Erkenn (2008) 68:331–344
DOI 10.1007/s10670-008-9109-y

ORIGINAL ARTICLE

Experimental Mathematics

Alan Baker

Received: 16 December 2006 / Accepted: 26 October 2007 / Published online: 30 May 2008

Abstract The rise of the field of "experimental mathematics" poses an apparent challenge to traditional philosophical accounts of mathematics as an a priori, non-empirical endeavor. This paper surveys different attempts to characterize experimental mathematics. One suggestion is that experimental mathematics makes essential use of electronic computers. A second suggestion is that experimental mathematics involves support being gathered for an hypothesis which is inductive rather than deductive. Each of these options turns out to be inadequate, and instead a third suggestion is considered according to which experimental mathematics involves calculating instances of some general hypothesis. The paper concludes with the examination of some philosophical implications of this characterization.

1 Introduction

To an increasing extent, significant work in contemporary philosophy of mathematics has focused on confronting the traditional or received view of mathematics with actual mathematical practice. Consider the issue of justification. According to the traditional picture, justification in mathematics is a priori and deductive. This is in sharp contrast with the canonical pattern of justification in empirical science, which is a posteriori and inductive. Fleshing out this contrast, what makes empirical science empirical is the crucial role played by observation, and—in particular—by experiment.

Given this picture, what are we supposed to make of the rise of a genre known as "experimental mathematics?" The past 15 years or so have seen the appearance of

A. Baker (✉)
Department of Philosophy, Swarthmore College, Swarthmore, PA 19081, USA
e-mail: abaker1@swarthmore.edu

journals,[1] institutes,[2] colloquia,[3] and books.[4] devoted to this theme, and there has also been a 'Short Course on Experimental Mathematics' held in conjunction with the annual meeting of the Mathematical Association of America in 2006. Against the background of the traditional dichotomy between mathematical and empirical routes to knowledge, the very term "experimental mathematics" seems at best oxymoronic and at worst downright paradoxical.

My aim in this paper is twofold. Firstly, I want to discuss how to characterize experimental mathematics. Secondly, I want to consider how—if at all—experimental mathematics is philosophically interesting. In thinking about this latter issue I shall investigate the impact of experimental mathematics not just on traditional epistemological and ontological questions in the philosophy of mathematics but also its impact on issues of mathematical methodology.

2 Characterizing Experimental Mathematics

Presumably the key to understanding the nature of experimental mathematics lies in getting clear about what is implied in this context by the term "experimental." One natural suggestion is that experimental mathematics involves performing *mathematical experiments*, where the term "experiment" here is construed as literally as possible. This is the approach adopted by van Bendegem in his important 1998 paper, "What, If Anything, Is an Experiment in Mathematics?" According to van Bendegem, an experiment involves "the manipulation of objects, ... setting up processes in the 'real' world and ... observing possible outcomes of these processes."[5] His suggestion is that the natural way to get an initial grip on what a mathematical experiment might be is to consider how an experiment in this paradigmatic sense might have mathematical ramifications.

One example that van Bendegem cites dates back to work done by the 19th-century Belgian physicist Plateau on minimal surface area problems. By building various geometrical shapes out of wire and dipping these wire frames into a soap solution, Plateau was able to answer specific questions about the minimum surface bounding various particular shapes, and—eventually—to formulate some general principles governing the configurations of such surfaces.[6] One way of understanding what is going on in this example is that a physical experiment—the dipping of a wire frame into a soap solution—is producing results that are directly relevant to a certain class of mathematical problem.

The main drawback of this way of characterizing experimental mathematics is that it is too restrictive. Examples of the sort van Bendegem cites are extremely

[1] E.g., *The Journal of Experimental Mathematics* (established 1992).

[2] E.g., the Institute for Experimental Mathematics at the University of Essen.

[3] E.g., the Experimental Mathematics Colloquium at Rutgers University.

[4] E.g., Borwein and Bailey (2004a, b).

[5] Van Bendegem (1998, p. 172).

[6] A good source for more details on Plateau's experiments is Courant and Robbins (1941, pp. 386–397).

rare (a point which van Bendegem himself concedes), hence the impact of mathematical experiments of this sort on actual mathematical practice can only be very limited at best. Moreover, it cannot be only this, literal sense of experiment that mathematicians have in mind when they talk about—and do—experimental mathematics.

It is worth noting that this notion of mathematical experiment is not without philosophical interest. For example, is any purely mathematical hypothesis being tested in such cases? However, the cited example involving soap bubbles is explicitly confined to discovering putative answers to questions about minimum surfaces. If mathematical experiments of this sort are restricted to the context of discovery then it may be possible to retain the traditional picture of a priori mathematical knowledge in the context of justification. If so then the occurrence of mathematical experiments would be compatible with mainstream philosophies of mathematics.

So much for the most literal reading of "mathematical experiment." What about taking a diametrically opposite stance and regarding the meaning of "experiment" in this mathematical context as prima facie quite distinct from its normal usage. One example of this sort is the term "mathematical induction." As is well known, (correct) proofs by mathematical induction are deductively valid. As a method, mathematical induction is not inductive in the most common sense of the term "induction." It may turn out that we are driven to this position by lack of plausible alternatives. But if there is nothing distinctively experimental about mathematical experiments then the philosophical interest of this field is unclear. So it seems worthwhile to consider some alternative characterizations of experimental mathematics before taking this route.

One way of trying to find a middle ground between the literal reading and the dismissive reading of "mathematical experiment" is to think more in analogical or functional terms. In other words, perhaps "experimental mathematics" is being used to label activities which function within mathematics in a way analogous to the role of experiment in empirical science. Thus mathematical experiments may share some features with literal experiments, but not other features. Before proceeding with this line of analysis, let us look briefly at a case study.

A nice example of current work in experimental mathematics appears in one of the two recent books by Borwein and Bailey.[7] A real number is said to be *normal in base n* if every sequence of digits (of any given length) occurs equally often in its base-*n* expansion. A number is *absolutely normal* if it is normal in every base. Consider the following hypothesis:

Conjecture Every non-rational algebraic number is absolutely normal.

Borwein and Bailey used a computer to compute to 10,000 decimal digits the square roots and cube roots of the positive integers smaller than 1,000, and then they subjected these data to certain statistical tests.

[7] Borwein and Bailey (2004b, Chap. 4).

There are a couple of striking features of this example that may point to a more general characterization of experimental mathematics. Firstly, it involves the use of computers. Secondly, the results do not deductively entail the conjecture. Each of these features can be thought of as linking to features typically possessed by experiments in science. Thus most scientific experiments involve apparatus or instruments which augment our observational powers in some way, as happens with an electron microscope, an MRI scanner, or a telescope. Correspondingly, computers can be thought of as augmenting a mathematician's calculational powers. (Popular accounts of experimental mathematics often talk of computers allowing mathematicians to "peer into new mathematical realms" but it is fairly clear that this is meant only metaphorically.) Another standard feature of scientific experiments is that they produce results which may support some general hypothesis but never prove such an hypothesis. This closely parallels the second feature of the example presented above. The question that needs to be addressed, therefore, is whether one or both of these 'analogical' features—computer use, and non-deductive support—are essential to experimental mathematics. I shall consider these two features in turn.

2.1 Using Computers

A striking feature of contemporary work in experimental mathematics is that it is done *using computers*. Is this reliance on complex pieces of electronics what makes the field 'experimental?' If so, then this suggests the following characterization:C*

(C∗) X is a piece of experimental mathematics
$\Leftrightarrow$ a computer was used to generate X

Note, however, that even defenders of characterizing experimental mathematics as something along the lines of 'computer-aided mathematics' will probably not want *all* uses of computers to count. For example, a mathematician who composes her (traditional) proof using the word-processing software of her desktop computer is presumably not thereby engaging in experimental mathematics! So perhaps the above characterization should be modified as follows: C

(C) X is a piece of experimental mathematics
$\Leftrightarrow$ a computer was essential in generating X

There remains the issue of exactly what counts as an "essential" use of a computer. (For example, is using a computer to do 1,000,000 tedious but elementary calculations an essential use? At 1 min per calculation, it would take a human mathematician about a year of full-time work to complete the task.) Nonetheless, let us set aside further questions about how one might narrow down the relevant kinds of computer use so as to make (C) a plausible sufficient condition, and focus for the moment on the thesis that computer use is a *necessary* feature of experimental mathematics.

The first question that needs to be asked is how well this proposed characterization fits with actual mathematical practice. The answer to this question depends in large part on which aspects of mathematical practice are focused on. If one looks at what gets published in contemporary journals, books, and conferences devoted to experimental mathematics, the impression is that all the items are closely bound up with computers. For example, I have not managed to find a single paper published in more than a decade's worth of issues of *Experimental Mathematics* that does not involve the use of computers. What about the kinds of examples which mathematicians tend to offer as paradigms of experimental mathematics? Here the data is less clear. On the one hand, an informal survey suggests that the majority of such exemplars do involve the explicit use of computers. On the other hand, it is not uncommon for mathematicians also to cite one or more historical examples, from well before the computer age, to illustrate the purported pedigree of the subdiscipline.

The biggest practice-based challenge to equating experimental mathematics with computer-based mathematics comes from what self-styled experimental mathematicians say about their nascent discipline. For when mathematicians self-consciously reflect on the notion of experimental mathematics, they tend to reject the claim that computer use is a necessary feature. For example, the editors of the journal *Experimental Mathematics*—in their "statement of philosophy" concerning the scope and nature of the journal—make the following remarks:

> The word "experimental" is conceived broadly: many mathematical experiments these days are carried out on computers, but others are still the result of pencil-and-paper work, and there are other experimental techniques, like building physical models.[8]

And here is another passage with a similar flavor from mathematician Doron Zeilberger:

> [T]raditional experimental mathematics ... has been pursued by all the great, and less-great, mathematicians through the centuries, using pencil-and-paper.[9]

It seems fair to say that linking experimental mathematics to computer use fits well with what contemporary experimental mathematicians do but not so well with what they say.

I now want to highlight a second problem with the proposed characterization that is more philosophical in nature. Consider another widely-cited example of experimental mathematics which arises in connection with the Goldbach Conjecture (GC). This conjecture, which states that every even number greater than 2 is expressible as the sum of two primes, has no known proof. However, many individual instances of the conjecture have been checked, and—over the past couple of decades or so—the vast majority of this checking has been done using computers. As of April 2007, all even numbers up to 10^{18} had been verified to conform to GC, and this project (under the direction of Oliveira e Silva) is ongoing. This massive

8 http://www.expmath.org/expmath/philosophy.html.

9 Gallian and Pearson (2007, p. 14).

computation task is generally considered to be a paradigm example of experimental mathematics. And it seems clear that computers are playing an essential role here: no mathematician, or group of mathematicians, could hope to duplicate 10^{18} calculations by hand.

Note, however, that checking a particular even number—especially a relatively small even number—does not require computer assistance. Indeed tables of checked values of GC began appearing in the late 19th Century, well before the development of the electronic, digital computer. So if our characterization of experimental mathematics involves the essential use of computers, the checking of any individual instance of GC does not count as experimental mathematics. But this puts us in the rather odd position of having to say that some collections of pieces of non-experimental mathematics count (collectively) as experimental! Turning this around, a piece of experimental mathematics is not *decomposable*: it may be made up of procedures none of which are themselves experimental. Perhaps the consequence of non-decomposability is not a fatal flaw of the computer-based characterization of experimental mathematics, but it is certainly counterintuitive.[10]

2.2 Inductive Support

A second feature of both the Borwein and Bailey example and the Goldbach Conjecture example which has connections to experiments in the empirical sciences is the prima facie relation of *inductive support* between the results collected and the conjecture under consideration. The paradigm case here is the calculation of specific instances of some universal hypothesis for which a deductive proof is lacking. For example, a calculation which shows how some specific even number breaks down into two primes seems to bear a relation of inductive support to the Goldbach Conjecture, a relation that is analogous to the support that individual experimental results may impart to a universal scientific hypothesis. Focusing on this aspect suggests the following characterization: I

(1) X is a piece of experimental mathematics

$\Leftrightarrow$ X involves a relation of inductive support between evidence and a conjecture

Tempting though this approach may seem, like the computer-based approach it raises concerns with respect to both mathematical practice and philosophical cogency. I shall start by considering how well an inductive-based characterization fits with actual mathematical practice.

The main difficulty with (I) stems from an ambiguity over how to understand the core condition concerning inductiveness. One natural way to read it is as requiring that the evidence *be presented as providing inductive support* for a given conjecture. On this interpretation, however, there seem to be many cases of experimental mathematics which do not meet this condition. The most obvious class of examples are those which occur in the context of discovery. Various computations are carried out with the goal of suggesting some 'new' general conjecture. This conjecture may

[10] Note that this objection is independent of the issue of whether the boundary between experimental and non-experimental mathematics is vague.

then be proved along traditional lines. At no point in this process need the computations be offered as providing nondeductive support for the conjecture.

One response to this objection is that, even in the context of discovery, to interpret the accumulated evidence as suggesting a particular conjecture for attempted proof is already to implicitly see a relation of inductive support as obtaining. Franklin makes just this sort of argument in his 1987 paper on nondeductive logic in mathematics:

> It is obvious on reflection that a mathematician must use non-deductive logic in the first stages of his work on a problem. Mathematics cannot consist just of conjectures, refutations and proofs. Anyone can generate conjectures, but which ones are worth investigating? Which ones are relevant to the problem at hand? Which can be confirmed or refuted in some easy cases, so that there will be some indication of their truth in a reasonable time? Which might be capable of proof by a method in the mathematician's repertoire? ... [T]o say that some answers are better than others is to admit that some are, on the evidence he has, more reasonable than others, that is, are rationally better supported by the evidence.[11]

If this is right, then the role of experimental mathematics in the context of discovery need not be in conflict with the inductive characterization, (I), given above.

So much for mathematical practice. What about the philosophical ramifications of (I)? Central here is the issue of whether there *is* any genuine relation of inductive support in typical examples of experimental mathematics. Consider again the computation of instances of the Goldbach Conjecture. There are at least a couple of reasons for thinking that inductive support is more problematic in the mathematical context than in the empirical context. One reason is that we are never definitively 'stuck' with induction in mathematics: there is always the possibility that a watertight, deductive proof of the given result will be found. This is in contrast to the empirical case, where it seems as if the accumulation of instances of a general hypothesis is the best we can do in terms of support. A second, more acute, reason is that the domain over which induction is carried out is often demonstrably infinite. This raises serious problems when claims of inductive support are cashed out in probabilistic terms. How can any finite collection of instances of a universal hypothesis lend support to it when there are still an infinite number of instances that have been left unchecked?[12]

These philosophical issues concerning the status of inductive claims in mathematics raise a challenge to the adequacy of (I), but they may not be insuperable. Even if they can be dealt with, and even if induction in mathematics can be put on a firm philosophical footing, problems with the inductive characterization of experimental mathematics remain. One worry concerns putative examples which meet the criterion in (I) but do not seem to be experimental in any

[11] Franklin (1987, p. 1).

[12] For more on issues concerning the role and status of enumerative induction in mathematics, see Baker (2007).

interesting sense. For example, consider some general hypothesis, H, of the form "All numbers have property F." Imagine that H has been proved (by traditional means) to be true of all even numbers, but that the truth of the hypothesis remains open for odd numbers. The proof of the even-numbered cases seems to provide inductive support for H, yet surely this is not thereby an instance of experimental mathematics.

A second worry is in a sense the mirror image of one of the worries discussed earlier in connection with the computer-based characterization, (C). The claim there was that (C) implies that a piece of experimental mathematics is not necessarily 'decomposable:' the whole may count as experimental while none of its constituent parts do. By contrast, (I) seems to imply that experimental mathematics is not *agglomerative*. What I mean by this is that combining two pieces of experimental mathematics may produce a whole which is not experimental (according to (I)). For example, let GC* be the claim that the Goldbach Conjecture holds for all numbers less than or equal to 100,000. Let EM_1 be the computational checking of all even numbers from 2 to 50,000, and EM_2 be the computational checking of all even numbers from 50,002 to 100,000. Each of EM_1 and EM_2, taken individually, provides inductive support for GC*. However, taken as a whole, $\{EM_1, EM_2\}$ deductively entail GC*, so the whole does not count as experimental mathematics according to (I). Indeed the transition from experimental to non-experimental mathematics can be made to seem even more arbitrary. A computational check of all even numbers less than 100,000 would be experimental (relative to GC*), but checking 100,000 as well would render the whole procedure non-experimental. Even when the domain of the conjecture is infinite the relation between the calculation of an instance and the conjecture can be deductive: after all, a single calculation of an instance which does not fit the universal hypothesis deductively entails its falsity. I conclude that, as with the computer-based characterization, the inductive characterization of experimental mathematics latches on to a feature that is commonly present in cases of experimental mathematics but which as a condition is neither necessary nor sufficient.

3 Computation of Instances

Thus far we have looked at literal and 'dismissive' readings of the term "experimental", and then at two ways of drawing an analogy with features of experiments in the empirical sciences, based on the presence of computers and of inductive reasoning. The first two approaches are clearly inadequate, and while there are close links between experimental mathematics and both computer use and inductive reasoning, neither of these latter two features seem to be essential to the field. This does not mean, however, that we need to give up on the idea of finding a defining feature of experimental mathematics that is shared with experiments in science. The thesis that I shall propose and defend in this section is that experimental mathematics is essentially bound up with the *computation of instances*.

It is worth noting, even in advance of defending it, that the above thesis has the resources to explain why the computer-based and inductive characterizations of experimental mathematics were so tempting. The category of 'computation-aided mathematics' subsumes the category of 'computer-aided mathematics,' and we can explain the prevalence of the latter in contemporary experimental mathematics by the simple fact that computers are (typically) the fastest and most reliable means that we have available to perform large numbers of computations.[13] Similarly, the computation of instances often does result in a situation where the relation between evidence and conjecture is (merely) inductive. In each case, according to this new view of the situation, the proposed characterizations are latching onto a symptom of the more central, underlying feature of experimental mathematics.

The idea that experimental mathematics is bound up with computation of instances is not new. However, what discussion there has been in the literature has focused on computation rather than on instances. And this has led philosophers to abandon—prematurely, I shall argue—this approach. Nor have the objections offered against computation as a criterion for experimental mathematics been particularly compelling. For example, van Bendegem argues that computations lack certain features that are essential to experiments, and that decisively undermines the experiment—computation link. In particular, computations are proofs. Since a formal proof can, at least in principle, be evaluated purely at the level of syntax, this marks an important difference with our intuitive notion of experiment. I think that van Bendegem is right in the specific point that he makes but that he is wrong in drawing such a pessimistic conclusion from it. I have already argued that a literal reading of "experiment," in the context of clarifying the nature of experimental mathematics, is unfruitful. Hence we can acknowledge differences between mathematical experiments and scientific experiments without undermining a given approach.

More serious, perhaps, for the computational characterization is that many uses of computation fall unproblematically under traditional mathematical methodology and seem to have nothing to do with experimental mathematics. Consider, for example, the use of computation to verify that $717 \times 82 = 58{,}794$ by performing a long multiplication with pencil and paper. This points to a more general problem with focusing on computation as the characteristic feature of experimental mathematics, which is that there is no clear distinction to be drawn between computation and proof. Consider the (perhaps apocryphal) story from Gauss's childhood in which the teacher of his class asked every child to find the sum of all the numbers from 1 to 100. While his classmates laboriously summed each integer in turn, Gauss quickly realized that the sum can be rearranged as 50 pairs of numbers, $(1 + 100)$, $(2 + 99)$, $(3 + 98)$, ..., each summing to 101. So the overall sum is 5050. It is tempting to characterize this latter approach by saying that Gauss *proved* that the sum is 5050, while his classmates *computed* this sum. But wasn't Gauss also performing a computation, at least in the initial stage of his proof, when he rearranged the original sum into pairs and noted the sum of each pair?

There has, of course, been plenty of attention and resources devoted to defining the associated notion of *computability*. Doesn't getting clear about this, which concerns what can be computed—whether this is cashed out in terms of Turing

[13] Mathematician George Andrews has described a computer as "a pencil with power-steering."

machines, primitive recursive functions, or whatever—also yield an answer to the question of what counts as a computation? The problem is that this general notion of computation still seems too broad for the purposes at hand. The simple arithmetical calculations mentioned above are clearly computations, but they are equally clearly not experimental in any interesting sense.

Given the difficulties associated with circumscribing a sense of computation which fits the purposes at hand, why not abandon the task of doing this in isolation and instead focus specifically on computation of *instances*? To highlight this shift of emphasis, we can also talk not of "computation" but (in less loaded terms) of "calculation." This leads, to a first approximation, to the following proposal: EM

(EM) X is a piece of experimental mathematics

$\Leftrightarrow$ X involves calculating instances of some general hypothesis

One immediate advantage of this characterization is that the notion of an instance can be defined precisely in purely logical terms. It also fits nicely with paradigm cases of experimental mathematics such as the checking of particular instances of the Goldbach Conjecture.

What issues are raised if we adopt the proposed characterization? One worry is whether all the central sorts of cases of experimental mathematics really are of this instance/general hypothesis form. Take the Borwein and Bailey example discussed earlier, where statistical tests were run on strings of digits of π to test for (absolute) normality. At first glance this does not seem to fit the template required by (EM). But perhaps it does have the right form if we reconceptualize what is going on. The procedure consists of randomly generating some n-digit string, say '30944,' and then testing to see how often it occurs in the decimal expansion of π. First it is compared with the first five digits of π, '14195', then the five digits beginning with the second place in π, '41952', and so on. There are 1 million possible combinations of 5 digits, so if '30944' matches approximately one every million times then this supports the hypothesis that π is normal. This hypothesis is a general one, in the sense that it applies to every string of digits of any finite length. The claim that some specific string, such as '30944,' appears at the expected frequency across the digits of π can then be seen as an instance of the general hypothesis. Thus even this sort of statistical example fits the instance/general hypothesis model.

Another potential problem for (EM) stems from the specificity of the notion of instance. Consider the celebrated computer-assisted proof of the Four-Color Theorem by Appel and Haken in 1976. Using mathematical rules and procedures based on properties of reducible configurations (e.g., the method of discharging, rings, Kempe chains) they found an unavoidable set of reducible configurations, thus proving that a minimal counterexample to the four-color conjecture could not exist. Their proof reduced the infinitude of possible maps to 1,936 reducible configurations which had to be checked individually by computer.This reducibility part of the work was independently double checked with different programs and computers.[14] This proof is generally considered to be a piece of experimental mathematics *par excellence*. Yet it might be objected that there is no calculation of instances involved

[14] See Appel and Haken (1978) for more details of their proof.

because each of the 1.936 reducible configurations represents a whole class of associated maps. On the other hand, while it is true that each of the reducible configurations represents a type, the actual configuration checked by the computer program is a fully specific, token map. Thus the overall procedure can still be thought of as checking instances, even if the conclusions drawn are more general.[15]

4 Philosophical Consequences

There is nothing philosophical controversial, or mathematically unfounded, about calculating instances of a general hypothesis. To this extent, if the thesis advanced in Sect. 3 is right and if something along the lines of (EM) correctly captures the essence of experimental mathematics, then there is nothing philosophically controversial about experimental mathematics per se. The same cannot be said, however, about the involvement of computers or about claims of inductive support in the mathematical context. Both phenomena often arise in conjunction with experimental mathematics, but neither is essential to it. Hence it makes sense to view these philosophically controversial aspects as freestanding issues in their own right.

One way of framing the situation is in terms of means and ends. Experimental mathematics, or so I have argued, involves the calculation of instances of some general hypothesis, and it is uncontroversial. Often the *means* chosen to carry out these calculations involves the use of computers. And sometimes the *ends* to which these calculations are put include claims of inductive support for the given general hypothesis. It is these means and ends which raise challenges to certain 'traditional' philosophical accounts of mathematics, and not the experimental mathematical core itself.

Consider the issue of computer-based mathematics. Some philosophers have argued that a proof that makes essential use of computers produces knowledge that is a posteriori, in contrast to the a priori knowledge produced by traditional mathematics.[16] This poses an obvious challenge to mainstream views in the philosophy of mathematics, however its impact on mathematical practice seems to be much more marginal. Mathematicians have certainly expressed concerns over computer-based proofs, but for the most part these concerns do not have to do with the kind of knowledge they produce. Instead they are connected to worries over reliability on the one hand and explanatoriness on the other.

The reliability of a given method of proof is clearly important, whether or not one is dealing with computer-based proofs. The other positive or negative features of a given proof are irrelevant if it does not in fact establish the result which it purports to prove. Mathematicians who use computers to help construct proofs tend to be sensitive to this concern. For example, the computer calculation of instances of the Goldbach Conjecture up to 10^{18} that is being carried out by a team led by Oliveira e

[15] Part of the ambiguity here between type and token stems from the fact that it is only the topological features of maps which matter for the purposes of the Four-Color Conjecture. Hence there are many features of individual maps which are irrelevant, so one such map can 'stand in' for a whole subclass.

[16] See, e.g., Tymoczko (1979).

Silva are each being double-checked. One indication that the worry here is one of reliability rather than of computer use per se is that the double-checking is also being carried out using computers (running different software and hardware). Clearly this procedure does nothing to break out of the a posteriori mode of reasoning—assuming it was a posteriori in the first place—though it does increase confidence in the accuracy of the overall result. Another area in which reliability issues are pushed to the fore is in the development of automated proof-checkers. Again, this may lead to one computer checking the accuracy of the computations of another computer. In general, the mathematical community will want to be assured that the computer hardware and software used to implement a particular proof is running properly. But beyond this there seems little anxiety, justification-wise, stemming from the bare fact that the proof is computer-based.

A second source of concern among mathematicians in connection with computer proofs pertains to understanding. There is a widespread view that computer proofs are unilluminating, that they offer little insight into *why* a given result is true. In other words, computer proofs are not *explanatory*. (Of course there are many unexplanatory non-computer-based proofs. The point is that there is some particular connection between computer methods and lack of explanatoriness.) The topic of explanation in mathematics has received scant attention from philosophers until quite recently.[17] There has been increasing realization, quite independently of issues concerning computer proofs and experimental mathematics, that mathematical proofs do more than simply justify their conclusions. If nothing else, taking into account the potential explanatory role of proof helps in itself to explain why mathematicians often produce many quite distinct proofs of one and the same theorem.

Why should the mere fact that a proof is produced using a computer render it less explanatory? One idea—to avoid the implausible conclusion that computers themselves somehow impede mathematical understanding—is that both recourse to computers and non-explanatoriness are symptoms of the same underlying feature. One candidate here is *disjunctiveness*. The most common reason for resorting to computers to help with proving a given result is because we are faced with an unfeasibly large number of particular cases that need to be worked through in order to verify a general claim. Appel and Haken's proof of the Four-Color Theorem, has this form. The checking of 1,476 different sub-cases of graphs, each involving the application of over 300 different "discharging rules," produced a proof which is highly disjunctive. This links to one recent approach to analyzing mathematical explanation, which borrows the unification model which Friedman and Kitcher developed for scientific explanation. If explanation is indeed tied closely to unification then it is not hard to see how the disjunctiveness characteristic of computer proof tends to yield proofs that are also considered relatively unexplanatory by mathematicians.

As has already been mentioned there has been a recent upsurge of interest among philosophers in the issue of mathematical explanation. It is worth distinguishing here between what might be termed the *internal* and the *external* explanatory roles of mathematics.[18] The internal question revolves around what makes one proof of a

[17] One exception is work by Mark Steiner in the late 1970s. See e.g., Steiner (1978).

[18] See Mancosu (2001) for a useful overview of contemporary work on mathematical explanation.

given result more (or less) explanatory than another. The suggestion canvased in Sect. 3 was that computer proofs may tend to be less explanatory than traditional proofs because they are more disjunctive, and disjunctiveness reduces explanatoriness. The external question concerns whether mathematics does (or can) ever play a genuine explanatory role, for example in explaining physical phenomena. Among those interested in this question are philosophers trying to strengthen the Quine-Putnam indispensability argument for mathematical platonism. This argument, that we ought to be committed to the existence of abstract mathematical entities because of their indispensable role in our best scientific theories, has been challenged by nominalists on the grounds that the actual role played by mathematics in science is very different from that played by concrete theoretical posits such as electrons and black holes. But if it can be shown that the role of mathematics—in at least some cases—is *explanatory*, then this opens the way for use of inference to the best explanation to argue for the existence of mathematical entities.[19]

We have now said a little bit about how 'traditional' philosophical worries about computer-based proofs contrast with the worries of practicing mathematicians. Roughly, philosophers have tended to worry about the type of epistemic warrant provided by computer proofs (a posteriori versus a priori), while mathematicians are typically more concerned about the reliability and the explanatory value of such proofs.

5 Conclusion

To summarize, I have argued in this paper that there is nothing intrinsically problematic about experimental mathematics, and that—despite initial appearances to the contrary—it poses no threat to traditional philosophical views of the nature of mathematics. Why not? Because the key feature that links experimental mathematics with experiments in science concerns the calculation of *instances* of some general hypothesis. And this is quite compatible with the view that mathematics is (say) a priori and deductive at its core.

This is not to say that the pursuit of experimental mathematics raises no interesting philosophical issues. The means by which the field is pursued typically involves reliance—sometimes almost total reliance—on electronic computers. And the claims made using the results that are gathered are often directed towards the end of claiming inductive support for a general hypothesis. Each of these raises interesting issues both philosophical and practical. Confusingly, both computer use and inductive reasoning also have links to aspects of experimentation in science. What I have argued, however, is that neither is an essential feature of experimental mathematics. There is experimental mathematics that makes no use of computers, and there is experimental mathematics that involves no inductive relations—claimed or actual—between evidence and hypothesis.

Another interesting aspect of these symptomatic (but non-essential) features of experimental mathematics is they can be hard to circumvent. More precisely,

[19] For more on this debate, see Baker (2005).

experimental mathematicians often face the situation where one feature can be avoided only by introducing the other. One such case is where several instances of a general hypothesis have been computed ‘by hand’, and are offered as providing inductive support for it. If the instances can be divided into a large but finite number of classes, then it may be possible to turn this inductive support into a deductive proof using the resources of a computer. Arguably, this is what took place with respect to the Four-Color Theorem. An inductive, ‘computer-free’ argument is turned into a deductive, computer-based proof. And one potentially problematic symptom of experimental mathematics is thereby replaced with the other.

The thesis that the essence of experimental mathematics lies in the calculation of instances is not intended to be the last word on the topic. There are further challenges to be raised and more work potentially to be done in order to refine and precisify this characterization. Relevant questions include:

(i) In the context of discovery, can the notion of ‘instance’ coherently be applied prior to any general hypothesis having been suggested?
(ii) Don’t traditional, deductively valid (and, presumably, non-experimental) proofs by mathematical induction involve the calculation of *one* instance, namely the base step?
(iii) Is *any* general hypothesis a candidate for having instances, even hypotheses with bizarre, gerrymandered predicates?

However, none of these challenges threaten the core claim of the essential *innocuousness* of experimental mathematics. It is innocuous because it is experimental in a way which is consistent with both a prioricity and deductive validity.

References

Appel, K., & Haken, W. (1978). The four-color problem. In L. Steen (Ed.), *Mathematics today: Twelve informal essays* (pp. 153–180). New York: Springer-Verlag.

Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind, 114*, 223–238.

Baker, A. (2007). Is there a problem of induction for mathematics? In M. Potter (Ed.), *Mathematical knowledge* (pp. 59–73). Oxford: Oxford University Press.

Borwein, J., & Bailey, D. (2004a). *Experimentation in mathematics: Computational paths to discovery*. Natick, MA: AK Peters.

Borwein, J., & Bailey, D. (2004b). *Mathematics by experiment: Plausible reasoning for the 21st century*. Natick, MA: AK Peters.

Courant, R., & Robbins, H. (1941). *What is mathematics?* Oxford: Oxford University Press.

Franklin, J. (1987). Non-deductive logic in mathematics. *British Journal for the Philosophy of Science, 38*, 1–18.

Gallian, J., & Pearson, M. (2007). An interview with Doron Zeilberger. *MAA Focus, May/June*, 14–17.

Mancosu, P. (2001). Mathematical explanation: Problems and prospects. *Topoi, 20*, 97–117.

Steiner, M. (1978). Mathematical explanation. *Philosophical Studies, 34*, 135–151.

Tymoczko, T. (1979). The four-color problem and its philosophical significance. *Journal of Philosophy, 76*, 57–83.

van Bendegem, J.-P. (1998). What, if anything, is an experiment in mathematics? In D. Anapolitanos, A. Baltas, & S. Tsinorema (Eds.), *Philosophy and the many faces of science* (pp. 172–182). London: Rowman & Littlefield.