Making Sense of Experimental Mathematics

J. Borwein, P. Borwein, R. Girgensohn and S. Parnes

June 9, 1995

1 Introduction

Philosophers have frequently distinguished mathematics from the physical sciences. While the sciences were constrained to fit themselves via experimentation to the 'real' world, mathematicians were allowed more or less free reign within the abstract world of the mind. This picture has served mathematicians well for the past few millennia but the computer has begun to change this. The computer has given us the ability to look at new and unimaginably vast worlds. It has created mathematical worlds that would have remained inaccessible to the unaided human mind, but this access has come at a price. Many of these worlds, at present, can only be known experimentally. The computer has allowed us to fly through the rarefied domains of hyperbolic spaces and examine more than a billion digits of π but experiencing a world and understanding it are two very different phenomena. Like it or not, the world of the mathematician is becoming experimentalized.

The computers of tomorrow promise even stranger worlds to explore. Today, however, most of these explorations into the mathematical wilderness remain isolated illustrations. Heuristic conventions, pictures and diagrams developing in one sub-field often have little content for another. In each sub-field unproven results proliferate but remain conjectures, strongly held beliefs or perhaps mere curiosities passed like folk tales across the Internet. The computer has provided extremely powerful computational and conceptual resources but it is only recently that mathematicians have begun to systematically exploit these abilities. It is our hope that by focusing on experimental mathematics today, we can develop a unifying methodology tomorrow.

1.1 Our Goals

The genesis of this article was a simple question: "How can one use the computer in dealing with computationally approachable but otherwise intractable problems in mathematics?" We began our current exploration of *experimental mathematics* by examining a number of very long–standing conjectures and strongly held beliefs regarding decimal and continued fraction expansions of certain elementary constants. These questions are uniformly considered to be hopelessly intractable given present mathematical technology. Unified field theory or cancer's "magic bullet" seem accessible by comparison. But like many of the most tantalizing problems in mathematics their statements are beguilingly simple. Since our experimental approach was unlikely to result in any new discoveries[1], we focused on two aspects of experimentation: systematization and communication.

For our attempted systematization of *experimental mathematics* we were concerned with producing data that were 'completely' reliable and insights that could be quantified and effectively communicated. We initially took as our model experimental physics. We were particularly interested in how physicists verified their results and the efforts they took to guarantee the reliability of their data. The question of reliability is undoubtedly central to mathematicians and here we believe we can draw a useful distinction between experimental physics and mathematics. While it is clearly impossible to extract perfect experimental data from nature such is not the case with mathematics. Indeed, reliability of raw mathematical data is far from the most vexing issue.

Let us turn to our second and primary concern: insight. All experimental sciences turn on the intuitions and insights uncovered through modeling and the use of probabilistic, statistical and visual analysis. There is really no other way to proceed, but this process even when applied to mathematics inevitably leads to some considerable loss of exactness.

The communication of insight, whether derived from mathematical experiment or not, is a complex issue. Unlike most experimentalized fields, Mathematics does not have a 'vocabulary' tailored to the transmission of condensed data and insight. As in most physics experiments the amount of raw data obtained from mathematical experiments is, in general, too large for anyone to grasp. The collected data needs to be compressed and compartmentalized. To make up for this lack of unifying vocabulary we have borrowed heavily from statistics and data analysis to interpret our results. For now we have used restraint in the presentation of our results in what we hope is an intuitive, friendly and convincing manner. Eventually what will probably be required is a multi-leveled hyper-textual presentation of mathematics, allowing mathematicians from diverse fields to quickly examine and interpret the results of others – without demanding the present level of specialist knowledge. [Not only do mathematicians have trouble communicating with lay audiences, but they have significant difficulty talking to each other. There are hundreds of distinct mathematical languages. The myth of a universal language of mathematics is just that. Many subdisciplines simply can not comprehend each other.]

[1]We will not discuss the computational difficulties here but there are many non-trivial mathematical and computer–related issues involved in this project.

1.2 Unifying Themes

We feel that many of these problems can be addressed through the development of a rigorous notion of experimental mathematics. In keeping with the positivist tradition, mathematics is viewed as the most exact of sciences and mathematicians have long taken pride in this. But as mathematics has expanded, many mathematicians have begun to feel constrained by the bonds placed upon us by our collective notion of proof. Mathematics has grown explosively during our century with many of the seminal developments in highly abstract seemingly non-computational areas. This was partly from taste and the power of abstraction but, we would argue, equally much from the lack of an alternative. Many intrinsically more concrete areas were, by 1900, explored to the limits of pre-computer mathematics. Highly computational, even "brute–force" methods were of necessity limited but the computer has changed all that. A re-concretization is now underway. The computer–assisted proofs of the four color theorem are a prime example of computer–dependent methodology and have been highly controversial despite the fact that such proofs are much more likely to be error free than, say, even the revised proof of Fermat's Last Theorem.

Still, these computerized proofs need offer no insight. The Wilf/Zeilberger algorithms for 'hypergeometric' summation and integration, if properly implemented, can rigorously prove very large classes of identities. In effect, the algorithms encapsulate parts of mathematics. The question raised is: " How can one make full use of these very powerful ideas?" Doron Zeilberger has expressed his ideas on experimental mathematics in a paper dealing with what he called 'semi-rigorous' mathematics. While his ideas as presented are somewhat controversial, many of his ideas have a great deal of merit.

The last problem is perhaps the most surprising. As mathematics has continued to grow there has been a recognition that the age of the mathematical generalist is long over. What has not been so readily acknowledged is just how specialized mathematics has become. As we have already observed, subfields of mathematics have become more and more isolated from each other. At some level, this isolation is inherent but it is imperative that communications between fields should be left as wide open as possible. As fields mature, speciation occurs. The communication of sophisticated proofs will never transcend all boundaries since many boundaries mark true conceptual difficulties. But experimental mathematics, centering on the use of computers in mathematics, would seem to provide a common ground for the transmission of many insights. And this requires a common meta–language[2]. While such a language may develop largely independent of any conscious direction on the part of the mathematical community, some focused effort on the problems of today will result in fewer growing pains tomorrow.

[2] This may not be a fanciful dream as the *Computer Algebra Systems* (CAS) of today are beginning to provide just that.

2 Experimental Mathematics

2.1 Journal of

A professor of psychology was exploring the creative process and as one of his subjects chose a mathematician who was world famous for his ability to solve problems. They gave him a problem to work on. He wrote something down and immediately scribbled it out. He wrote something else down and scribbled it out. The professor asked him to leave everything on the page. He explained that he was interested in the process, the wrong answers and the right answers. The mathematician sat down. Wrote something. The psychology professor waited in anticipation but the mathematician announced he could not proceed without erasing his mistakes. While the mathematician in this situation is undoubtedly fairly idiosyncratic in how he attacks problems there is a strongly felt separation between the creative process of mathematics and the published or finished product.

A current focal point for experimental mathematics is the journal called *Experimental Mathematics*. But does it really seek to change the way we do mathematics, or to change the way we write mathematics? We begin by attempting to extract a definition of 'experimental' from the Journal's introductory article ([**?**]) "About this Journal" by David Epstein, Silvio Levy and Rafael de la Llave.

> 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. ([**?**] p. 1)

It seems that almost anything can be conceived of as being experimental. Let us try again.

> Experiment has always been, and increasingly is, an important method of mathematical discovery. (Guass declared that his way of arriving at mathematical truths was 'through systematic experimentation'.) Yet this tends to be concealed by the tradition of presenting only elegant, well-rounded and rigorous results. ([**?**] p. 1)

Now we begin to get closer to the truth. Experimentation is still ill defined but is clearly an important part of the mathematical process. It is clearly not new but by implication must be inelegant, lopsided and lax. We, of course, dispute all three of these points and while we do not reply directly to these charges, we hope the reader will be convinced that there need be no compromises made with respect to the quality of the work.

But what is the journal interested in publishing? Their goal seems to be two-fold.

> While we value the theorem-proof method of exposition, and while we do not depart from the established view that a result can only become part of mathematical knowledge once it is supported by a logical proof, we consider it anomalous that an important component of the process of mathematical creation is hidden from public discussion. It is to our loss that most of the mathematical community are almost always unaware of how new results have been discovered. ([?] p. 1)

and

> The early sharing of insights increases the possibility that they will lead to theorems: an interesting conjecture is often formulated by a researcher who lacks the techniques to formalize a proof, while those who have the techniques at their fingertips have been looking elsewhere.

It appears that through the journal Experimental Mathematics the editors advocate a not undramatic change in writing style. So what does a paper published in that journal look like? A recent example is "Experimental Evaluation of Euler sums" by D.H. Bailey, J. Borwein and R. Girgensohn ([?]). The authors describe how their interest in Euler sums was roused by a surprising discovery:

> In April 1993, Enrico Au-Yeung, an undergraduate at the University of Waterloo, brought to the attention of one of us the curious fact that
>
> $$\begin{aligned} \sum_{k=1}^{\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right)^2 k^{-2} &= 4.59987\cdots \\ &\approx \frac{17}{4}\zeta(4) = \frac{17\pi^4}{360} \end{aligned}$$
>
> based on a computation to 500,000 terms. This author's reaction was to compute the value of this constant to a higher level of precision in order to dispel this conjecture. Surprisingly, a computation to 30 and later to 100 decimal digits still affirmed it. ([?] p. 17)

This type of serendipitous discovery must go on all the time, but it needs the flash of insight that will place it in a broader context. It is like a gold nugget waiting to be refined — without a context it would remain a curiosity. The authors now proceeded to provide a context by mounting a full-fledged assault on the problem. They systematically applied an integer relation detection algorithm to large classes of sums of the above type, trying to find evaluations of these sums in terms of zeta functions (see box for details). Some of the experimentally discovered evaluations were then proven rigorously, others remain conjectures. While Au-Yeung's insight may fill us with a sense of amazement, the experimenters' approach appears quite natural and systematic.

Serendipity and Experimentation

After Enrico Au-Yeung's serendipitous discovery, D. Bailey, J. Borwein and R. Girgensohn launched a full fledged assault on the problem. This is documented in *Experimental Detection of Euler Sums* (the material below was taken from David Bailey's slides).

Experimental Approach

1. Employ an advanced scheme to compute high-precision (100+ digit) numerical values for various constants in a class.
2. Conjecture the form of terms involved in possible closed-form evaluations.
3. Employ an integer relation finding algorithm to determine if an Euler sum value is given by a rational linear combination of the conjectured terms.
4. Attempt to find rigorous proofs of experimental results.
5. Attempt to generalize proofs for specific cases to general classes of Euler sums.

The editors of *Experimental Mathematics* are advocating a change in the way mathematics is written, placing more emphasis on the mathematical process. Imre Lakatos in his influential though controversial book *Proofs and Refutations* [?] advocated a similar change from what he called the deductivist style of proof to the heuristic style of proof. In the deductivist style, the definitions are carefully tailored to the proofs. The proofs are frequently elegant and short. But it is difficult to see what process led to the discovery of the theorem and its proof. The heuristic style maintains the mathematical rigor but again the emphasis is more on process. One does not merely give the definition but perhaps includes a comment on why this definition was chosen and not another. This is clearly an important shift if the editors wish to meet their second objective, the sharing of insights.

Some Experimental Results

Definitions

$$\zeta(s) := \sum_{k=1}^{\infty} k^{-s}$$

$$s_{\mathrm{h}}(m,n) := \sum_{k=1}^{\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right)^{m} (k+1)^{-n} \quad m \geq 1,\, n \geq 2$$

Some experimentally derived conjectures

$$\begin{aligned}
s_{\mathrm{h}}(3,2) &= \frac{15}{2}\zeta(5) + \zeta(2)\zeta(3) \\
s_{\mathrm{h}}(3,3) &= -\frac{33}{16}\zeta(6) + 2\zeta^2(3) \\
s_{\mathrm{h}}(3,4) &= \frac{119}{16}\zeta(7) - \frac{33}{4}\zeta(3)\zeta(4) + 2\zeta(2)\zeta(5) \\
s_{\mathrm{h}}(3,6) &= \frac{197}{24}\zeta(9) - \frac{33}{4}\zeta(4)\zeta(5) - \frac{37}{8}\zeta(3)\zeta(6) + \zeta^3(3) + 3\zeta(2)\zeta(7) \\
s_{\mathrm{h}}(4,2) &= \frac{859}{24}\zeta(6) + 3\zeta^2(3)
\end{aligned}$$

We are given the raw data with which to work, carefully organized to give us a glimpse into the investigators' insights on the problem. Note in the first formula for $s_{\mathrm{h}}(3,2)$, $3+2=5$, on the right hand side of the equation we have $\zeta(5)$ and $\zeta(3)\zeta(2)$.

Some Proven Euler Sums

$$\begin{aligned}
s_{\mathrm{h}}(2,2) &= \frac{3}{2}\zeta(4) + \frac{1}{2}\zeta^2(2) = \frac{11\pi^4}{360} \\
s_{\mathrm{h}}(2,4) &= \frac{2}{3}\zeta(6) - \frac{1}{3}\zeta(2)\zeta(4) + \frac{1}{3}\zeta^3(2) - \zeta^2(3) = \frac{37\pi^6}{22680} - \zeta^2(3)
\end{aligned}$$

The proven evaluation for $s_{\mathrm{h}}(2,2)$ above implies the truth of Au-Yeung's discovery.

2.2 The Deductivist Style

The major focus of this section is Imre Lakatos's description of the deductivist style in *Proofs and Refutations.* An extreme example of this style is given in the form of a computer generated proof of $(1+1)^n = 2^n$ in the box.

> Euclidean Methodology has developed a certain obligatory style of presentation. I shall refer to this as 'deductivist style'. This style starts with a painstakingly stated list of *axioms, lemmas* and/or *definitions.* The axioms and definitions frequently look artificial and mystifyingly complicated. One is never told how these complications arose. The list of axioms and definitions is followed by the carefully worded *theorems.* These are loaded with heavy-going conditions; it seems impossible that anyone should ever have guessed them. The theorem is followed by the *proof.* ([?] p. 142)

This is the essence of what we have called formal understanding. We know that the results are true because we have gone through the crucible of the mathematical process and what remains is the essence of truth. But the insight and thought processes that led to the result are hidden.

> In deductivist style, all propositions are true and all inferences valid. Mathematics is presented as an ever-increasing set of eternal, immutable truths. ([?] p. 142)

> Deductivist style hides the struggle, hides the adventure. The whole story vanishes, the successive tentative formulations of the theorem in the course of the proof-procedure are doomed to oblivion while the end result is exalted into sacred infallibility. ([?] p. 142)

Perhaps the most extreme examples of the deductivist style come out of the computer generated proofs guaranteed by Wilf and Zeilberger's algorithmic proof theory. It is important to note here that Wilf and Zeilberger transform the problem of proving identities to the more computer oriented problem of solving a system of linear equations with symbolic coefficients.

Shrinking or Encapsulating Mathematics

When one first learns to sum infinite series one is taught to sum geometric series

$$1+\rho+\rho^2+\cdots+\rho^k+\cdots=\frac{1}{1-\rho}$$

when $|\rho|<1$. Next one learns to sum telescoping series. For example if $f(i) := \frac{1}{i+1}-\frac{1}{i+2}$, it is not to hard to see that

$$\sum_{i=0}^{n} f(i) = (1-\frac{1}{2})+(\frac{1}{2}-\frac{1}{3})\cdots(\frac{1}{n+1}-\frac{1}{n+2}) = 1-\frac{1}{n+2}$$

and in particular that

$$\sum_{i=0}^{\infty} f(i) = 1.$$

The **Wilf-Zeilberger** algorithms employ "creative telescoping' to show that a sum or integral is zero. The algorithms really provide a meta-insight into a broad range of problems involving identities. Unfortunately the proofs produced by the computer, while understandable by most mathematicians are at the same time uninteresting. On the other hand, the existence of WZ proofs for large classes of objects gives us a global insight into these areas.

These WZ proofs (see next page) are perhaps the ultimate in the deductivist tradition. At present, knowing the WZ proof of an identity amounts to little more (We will discuss the importance of certificates later.) than knowing that the identity is true. In fact, Doron Zeilberger in [**?**] has advocated leaving only a QED at the end of the statement, the author's seal that he has had the computer perform the calculations needed to prove the identity. The advantage of this approach is that the result is completely encapsulated. Just as one would not worry about how the computer multiplied two huge integers together or inverted a matrix, one now has results whose proofs are uninteresting.

An Uninteresting(?) Proof

Below is a sample WZ proof of $(1+1)^n = 2^n$ (this proof is a modified version of the output of Doron Zeilberger's original Maple program, influenced by the proof in [?]).
Let $F(n,k) = \binom{n}{k} 2^{-n}$. We have to show that $l(n) := \sum_k F(n,k) = 1$. To do this we will show that $l(n+1) - l(n) = 0$ for every $n \geq 0$ and that $l(0) = 1$. The second half is trivial since for $n = 0$, $F(0,0)$ is equal to 1 and 0 otherwise. The first half is proved by the WZ algorithm.
We construct

$$G(n,k) := \frac{-1}{2^{(n+1)}} \binom{n}{k-1} \left(= \frac{-k}{2(n-k+1)} F(n,k) \right),$$

with the motive that

$$WZ := F(n+1,k) - F(n,k) = G(n,k+1) - G(n,k) \quad \text{(check!)}.$$

Summing WZ with respect to k gives

$$\sum_k F(n+1,k) - \sum_k F(n,k) =$$

$$\sum_k (G(n,k+1) - G(n,k)) = 0$$

(by telescoping). We have now established that $l(n+1) - l(n) = 0$ and we are done.
The proof gives little insight into this binomial coefficient identity. However, the algorithms give researchers in other fields direct access to the field of special function identities.

3 Zeilberger and the Encapsulation of Identity

3.1 Putting a price on reliability

In the last two sections we talked about the importance of communicating insights within the mathematical community. There we focused on the process of mathematical thought but now we want to talk about communicating insights that have not been made rigorous.

We have already briefly talked about Wilf and Zeilberger's algorithmic proof theory and its denial of insight. In this section we will discuss the implications of this theory and D. Zeilberger's philosophy of mathematics as contained in *Theorems for a Price: Tomorrow's Semi-Rigorous Mathematical Culture* ([?]).

It is probably unfortunate but perhaps necessary that the two voices most strongly advocating truly experimental math are also at times the most hyperbolic in their language. We will concentrate mostly on the ideas of Doron Zeilberger but G. J. Chaitin should not and will not be ignored.

We will begin with D. Zeilberger's "Abstract of the future"

> We show in a certain precise sense that the Goldbach conjecture is true with probability larger than 0.99999 and that its complete truth could be determined with a budget of 10 billion. ([?] p. 980)

Once people get over the shock of seeing probabilities assigned to truth in mathematics the usual complaint is that the 10 billion is ridiculous. Computers have been getting better and cheaper for years. What can it mean that "the complete truth could be determined with a budget of 10 billion?" What is clear from the article is that this is an additive measure of the difficulty of completely solving this problem. If we know that the Reimann hypothesis will be proven if we prove lemmas costing 10 billion, 2 billion and 2 trillion dollars respectively, we can tell at a glance not merely what it would 'cost' to prove the hypothesis but also where new ideas will be essential in any proof. (This assumes that 2 trillion is a lot of 'money'.)

The introduction of 'cost' leads immediately to consideration of a trend that has over taken the business world and is now intruding rapidly on academia: a focus on productivity and efficiency.

> It is a waste of money to get absolute certainty, unless the conjectured identity in question is known to imply the Riemann Hypothesis ([?] p. 980)

We have taken this quote out of its context (Wilf and Zeilberger's algorithmic proof theory of identities) [?] but even so We think it is indicative of a small but growing group of mathematicians who are asking us to to look at not just the benefits of reliability in mathematics but also the associated costs. See for example A. Jaffe and F. Quinn in [?] and G. Chaitin in [?]. Still, we have not dealt with the central question. Why does D. Zeilberger need to introduce probabilistic 'truths'? and how might we from a 'formalist' perspective not feel this to be a great sacrifice?

3.2 It's all about insight

Why is Zeilberger so willing to give up on absolute truths? The most reasonable answer is that he is pursuing deeper truths. In *Identities in Search of Identities*, Zeilberger advocates an examination of identities for the sake of studying identities. Still as Herb Wilf and others have pointed out it is possible to produce an unlimited number of identities. It is the context, the ability to use and manipulate these identities, that make them interesting. Why then might we think that studying identities for their own sake may lead us down the golden path rather than the garden path?

We are now looking for what might be called meta-mathematical structures. We remove the math from its original context and isolate it, trying to detect new structures. When doing this it is impossible to collect only the relevant information that will lead to the new discovery. One collects objects (theorems, statistics, conjectures, etc.) that have a reasonable degree of similarity and familiarity and then attempts to eliminate the irrelevant or the untrue (counter examples). We are preparing for some form of eliminative induction. There is a built in stage, where objects are subject to censorship. In this context, it is not unreasonable to introduce objects where one is not sure of their truth, since all the objects, whether proved or not, will be subject to the same degree of scrutiny. Moreover, if these probably true objects fall into the class of reliable (i.e., they fit the new conjecture) objects, it may be possible to find a legitimate proof in the new context. Recall that the fast WZ algorithms transform the problem of proving an identity to one of solving a system of linear equations with symbolic coefficients.

> It is very time consuming to solve a system of linear equations with symbolic coefficients. By plugging in specific values for n and other parameters if present, one gets a system with numerical coefficients, which is much faster to handle. Since it is unlikely that a random system of inhomogeneous linear equations with more equations than unknowns can be solved, the solvability of the system for a number of special values of n and the other parameters is a very good indication that the identity is indeed true. ([**?**] p. 980)

Suppose we can solve the system above for ten different assignments for n and the other parameters but cannot solve the general system. What do we do if we really need this identity? We are in a peculiar position. We have reduced the problem of proving identities involving sums and integrals of proper-hypergeometric terms to the problem of solving a possibly gigantic system of inhomogeneous linear equations with more equations than unknowns. We have an appropriately strong belief that this system has a solution but do not have the resources to uncover this solution.

What can we do with our result? If we agree with G. J. Chaitin, we may want to introduce it as an 'axiom'.

> I believe that elementary number theory and the rest of mathematics should be pursued more in the spirit of experimental science, and that you should be willing to adopt new principles. I believe that Euclid's statement that an axiom is a self-evident truth is a big mistake[3]. The Schrödinger equation certainly isn't a self-evident

[3]There is no evidence that Euclid ever made such a statement. However, the statement does have an undeniable emotional appeal.

> truth! And the Riemann hypothesis isn't self-evident either, but it's very useful. A physicist would say that there is ample experimental evidence for the Riemann hypothesis and would go ahead and take it as a working assumption. ([?] p. 24)

In this case, we have ample experimental evidence for the truth of our identity and we may want to take it as something more than just a working assumption. We may want to introduce it formally into our mathematical system.

4 Experiment and 'Theory'

We have now examined two views of experimental mathematics but we appear to be no closer to a definition than when we began. However, we are now ready to begin in full our exploration of experiment. In *Advice to a Young Scientist*, P.B. Medawar defines four different kinds of experiment: the Kantian, Baconian, Aristotelian, and the Galilean. Mathematics has always participated deeply in the first three categories but has somehow managed to avoid employing the Galilean model. In developing our notion of experimental mathematics we will try to adhere to this Galilean mode as much as possible.

We will begin with the Kantian experiments. Medawar gives as his example:

> generating 'the classical non-Euclidean geometries (hyperbolic, elliptic) by replacing Euclid's axiom of parallels (or something equivalent to it) with alternative forms.' ([?] pp. 73-74)

It seems clear that mathematicians will have difficulty escaping from the Kantian fold. Even a Platonist must concede that mathematics is only accessible through the human mind and thus at a basic level all mathematics might be considered a Kantian experiment. We can debate whether Euclidean geometry is but an idealization of the geometry of nature (where a point has no length or breadth and a line has length but no breadth?) or nature an imperfect reflection of 'pure' geometrical objects, but in either case the objects of interest lie within the minds eye.

Similarly, we cannot escape the Baconian experiment. We Medawar's words this

> is a contrived as opposed to a natural happening, it "is the consequence of 'trying things out' or even of merely messing about." ([?] p. 69)

Most of the research described as experimental is Baconian in nature and in fact one can argue that all of mathematics proceeds out of Baconian experiments. One tries out a transformation here, an identity there, examines what happens when one weakens this condition or strengthens that one. Even the application of probabilistic arguments in number theory can be seen as a Baconian experiment. The experiments may be well thought out and very likely to succeed but the ultimate criteria of inclusion of the result in the literature is success or failure. If the 'messing about' works (e.g., the theorem is proved, the counterexample found) the material is kept; otherwise, it is relegated to the scrap heap.

The Aristotelian experiments are described as demonstrations:

> apply electrodes to a frog's sciatic nerve, and lo, the leg kicks; always precede the presentation of the dog's dinner with the ringing of a bell, and lo, the bell alone will soon make the dog dribble. ([**?**] p. 71)

The results are tailored to demonstrate the theorems, as opposed to the experiments being used to devise and revise the theorems. This may seem to have little to do with mathematics but it has everything to do with pedagogy. The Aristotelian experiment is equivalent to the concrete examples we employ to help explain our definitions, theorems, or the problems assigned to students so they can see how their newly learned tools will work.

The last and most important is the Galilean experiment:

> (the) Galilean Experiment is a critical experiment – one that discriminates between possibilities and, in doing so, either gives us confidence in the view we are taking or makes us think it in need of correction. ([**?**] p.)

Ideally one devises an experiment to distinguish between two or more competing hypotheses. In subjects like medicine the questions are in principal more clear cut (the Will Roger's phenomenon or Simpson's paradox complicates matters). Does this medicine work (longevity, quality of life, cost effectiveness,etc.)? Is this treatment better than that one? Unfortunately, these questions are extremely difficult to answer and the model Medawar presents here does not correspond with the current view of experimentation. Since the spectacular 'failure' (i.e., it worked beautifully but ultimately was supplanted see [**?**]) of Newtonian physics it has been widely held that no amount of experimental evidence can prove or disprove a theorem about the world around us and it is widely known that in the real world the models one tests are not true. Medawar acknowledges the difficulty of proving a result but has more confidence than modern philosophers in disproving hypotheses. If experiment cannot distinguish between hypotheses or prove theorems, what can it do? What advantages does it have? Is it necessary?

5 'Theoretical' Experimentation

While there is an ongoing crisis in mathematics, it is not as severe as the crisis in physics. The untestability of parts of theoretical physics (e.g., string theory) has led to a greater reliance on mathematics for 'experimental verification'. This may be in part what led Arthur Jaffe and Frank Quinn to advocate what they have named 'Theoretical Mathematics' (note that many mathematicians think they have been doing theoretical mathematics for years) but which we like to think of as 'theoretical experimentation'. There are certainly some differences between our ideas and theirs but we believe they are more of emphasis than substance.

Unlike our initial experiment where we are working with and manipulating floating point numbers, 'theoretical experimentation' would deal directly with theorems, conjectures, the consequences of introducing new axioms.... Note that by placing it in the realm of experimentation, we shift the focus from the more general realm of mathematics, which concerns itself with the transmission of both truth and insight, to the realm of experimentation, which primarily deals with the establishment of and transmission of insight. Although it was originally conceived outside the experimental framework, the central problems Jaffe and Quinn need to deal with are the same. They must attempt to preserve the rigorous core of mathematics, while contributing to an increased understanding of mathematics both formally and intuitively.

As described in Arthur Jaffe and Frank Quinn's *"Theoretical Mathematics": Toward a Cultural Synthesis of Mathematics and Theoretical Physics* it appears to be mainly a call for a loosening of the bonds of rigor. They suggest the creation of a branch of theoretical (experimental) mathematics akin to theoretical physics, where one produces speculative and intuitive works that will later be made reliable through proof. They are concerned about the slow pace of mathematical developments when all the work must be rigorously developed prior to publication. They argue convincingly that a haphazard introduction of conjectorial mathematics will almost undoubtedly result in chaos.

Their solution to the problems involved in the creation of theoretical (experimental) mathematics comes in two parts. They suggest that

> theoretical work should be explicitly acknowledged as theoretical and incomplete; in particular, a major share of credit for the final result must be reserved for the rigorous work that validates it. (**[?]** p.10?)

This is meant to ensure that there are incentives for following up and proving the conjectured results.

To guarantee that work in this theoretical mode does not affect the reliability of mathematics in general, they propose a linguistic shift.

> Within a paper, standard nomenclature should prevail: in theoretical material, a word like "conjecture" should replace "theorem"; a word like "predict" should replace "show" or "construct"; and expressions such as "motivation" or "supporting argument" should replace "proof." Ideally the title and abstract should contain a word like "theoretical", "speculative", or "conjectural". (**[?]** p.10)

Still, none of the newly suggested nomenclature would be entirely out of place in a current research paper. Speculative comments have always had and will always have a place in mathematics.

This is clearly an exploratory form of mathematics. But is it truly experimental in any but the Baconian sense? The answer will of course lie in its application. If we accept the description at face value, all we have is a lessening of rigor, covered by the introduction of a new linguistic structure. More 'mathematics' will be produced but it is not clear that this math will be worth more, or even as much as, the math that would have been done without it.

It is not enough to say that mathematical rigor is strangling mathematical productivity. One needs to argue that by relaxing the strictures temporarily one can achieve more. If we view theoretical (experimental) mathematics as a form of Galilean experimentation then in its idealized form 'theoretical' (experimental) mathematics should choose between directions (hypotheses) in mathematics. Like any experimental result the answers will not be conclusive, but they will need to be strong enough to be worth acting on.

Writing in this mode, a good theoretical paper should do more than just sketch arguments and motivations. Such a paper should be an extension of the survey paper, defining not what has been done in the field but what the author feels can be done, should be done and might be done, as well as documenting what is known, where the bottlenecks are, etc. In general, we sympathize with the desire to create a 'theoretical' mathematics but without a formal structure and methodology it seems unlikely to have the focus required to succeed as a separate field.

One final comment seems in order here. 'Theoretical' mathematics, as practiced today, seems a vital and growing instititution. Mathematicians now routinely include conjectures and insights with their work (a trend that seems to be growing). This has expanded in haphazard fashion to include algorithms, suggested algorithms and even pseudo algorithms. We would distinguish our vision of 'experimental' mathematics from 'theoretical' mathematics by an emphasis on the constructive/algorithmic side of mathematics. There are well established ways of dealing with conjectures but the rules for algorithms are less well defined. Unlike most conjectures, algorithms if sufficiently efficacious soon find their way into general use.

While there has been much discussion of setting up standardized data bases to run algorithms on, this has proceeded even more haphazardly. Addressing these issues of reliability would be part of the purview of experimental mathematics. Not only would one get a critical evaluation of these algorithms but by reducing the problems to their algorithmic core, one may facilitate the sharing of insights both within and between disciplines. At its most extreme, a researcher from one discipline may not need to understand anything more than the outline of the algorithm to make important connections between fields.

6 A mathematical experiment

6.1 Experimentation

We now turn to a more concrete example of a mathematical experiment. Our meta-goal in devising this experiment was to investigate the similarities and differences between experiments in mathematics and in the natural sciences, particularly in physics. We therefore resolved to examine a conjecture which could be approached by collecting and investigating a huge amount of data: the conjecture that every non-rational algebraic number is normal in every base (see box). It is important to understand that we did not aim to prove or disprove this conjecture; our aim was to find evidence pointing in one or the other direction. We were hoping to gain insight into the nature of the problem from an experimental perspective.

Background

Definition. A real number is normal to the base 10 if every block of digits of length k occurs with frequency $1/10^k$.

Example: the Champernowne number

$$.0123456789101112131 4\ldots 99100101\ldots$$

is known to be normal base 10.
Except for artificially created examples no numbers have been proven normal in any particular base. If we allow artificial numbers there are no explicit numbers known to be normal in every base[a]

Questions

- Are all non-rational algebraic numbers normal base 10?
- Do all non-rational algebraic numbers have uniformly distributed digits?

[a] J. G. Chaitin in *Randomness and Complexity in Pure Mathematics* has a number he calls $\Omega := \sum_{p \text{ halts}} 2^{-|p|}$, the halting probability, which he notes is "sort of a mathematical pun", but is normal to all bases. He does this by identifying integers with binary strings representing Turing machines and summing over the programs that stopped (non-trivially, see [?] p.12).

The actual experiment consisted of computing to 10,000 decimal digits the square roots and cube roots of the positive integers smaller than 1000 and then subjecting these data to certain statistical tests (again, see box). Under the hypothesis that the digits of these numbers are uniformly distributed (a much weaker hypothesis than normality of these numbers), we expected the probability values of the statistics to be distributed uniformly between 0 and 1. Our first run showed fairly conclusively that the digits were distributed uniformly. In fact, the Anderson-Darling test, which we used to measure how uniformly distributed our probabilities were suggested that the probabilities might have been 'too uniform' to be random. We therefore ran the same tests again, only this time for the first 20,000 decimal digits, hoping to detect some non-randomness in the data. The data were not as interesting on the second run.

6.2 Verification

It is even more important in mathematics than in the physical sciences that the data under investigation are completely reliable. At first glance it may seem that the increasing reliance of mathematicians on programs such as Maple and Mathematica has decreased the need for verification. Computers very rarely make arbitrary mistakes in arithmetic and algebra. But all the systems have known and unknown bug in there programming. It is therefore imperative that we that we check our results. So what efforts did we take to verify our findings?

First of all, we had to make sure that the roots we computed were accurate to at least 10,000 (resp. 20,000) digits. We computed these roots using Maple as well as Mathematica, having them compute the roots to an accuracy of 10,010 digits. We then did two checks on the computed approximation s_n to $\sqrt{n}$. First, we tested that $\sqrt{n} \in [s_n - 10^{-10005}, s_n + 10^{-10005}]$ by checking that $(s_n - 10^{-10005})^2 < n < (s_n + 10^{-10005})^2$. Second, we tested that the 10,000th through 10,005th digits were not all zeros or nines. This ensures that we actually computed the first 10,000 digits of the decimal expansion of $\sqrt{n}$. (We note that Maple initially did not give us an accuracy of 10,000 digits for all of the cube roots, so that we had to increase the precision here.)

We then had to make sure that we computed the statistics and probability values accurately—or at least to a reasonable precision, since we used asymptotic formulas anyway. We did this by implementing them both in Maple and in Mathematica and comparing the results. We detected no significant discrepancy.

We claim that these measures reasonably ensure the reliability of our experimental results.

6.3 Interpretation

Data and Statistics

We looked at the first 10,000 digits after the decimal point of the $\sqrt{n}$ where $n < 1000$ is not a perfect square and of $\sqrt[3]{n}$ where $n < 1000$ is not a perfect cube.

Tests used

- χ^2 — to check that each digit occurs 1/10 of the time (discrete uniform distribution base 10).
- Discrete Cramér-von Mises — to check that all groups of 4 consecutive digits occurs 1/10,000 of the time (discrete uniform distribution base 10,000).
- Anderson-Stephens — to check that the power spectrum of the sequence matches that of white noise (periodicity).
- Anderson-Darling — continuous uniform distribution.

Important point. In order for us to claim we have generated any evidence at all either for or against we have made two fairly strong assumptions.

- The first 10,000 digits are representative of the remaining digits.
- These digits behave as far as our statistical tests go like independent random variables.

In fact, for the first and second 10,000 digits our final conclusions are identical. The second assumption is problematic. Since we have beautiful algorithms to calculate these numbers. By most reasonable definitions of independent and random, these digits are neither.

Our experimental results support the conjecture that every non-rational algebraic number is normal; more precisely, we have found no evidence against this conjecture. In this section we will describe how we looked at and interpreted the experimental data to arrive at this conclusion. We include only a few examples of how we looked at the data here. In fact, we have only looked at certain aspects of normality and randomness in decimal expansions. Thus our results may be interpreted more narrowly to support the hypothesis that algebraic numbers our normal base 10. A full description will be found in [?].

Our main goal here is to give a quick visual summary that is at once convincing and data rich. These employ some of the most basic tools of visual data analysis and should probably become form part of the basic vocabulary of an experimental mathematician. Note that traditionally one would run a test such as the Anderson-Darling test (which we have done) for the continuous uniform distribution and associate a particular probability with each of our sets of probability, but unless the probability values are extremely high or low it is difficult to interpret these statistics.

Experimentally, we want to test graphically the hypothesis of normality and randomness (or non-periodicity) for our numbers. Because the statistics themselves do not fall into the nicest of distributions, we have chosen to plot only the associated probabilities. We include two different types of graphs here. A quantile-quantile plot is used to examine the distribution of our data and scatter plots are used to check for correlations between statistics.

The first is a quantile-quantile plot of the chi square base 10 probability values versus a a discrete uniform distribution. For this graph we have placed the probabilities obtained from our square roots and plotted them against a perfectly uniform distribution. Finding nothing here is equivalent to seeing that the graph is a straight line with slope 1. This is a crude but effective way of seeing the data. The disadvantage is that the data are really plotted along a one dimensional curve and as such it may be impossible to see more subtle patterns.

The other graphs are examples of scatter plots. The first scatter plot shows that nothing interesting is occurring. We are again looking at probability values this time derived from the discrete Cramer-von Mises (CVM) test base 10,000. For each cube root we have plotted the point (f_i, s_i), where f_i is the CVM base 10,000 probability associated with the first 2500 digits of the cube root of i and s_i is the probability associated with the next 2500 digits. A look at the graph reveals that we have now plotted our data on a two dimensional surface and there is a lot more 'structure' to be seen. Still, it is not hard to convince oneself that there is little or no relationship between the probabilities of the first 2500 digits and the second 2500 digits.

The last graph is similar to the second. Here we have plotted the probabilities associated with the Anderson-Stephens statistic of the first 10,000 digits versus the first 20,000 digits. We expect to find a correlation between these tests since there is a 10,000 digit overlap. In fact, although the effect is slight, one can definitely see the thinning out of points from the upper left hand corner and lower right hand corner.

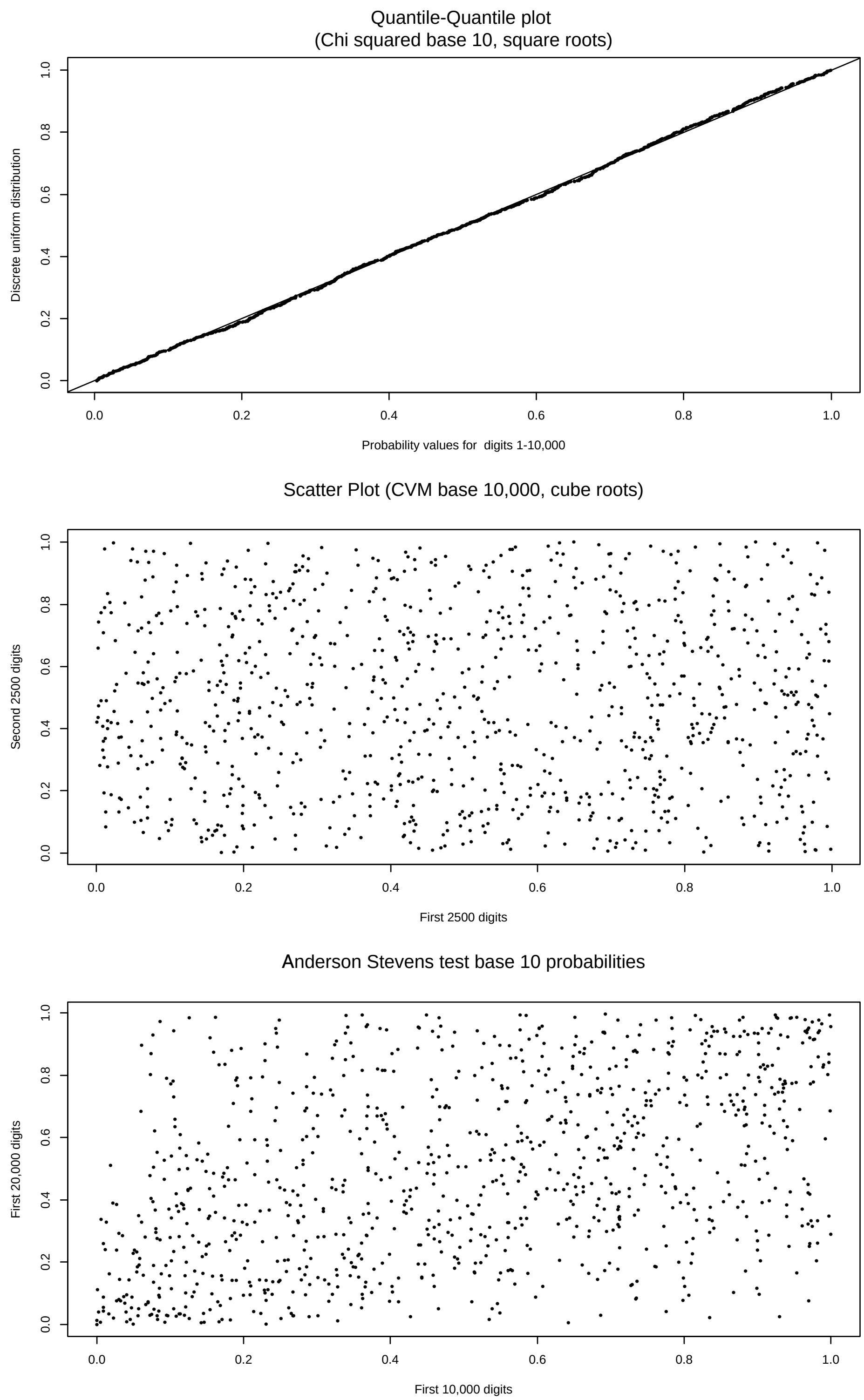

Figure 1: Graphs 1-3

7 Conclusion

All the versions of experimental mathematics that we have dealt with so far have two characteristics: their main interest is in expanding our mathematical knowledge as rapidly as possible and none of them stray too far from the mainstream. In many cases this urgency leads to a temporary relaxation of rigor, a relaxation that is well documented and hopefully can be cleaned up afterwords. In other cases it may be intrinsic to the mathematics they wish to explore. When a field has been as wildly successful as mathematics has been in the past few centuries there is a reluctance to change. We have hoped to convince some of the readers that these changes are revolutionary only in the same sense that the earth revolves around the sun.

We conclude with a definition of experimental mathematics.

Experimental Mathematics *is that branch of mathematics that concerns itself ultimately with codification and transmission of insights within the mathematical community through the use of experimental exploration of conjectures and more informal beliefs and a careful analysis of the data acquired in this pursuit.*

Results discovered experimentally will, in general, lack some of the rigor associated with mathematics but will provide general insights into mathematical problems to guide further exploration, either experimental or traditional. We have restricted our definition of experimental mathematics to methodological pursuits that in some way mimic Medawar's views of Gallilean experimentation. However, our emphasis on insight also calls for the judicious use of examples (Aristotelian experimentation).

If the mathematical community as a whole, was less splintered, we would probably remove the word 'codification' from the definition. Since there are real communications problems between fields and since the questions to be explored will be difficult, it seems imperative that experimental investigators make every effort to organize their insights and present there data in a manner that will be as widely accessible as possible[4].

With respect to reliability and rigor, the main tools here are already in place. We need to stress systematization of our exploration. As in our experimental project on normality, it is important to clearly define what has been looked at, how things have been examined, and what confidence the reader should have in the data. Although mathematicians may not like to admit it, ease of use will have to be a primary consideration if experimental results are to be of widescale use. As such, visualization and hypertextual presentations of material will become increasingly important in the future. We began by stealing some of the basic tools of scientific analysis and laying claim to them. As the needs of

[4]It is clear that mechanisms are developing for transmitting insights within fields, even if this is only through personal communications.

the community become more apparent one would expect these tools and others to evolve into a form better suited to the particular needs of the mathematical community. Someday, who knows, first year graduate students may be signing up for Experimental Methods in Mathematics I.

References

[1] Atiyah, M., Borel, A., et al. (1994). Responses to "Theoretical Mathematics: Towards a Cultural Synthesis of Mathematics and Theoretical Physics" by A. Jaffe and F. Quinn. *Bulletin (New Series) of the American Mathematical Society* **30** No. 2, 178-207.

[2] Bailey, D. H., Borwein, J. M. , Girgensohn, R. (1994). Experimental Evaluation of Euler Sums. *Experimental Math.* **3, No. 1**, 17-30.

[3] Borwein, J. M., Borwein, P., Girgensohn, R., Parnes, S. (in preparation). Experimental Mathematical Investigation of Decimal and Continued Fraction Expansions of Select Constants.

[4] Calude, C., Jürgensen, H., Zimand, M. (1994). Is Independence and Exception? *Applied Mathematics and Computation* **66**, 63-76.

[5] Chaitin, G. J. (1994). Randomness & Complexity in Pure Mathematics. *International Journal of Bifurcation and Chaos* **4**, 3-15.

[6] Cleveland, W. S. (1993). *Visualizing Data.* USA: Hobart Press.

[7] D'Agostino, R. B., Stephens, M. A. (Ed.) (1986). *Goodness-of-Fit Techniques.* USA: Marcel Dekker.

[8] Epstein, D., Levy, S., Llave, de la, R. (1992). About This Journal. *Experimental Mathematics* **1, No. 1**.

[9] Jaffe, A., Quinn, F. (1993) Theoretical Mathematics: Towards a Cultural Synthesis of Mathematics and Theoretical Physics. *Bulletin (New Series) of the American Mathematical Society*

[10] Jaffe, A., Quinn, F. (1994) Response to comments on "Theoretical Mathematics". *Bulletin (New Series) of the American Mathematical Society* **30** No. 2, 208-211.

[11] Lakatos, I. (1979). *Proofs and Refutations.* Great Britain. Cambridge University Press.

[12] Lakatos, I. (1978). *The Methodology of Scientific Research Programmes: Philosophical Papers Volume 1.* Great Britain: Cambridge University Press.

[13] Medawar, P.B. (1981). *Advice to a Young Scientist.* New York: Harper Colophon.

[14] Thurston, W. P. (1994). On Proof and Progress in Mathematics. *Bulletin (New Series) of the American Mathematical Society* **30** No. 2, 161-177.

[15] Zeilberger, D. (199?). Identities in Search of Identities. Preprint.

[16] Zeilberger, D. (1993). Theorems for a Price: Tomorrow's Semi-Rigorous Mathematical Culture. *Notices of the American Mathematical Society* **40, No. 8**, 978-981.