical truths' [Fallis, 1997, p. 166]. He leaves open the possibility of other goals for proof, such as providing good *explanations* for conclusions, which may very well favor deductive proofs over probabilistic proofs. But in response to a passage of Wittgenstein advocating deductive proofs for just this reason, he says, 'while providing understanding is nice, it is not required' [Fallis, 1997, p. 170]. After all, many deductive proofs provide very little understanding or explanation, but they are still published if they are the first proofs of some interesting result.

The first criterion Fallis considers is that of providing absolute certainty. One might think that a deductive proof from the axioms gives absolute Cartesian certainty in a conclusion, while probabilistic proofs cannot. But as I think Fallis is right to point out, deductive proofs do not in general provide absolute certainty either—when a proof is exceedingly long and complicated, one is in general not absolutely certain that the conclusions in fact follow from the premises. Even when it has passed the refereeing process at a journal, one should not be completely certain—journals often have to withdraw former publications that turn out to contain serious flaws, and thus we ought to be open to the possibility that a particular very complicated proof is one of these cases.

He also suggests that the particular *degree* of certainty cannot be relevant—after all, probabilistic methods can in many cases be made arbitrarily reliable, while acceptable proofs as large as the classification theorem for finite simple groups[10] almost certainly contain some invalid steps. In this case, there is good reason to believe that any invalidities in the argument can be fixed, but this does not clearly justify certainty beyond one chance in a billion of error, which one can easily achieve with the Miller-Rabin primality test.[11] Since any threshold of certainty will either exclude some long deductive proofs or include some probabilistic proofs (or both),

I will call 'transferability') that traditionally acceptable proofs have but probabilistic proofs lack.

[10] A group is a certain type of abstract algebraic structure, and ones with a particular property are known as 'simple groups'. Around 1980, it was established that all finite simple groups fell into one of fifteen well-defined infinite classes, except for 26 particular 'sporadic' groups. The proof proceeded by a huge enumeration of cases, and was carried out by dozens of mathematicians in hundreds of published papers and books, totaling around 10,000 pages. It has all been refereed, but no one person has been able to follow all of it.

[11] Technically, we should worry about the uncertainty involved in Rabin's proof that the method works, as well as the one in a billion chance associated with the particular random process, but Rabin's proof is much more straightforward and simple than the group classification theorem; so the difference in confidence levels is already probably more than one in a billion.

it cannot provide the criterion for acceptability used by the mathematical community.

The next property Fallis considers is that of providing 'conditional certainty'. Although there is a chance of invalidities in an argument, or failure with computer software, it seems that the conclusion is absolutely certain, *conditional* on the claim that 'nothing went wrong'. Probabilistic proofs do not provide this guarantee—even with no errors in the calculation, it is still possible to get unlucky with the random numbers one chooses, and mistakenly conclude that a number is prime. However, as Fallis points out (especially in [Fallis, 2003]), proof 'sketches' are often considered acceptable—very few published proofs actually cover all the relevant steps, instead relying on the reader to fill some in based on her familiarity with the material. In particular, Fallis discusses the first publication of Gödel's second incompleteness theorem—Gödel relies essentially on the claim that the proof of the first incompleteness theorem can be carried out in a formal system in order to prove the second, though he only ever actually gives an *informal* proof of the first. No full proof of the second incompleteness theorem was published until decades later. In [Fallis, 2003], he discusses other examples as well, where mathematicians rely on 'folk theorems' for which no complete and correct proof has been published, and examples in real analysis and category theory where slight modifications of standard theorems are cited, even though the modified version has never actually been verified. But however one makes precise the notion of conditional certainty, these sorts of arguments cannot give it.[12] Therefore, conditional certainty cannot be the criterion for acceptability that excludes probabilistic proofs.

The final property Fallis considers is that of giving *a priori* warrant. However, as he points out, since long deductive proofs require checking, and can be discovered to be invalid, they are not *a priori* under many analyses of this notion. In the case of many very long calculations, mathematicians also rely on the fact that computers (and paper, and blackboards) reliably preserve their data during the calculation, since the entire calculation cannot be internalized at once. But if these methods count as *a priori*, then the Miller-Rabin primality test should as well.

Thus, he has surveyed a range of potential properties that could separate probabilistic proofs from acceptable means of establishing mathematical

[12] Technically, one might define a 'mistake' as moving from a set of premises to a conclusion that they do not entail—but if we conditionalize on the lack of mistakes of this form, then it looks as if one could acceptably just jump from the premises to the conclusion with absolutely no argument and count as having a proof that gives this sort of conditional certainty. It looks very difficult to come up with a notion of conditional certainty such that acceptable gappy proofs guarantee it, while a 'proof' that skips *every* step in the reasoning does not.

conclusions. Assuming that long deductive proofs and incomplete proof sketches are all acceptable, Fallis seems to have argued that probabilistic proofs should be accepted as well. Although he also used computer proofs to provide more counterexamples to these criteria, I think they are not essential to his arguments—extremely long and complicated proofs, and incomplete proofs will suffice. And this is good—the acceptability of computer proofs is not as clear-cut. Although the Four-Color Theorem of Haken and Appel has eventually been nearly universally accepted by mathematicians, the computerized proof by Thomas Hales of the Kepler Conjecture was under referee for several years at the *Annals of Mathematics*, and only the non-computerized part was eventually accepted for publication. The *Annals* now has a policy on 'computer-assisted proofs of exceptionally important mathematical theorems' to help avoid some of the controversy that arose around this case, but it seems that this is a proof method they will accept only in 'exceptionally important' cases—suggesting that computer proofs are still in the borderline area.

3. Transferability

In discussion of this project, many people have sought to raise objections to some parts of Fallis's argument as I have summarized it. There may well be parts of the argument that are somewhat problematic. (In particular, the discussion of *a priori* warrant seems to raise many further issues that I do not have space to address.) However, I will concede all of Fallis's points here about various criteria not drawing the line at the same point as acceptability, and show that even so there is a property, which I call 'transferability', that draws the appropriate distinction. If Fallis is wrong and one of the other criteria also draws the line in the appropriate place, then there will still be an interesting question of whether this criterion is coextensive with transferability, and whether the two criteria illuminate different aspects of mathematical practice. It may turn out that thinking of things in terms of transferability may help avoid some of the conceptual problems of understanding what can be conditioned on in conditional certainty, or what methods count as relevantly *a priori*.

To motivate the property of transferability, I will start with an argument, which David Corfield gives, that some notion like probability must be used to understand reliability and uncertainty in mathematics.

> To contemplate the reliability of a result in a particular field we should think of someone from outside the field asking a specialist for their advice. If the trustworthy expert says she is very certain that the result may be relied upon, does it matter to the enquirer how the specialist's confidence arises? [Corfield, 2003, p. 110]

For the outsider, confidence in the result will be based entirely on the specialist's confidence. He cannot worry about how she got the result, because he is not qualified to decide between methods. So at least for the outsider, some single probabilistic scale of certainty seems to be the right notion of partial belief, just as it is for other non-mathematical areas.

However, if it is another mathematician asking the insider, then Corfield's argument makes a different suggestion: an insider may be very reliable at recognizing solid proofs, and even outlines of proofs, but very bad at making conjectures, or vice versa. A friend working in model theory once said that the great model theorist Boris Zilber is 'everywhere locally wrong but globally right', because he has made a series of conjectures that have each turned out to be false, but have motivated exactly the right sort of thinking to prove interesting related statements. Whether or not this is the right way to characterize him, it seems plausible that some mathematicians may have this sort of track record; so a specialist would not just want to adopt their credences as her own.

While an outsider may not be able to judge a mathematician's reliability, an insider may. Thus, this reliability will be an important consideration for a mathematician that wants to get involved in a field. A mathematician reading a paper may want to base her credences on evidence without having to rely on the testimony of mathematicians she thinks might be unreliable on a particular topic. If I am right, this suggests that appeals to authority are to be avoided—a mathematician reading a paper might not want her justification to consist just in the fact that it was in a prestigious journal, but wants to be confronted directly with evidence of a non-testimonial sort that will raise her credences.

The position that such a mathematician will find herself in is a strange one—she wants to gain true beliefs about mathematics, but wants to do so in a way that does not depend essentially on the reliability of other people. This position is obviously untenable in one's ordinary life. If I did not believe street signs that said 'road work ahead', or friends that told me they would meet me for dinner, my life would be very difficult indeed. Surprisingly, in mathematics (unlike most areas of life) this may actually be a tenable position. If someone presents a sequence of propositions for my consideration, and each proposition is such that mere *consideration* of it in light of my current beliefs leads me to believe it, then I can learn quite a bit from this person, even if I do not trust him. For instance, if he presents a deductive proof of some conclusion, I do not have to believe anything he says, as long as I independently have a high credence in the premises, and see independently that each step follows from previous ones.

As Don Fallis points out:

> [A] mathematician does not just want to communicate a sequence of propositions to other mathematicians. He also wants to convince the

> other mathematicians that the sequence of propositions is a proof. However, the other mathematicians can do this for themselves. [2003, p. 55]

Of course, Fallis acknowledges that most mathematicians do not *actually* do this most of the time when reading a paper, but just check certain steps to convince themselves that the argument works. But what I claim is important is that a relevant specialist can in principle (and in most cases even in practice) convince herself that the entire proof is correct, without having to construct independently much of anything beyond what was provided by the author.

Of course, a mathematician cannot maintain this position for her entire mathematical life because she sometimes needs to use results from outside her specialty. For instance, a real analyst who is told that a certain tangential claim is equivalent to a large-cardinal axiom in set theory will stop working to prove it—she has been told that these axioms are provably independent of ZFC, and does not need to work through this whole proof herself. Similarly, a topologist might reduce some claim to an algebraic one, and then just appeal to outside sources to convince herself that this algebraic claim is true. However, directly in the core parts of her own research, she will want to convince herself of everything and avoid trusting testimony.

If this is an important goal for mathematicians—to be justified in their conclusions through non-testimonial means[13]—then it may be relevant for journals to require the relevant sort of exposition. Papers will rely only on premises that the competent reader can be assumed to antecedently believe, and only make inferences that the competent reader would be expected to accept on her own consideration. Arguments of this form I will call 'transferable', following Rabin's terminology. On full consideration of a transferable proof, the reader will (if she has the right mathematical expertise) come to believe the conclusion of the argument. With a non-transferable proof, the reader may be competent, but may doubt the competence of the author to justify particular asserted claims, and thus remain uncertain of the claim. A proof is transferable just in case the sequence of propositions itself constitutes the proof—nothing about the method by which the propositions were generated is essential. That is, mere consideration of the proposition suffices for a relevant expert to become convinced of the conclusion, unlike arguments in which one needs to know that certain propositions were generated in a suitably random manner, or were generated by a reliable source.

[13] I really mean this 'if'—I have only provided here a *prima facie* argument that transferability is a good normative rule for mathematical acceptability. My main focus is to argue that transferability is a good *descriptive* account of mathematical acceptability, leaving the normative question for further inquiry.

Note that this standard does not necessarily require complete deductive proofs—in many cases, mathematicians can be relied upon to be familiar with certain modes of argumentation, so that a presentation of a sequence of propositions in a proof sketch can in many cases be sufficient for the reader to convince herself of the result (whether by mentally filling in the missing steps or just being familiar enough with the domain to see that the claim follows).[14] When other papers are cited, if they have all been published to this sort of standard, then a mathematician can in principle go through and convince herself in each case of the relevant result without relying on testimony. If she decides that certain results are far enough afield that she does not care to check, then she can rely on testimony. For exceptionally long and complicated proofs, or proofs that involve expertise in multiple areas of mathematics, *every* mathematician may find it necessary to accept *some* parts based on testimony alone. However, as long as all the relevant proofs are transferable, there is no *particular* step for which testimonial justification is essential.

Philosophers of mathematics have long focused on individual epistemic norms for mathematics, and deductive logic has been very useful in helping to understand these norms. However, transferability is a social norm—it can help the community develop a better grasp on the knowledge of its members, even though it may not have any advantages for the individual

[14] This is related to some counterexamples to Grice's thesis about speaker meaning. Grice claimed that for a speaker to mean something, she must intend that her intentions play a role in generating a belief in the listener. However, with a deductive proof from shared premises, the intentions are irrelevant. This seems to be the case with many sorts of arguments beyond strict deductive proofs from shared premises:

> If Grice's account of what it is for someone to mean something were correct, an unwelcome and somewhat ironic consequence would be that although Grice will have written and published an article of several pages on what it is for someone to mean something, Grice will have meant almost nothing by what he wrote. [Schiffer, 1972, p. 42]

There are important questions about exactly what sorts of gaps can be left in a proof while still meeting this standard. In general I think this can be answered in a somewhat sociological or psychological way—the sorts of gaps that are acceptable are the ones that relevant experts can see and still be convinced. This may end up making the criterion somewhat community-dependent, but that might be reasonable—while a set theorist might be quite comfortable with just a proof sketch that says 'a straightforward application of the Axiom of Choice shows that . . .', an algebraist or number theorist might want the argument to be further spelled out.

Of course, whatever gaps count here, it seems clear that they should not only be convincing steps in reasoning but should also be valid—if all the experts in a field are mistaken when they reason in certain ways (as has arguably been the case in some brief periods of mathematical history such as early nineteenth-century real analysis before Cauchy and Riemann) then apparently acceptable gaps may turn out to be unacceptable.

considered in isolation. Although it may be possible for both transferable and non-transferable proofs to give mathematical knowledge, when proofs are transferable, the community has better access to them, and there is no need for the community to acknowledge any special authority for particular individuals. If every proof is published in a transferable form, then the arguments for any conclusion are always publicly available for the community to check. Thus, if standards of rigor change, then earlier mistakes may well be detected. However, 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.

Thus, transferability seems to be a criterion that is properly met by standard deductive proofs, including very long ones, and importantly also by ones that skip steps, as is common in mathematics. The other type of acceptable proof Fallis points to in judging criteria is computer proof. I believe that computer proofs are in principle transferable as well, though the situation here is less clear. If the author provides the code for the relevant computer program, then the reader can presumably see just as well as the author that it does what is claimed. Actually running the program can then be done in the same way. This way, we can see that it really is the same proof that is being transferred from the author to the reader, and not a different one.[15] But the transfer here is less explicit—the reader does not actually go through a sufficient proof of the result, but must still rely on a computer. But since computer proofs are (as mentioned above) at least sometimes borderline cases of acceptable proofs, the fact that they are also borderline cases of transferability looks good for the descriptive adequacy of this criterion of acceptability. There may also be a way to assimilate computer proofs to extremely long deductive proofs. If the computer program produces a long chain of reasoning that leads to the relevant conclusion, a mathematician can verify for herself any segment of that reasoning she cares to check.[16] But since this computer reasoning is not normally published along with the proof, there is some sort of difference between this sort of checkability and the checkability of a proof such as the classification theorem for finite simple groups, which has been published in its entirety (scattered across many journals and volumes).

[15] There are still worries about the fact that the same program can be implemented in different operating systems or computer architectures, in which case a very different series of actual calculations might take place. At this point questions about the identity of proofs, calculations, and computer programs might come into play.

[16] This seems as if it may be an important contrast between computer reasoning and experiment. Although both are in some sense empirical sources of knowledge, in most experiments there is nothing like an inference process that can be followed and checked.

Non-transferability of Probabilistic Proofs

Returning to the probabilistic proofs described earlier, I suggest that in each case, though the test satisfies Fallis's criteria, it does not satisfy this criterion of transferability. In the case of probabilistic DNA proof, the published paper can say that the test was run, and say what the results were, but the reader has no independent way of convincing herself that this is actually true. She must rely on the testimony of the author.

In the case of the primality test, there is a bit more that can be done, because the author can publish the sequence of integers less than n that are checked, and the details of the calculations showing for each that it is not a witness of non-primality. However, unless the reader believes that these numbers were selected in a manner independent of the primality of n, she has no reason to be convinced. Miller's and Rabin's initial proof only shows that at most $n/4$ such integers fail to be witnesses—so a sequence of 100 non-witnesses can often be found, even for non-prime n. The author can be convinced, because she selects the values to check 'at random' (that is, in some manner independent of her selection of n as the number to consider).[17] But the reader just has to trust that the author has not cherry-picked the sequence of ks to fit the n, or cherry-picked the n to fit the sequence of ks. Thus, she cannot convince herself without relying in some sense on the testimony of the author. The proof is non-transferable.

One might seek to make the proof transferable by fixing some canonical list of 'random' values to check for each problem. If such a list is 'random

[17] Fallis [2000] spends a long time discussing worries about the impossibility of using a random-number generator that has an equal chance of producing each integer in the relevant range. However, the relevant notion of probability here is not chance, but rather uncertainty. Sampling the least significant bits of the system clock at the moment the author decides to run the algorithm may not have a well-defined *chance* of producing every number in the relevant range (especially if it turns out that the author's decision-making process is deterministic), but it still seems rational for the author to believe to an equal degree the proposition that this process will produce any number as opposed to any other. Thus, when updating her subjective certainties in light of the outcome of the process, she will rationally be using equal probabilities, regardless of what the underlying chances actually are.

If it turns out that there is in fact a systematic bias in this process, so that it leads to false diagnoses of primality relatively often, then although the author has been misled, I think she has still been perfectly rational. Only if she has reason to believe that something like this is likely would she seem irrational for assigning an equal degree of belief to each number being picked by this process.

Fallis claims, 'unpredictability is not the important issue for our purposes', but I think this is not quite correct. What is important is that the mathematician rationally have an extremely low degree of belief that a series of non-witnesses will be picked given that the number is not prime, and certainty that such a series will be picked given that the number is prime. Then, updating by Bayes' theorem guarantees that if a series of non-witnesses is picked, her degree of belief in primality will greatly increase, and if at least one number is a witness, then she will become certain of compositeness.

enough' it can serve the same purpose as randomly selected numbers, and because it is canonical, the reader can be sure that the author has not manipulated things. One way to do so might be to use something like the successive strings of appropriate length chosen from the decimal expansion of π. Of course, if the particular number whose primality is being checked is connected to π in some way, we can no longer be sure of the randomness here; so we may have to do something more convoluted, like using some particular published table of the distances between pairs of stars in our galaxy in alphabetical order.

However, this will not always work for all purposes, because once the sequence of potential witnesses to check is fixed, one might be able to cherry-pick the number n to be checked for primality. The claim that some particular number n is prime is almost never publishable (exceptions generally occur only for values larger than any currently known prime). However, one may be able to prove, say, that a certain equation has no solutions, provided that a particular number is prime. If there is only one relevant number, then the canonical list of witnesses to check may work, provided that the list is long enough (and the reader's prior betting odds on the primality of this number are no worse than the author's). But if the author just needs to prove that *at least one member* of some set is prime, and she does not say how she picked the one whose primality is to be demonstrated, then she could search for one in this set that might be 'easier' to claim to be prime, given the particular canonical sequence.

Thus, there seems to be no way to use a canonical list of 'random' witnesses to check. Instead, mathematicians will have to generate this list themselves each time, ensuring that the process used to generate potential witnesses is independent of the process used to generate the number whose primality is being tested. Because this list must be generated each time, the reader must trust the author's testimony as to the source (and independence) of the list, and must share the author's prior credence of primality, in order for the proof to transfer. To ensure independence, this source will most likely have to be something like the separation in nanoseconds between consecutive keystrokes on the author's computer, rather than some pseudo-random-number-generating algorithm that can be directly transferred to the reader. The author can describe the process used to generate the numbers, but the reader will not be able to verify that this process does in fact produce these particular numbers.

Another attempt to make these proofs transferable is to compare both DNA and Miller-Rabin proofs to deterministic computer proofs. The author can publish the method used, and allow the reader to repeat the computation herself. Presumably the reader will assign the same conditional credences as the author in the case of DNA calculation, and can also find some source she trusts to generate integers uniformly at random in some range for the Miller-Rabin test. However, in both cases, the actual series of steps

generated by the reader will be different from the ones generated by the author. In deductive proofs, proof sketches, and deterministic computer proofs by contrast, the sequence of steps is the same. Thus, although there is a way to transfer *something*, it is not the same proof that gets transferred.

Additionally, the thing that is transferred in the probabilistic cases is not an *argument*, but merely a piece of *evidence*. There is a superficial similarity between re-running a computer program, and re-running an experiment, but the computer program has intermediate steps that can actually be checked. We can view the operation of a computer program (in many cases) as a kind of generalized inference, whereas the same is not true of the DNA operations. In many automated theorem-proving systems, not only do the initial and final states of the system contain representations of a mathematical sentence, but the intermediate states of the system do as well. Additionally, the sentences so represented follow from previous sentences by valid rules of inference. In the case of the DNA calculation, the initial state contains representations of vertices and edges, and the final state contains representations of various paths. But the operations in between do not in any sense correspond to inference rules—they are just chemical processes.

4. Is Transferability Desirable?

Thus, I claim that Fallis's conclusion was not quite right; there is one epistemic property that plays a role in establishing the truth of mathematical claims that traditional methods have but probabilistic methods lack. However, there still remains a question as to whether this property actually *should* be a consideration in the acceptability of mathematical proofs, and whether it is part of the concept of proof underlying mathematical knowledge.

Transferability is clearly not a criterion for scientific arguments in general. As long as the conclusions depend at least in part on the results of some experiment, the reader must rely on the author's (and perhaps referee's) testimony that the author really performed the experiment exactly as claimed, and that it worked out as reported. This practice is standard in the physical sciences, and could easily be adopted by mathematicians as well, as suggested at the end of [Fallis, 2000]. It is important to note though that scientists standardly take great care to include in their papers a description of the methodology of the experiments, so that the reader can attempt to reproduce them.[18] While experimental results are not

[18] Actually, in some cases, the experiment may be essentially unreproducible because it involves an analysis of a particular object where that object is damaged or destroyed, perhaps as in analyzing the chemical composition of a particular meteorite. In other cases, an experiment will be *practically* unreproducible, because it is so expensive or difficult

transferable in my sense, they are certainly very useful things to publish for the reason that they are reproducible. It seems that probabilistic proofs are just as good in this sense.[19]

Although transferability is not required in the physical sciences, it seems that it might be expected in parts of the humanities, and especially certain types of philosophical papers.[20] In a paper like Gettier's famous one [Gettier, 1963] on justified true belief, the reader does not need to trust or agree with the author in order to become convinced of the conclusions. As long as she shares the intuitions about the specific cases at hand, she will come away believing that knowledge is not justified true belief. Thus this argument is just as transferable as any mathematical argument (and in fact can probably be transferred to a far wider audience than a typical cutting-edge research paper in mathematics).

If transferability is a *desideratum* for a certain type of philosophical argument, as well as for mathematical ones, then this may explain some of the resistance many philosophers feel to 'experimental philosophy'. On the one hand it seems strange to be willing to accept the intuitions of one person when Gettier puts them forward, and not the collected intuitions of hundreds of people when put forward by experimental philosophers trying to show that intuitions in Gettier cases are in part culturally determined, and therefore unreliable in philosophical reasoning [Nichols *et al.*, 2003].

to carry out, as with experiments using the Hubble Space Telescope, or the Large Hadron Collider. An experiment may also be *ethically* unreproducible—if it gave good reason to believe that one treatment for a disease is far better than another, it would now be immoral to subject one group of patients to the worse treatment. I have also heard that it is also not uncommon in certain sciences for descriptions of experimental methodology to be intentionally incomplete, to avoid giving away secrets that might allow other labs to beat the author to future discoveries. But these are all exceptions to the broad guideline of making experiments as reproducible as possible.

[19] Cory Juhl (private communication) has suggested that one way to understand the distinction between transferability and reproducibility is that transferability is a property of arguments whose goodness can be verified in an *a priori* manner, while reproducibility is the same property, where verification may require empirical methods. Again, whether this is correct will depend on the proper understanding of the *a priori*. This may seem to suggest a return to one of the criteria that Fallis has apparently refuted. However, Fallis seems to have been considering whether the proof can be *generated* in a manner that is *a priori*, while I am considering whether the proof can be *verified* in an *a priori* manner. Probabilistic proofs may well be able to be generated in an *a priori* manner, but verifying them requires recognizing a sequence of numbers as having been generated at random, which should be impossible, because every sequence of numbers is just as likely to have been generated at random.

[20] I do not claim that this is true of all philosophical papers, or even all parts of many. For instance, in philosophy a greater emphasis is placed on describing historical dialectics and asserting the importance of one's claims than in mathematics. This part of the discussion may well involve claims that are neither obvious nor supported by clear arguments from shared presuppositions, but instead may depend on historical references.

But on the other hand, if the reader shares Gettier's intuitions, then the former argument is transferable, while the latter relies crucially on the way various experiments were actually conducted, and what the results were. Of course, there are many other reasons one might be worried about this sort of argument, but the lack of transferability emphasizes the fact that intuitions enter the argument only as *evidence* for a conclusion, rather than as *premises* to reason from in Gettier's case. In order for Gettier's argument to be transferable, he must not use the fact that the reader shares his intuition as *evidence* for the claim that the agent has justified true belief without knowledge—rather, Gettier must just state this claim and assume that the reader already believes it. Gettier uses the *content* of intuitions as a premise, while Nichols, Stich, and Weinberg [2003] use the *existence* of intuitions as a premise.

Thus, transferability may have some analogies in disciplines other than mathematics, but it is not clear just what role it plays elsewhere. To work out the details of probabilistic proof, we will need a better-developed account of Bayesianism in mathematics (or perhaps some other sort of notion of partial belief, which will presumably interact appropriately with the probabilistic methods mentioned here). There is a difference between probabilistic proofs and traditional mathematical proofs, and it may well be a distinction that mathematicians care about for irrational reasons—after all, the other sciences seem not to insist on it. But insisting on transferability is a way to improve the quality of a body of knowledge in some minor way (by making all the evidence for the knowledge publicly available on non-testimonial grounds); allowing for reproducibility instead of transferability trades off this quality for the much greater quantity of knowledge it can lead to. It seems that transferability is not an essential part of an argument in producing scientific knowledge, but there is still a question as to whether it is required for distinctively mathematical knowledge. Thus, although I disagree with Fallis about the particular point about the non-existence of a property to distinguish probabilistic proofs from traditional ones, I agree with him about the broader picture. Mathematicians will surely need to embark at some point on this debate about whether non-transferable proofs that are easily replicated should be allowed.

References

Benacerraf, P. [1973]: 'Mathematical truth', *The Journal of Philosophy* **70**, 661–679.

Corfield, D. [2003]: *Towards a Philosophy of Real Mathematics*. Cambridge: Cambridge University Press.

Fallis, D. [1997]: 'The epistemic status of probabilistic proof'. *The Journal of Philosophy* **94**, 165–186.

_____ [2000]: 'The reliability of randomized algorithms', *British Journal for the Philosophy of Science* **51**, 255–271.

FALLIS, D. [2002]: 'What do mathematicians want? Probabilistic proofs and the epistemic goals of mathematicians', *Logique et Analyse* **45**, 1–16.

_____ [2003]: 'Intentional gaps in mathematical proofs', *Synthese* **134**, 45–69.

GAIFMAN, H. [2004]: 'Reasoning with limited resources and assigning probabilities to arithmetical statements', *Synthese* **140**, 97–119.

GARBER, D. [1983]: 'Old evidence and logical omniscience in Bayesian confirmation theory', in J. Earman, ed., *Testing Scientific Theories*, pp. 99–131. Minnesota Studies in the Philosophy of Science; 10. Minneapolis: University of Minnesota Press.

GETTIER, E. [1963]: 'Is justified true belief knowledge?', *Analysis* **23**, 121–123.

HORSTEN, L. [2001]: 'Platonistic formalism', *Erkenntnis* **54**, 173–194.

MADDY, P. [1997]: *Naturalism in Mathematics*. Oxford: Oxford University Press.

MILLER, GARY L. [1976]: 'Riemann's hypothesis and tests for primality', *Journal of Computer and System Sciences* **13**, 300–317.

NICHOLS, S., S. STICH, and J. WEINBERG [2003]: 'Metaskepticism: Meditations in ethno-epistemology', in Steven Luper, ed., *The Skeptics*, pp. 227–248. Aldershot, Hampshire: Ashgate.

RABIN, MICHAEL O. [1980]: 'Probabilistic algorithm for testing primality', *Journal of Number Theory* **12**, 128–138.

ROUSH, S. [2005]: *Tracking Truth: Knowledge, Evidence, and Science*. Oxford: Oxford University Press.

SCHIFFER, S. [1972]: *Meaning*. Oxford: Clarendon Press.