Chapter 1
Non-Deductive Logic in Mathematics: The Probability of Conjectures

James Franklin

1.1 Introduction

Mathematicians often speak of conjectures as being confirmed by evidence that falls short of proof. For their own conjectures, evidence justifies further work in looking for a proof. Those conjectures of mathematics that have long resisted proof, as Fermat's Last Theorem did and the Riemann Hypothesis still does, have had to be considered in terms of the evidence for and against them. It is not adequate to describe the relation of evidence to hypothesis as "subjective", "heuristic" or "pragmatic" there must be an element of what it is objectively rational to believe on the evidence, that is, of non-deductive logic. Mathematics is therefore (among other things) an experimental science.

The occurrence of non-deductive logic, or logical probability, or the rational support for unproved conjectures, in mathematics is however an embarrassment. It is embarrassing to mathematicians, used to regarding deductive logic as the only real logic. It is embarrassing for those statisticians who wish to see probability as solely about random processes or relative frequencies: surely there is nothing probabilistic about the truths of mathematics? It is a problem for philosophers who believe that induction is justified not by logic but by natural laws or the "uniformity of nature": mathematics is the same no matter how lawless nature may be. It does not fit well with most philosophies of mathematics. It is awkward even for proponents of non-deductive logic. If non-deductive logic deals with logical relations weaker than entailment, how can such relations hold between the necessary truths of mathematics?

Work on this topic was therefore rare in the mid-twentieth century "classical" period in the philosophy of science and mathematics. The recent turning of attention in philosophy of mathematics towards mathematical practice has produced a number

James Franklin
University of New South Wales, Australia
e-mail: j.franklin@unsw.edu.au

of examinations of experimental mathematics (Franklin, 1987; Fallis, 1997; Brown, 1999, ch. 10; Fallis, 2000; Corfield, 2003, ch. 5; Lehrer Dive, 2003; Van Kerkhove and Van Bendegem, 2008; Baker, 2009; Dove, 2009; brief earlier remarks in Kolata, 1976) but these have mostly not discussed in depth the theoretical issues raised. Many of these works were inspired by one important earlier contribution, the pair of books by the mathematician George Pólya on *Mathematics and Plausible Reasoning* (1954). Despite their excellence, these books of Pólya's had been little noticed by mathematicians, and even less by philosophers. Undoubtedly that is largely because of Pólya's unfortunate choice of the word "plausible" in his title—"plausible" has a subjective, psychological ring to it, so that the word is almost equivalent to "convincing" or "rhetorically persuasive". Arguments that happen to persuade, for psychological reasons, are rightly regarded as of little interest in mathematics and philosophy. Pólya made it clear, however, that he was not concerned with subjective impressions, but with what degree of belief was *justified* by the evidence (Pólya, 1954, I, 68).

Non-deductive logic deals with the support, short of entailment, that some propositions give to others. If a proposition has already been proved true, there is of course no longer any need to consider non-conclusive evidence for it. Consequently, non-deductive logic will be found in mathematics in those areas where mathematicians consider propositions which are not yet proved. These are of two kinds. First there are those that any working mathematician deals with in his preliminary work before finding the proofs he hopes to publish, or indeed before finding the theorems he hopes to prove. The second kind are the long-standing conjectures which have been written about by many mathematicians but which have resisted proof.

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? Which might follow from someone else's theorem? Which are unlikely to yield an answer until after the next review of tenure? The mathematician must answer these questions to allocate his time and effort. But not all answers to these questions are equally good. To stay employed as a mathematician, he must answer a proportion of them well. But to 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. That is to accept a role for non-deductive logic.

The area where a mathematician must make the finest discriminations of this kind—and where he might, in theory, be guilty of professional negligence if he makes the wrong decisions—is as a supervisor advising a prospective Ph.D. student. It is usual for a student beginning a Ph.D. to choose some general field of mathematics and then to approach an expert in the field as a supervisor. The supervisor then selects a problem in that field for the student to investigate. In mathematics, more

than in any other discipline, the initial choice of problem is the crucial event in the Ph.D.-gathering process. The problem must be

1. unsolved at present
2. not being worked on by someone who is likely to solve it soon

but most importantly

3. tractable, that is, probably solvable, or at least partially solvable, by three years' work at the Ph.D. level

It is recognised that of the enormous number of unsolved problems that have been or could be thought of, the tractable ones form a small proportion, and that it is difficult to discern which they are. The skill in non-deductive logic required of a supervisor is high. Hence the advice to Ph.D. students not to worry too much about what field or problem to choose, but to concentrate on finding a good supervisor.

It is also clear why it is hard to find Ph.D. problems that are also

4. interesting

It is not possible to dismiss these non-deductive techniques as simply "heuristic" or "pragmatic" or "subjective". Although those are correct descriptions as far as they go, they give no insight into the crucial differences among techniques, namely, that some are more reasonable and consistently more successful than others. "Successful" can mean "lucky", but "consistently successful" cannot. "If you have a lot of lucky breaks, it isn't just an accident", as Groucho Marx said (Chandler, 1999, 560). Many techniques can be heuristic, in the sense of leading to the discovery of a true result, but we are especially interested in those which give reason to believe the truth has been arrived at, and justify further research. Allocation of effort on attempted proofs may be guided by many factors, which can hence be called "pragmatic", but those which are likely to lead to a completed proof need to be distinguished from those, such as sheer stubbornness, which are not. Opinions on which approaches are likely to be fruitful in solving some problem may differ, and hence be called "subjective", but the beginning graduate student is not advised to pit his subjective opinion against the experts' without good reason. Damon Runyon's observation on horse-racing applies equally to courses of study: "The race is not always to the swift, nor the battle to the strong, but that's the way to bet" (Fadiman, 1955, 794). An example where the experts agreed on their opinion and were eventually proved right is the classification of finite simple groups, described in section 4 below.

It is true that similar remarks could be made about any attempt to see rational principles at work in the evaluation of hypotheses, not just those in mathematical research. In scientific investigations, various inductive principles obviously produce results, and are not simply dismissed as pragmatic, heuristic or subjective. Yet it is common to suppose that they are not principles of *logic*, but work because of natural laws (or the principle of causality, or the regularity of nature). This option is not available in the mathematical case. Mathematics is true in all worlds, chaotic or regular. Any principles governing the relationship between hypothesis and evidence in mathematics can only be logical.

1.2 Evidence for (and against) the Riemann Hypothesis

In modern mathematics, it is usual to cover up the processes leading to the construction of a proof, when publishing it—naturally enough, since once a result is proved, any non-conclusive evidence that existed before the proof is no longer of interest. That was not always the case. Euler, in the eighteenth century, regularly published conjectures which he could not prove, with his evidence for them. He used, for example, some daring and obviously far from rigorous methods to conclude that the infinite sum

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \ldots \tag{1.1}$$

(where the numbers on the bottom of the fractions are the successive squares of whole numbers) is equal to the *prima facie* unlikely value $\frac{\pi^2}{6}$. Finding that the two expressions agreed to seven decimal places, and that a similar method of argument led to the already proved result

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \ldots = \frac{\pi}{4} \tag{1.2}$$

Euler concluded, "For our method, which may appear to some as not reliable enough, a great confirmation comes here to light. Therefore, we shall not doubt at all of the other things which are derived by the same method" (Pólya, 1954, I, 18–21). He later proved the result. A translation of another of Euler's publications devoted to presenting "such evidence . . . as might be regarded as almost equivalent to a rigorous demonstration" of a proposition is given as a chapter in Pólya's books (Pólya, 1954, I, 91–98).

Even today, mathematicians occasionally mention in print the evidence that led to a theorem. Since the introduction of computers, and even more since the recent use of symbolic manipulation software packages, it has become possible to collect large amounts of evidence for certain kinds of conjectures. (Many examples in Borwein and Bailey, 2004; Borwein et al., 2004; Müller and Neunhöffer, 1987; some comments on experimental mathematics of this kind in Epstein et al., 1992; philosophical examination in Baker, 2008). A few mathematicians argue that in some cases, it is not worth the excessive cost of achieving certainty by proof when "semi-rigorous" checking will do (Zeilberger, 1993).

At present, it is usual to delay publication until proofs have been found. This rule is broken only in work on those long-standing conjectures of mathematics which are believed to be true but have so far resisted proof. The most notable of these, which stands since the proof of Fermat's Last Theorem as the Everest of mathematics, is the Riemann Hypothesis.

Riemann stated in a celebrated paper of 1859 (Riemann, 1974) that he thought it "very likely" that

> All the roots of the Riemann zeta function (with certain trivial exceptions) have real part equal to $\frac{1}{2}$.

This is the still unproved Riemann Hypothesis. The Riemann zeta function is defined on positive whole numbers $s > 1$ by the formula

$$\zeta(s) = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \ldots \qquad (1.3)$$

(Thus for example $\zeta(2) = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \ldots$, which is $\frac{\pi^2}{6}$, as mentioned above.) The definition can be extended to the entire complex plane: $\zeta(s)$ is the unique complex function, analytic except at $s = 1$, which agrees with the above formula on the positive integers greater than 1. It is found that $\zeta(s)$ has obvious ("trivial") zeros at the negative even integers. The Riemann hypothesis is that all the (infinitely many) other zeros have real part equal to $\frac{1}{2}$. For the present purpose an understanding of complex functions is not necessary: it is only important that this is a simple universal proposition like "all ravens except Texan ones are black". It is also true that the infinitely many non-trivial roots of the Riemann zeta function have a natural order, so that one can speak of "the first million roots". (Accounts in Edwards, 1974; Derbyshire, 2003, ch. 5; Sabbagh, 2002; du Sautoy, 2003.)

Once it became clear that the Riemann Hypothesis would be very hard to prove, it was natural to look for evidence of its truth (or falsity). The simplest kind of evidence would be ordinary induction: Calculate as many of the roots as possible and see if they all have real part $\frac{1}{2}$. This is in principle straightforward (though in practice computational mathematics is difficult, since one needs to devise subtle algorithms which save as much calculation as possible, so that the results can go as far as possible). Such numerical work was begun by Riemann and was carried on later with the results in Table 1.1.

Table 1.1 Hand calculations of roots of the Riemann zeta function

Date	Worker	Number of roots found to have real part $\frac{1}{2}$
1903	Gram	15
1914	Backlund	79
1925	Hutchinson	138
1935/6	Titchmarsh	1,041

"Broadly speaking, the computations of Gram, Backlund and Hutchinson contributed substantially to the plausibility of the Riemann Hypothesis, but gave no insight into the question of why it might be true" (Edwards, 1974, 97). The next investigations were able to use electronic computers; the results are shown in Table 1.2 (Brent et al., 1982; Gourdon, 2004).

It is one of the largest inductions in the world.

Besides this simple inductive evidence, there are some other reasons for believing that Riemann's Hypothesis is true (and some reasons for doubting it). In favour, there are:

Table 1.2 Computer calculations of roots of the Riemann zeta function

Date	Worker	Number of roots found to have real part $\frac{1}{2}$
1956	Lehmer	25,000
1958	Meller	35,337
1966	Lehman	250,000
1968	Rosser, Yohe & Schoenfeld	3,500,000
1979	Brent	81,000,001
1986	Te Riele, van de Lune et al	1,500,000,001
2004	Gourdon	10^{13}

1. Hardy proved in 1914 that infinitely many roots of the Riemann zeta function have real part $\frac{1}{2}$ (Edwards, 1974, 226–9). This is quite a strong consequence of Riemann's Hypothesis, but is not sufficient to make the Hypothesis highly probable, since if the Riemann Hypothesis is false, it would not be surprising if the exceptions to it were rare.

2. Riemann himself showed that the Hypothesis implied the "prime number theorem", then unproved. This theorem was later proved independently. This is an example of the general non-deductive principle that non-trivial consequences of a proposition support it.

3. Also in 1914, Bohr and Landau proved a theorem roughly expressible as "Almost all the roots have real part very close to $\frac{1}{2}$." More exactly, "For any $\delta > 0$, all but an infinitesimal proportion of the roots have real part within δ of $\frac{1}{2}$." This result "is to this day the strongest theorem on the location of the roots which substantiates the Riemann hypothesis" (Edwards, 1974, 193).

4. Studies in number theory revealed areas in which it was natural to consider zeta functions analogous to Riemann's zeta function. In some famous and difficult work, André Weil proved that the analogue of Riemann's Hypothesis is true for these zeta functions (Weil, 1948), and his related conjectures for an even more general class of zeta functions were proved to widespread applause in the 1970s. "It seems that they provide some of the best reasons for believing that the Riemann hypothesis is true—for believing, in other words, that there is a profound and as yet uncomprehended number-theoretic phenomenon, one facet of which is that the roots ρ all lie on Re $s = \frac{1}{2}$" (Edwards, 1974, 298).

5. Finally, there is the remarkable "Denjoy's probabilistic interpretation of the Riemann hypothesis" (Edwards, 1974, 268–9). If a coin is tossed n times, then of course we expect about $\frac{1}{2}n$ heads and $\frac{1}{2}n$ tails. But we do not expect *exactly* half of each. We can ask, then, what the average deviation from equality is. The answer, as was known by the time of Bernoulli, is $\sqrt{n}$. One exact expression of this fact is:

   For any $\varepsilon > 0$, with probability one the number of heads minus the number of tails in n tosses grows less rapidly than $n^{\frac{1}{2}+\varepsilon}$

Now we form a sequence of "heads" and "tails" by the following rule: Go along the sequence of numbers and look at their prime factors. If a number has two or more prime factors equal (i.e., is divisible by a square), do nothing. If not, its prime factors must be all different; if it has an even number of prime factors, write "heads". If it has an odd number of prime factors, write "tails". Table 1.3 shows the beginning of the sequence.

Table 1.3 "Head" and "tail" sequence from the factors of integers

2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	···
2	3	2^2	5	2×3	7	2^3	3^2	2×5	11	$2^2\times3$	13	2×7	3×5	2^4	17	···
T	T		T	H	T			H	T		T	H	H		T	···

The resulting sequence is of course not "random" in the sense of "probabilistic", since it is totally determined. But it is "random" in the sense of "patternless" or "erratic"; such sequences are common in number theory, and are studied by the branch of the subject called misleadingly "probabilistic number theory" (Tenenbaum, 1995). From the analogy with coin tossing, it is likely that

> For any $\varepsilon > 0$, the number of heads minus the number of tails in the first n "tosses" in this sequence grows less rapidly than $n^{\frac{1}{2}+\varepsilon}$.

This statement is equivalent to Riemann's Hypothesis. Edwards comments, in his book on the Riemann zeta function,

> One of the things which makes the Riemann hypothesis so difficult is the fact that there is no plausibility argument, no hint of a reason, however unrigorous, why it should be true. This fact gives some importance to Denjoy's probabilistic interpretation of the Riemann hypothesis which, though it is quite absurd when considered carefully, gives a fleeting glimmer of plausibility to the Riemann hypothesis (Edwards, 1974, 268).

Not all of the probabilistic arguments bearing on the Riemann Hypothesis are in its favour. In the balance against, there are the following arguments:

1. Riemann's paper is only a summary of his researches, and he gives no reasons for his belief that the Hypothesis is "very likely". No reasons have been found in his unpublished papers. Edwards does give an account, however, of facts which Riemann knew, which would naturally have seemed to him evidence of the Hypothesis. But the facts in question are true only of the early roots; there are some exceptions among the later ones. Edwards concludes:

   > The discoveries ... completely vitiate any argument based on the Riemann-Siegel formula and suggest that, unless some basic cause is operating which has eluded mathematicians for 110 years, occasional roots ρ off the line [i.e., with real part not $\frac{1}{2}$] are altogether possible. In short, although Riemann's insight was stupendous it was not supernatural, and what seemed "probable" to him in 1859 might seem less so today (Edwards, 1974, 166).

This is an example of the non-deductive rule given by Pólya, "Our confidence in a conjecture can only diminish when a possible ground for the conjecture is exploded" (Pólya, 1954, II, 20).

2. Although the calculations by computer did not reveal any counterexamples to the Riemann Hypothesis, Lehmer's and later work did unexpectedly find values which it is natural to see as "near counterexamples" (Edwards, 1974, 175–9, further in Ivić, 2003). An extremely close one appeared near the 13,400,000th root. It is partly this that prompted the calculators to persevere in their labours, since it gave reason to believe that if there were a counterexample it would probably appear soon. So far it has not, despite the distance to which computation has proceeded, so the Riemann Hypothesis is not so undermined by this consideration as appeared at first.

3. Perhaps the most serious reason for doubting the Riemann Hypothesis comes from its close connections with the prime number theorem. The theorem states that the number of primes less than x is (for large x) approximately equal to the integral

$$\int_2^x \frac{dt}{\log t} \tag{1.4}$$

If tables are drawn up for the number of primes less than x and the values of this integral, for x as far as calculations can reach, then it is always found that the number of primes less than x is actually *less* than the integral. On this evidence, it was thought for many years that this was true for all x. Nevertheless Littlewood proved that this is false. While he did not produce an actual number for which it is false, it appears that the first such number is extremely large—well beyond the range of computer calculations. Edwards comments

> In the light of these observations, the evidence for the Riemann hypothesis provided by the computations of Rosser *et al.* ... loses all its force.

That seems too strong a conclusion, since the degree of relevance of Littlewood's discovery to the Riemann Hypothesis is far from clear. But it does give some reason to suspect that there may be a very large counterexample to the Hypothesis even though there are no small ones.

It is plain, then, that there is much more to be said about the Riemann Hypothesis than, "It is neither proved nor disproved." Without non-deductive logic, though, nothing more can be said.

1.3 Goldbach's Conjecture

The situation with Goldbach's Conjecture, possibly the easiest to state of the classic unsolved problems of mathematics, is similar. Based on a letter of 1742 from Goldbach, Euler conjectured that every even number (except 2) is the sum of two primes.

The conjecture is still neither proved nor disproved and it is believed that a proof is not close. There is a simple heuristic argument that the larger the number, the more ways it can be made up of smaller numbers, so the easier it should be to write it as the sum of two primes; but there seems to be no way of converting that into a deductive argument. Computer verification for individual numbers is possible and there is a distributed computing project that has checked the Conjecture for even numbers up to and beyond 10^{18} (Wang, 2002; discussed from the point of view of experimental methods in Echeverría, 1996 and Baker, 2007).

Various consequences of it have been proved (Pólya, 1954, II, 210), and, remarkably, connections have appeared between Goldbach's Conjecture and the Riemann Hypothesis. Hardy and Littlewood proved in 1924 that a generalisation of the Riemann Hypothesis and a certain estimate implied that *most* even integers are the sum of two primes. Vinogradov in 1937 showed that every sufficiently large odd integer is the sum of three primes, and these methods were soon adapted to show Hardy and Littlewood's result without any assumptions. In 1948 Renyi found that every even number is the sum of a prime and an "almost prime" (a number with few prime factors) (Renyi, 1962). Linnik showed in 1952 that the Riemann Hypothesis itself implied a proposition relevant to Goldbach's Conjecture (Linnik, 1952). Results on the problem are still sometimes found, but there do not seem to have been dramatic advances in the last fifty years.

1.4 The Classification of Finite Groups

A last mathematical example of the central role of non-deductive inference is provided by the classification of finite simple groups, one of the great co-operative efforts of modern pure mathematics. As a case study, it has the merit that the non-deductive character of certain aspects was admitted rather explicitly by the principals. That was so because of the size of the project. Since so many people were involved, living in different continents and working over some years, it was necessary to present partial findings in print and at conferences, with explanations as to how these bore on the overall results hoped for.

Groups are one of the basic abstract entities of mathematics, having uses in describing symmetry, in classifying the kinds of curved surfaces and in many other areas. To read the following it is only necessary to know:

1. A group consists of finitely or infinitely many members; the number of members of a finite group is called its order.
2. Any group is composed, in a certain sense, of "simple" groups. ("Simple", like "group", is a technical term; "simple" groups are not in any sense uncomplicated or easy to understand but are so-called because they are not composed of smaller groups.)

A fundamental question is then: how many different finite simple groups are there? And what is the order of each? It is these questions that were attacked by the classification of finite groups project.

The project proper covered the twenty years from 1962 to 1981 inclusive. Groups had been studied in the nineteenth and early twentieth centuries, and various finite simple groups were found. It was discovered that most of them fell into a number of infinite families. These families were quite well described by the mid-1950s, with some mopping-up operations later. There were, however, five finite simple groups left over from this classification, called the Mathieu groups after their discoverer in the 1860s. Around 1960 it was not known whether any more should be expected, or, if not, how much work it might take to prove that these were the only possible simple groups.

The field was opened up by the celebrated theorem of Feit and Thompson in 1963 ("a moment in the evolution of finite group theory analogous to the emergence of fish onto dry land" Solomon, 2001). The theorem stated:

> The order of any finite simple group is an even number.

Though the result is easy to state and understand, their proof required an entire 255-page issue of the *Pacific Journal of Mathematics*. This theorem is a consequence of the full classification result (since if one knew all the finite simple groups, one could easily check that the order of each of them was even). It thus appeared that if the full classification could be found at all it would be a vast undertaking.

The final step in the answer was announced as completed in February, 1981. The full proof is spread over some 300 to 500 journal papers, taking up somewhere between 5,000 and 10,000 pages (Gorenstein, 1982, 1; "cleaned-up" version in Gorenstein et al., 2005). Of interest is the logical situation as the proof developed, particularly the increasing confidence—justified as it happened—that the workers in the field had in the answer long before the end was reached.

It turned out that the five Mathieu groups were not the only "sporadic" groups, as groups outside the infinite families came to be called. The first new one was discovered by Zvonimir Janko in Canberra (Janko, 1966), and excitement ran high as researchers applied many methods and discovered more. The final tally of sporadic groups stands at 26. These "discoveries" had in many cases a strong non-deductive aspect, as explained by Daniel Gorenstein of Rutgers, who became the father figure of the project and leading expert on how it was progressing:

> Another aspect of sporadic group theory makes the analogy with elementary particle theory even more apt. In a number of cases (primarily but not exclusively those in which computer calculations were ultimately required) "discover" did not include the actual construction of a group—all that was established was strong evidence for the existence of a simple group G satisfying some specified set of conditions X. The operative metamathematical group principle is this: if the investigation of an arbitrary group G having property X does not lead to a contradiction but rather to a "compatible" internal subgroup structure, then there exists an actual group with property X. In all cases, the principle has been vindicated; however, the interval between discovery and construction has varied from a few months to several years (Gorenstein, 1982, 3–4).

Michael Aschbacher, another leader of the field in the 1970s, distinguished three stages for any new group: discovery, existence and uniqueness.

> I understand a sporadic group to be discovered when a sufficient amount of self-consistent information about the group is available ... Notice that under this definition the group can be discovered before it is shown to exist ... Of course the group is said to exist when there is a proof that there exists some finite simple group satisfying P ... (Aschbacher, 1980, 6–7).

Some groups attracted more suspicion than others; for example that discovered by Richard Lyons was for some time habitually denoted Ly? and spoken of in such terms as, "If this group exists, it has the following properties ..." (Tits, 1971, 204). Lyons entitled his original paper 'Evidence for the existence of a new finite simple group' (Lyons, 1972). A similar situation arose with another of the later groups, discovered by O'Nan. His paper, 'Some evidence for the existence of a new simple group', was devoted to finding "some properties of the new simple group G, whose existence is pointed at by the above theorems" (O'Nan, 1976, 422).

The rate of discovery of new sporadic groups slowed after 1970 and attention turned to the problem of showing that there were no more possible. At a conference at the University of Chicago in 1972 Gorenstein laid out a sixteen-point program for completing the classification (Gorenstein, 1979). It was thought over-optimistic at the time but immense strides were soon made by Aschbacher, Glauberman and others, more or less following Gorenstein's program.

> The turning point undoubtedly occurred at the 1976 summer conference in Duluth, Minnesota. The theorems presented there were so strong that the audience was unable to avoid the conclusion that the full classification could not be far off. From that point on, the practicing finite group theorists became increasingly convinced that the "end was near"—at first within five years, then within two years, and finally momentarily. Residual skepticism was confined largely to the general mathematical community, which quite reasonably would not accept at face value the assertion that the classification theorem was "almost proved" (Gorenstein, 1982, 5–6).

Notice that "almost proved" indeed does not mean anything in deductive logic. With hindsight, one can say that a theorem was almost proved when most of the steps in the proof were found; but before a proof is complete, there can only be good non-deductive reason to believe that a sequence of existing steps will constitute most of a future proof.

By the time of the conference at Durham, England in 1978 (described in its Proceedings as on "the classification of simple groups, a programme which is now almost complete") optimism ran even higher. At that stage existence and uniqueness had been proved for 24 of the sporadic groups, leaving two "for which considerable evidence exists" (Collins, 1980, 21). One of these was successfully dealt with in 1980 ("four years after Janko's initial evidence for such a sporadic group" Gorenstein, 1982, 110), and attention focussed on the last one, known as the "Monster" because of its immense size (order about 10^{54}).

> That the search for sporadic groups was not totally haphazard can be seen from the remarkable simultaneous realization by Fischer in West Germany and Griess in the United States in 1974 that there might be a simple group having a covering group ... (Gorenstein, 1982, 92).

Consequences of the existence of this group were then studied:

> Soon after the initial "discovery", Griess, Conway and Norton noticed that every nontrivial irreducible character of a group G of type F, has degree at least 196,883 and very likely such a group G must have a character of this exact degree. Indeed, on this assumption, Fischer, D. Livingstone and Thorne eventually computed the full character table of such a group G (Gorenstein, 1982, 126–7).

Aschbacher, lecturing at Yale in 1978, said:

> When the Monster was discovered it was observed that, if the group existed, it must contain two new sporadic groups (the groups denoted by F_3 and F_5 in Table 2) whose existence had not been suspected up to that time. That is, these groups were discovered as subgroups of the Monster. Since that time the groups F_3 and F_5 have been shown to exist. This is analogous to the situation in the physical sciences where a theory is constructed which predicts certain physical phenomena that are later verified experimentally. Such verification is usually interpreted as evidence that the theory is correct. In this case, I take the existence of F_3 and F_5 to be very good evidence that the Monster exists ... My belief is that there are at most a few groups yet to be discovered. If I were to bet, I would say no more (Aschbacher, 1980, 13–15).

Gorenstein's survey article of 1978 contains perhaps the experts' last sop to deductivism, the thesis that all logic is deductive. He wrote:

> At the present time the determination of all finite simple groups is very nearly complete. Such an assertion is obviously presumptuous, if not meaningless, since one does not speak of theorems as "almost proved" (Gorenstein, 1979, 50–51).

To the deductivist, the fact that most steps in a proposed proof are completed is no reason to believe that the rest will be. Undeterred, however, Gorenstein went on to say:

> The complete proof, when it is obtained, will run to well over 5,000 journal pages! Moreover, it is likely that at the present time more than 80% of those pages exist ... The assertion that the classification is nearly complete is really a prediction that the presently available techniques will be sufficient to deal with the problems still outstanding. In its support, we cite the fact that, with two exceptions, all open questions are open because no one has yet examined them and not because they involve some intrinsic difficulty.

A year after the Durham conference, the experts assembled again at Santa Cruz, California, in a mood of supreme confidence. Gorenstein's survey opened with the remark:

> My aim here is to present a brief outline of the classification of the finite simple groups, now rapidly nearing completion (Gorenstein, 1980, 3).

Another contributor to the conference began his talk:

> Now that the problem of classifying finite simple groups is probably close to completion ... (Hunt, 1980).

What concern remained was less about the completion of the project than about what to do next; the editor of the conference proceedings began by commenting, "In the last year or so there have been widespread rumors that group theory is finished,

that there is nothing more to be done" (Mason, 1980, xii). The *New York Times Week in Review* (June 22, 1980) headlined an article 'A School of Theorists Works Itself Out of a Job.'

The confidence proved justified. Griess was able to show the existence of the Monster, and finally, in 1981, Simon Norton of Cambridge University completed the proof of the uniqueness of the Monster (Gorenstein, 1982, 1).

At least, that was claimed at the time. In the late 1980s it was discovered that a part of the proof, on "quasithin" groups, was not quite as complete as had been thought. One gap proved hard to fill in, but was completed by Aschbacher and others in 2001 (Aschbacher, 2001).

1.5 Probabilistic relations between necessary truths?

The most natural conceptualization of the non-deductive relations between evidence and conclusion is that of objective Bayesianism. The (objective) Bayesian theory of evidence (also known as the logical theory of probability) aims to explain what the nature of evidence is. It holds that the relation of evidence to conclusion is a matter of strict logic, like the relation of axioms to theorems in mathematics but less conclusive—a kind of partial implication. Given a fixed body of evidence—say in a trial, or in a dispute about a scientific theory—and given a conclusion, there is a fixed degree to which the evidence supports the conclusion. It was defended in Keynes' *Treatise on Probability* (Keynes, 1921) and more recently by E.T. Jaynes (Jaynes, 2003; a slightly less objective version in Williamson, 2010; introductions in Franklin, 2001; Franklin, 2009, ch. 10). It says, for example, that if we could establish just what the legal standard of "proof beyond reasonable doubt" is, then, in a given trial, it is an objective matter of logical fact whether the evidence presented does or does not meet that standard, and so a jury is either right or wrong in its verdict on the evidence.

It is not essential to the Bayesian perspective that the relation of evidence to conclusion should be given a precise number, nor that it be possible to compute the logical relation between evidence and conclusion in typical cases. It is sufficient for objective Bayesianism that it is sometimes intuitively evident that some hypotheses, on some bodies of evidence, are highly likely, or almost certain, or virtually impossible (Franklin, 2011). Keynes certainly believed that it was not always possible even in principle to compute an exact number expressing the relation between an arbitrary body of evidence and a conclusion. Nevertheless, it is usual as an idealization to suppose that for any body of evidence e and any conclusion h, there is as number $P(h|e)$, between 0 and 1, expressing the degree to which e supports h; and that that number satisfies the usual axioms of conditional probability:

$P(\text{not-}h|e) = 1 - P(h|e)$
$P(h_1 \text{ and } h_2|e) = P(h_1|e) \times P(h_2|h_1 \text{ and } e)$

Pólya's qualitative principles of evidence, such as the confirmation of hypotheses by their non-trivial consequences, are then easy deductions from those axioms.

The logical nature of the relation makes it particularly suitable for application to the necessary subject matter of pure mathematics. Conversely, its intuitive agreement with actual evaluation of conjectures supports it as a possible meaningful interpretation of probability (not necessarily the only valid one, as stochastic outcomes or idealized degrees of belief or idealized relative frequencies may also turn out to satisfy the same axioms.)

There is one point that needs to be made precise especially in applying the theory of logical probability or non-deductive logic in *mathematics*. If evidence e entails hypothesis h, then $P(h|e)$ is 1. But in mathematics, the typical case is that e *does* entail h, though that is perhaps as yet unknown. If, however, $P(h|e)$ is really 1, how is it possible in the meantime to discuss the (non-deductive) support that e may give to h, that is, to treat $P(h|e)$ as not equal to 1? In other words, if h and e are necessarily true or false, how can $P(h|e)$ be other than 0 or 1? How can there be probabilistic relations between necessary truths?

The answer is that, in both deductive and non-deductive logic, there can be *many* logical relations between two propositions. Some may be known and some not. To take an artificially simple example in deductive logic, consider the argument

> If all men are mortal, then this man is mortal
> All men are mortal
> ───────────────────
> Therefore, this man is mortal

The premises entail the conclusion, certainly, but there is more to it than that. They entail the conclusion in two ways: firstly, by *modus ponens*, and secondly by instantiation from the second premise alone. That is, there are two logical paths from the premises to the conclusion.

More complicated and realistic cases are common in the mathematical literature. Feit and Thompson's proof that all finite simple groups have even order, occupying 255 pages, was simplified by Bender (1970). That means that Bender found a different and shorter logical route from the definition of "finite simple group" to the proposition, "All finite simple groups have even order" than the one known to Feit and Thompson.

Now just as there can be two deductive paths between premises and conclusion, so there can be a deductive and non-deductive path, and it may be that only the latter is known. Before the Greeks' development of deductive geometry, it was possible to argue

> All equilateral (plane) triangles so far measured have been found to be equiangular
> This triangle is equilateral
> ──────────────────────────────
> Therefore, this triangle is equiangular

There is a non-deductive logical relation between the premises and the conclusion; the premises inductively support the conclusion. But when deductive geometry appeared, it was found that there was also a deductive relation, since the second premise alone entails the conclusion. This discovery in no way vitiates the correctness of the previous non-deductive reasoning or casts doubt on the existence of the

non-deductive relation. That relation cannot be affected by discoveries about any other relation.

So the answer to the question, "How can there be probabilistic relations between necessary truths?" is simply that those relations are additional to any deductive relations (and may be known independently of them).

1.6 The problem of induction in mathematics

That non-deductive logic is used in mathematics is important first of all to mathematics. But there is also some wider significance for philosophy, in relation to the problem of induction, or inference from the observed to the unobserved.

It is common to discuss induction using only examples from the natural world, such as, "All observed flames have been hot, so the next flame observed will be hot" and "All observed ravens have been black, so all ravens are black". That has encouraged the view that the problem of induction should be solved in terms of natural laws (or causes, or dispositions, or the regularity of nature, or some other contingent principle) which provide a kind of "cement of the universe" to bind the observed to the unobserved.

The difficulty for such a view is that it does not apply to mathematics, which deals in necessary matter. Yet induction works just as well in mathematics as in natural science.

Examples were given above in the second section in connection with the calculation of roots for the Riemann Hypothesis, but let us take a particularly straightforward case:

> The first million digits of π are random
> _____
> Therefore, the second million digits of π are random

("Random" here means "without pattern", "passes statistical tests for randomness", not "probabilistically generated", "stochastic": Ruhkin, 2001; Franklin, 2009, 162–3.) The number π has the decimal expansion

$$3.14159265358979323846264338327950288419716939937\ldots$$

There is no apparent pattern in these numbers. The first million digits have long been calculated (calculations have reached beyond one trillion). Inspection of these digits reveals no pattern, and computer calculations can confirm this impression. It can then be argued inductively that the second million digits will likewise exhibit no pattern. This induction is a good one (indeed, everyone believes that the digits of π continue to be random indefinitely, though there is no proof, Marsaglia, 2005).

It is true, as argued by Baker (2007), that there is a special problem with inductive arguments in mathematics in that all the observed cases are of small numbers. Any number that can be calculated with is very small, compared to numbers in general. That bias in the evidence could raise a question as to whether any induction of

the form "All observed numbers have property X, therefore all numbers have property X" could have high probability. That does not imply, however, that inductive arguments in mathematics are generally poor. Firstly, a bias in the evidence towards small numbers does not affect inductive arguments with more modest conclusions, such as "All observed numbers have property X, so the next number calculated will have property X." (The argument above about the randomness of the digits of π only extrapolated a finite distance, keeping to small numbers.) Secondly, many other inductive arguments have a bias in the evidence, without thereby becoming worthless (though they may become less secure). For example, extrapolative inductive inference like "All observed European swans are white, therefore all swans are white" is a worthwhile inductive argument, although the extrapolation beyond the observed range weakens it.

Now there seems to be no reason to distinguish the logic involved in such mathematical arguments from that used in inductions about flames or ravens. But the digits of π are the same in all possible worlds, whatever natural laws may hold in them or fail to. Any reasoning about π is also rational or otherwise, regardless of any empirical facts about natural laws. Therefore, induction can be rational independently of whether there are natural laws (or any other such contingent principle).

This argument does not show that natural laws have no place in discussing induction. It may be that mathematical examples of induction are rational because there are *mathematical* laws or regularities, and that the aim in natural science is to find some substitute, such as natural laws, which will take the place of mathematical laws in accounting for the continuance of regularity. But if this line of reasoning is pursued, it is clear that simply making the supposition, "There are laws", is of little help in making inductive inferences. No doubt mathematics is completely lawlike, but that does not help at all in deciding whether the digits of π continue to be random. In the absence of any proofs, induction is needed to support the law (if it is a law), "The digits of π are random", rather than the law being able to give support to the induction. Either "The digits of π are random" or "The digits of π are not random" is a law, but in the absence of knowledge as to which, we are left only with the confirmation that the evidence gives to the first of these hypotheses. Thus consideration of a mathematical example reveals what can be lost sight of in the search for laws: laws or no laws, non-deductive logic is needed to make inductive inferences.

It is worth noting that there are also mathematical analogues of Goodmans "grue" paradox. Let a number be called "prue" if its decimal expansion is random for the first million digits and 6 thereafter. The predicate "prue" is like "grue" in not being projectible. "π is random for the first million digits" is logically equivalent to "π is prue for the first million digits", but this proposition supports "π is random always", not "π is prue". Any solutions to the "grue" paradox must allow projectible or "natural" properties to be found not only in nature but also in mathematics.

These examples illustrate Pólya's remark that non-deductive logic is better appreciated in mathematics than in the natural sciences (Pólya, 1954, II, 24). In mathematics there can be no confusion over natural laws, the regularity of nature, approximations, propensities, the theory-ladenness of observation, pragmatics, scientific

revolutions, the social relations of science or any other red herrings. There are only the hypothesis, the evidence and the logical relations between them.

References

Aschbacher, M. (1980). *The Finite Simple Groups and Their Classification*. Yale University Press, New Haven, CT.

Aschbacher, M. (2001). The status of the classification of the finite simple groups. *Notices of the American Mathematical Society*, 51:736–740.

Baker, A. (2007). Is there a problem of induction for mathematics? In Leng, M., Paseau, A., and Potter, M., editors, *Mathematical Knowledge*, pages 59–73. Oxford University Press, Oxford.

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

Baker, A. (2009). Non-deductive methods in mathematics. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/mathematics-nondeductive/.

Bender, H. (1970). On the uniqueness theorem. *Illinois Journal of Mathematics*, 14:376–384.

Borwein, J. M. and Bailey, D. H. (2004). *Mathematics by Experiment: Plausible Reasoning in the 21st Century*. A.K. Peters, Natick, MA.

Borwein, J. M., Bailey, D. H., and Girgensohn, R. (2004). *Experimentation in Mathematics: Computational Paths to Discovery*. A.K. Peters, Natick, MA.

Brent, R., van de Lune, J., te Riele, H., and Winter, D. (1982). On the zeros of the Riemann Zeta Function in the critical strip. II. *Mathematics of Computation*, 39:681–688.

Brown, J. R. (1999). *Philosophy of Mathematics: An Introduction to the World of Proofs and Pictures*. Routledge, London.

Chandler, C. (1999). *Hello, I Must Be Going: Groucho Marx and His Friends*. Doubleday, Garden City, NY.

Collins, M. J. (1980). *Finite Simple Groups, II*. Academic Press, London.

Corfield, D. (2003). *Towards a Philosophy of Real Mathematics*. Cambridge University Press, Cambridge.

Derbyshire, J. (2003). *Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics*. Joseph Henry Press, Washington, DC.

Dove, I. J. (2009). Towards a theory of mathematical argument. *Foundations of Science*, 14(1–2):137–152.

du Sautoy, M. (2003). *The Music of the Primes: Searching to Solve the Greatest Mystery in Mathematics*. Harper Collins, New York, NY.

Echeverría, J. (1996). Empirical methods in mathematics. In Munévar, G., editor, *Spanish Studies in the Philosophy of Science*, volume 86 of *Boston Studies in the Philosophy of Science*, pages 19–55. Kluwer, Dordrecht.

Edwards, H. M. (1974). *Riemann's Zeta Function*. Academic Press, New York.

Epstein, D., Levy, S., and de la Llave, R. (1992). About this journal. *Experimental Mathematics*, 1:1–13.

Fadiman, C. (1955). *The American Treasury*. Harper, New York, NY.

Fallis, D. (1997). The epistemic status of probabilistic proof. *Journal of Philosophy*, 94(165-186).

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

Feit, W. and Thompson, J. G. (1963). Solvability of groups of odd order. *Pacific Journal of Mathematics*, 13:775–1029.

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

Franklin, J. (2001). Resurrecting logical probability. *Erkenntnis*, 55:277–305.

Franklin, J. (2009). *What Science Knows: And How It Knows It*. Encounter Books, New York, NY.

Franklin, J. (2011). The objective Bayesian conceptualisation of proof and reference class problems. *Sydney Law Review*, 33:545–561.

Gorenstein, D. (1979). The classification of finite simple groups (I). *Bulletin of the American Mathematical Society*, New Series 1:43–199.

Gorenstein, D. (1980). An outline of the classification of finite simple groups. In Cooperstein, B. and Mason, G., editors, *The Santa Cruz Conference on Finite Groups*, volume 37 of *Proceedings of Symposia in Pure Mathematics*, pages 3–28. American Mathematical Society.

Gorenstein, D. (1982). *Finite Simple Groups*. Plenum, New York, NY.

Gorenstein, D., Lyons, R., and Solomon, R. (1994–2005). *The Classification of the Finite Simple Groups*. American Mathematical Society, Providence RI. 6 Vols.

Gourdon, X. (2004). The 10^{13} first zeros of the Riemann Zeta Function, and zeros computation at very large height. `http://numbers.computation.free.fr/Constants/Miscellaneous/zetazeros1e13-1e24.pdf`.

Hunt, D. (1980). A computer-based atlas of finite simple groups. In Cooperstein, B. and Mason, G., editors, *The Santa Cruz Conference on Finite Groups*, volume 37 of *Proceedings of Symposia in Pure Mathematics*, pages 507–510. American Mathematical Society.

Ivić, A. (2003). On some reasons for doubting the Riemann Hypothesis. `http://arxiv.org/abs/math/0311162`.

Janko, Z. (1966). A new finite simple group with abelian 2-Sylow subgroups and its characterization. *Journal of Algebra*, 3:147–186.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge.

Keynes, J. M. (1921). *A Treatise on Probability*. Macmillan, London.

Kolata, G. B. (1976). Mathematical proofs: The genesis of reasonable doubt. *Science*, 192:989–990.

Lehrer Dive, L. (2003). *An Epistemic Structuralist Account of Mathematical Knowledge*. PhD thesis, University of Sydney.

Linnik, Y. V. (1952). Some conditional theorems concerning the binary Goldbach problem. *Izvestiya Akademii Nauk SSSR*, 16:503–520.

Lyons, R. (1972). Evidence for a new finite simple group. *Journal of Algebra*, 20:540–569.

Marsaglia, G. (2005). On the randomness of pi and other decimal expansions. `http://www.yaroslavvb.com/papers/marsaglia-on.pdf`.

Mason, G. (1980). Preface. In Cooperstein, B. and Mason, G., editors, *The Santa Cruz Conference on Finite Groups*, volume 37 of *Proceedings of Symposia in Pure Mathematics*, page xiii. American Mathematical Society.

Müller, J. and Neunhöffer, M. (1987). Some computations regarding Foulkes' conjecture. *Experimental Mathematics*, 14:277–283.

O'Nan, M. (1976). Some evidence for the existence of a new finite simple group. *Proceedings of the London Mathematical Society*, 32:421–479.

Pólya, G. (1954). *Mathematics and Plausible Reasoning*. Two Volumes. Princeton University Press, Princeton, N.J.

Renyi, A. (1962). On the representation of an even number as the sum of a prime and an almost prime. *American Mathematical Society Translations, 2nd series*, 19:299–321.

Riemann, B. (1859 [1974]). On the number of primes less than a given magnitude. In Edwards, H., editor, *Riemann's Zeta Function*, pages 299–305. Academic Press, New York, NY.

Ruhkin, A. (2001). Testing randomness: a suite of statistical procedures. *Theory of Probability and its Applications*, 45:111–132.

Sabbagh, K. (2002). *Dr Riemann's Zeros*. Atlantic Books, London.

Solomon, R. (2001). A brief history of the classification of the finite simple groups. *Bulletin of the American Mathematical Society*, 38:315–352.

Tenenbaum, G. (1995). *Introduction to Analytic and Probabilistic Number Theory*. Cambridge University Press, Cambridge.

Tits, J. (1971). Groupes finis simples sporadiques. In *Séminaire Bourbaki*, volume 180 of *Springer Lecture Notes in Mathematics*, pages 187–211. Springer.

Van Kerkhove, B. and Van Bendegem, J. P. (2008). Pi on earth, or mathematics in the real world. *Erkenntnis*, 68:421–435.

Wang, Y., editor (2002). *Goldbach Conjecture*. World Scientific, River Edge, NJ.

Weil, A. (1948). *Variétés abéliennes et courbes algébriques*. Hermann, Paris.

Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford University Press, Oxford.

Zeilberger, D. (1993). Theorems for a price: Tomorrow's semi-rigorous mathematical culture. *Notices of the American Mathematical Society*, 46:978–981.