Benjamin Peirce and the Howland Will

PAUL MEIER and SANDY ZABELL*

The Howland will case is possibly the earliest instance in American law of the use of probabilistic and statistical evidence. Identifying 30 downstrokes in the signature of Sylvia Ann Howland, Benjamin Peirce attempted to show that a contested signature on a will had been traced from another and genuine signature. He argued that their agreement in all 30 downstrokes was improbable in the extreme under a binomial model. Peirce supported his model by providing a graphical test of goodness of fit. We give a critique of Peirce's model and discuss the use and abuse of the "product rule" for multiplying probabilities of independent events.

KEY WORDS: Charles S. Peirce; 19th-century mathematical statistics; Law and statistics; Handwriting analysis; Independence; Product rule.

1. INTRODUCTION

On July 2, 1865, Sylvia Ann Howland of New Bedford, Massachusetts, died, leaving an estate of $2,025,000, and a single heir—her niece Hetty H. Robinson, who had lived with her for some years. Two million dollars is a respectable sum, even by today's inflated standards, and in 1865 it was certainly a prize worth contesting.

The Howland will, dated September 1, 1863, and a codicil of November 18, 1864, left half the estate to a number of individuals and institutions. It provided that the residue was to be held in trust for the benefit of Hetty Robinson and, at the time of Hetty's death, to be distributed to the lineal descendents of Hetty's paternal grandfather, Gideon Howland, Sr. Hetty's father had died a month earlier, leaving her $910,000 outright and the income from a trust of $5,000,000. Although Hetty Robinson was thus already a wealthy woman, she claimed a right to inherit outright the entire estate of her aunt, providing as evidence for this claim an earlier will, dated January 11, 1862, which left the entire estate to Hetty with instructions that no later will was to be honored.

Family feuds are relevant here, as is a claimed agreement between Hetty and her aunt to leave mutual wills excluding Hetty's father. For our purposes, it is enough to observe that Hetty Robinson had an arguable claim that the earlier will should be recognized. But this the Executor, Thomas Mandell, declined to do. Not only did he claim that the later will governed, but he also contended that two of the three signatures on the earlier will were tracings from the third, thus providing an independent ground for invalidating that will (see Appendix 1).

Hetty Robinson retained distinguished counsel and sued.[1] Thomas Mandel retained equally distinguished counsel and defended. Between them, both sides called as witnesses some of the most prominent academicians of the day. Oliver Wendell Holmes, Sr., Parkman Professor of Anatomy and Physiology in the Medical School of Harvard University, examined the contested signatures under a microscope, as did Louis Agassiz, and testified (for Robinson) that he could find no evidence of pencil marks such as might have appeared in tracing. (The cross-examination of Professor Holmes was both brief and painful; see Appendix 2.) Among those testifying for the Executor were Benjamin Peirce, Professor of Mathematics at Harvard, and his son, Charles Sanders Peirce, who was later to do important work in several branches of philosophy. The Peirces were among the first to contribute to the development of mathematical statistics in the United States; a recent survey of their work in this area is given in Stigler (1978, pp. 244–251).

Professor Peirce undertook to demonstrate by statistical means that the disputed signatures on the second page of the will were indeed forgeries. His method was to contrast the similarities between one of the disputed signatures (referred to in the trial record as signature #10) and the unquestioned original (referred to as signature #1) with the lesser degree of similarity to be found in pairing 42 other signatures penned by Sylvia Ann Howland in her later years on other documents. (See Figures A and B.) It is this analysis with which we shall be concerned.

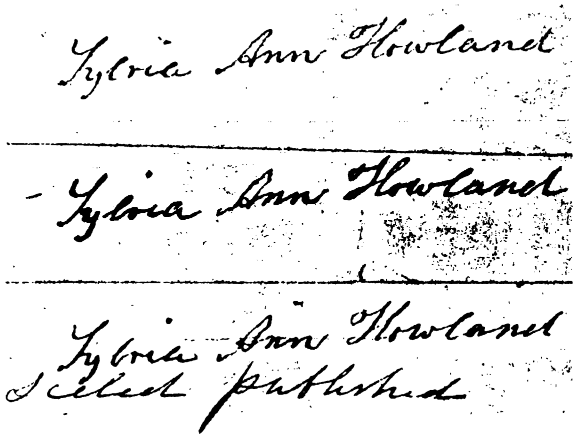
The Peirce analysis, as we shall see, is at once ingenious and in some respects naive. It leads to a distributional problem, discussed in Appendix 4, which is nontrivial and has some interest in its own right. The data available

* Paul Meier is Professor of Statistics, The University of Chicago, Chicago, IL 60637, and Sandy Zabell is Associate Professor of Mathematics, Northwestern University, Evanston, IL 60201. Support for this research was provided in part by NSF Grant SOC 76-80389. The authors are grateful to a number of friends and colleagues with whom they discussed aspects of this case—P. Diaconis, S. Fienberg, J. Langbein, P. McCullagh, J. Tukey, and H. Zeisel, among others. Special thanks are due to David Wallace for extensive discussion concerning the inference problem that arose, and Stephen Stigler for helpful correspondence on the historical background. The staff of the New Bedford Public Library were especially accommodating in locating and arranging for the copying of unique source documents and exhibits, and to them the authors express their special gratitude.

[1] December, 1865. Testimony was taken in 1867, decision rendered October 1868 (*Robinson* v. *Mandell*, 20 *Fed. Cas.* 1027). A copy of the trial transcript is available in the New Bedford Public Library and the testimony in the text and Appendixes 2 and 3 is extracted from that source.

September 1980, Volume 75, Number 371
Applications Section

A. The Unquestioned Original Signature (no. 1) and the Two Disputed Signatures (nos. 10 and 15)



in the record is sufficient to allow some interesting analyses, but tantalizing for its failure to include details of major relevance for the problem at issue. However, as we proceed, it is well to keep in mind that no organized theory of statistical inference existed at the time, and that, for example, casual imputations of independence were made almost automatically by those seeking to bring considerations of probability to bear on problems of human affairs. We should not be surprised, therefore, to find such imputations as a matter of course in the present problem.

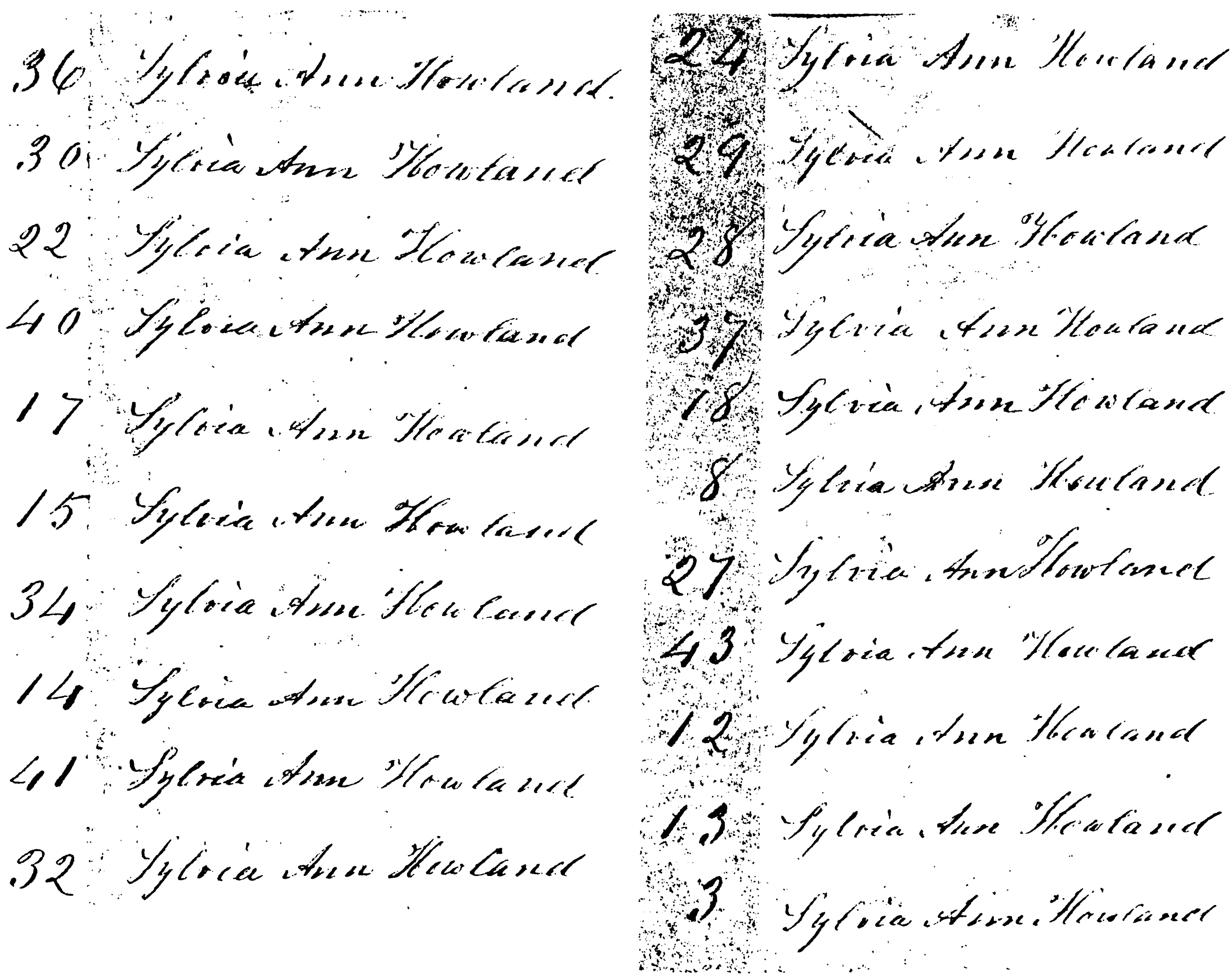
2. THE PROBLEM

Let us start with the evidence as given, making assumptions as needed, and return later to judge which seem plausible and which not. The evidence consisted of photographic copies of 42 signatures, in addition to the original (#1) and one of the purported copies (#10).

Suppose we are given some index of agreement between signatures. In the present case Professor Peirce chose the number of coincidences—in length and position—among the 30 downstrokes in each of a pair of signatures, when the two were superimposed. We might then let $X_1, \ldots, X_{42}$ be the values of that index obtained in matching each of the 42 uncontested and uninvolved signatures with #1, and let Y be the value of that index in matching #10 with #1. We could then pose the problem as one of judging whether Y could reasonably have been drawn at random from the same population that gave rise to the sample of size 42.

Since, in this case, $Y = 30$, that is, #1 and #10 coincide in all 30 downstrokes, and the X_i are presumably

B. Some of the 42 Comparison Signatures



all smaller than 30, we might assign a significance probability of $1/43 = 2.3\%$ to this outcome. This is a small enough probability to raise suspicion, but far from overwhelming.[2] (If we had available the individual X_i values (which we do not) we might be willing to proceed a little more parametrically. We can determine (see Table 1) that the average index in random pairings of the 42 signatures is near 6 with a standard deviation of about 2.5. Thus Y is nearly 10 sample standard deviations away from $\bar{X}$ and, if we believe that the population sampled is at least approximately normal, the discrepancy would be judged significant at an extreme level.)

Professor Peirce proceeded in a different way and came to the conclusion that the odds against such a coincidence as that between signatures # 1 and # 10 were astronomical (of order 10^{20}). To see how he arrived at this striking result, let us follow the order of testimony in the case.

2.1 Testimony of Charles Sanders Peirce

The son, Charles Sanders Peirce—then a member of the staff of the United States Coast Survey—testified on the work he carried out under the direction of his father. Charles was provided with photographic copies of signatures #1 and #10 (the third signature is not mentioned in the testimony) and with copies of 42 signatures of Sylvia Ann Howland on leases and other documents. Following instructions given to him by his father, Charles undertook to match every possible pair of the 42 uncontested signatures and to observe the agreement between each pair in respect to the number out of 30 identified "downstrokes" in which they were in agreement. There being $\binom{42}{2} = 861$ possible pairs, and 30 downstrokes to be checked for each, this meant a total of 25,830 comparisons, for which Charles reported agreement in 5,325 or 20.6 percent of them—very nearly 1 in 5. More particularly, he gave the distribution of number of agreements as shown in Table 1. Thus, for example, 15 pairs each had two coincidences out of 30, for a total of $2 \times 15 = 30$ strokes coinciding.

The precise details of the matching process and of the criteria for judgment of agreement were only briefly touched on in C.S. Peirce's testimony:

> *Q*. What was the standard . . . of coincidence?
> *A*. I considered that two lines coincided whenever they coincided as well as . . . nos. 1 and 10, in those same lines.

Later, under cross-examination, Charles added some further details:

> I made it a condition that in shifting one photograph over another in order to make as many lines as possible coincide,—that the line of writing should not be materially changed. By materially changed, I mean so much changed that there could be no question that there was a difference in the general direction of the two signatures.

1. C. S. Peirce's Table of Coincidences

Coincidences	*Number of Pairs*	*Number of Strokes*
0, 1	0	0
2	15	30
3	97	291
4	131	524
5	147	735
6	143	858
7	99	693
8	88	704
9	55	495
10	34	340
11	17	187
12	15	180
13 . . . 30	20	288
	861	5325

2.2 Testimony of Benjamin Peirce

C.S. Peirce was followed to the stand by his father, who testified as to his own examination and comparison of #1 and #10.

> The coincidence is extraordinary and of such a kind as irresistably to suggest design, and especially the tracing of 10 over 1. There are small differences . . . but the differences are such as to strengthen the argument for design. . . . The coincidences are manifestly those of position; while the differences are those of form. Coincidence of position is easily effected by design, and can be tolerably well accomplished by an unskilful hand. It especially belongs to the downward stroke. . . . It involves a close agreement in the average position of the stroke, an agreement in its slant, and no extravagant diversity of curvature. But coincidence of form is exceedingly difficult. . . .
>
> In a long signature, complete coincidence of position is . . . an infallible evidence of design. The mathematical discussion of this subject has never, to my knowledge, been proposed, but it is not difficult; and a numerical expression applicable to this problem, the correctness of which would be instantly recognized by all the mathematicians of the world, can be readily obtained.

Professor Peirce went on to describe his son's work and, without qualification, asserted that since the overall proportion of matches was very nearly one-fifth, the probability of finding 30 matches in a given pair of signatures was one in 5^{30} or "once in 2,666 millions of millions of millions." (Note: There is some error here, since $5^{30} = 9.31 \times 10^{20}$.) Lest the point be missed, Peirce added:

> This number far transcends human experience. So vast an improbability is practically an impossibility. Such evanescent shadows of probability cannot belong to actual life. They are unimaginably less than those least things which the law cares not for.

When asked the purpose of his son's display of the distribution of the number of coincidences, Peirce ex-

[2] One might try to improve on this a little by considering the probability that out of 44 signatures, a given pair will have the greatest index of coincidences, that is, $\binom{44}{2}^{-1} = .0011$. (Strictly speaking, we are not given the number of coincidences between either signature #1 or #10 on the one hand, and any of the 42 other signatures on the other, but presumably none of these even approaches the 30 of #1 and #10.) The appropriateness of such a significance test is an interesting issue, but one that we will not touch on.

2. *Benjamin Peirce's Table (and chi-squared computations)*

Number of Agreements	Obs.	Exp.	(O − E)	$(O-E)^2/E$
0, 1	0	9.1	−9.1	9.1
2	15	29.0	−14.0	6.7
3	97	67.6	29.4	12.8
4	131	114.1	16.9	2.5
5	147	148.3	−1.3	0.0
6	143	154.5	−11.5	0.9
7	99	132.4	−33.4	8.4
8	88	95.2	−7.2	0.5
9	55	58.2	−3.2	0.2
10	34	30.5	3.5	0.4
11	17	13.9	3.1	0.7
12	15	5.5	9.5	16.4
13 . . . 30	20	2.7	17.3	112.1
	861	861	0.0	170.7

plained that

> It was ordered by me in order to obtain as severe a test as possible from observation for the criticism of my mathematical deductions.

To facilitate comparison, Peirce provided a table, recapitulating his son's counts, together with the expected number of counts under the binomial model

$$E\{Y_j\} = 861\binom{30}{j}\left(\frac{1}{5}\right)^j\left(\frac{4}{5}\right)^{30-j}.$$

(See Table 2. The table given by Peirce omits the 15 cases with two or fewer coincidences, and the 20 with more than 12.)[3] Peirce commented:

> This is a species of comparison familiar to mathematicians and such a coincidence as above would be regarded as quite extraordinary, and a sufficient demonstration that the true law of the case was embodied in the formula.

A graphical display of the comparison was also provided, which is reconstructed in Figure C according to the description given in the trial record (the actual display was omitted from the trial record).

3. CRITIQUE OF PEIRCE'S MODEL

3.1 Goodness of Fit

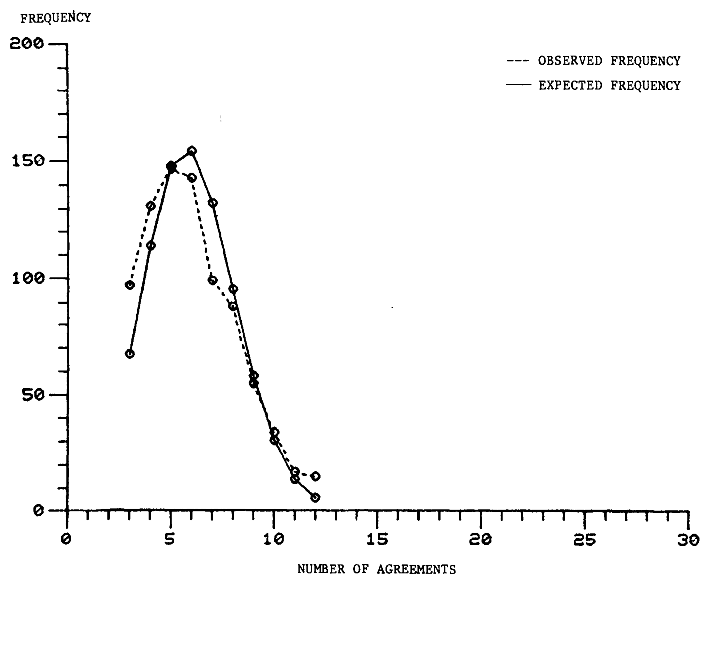
How convincing is Peirce's model and his evidence for it? No formal theory of goodness-of-fit testing existed at the time, and Peirce can scarcely be faulted for failing to provide one. Still, the quality of fit can reasonably be questioned. Peirce's graph shows only the middle range, and even here there appears to be a consistent trend in the deviations. What the graph does not show is the substantial absence of really poor matches—no cases of 0 or 1 agreement, with 9 expected—and the considerable excess of outstandingly good matches—20 (with 13 or more agreements), with less than 3 expected.

If we calculate the usual goodness-of-fit chi-squared for this frequency table we get $X^2 = 170.7$ on 12 df. We can improve this a little by choosing p to minimize chi-squared, but the minimizing $p = .211$ actually worsens things in the middle range of the graph and gives a $X^2 = 135.3$, which remains an exceedingly poor fit. (The appropriateness of the chi-squared statistic is discussed in Sec. 3.2.) If one compares Peirce's graphical test of fit with the chi-squared statistic, it is obvious that the visual impact of the graphical display is to weight all deviations from the expected values equally, while chi-squared weights such deviations relative to the magnitude of each expected value. For this reason, discrepancies between the observed and expected values in the tails, even if they had been plotted, might not have appeared to be more serious than those in the middle of the distribution, while it is precisely the contribution of the tails to chi-squared that makes the latter significant.

Clearly the most frustrating aspect of this data set is its lack of detail with regard to the 20 good matches. Were even one of them to show 30 agreements, or even 29, the persuasiveness of the mathematical evidence would evaporate. Indeed, as shown in the earlier table, we can calculate from the data given that the total number of agreements in these 20 cases was 288, or an average of 14.4 for those signatures with 13 or more agreements. Thus, although one or two very high values remains a numerical possibility, it seems likely that few if any values were as high as 20.

Unfortunately for the record, Professor Peirce's demeanor and reputation as a mathematician must have been sufficiently intimidating to deter any serious mathematical rebuttal. He was made to confess a lack of any general expertise in judging handwriting, but he was

C. *Distribution of Coincidences*



[3] The table in the trial record gives 41 cases as "undistributed," but it is clear from Peirce's testimony immediately following that 35 was meant. ("My son had also 35 undistributed cases . . . .")

not cross-examined at all on the numerical and mathematical parts of his testimony, let alone made to list the 20 high values (see Appendix 3).

3.2 Model Assumptions

The frequent paralysis of the law before a demonstration of mathematics has long been noted. An extreme view, expressed by Laurence Tribe in his 1971 *Harvard Law Review* article "Trial by Mathematics" (Tribe 1971) is that laymen are too readily impressed and misled by such presentations and that they should therefore be excluded—at least in jury trials. Whatever one thinks of this position as a generality, Peirce's model, as well as his data analysis, deserves closer examination than it received during the trial.

One might, for example, question the assumption—required for the binomial model—that the probability of a match should be the same at each of the 30 positions. One can easily imagine that certain strokes—perhaps those early in the signature—would have a higher probability of matching than do others. The assumption of equal probability could readily have been checked in the course of C.S. Peirce's work, but it seems never to have been questioned. (Of course, the effect of such inequality would be to *reduce* the variance in the number of matches, and thus to make the probability of 30 matches even smaller than that quoted.)

Far more serious is the implicit assumption of independence. One would expect that agreement in, say, positions 1 and 3, would make far more likely an agreement in position 2. The effect of such dependence would be to *increase* the variance and thus increase the probability of 30 matches over that quoted, quite possibly by orders of magnitude. To the extent that the data bears on this point, it does indeed suggest the possibility of dependence by exhibiting what appear to be far too many cases with over 12 matches to agree with the binomial model.

A final critique of Peirce's discussion is the absence of any comment on correlation between signatures made close in time. It is at least plausible that signatures late in life would differ substantially from those made at a younger age, and also plausible that signatures made at a single sitting would be more similar than those made days or weeks apart. (Similar problems arise in the statistical analysis of authorship, where both style and vocabulary may depend on the writer's subject matter and may vary over his lifetime; see Mosteller and Wallace 1964, pp. 18–22, 195–199, 265–266.) Since many of the 42 signatures used in the study were on leases and other dated documents, this question could also have been examined, but there is no evidence that it was considered.

In part these challenges to the fairness of Peirce's test can be met by the nonparametric procedure suggested earlier. The questions of uniform probability of matching in each of the 30 positions, and independence between them, would then be irrelevant. (The problem of time dependence, on the other hand, remains in full force.) *That* test, however, as noted before, yields only a strongly suggestive significance probability, but nothing at all as compelling as Peirce's calculation.

For all that one can raise questions about Peirce's model and analysis, it remains to consider the extent to which the evidence available really contradicts it. Although the conventional chi-squared test clearly rejects the binomial model, it is not at all clear that the chi-squared statistic should have even approximately the familiar distribution. Under Peirce's model the probability that a random pair of true signatures should match in k stokes is indeed binomial, as claimed, and thus the expected fraction of pairs with k agreements is also binomial. However, the $\binom{42}{2} = 861$ pairs are generated from only 42 signatures and are not at all the same as 861 independent trials from a binomial population. It is to be expected, therefore, that the relative frequencies will show variability greater than multinomial and that the chi-squared statistic will also exhibit greater variability. To assess more rigorously the possible lack of fit of Peirce's model we turn to consideration of the rather large excess of pairs with over 12 matches (20 observed, only 2.7 expected). The Markov inequality (e.g., see Billingsley 1979, p. 65), of which the familiar Chebyshev inequality is a direct corollary, states that for any nonnegative random variable Z, one has

$$\Pr\{Z \geq t\} \leq t^{-1}E\{Z\} \ .$$

Taking Z = the number of pairs out of the 861 showing 13 or more agreements, and $t = 20$, we find $\Pr\{Z \geq t\} \leq 2.7/20 = .14$—not in itself sufficient reason to reject the Peirce model. A more refined analysis, however, gives fairly convincing evidence that the Peirce model does not fit the data (see Appendix 4).

3.3 Independence and the Product Rule

Peirce's unargued and possibly unwarranted assumption of independence should occasion no surprise. It was often taken for granted by thoughtful and astute scientists of that era and later ones—not entirely excepting the scientific community of today. For example, Simon Newcomb wrote in *The Universal Cyclopedia* (1900, s.v. "Probability, Theory of"):

> The mathematical rule for determining probability in such a case [when the concurrence of a large number of circumstances is necessary to the production of an event] is that the probability of the concurrence of all the events is equal to the continued product of the probabilities of all the separate events. As one example, suppose that a law requiring the concurrence of the two houses of Congress and the President were as likely as not to be rejected by any one of them, and that each one of the three authorities formed his own opinion independently of the other two. Then the probability of its passing all three would be $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$.

The product rule is misstated here, independence being only introduced subsequently in the context of an example. And while Newcomb has the *mathematics* right in his example, the *application*, although it can be given the correct interpretation, is exceedingly apt to be given an incorrect one. For let us agree that each branch of government arrives at its own opinion independently of the other two on each law, where "independently" is given the common meaning of deciding in ignorance of the deliberations or decision of the others. It does not at all follow from that that the decisions will be statistically independent. Thus, if a law on capital punishment is under consideration, all might be influenced by public opinion, a variety of lobbies, or the recent report of some heinous crime. The failure of each to directly influence one another does not protect them from being simultaneously influenced by external events.

The confusion of this ordinary meaning of "independence" with "statistical independence" remains a common error today and, although Newcomb himself may have been clear about the matter, his example here seems especially unfortunate.

For while Newcomb may just have been careless, skimming over a point obvious to himself, the statistical layman who read him could easily go astray. This was certainly the case for a noted expert on forgery, Albert S. Osborn (1929, pp. 225–234), whose text *Questioned Documents* cites "Professor Newcomb's rule" with approbation, but whose illustrations of it show no awareness of the need for independence. In one instance, Osborn supposed an individual sought who was 5 feet $11\frac{1}{4}$ inches tall, with blue eyes and brown hair, had lost his left thumb and the lower part of his right ear, and had a distinctive mole, tattoo, and scar. Osborn assumed for the sake of argument that the first of these traits had a probability of $\frac{1}{10}$, the second $\frac{1}{3}$, the third $\frac{1}{4}$, and all the rest 1/200, and concluded that the probability of their joint occurrence was 1 in 38,400,000,000,000. There is, of course, no reason to suppose a priori that height, eye color, and hair color are independent traits, although in this instance possible mutual dependence might not be too serious. Surely, however, the occurrences of a tattoo, a scar, and a missing thumb and ear would be highly dependent. (A more recent and notorious instance is the celebrated *Collins* case.[4])

3.4 "Outrageous" Events

There are some general criticisms that can be leveled at the use of an extremely small significance probability to exclude one hypothesis in favor of another.

Significance probabilities are frequently misconstrued to be statements of posterior probability when only two alternatives are under consideration (here, "Sylvia Ann Howland wrote both signatures independently" vs. "Hetty Robinson forged the second signature"); that is, a significance probability of ϵ against the null hypothesis comes to be regarded (erroneously) as effectively a posterior probability of $1 - \epsilon$ in favor of the alternative.

[4] 68 Cal. 2d 319, 438 P.2D 33, 66 Cal. Rptr. 497 (1968). Fairley and Mosteller (1977, pp. 355–379) give the decision of the California Supreme Court and a critical evaluation.

This intuitive but incorrect interpretation is not entirely off the mark, however, in a case—such as the present one—in which the probability of the evidence, given the alternative (i.e., close agreement in the signatures, given a deliberate forgery), is taken to be very near one. For in that case the likelihood ratio

$$\frac{\Pr(\text{evidence} \mid \text{null hypothesis})}{\Pr(\text{evidence} \mid \text{alternative})} \approx \frac{\text{significance probability}}{1}$$

and the posterior odds for the null hypothesis thus reduce to

$$(\text{significance probability}) \times \text{prior odds} .$$

Hence, in this case, if the prior odds are not extreme, the posterior odds will indeed be of the same order of magnitude as the significance probability.

Another problem affecting the apparent strength of the evidence, however, comes from a very different direction. In framing the problem Peirce quite reasonably divided the possibilities into two categories—either Sylvia Ann Howland wrote both signatures, or she didn't. In the latter case we are left to presume that Hetty Robinson had a hand in the matter. However, there are in fact many alternatives, which, although not very plausible, still have prior probabilities well in excess of 5^{-30}. For example, Sylvia Ann Howland might have been in the habit of tracing her own signature. Or perhaps the Executor, devoted to the destruction of Hetty's claim (and her reputation), might have stolen the original "second page" and replaced it with a forgery. Or we could imagine that C.S. Peirce, in his enthusiasm for the mathematical method, might have been overgenerous in counting agreements when comparing signatures #1 and #10, and more particular when comparing other signatures. Indeed, almost any outrageous possibility that one could conceive of, provided only that it be physically possible, would have a prior probability many orders of magnitude greater than 5^{-30}. Without going into the calculations, it should be clear that the odds against Hetty's guilt cannot be assessed as anything less than the prior odds in favor of these "outrageous" possibilities. (The authors' intuitive assessment puts the odds in favor of some "outrageous" explanation in the present case as at least 10^{-4} or 10^{-5}, and thus—even if we accept Peirce's model in its entirety—the strength of the evidence against Hetty is bounded in this way.) The problem of "outrageous" alternatives and their effect on limiting extreme posterior probabilities is discussed by Mosteller and Wallace (1964, pp. 90–91).

Despite the criticism made here of Peirce's model and data analysis, it stands as an example, in many ways excellent, of the early use of formal statistical methods in a scientific and social setting. Not only did Peirce formulate the problem in a plausible probabilistic framework, and calculate a proper significance probability for that model, but he provided a test of the model itself. That the test was solely graphical is in no way to his discredit, since no more incisive testing procedures had been developed, nor would be for many years to come.

Considered against the background of his era, Peirce's analysis must be judged unusually clear and complete, even though—from a modern viewpoint—we may regard his conclusion about the significance of his findings less compelling than he did.

4. AFTERMATH

Peirce was only one among many witnesses for both sides, but from our point of view the next most interesting witness was J.C. Crossman, an engraver. Crossman testified for the plaintiff, essentially in rebuttal to Peirce. Peirce had said that such agreement as found between signature # 1 and # 10 could not occur except by design, no matter how regular the habits of the writer. Crossman, however, gave testimony about the regularity of the signatures of several individuals, most notably those of the sixth President of the United States, John Quincy Adams (who signed his name "J.Q. Adams"). In each of these cases, Crossman testified, pairs of signatures had been found in which the matching was as good as or better than the matching of # 1 and # 10. Unfortunately, the testimony does not suggest the corresponding value of p (.2 for Sylvia Ann Howland), which powerfully influences the final probability—especially since the number of strokes to be matched in this case is considerably less than 30.

How convincing was the Peirce testimony at the time? Or what of Crossman's, supplementing the microscopic examination of Professors Holmes and Agassiz? The court did not say directly, but its decision can be viewed as suggesting that the court was at least left in doubt about the validity of the contested signatures. In brief, the legal situation was as follows.

As a general matter, a will may be revoked by the testator at any time until his death, although he can by contract bind himself not to exercise his power of revocation. Thus one must look for the grounds on which Hetty Robinson asked that the earlier will and not the later one prevail. She might have sought to discredit the later will on the ground of undue influence. (Indeed, in an earlier proceeding, Hetty had done exactly that, but then withdrew the case.) Alternatively, if the terms of a will do evidence a contractual obligation—as if A puts B's children through college, and B agrees to leave A his house in return—a situation may arise in which a testator is not free unilaterally to change the terms. It was Hetty Robinson's claim, supported with the 1862 will and other evidence, that she and her aunt had contracted to make mutual wills, specifically excluding Hetty's father Edward Robinson, so that in no case could he inherit the Howland fortune. However, Massachusetts law at the time treated a beneficiary under a will as disqualified on account of self-interest from testifying about the circumstances surrounding the will, unless called to testify by the opposite party or by the court.[5] Without Hetty's testimony, the evidence was judged insufficient to support her claim that the 1862 will was in fulfillment of a contract. On this ground the court found for the defendant, Thomas Mandell, and it can only be conjectured whether it might have ruled otherwise had it considered the contested signatures to be valid.

4.1 Did She or Didn't She?

So ends the case—without a finding on the question of most interest to the outsider. Clearly if the contested signatures were forgeries, Hetty Robinson was responsible for them. Did she trace her aunt's signature or didn't she? Here the amateur finds himself on shifting sands. Peirce includes in his testimony some further bits of evidence that point toward forgery, such as the position of the signature on the page, and the level of each letter. Osborn, a specialist in the field of questioned documents and a student of this case in particular, had no doubt whatever that the codicil signature was a forgery by tracing. However, he makes no comment on the J.Q. Adams signatures that appear to defy the usual tests of forgery. Examination of those of the 42 comparison signatures still extant shows a great deal of regularity, but one does not find the same sort of identity as between # 1 and # 10. But the original and both the contested signatures were said to have been written on the same day, and thus might well be more alike than signatures written many days apart. The tendency of all three of the signatures at issue to follow one line from S through w, and then slope off for the last three letters in a slightly different direction, looks at first compelling—until we notice that precisely this pattern is characteristic of most of the 42 comparison signatures. This is, perhaps, as far as an amateur can go, and no unequivocal conclusion seems possible.

5. EPILOGUE

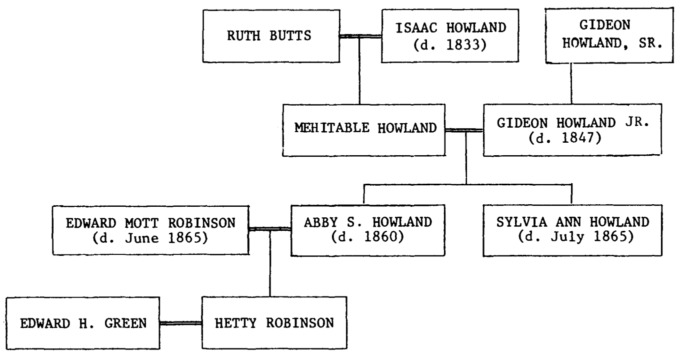
Hetty Robinson Green—she had recently married—lost this case, and other legal actions subsequently. However, far from being weak in those skills necessary to the management of so large a sum—the ostensible reason for putting the money in trust in the first place—Hetty Green turned out to be a dynamic and effective operator in finance. According to contemporary newspaper accounts she devoted herself to the multiplication of her fortune and became one of the most colorful figures Wall Street had known for many years. Despite her enormous wealth, she lived for many years in an inexpensive flat in Hoboken and took the ferry to work in New York City each morning. All this was part of a consciously projected image of penury, which even extended to her wearing old and worn clothing. Viewed as eccentric by many, she had a complex personality

[5] The *American Law Review* (1870, Vol. 4, pp. 656–663), discusses this aspect of the case in considerable detail.

that few—if any—understood. When she died on July 3, 1916, the *New York Times* ran her front-page obituary the next day with the simple headline, "Hetty Green dies, worth $100,000,000."[6]

APPENDIX 1: THE HOWLAND GENEALOGY, AND CHRONOLOGY OF EVENTS CONNECTED WITH THE WILL

Hetty's father, Edward Mott Robinson, acquired his wealth by marrying into the Howland family.



When Hetty's mother died in 1860, a settlement resulted in only $8,000 of the estate passing directly to the daughter, Hetty, the rest (over $100,000) going to the husband, Edward Robinson. This was said to have resulted in ill feelings toward Robinson both on the part of his daughter, Hetty, and her aunt, Sylvia Ann. Hetty alleged that as a result she and her aunt agreed to make and to exchange mutual wills, excluding Hetty's father from any inheritance, should he survive either or both of them. Because of the contractual nature of this agreement, subsequent wills by either of the parties would not necessarily take precedence.

To this end, according to Hetty, Sylvia Ann Howland had dictated a two-page will to her. Miss Howland had also dictated an additional one-page insert—to be brought forward only if necessary—explaining that she feared pressure from her caretakers to make a later will and imploring the judge to recognize this will only and no later one. This "second page" and a copy of it to be kept by Hetty—both signed by Sylvia Ann—were attached to the first page, but not shown to witnesses when the will was executed and signed by Miss Howland later in the day.

[6] *Notable American Women* (1971) gives a brief, excellent, and objective account of Hetty Green's life with further bibliography; a somewhat more sensationalist account appears in John T. Flynn's inappropriately titled *Men of Wealth* (1941). Much better than its title would suggest is Boyden Sparkes and Samuel T. Moore's book-length biography *Hetty Green: A Woman Who Loved Money* (1931), reprinted in 1948 as *Hetty Green: The Witch of Wall Street.* This case, along with other instances of the use of probability in legal cases, is described by Elmer Mode, *Journal of the American Statistical Association* (1963). Unfortunately, his presentation is compressed and erroneous in important particulars. The later vicissitudes and ultimate dissolution of the Howland fortune after Hetty's death are described in Arthur H. Lewis's *The Day They Shook the Plum Tree* (1963).

Evidence at the trial was introduced suggesting that relations between Hetty and her aunt were by no means as cordial as Hetty made out, thus casting doubt on the credibility of Hetty's testimony. For further details of the case, see the *American Law Review* (1870, Vol. 4, pp. 625–664).

APPENDIX 2: CROSS-EXAMINATION OF OLIVER WENDELL HOLMES, SR.

Q. What was that color whose different degrees of intensity you have spoken of?
A. Ink color.
Q. Is that the best answer you can give to that question?
A. Black ink color would be a better.
Q. That answer applies to the three signatures, does it?
A. To the three signatures, yes.
Q. What component materials did you discover, by the aid of the microscope in the ink?
A. I did not discover any definable particles.
Q. Any knowledge which you acquired by your examination of the materials of which the ink was composed, you may give us.
A. That it was a deposit from a fluid.
Q. That all?
A. That's all. I have spoken of the color.
Q. The materials of which the paper was composed, and the difference of their absorbent powers, please state.
A. Fibers. I did not experiment on their absorbent powers.
Q. State the examination which you made of the writing in the body of those instruments, under the microscope.
A. My attention was not particularly directed to that.
Q. Then if I understand you rightly, you did not examine under the microscope, the writing of the body of the instrument.
A. I made various examinations of the writing of the body of these instruments, but not with special reference to evidence.
Q. When did you first see these papers?
A. I first saw these papers yesterday.

APPENDIX 3: CROSS-EXAMINATION OF BENJAMIN PEIRCE

Professor Peirce was an assured and effective witness for his side, at times perhaps too assured. It is instructive to contrast the first part of his testimony with the final portion of his cross-examination.

Q. Your life has been spent, has it not, for the most part in mathematical studies, and in teaching the same, has it not?
A. It has.
Q. You are not an instructor in writing and do not profess any peculiar skill in handwriting?
A. I am not, and do not.

Q. You are not and never have been, have you, an engraver or lithographer?
A. I am not and never have been.
Q. You have not practised the art of tracing, have you?
A. A very little, to find out what it consisted in, and what could be accomplished in tracing by an inexperienced hand.
Q. When did you so experiment?
A. The last time within about ten days; a long time ago I had done something of the kind, so that I had an idea of it.
Q. Have you ever been called as a witness to testify in regard to the genuineness of handwriting?
A. Only where there was no doubt.
Q. What do you mean by that?
A. In the ordinary course of business to testify to a signature that was not known to persons who had to use it in the way of business.
Q. But not in the case of questioned or suspected handwriting?
A. No.

APPENDIX 4: ANALYSIS OF PEIRCE'S MODEL

Under Peirce's model, with $p = .2$, the probability of finding 13 or more agreements out of 30 trials is $Q = B(13; 30, .2) = .0031$. The observed tail frequency provides an estimate $\hat{Q} = 20/861 = .0232$ and, calculating a binomial contrast based on $n = \binom{42}{2} = 861$ pairs, we find that

$$\frac{\hat{Q} - Q}{\left(\frac{Q(1-Q)}{n}\right)^{\frac{1}{2}}} = \frac{.0232 - .0031}{\left(\frac{(.0031)(.9969)}{861}\right)^{\frac{1}{2}}} = 10.61$$

which is highly significant. However, as noted before, the situation is *not* binomial, since the 861 pairs are not independent. To assess the discrepancy more precisely, we note that, properly viewed, $\hat{Q}$ is a U statistic. For let X be a random variable denoting the "configuration" of a randomly chosen signature, so that $X_1, \ldots, X_n$ $(n = 42)$ are independent and identically distributed copies of X, and let

$$\psi(X_i, X_j) = \begin{cases} 1, & \text{if the } i\text{th and } j\text{th signatures agree in 13 or more downstrokes} \\ 0, & \text{if not.} \end{cases}$$

Then

$$\hat{Q} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \psi(X_i, X_j)$$

is a U statistic with variance under Peirce's model (i.e., $Q = B(13; 30, .2)$) equal to

$$\operatorname{var}(\hat{Q}) = \frac{Q(1-Q)}{\binom{n}{2}} + \frac{2(n-2)}{\binom{n}{2}} \sigma_1^2 ,$$

where

$$\sigma_1^2 = \operatorname{var}[E(\psi(X_1, X_2) \mid X_1)]$$

(or equivalently,

$$\sigma_1^2 = \operatorname{cov}(E(\psi(X_1, X_2) \mid X_1) E(\psi(X_1, X_3) \mid X_1)) .$$

(See, e.g., Lehmann 1975, p. 368). Note that

$$E(\psi(x_1, X_2) \mid X_1 = x_1)$$

is the probability, given that one signature is in configuration x_1, that another signature matches it in 13 or more downstrokes.

If $\sigma_1^2 = 0$, then Chebyshev's inequality gives

$$P\{|\hat{Q} - Q| \ge .0201\} \le \frac{\operatorname{var}(Q)}{(.0201)^2} = \frac{(.0031)(.9969)}{(861)(.0201)^2} = .0089$$

and Peirce's model is rejected at the one percent significance level. For example, we might suppose that each one of the 30 downstrokes in Sylvia Ann Howland's signature can be in one of five equally possible states. Then

$$E(\psi(X_1, X_2) \mid X_1) \equiv B(13; 30, .2)$$

is constant, independent of X_1. However, a more realistic model might suppose that each downstroke can be in one of a variable number of states, say $S_j^{(i)}$ $(1 \le i \le 30,$ $1 \le j \le n_i)$, with possibly unequal probabilities $p_j^{(i)}$. In this case,

$$Q(X_1) \equiv E(\psi(X_1, X_2) \mid X_1)$$

is a nonconstant random variable, with $E(Q(X_1)) = B(30; 13, .2)$, so that $\sigma_1^2 > 0$, and $\operatorname{var}(\hat{Q}) = 0(n^{-1})$ (instead of $0(n^{-2})$). Whether 20 signatures with 13 or more downstroke agreements is excessive under Peirce's model thus depends on the value of σ_1^2. Specifically, since U statistics are asymptotically normal, we can calculate that

$$\frac{|\hat{Q} - Q|}{(\operatorname{var}(\hat{Q}))^{\frac{1}{2}}} \le 1.96 \Leftrightarrow \frac{|.0232 - .0031|}{\left(3.59 \times 10^{-6} + \frac{80\sigma_1^2}{861}\right)^{\frac{1}{2}}} \le 1.96$$

$$\Leftrightarrow \sigma_1 \ge .033 .$$

Thus, unless the standard deviation of $E(\psi(X_1, X_2) \mid X_1)$ is at least ten times as large as its expected value (.0031), Peirce's model is rejected at the five percent significance level.

[*Received March 1979. Revised October 1979.*]

REFERENCES

Billingsley, Patrick (1979), *Probability and Measure*, New York: John Wiley & Sons.
Fairley, William B., and Mosteller, Frederick (1977), *Statistics and Public Policy*, New York: Addison-Wesley Publishing Co.
Flynn, John T. (1941), *Men of Wealth*, New York: Simon & Schuster.

Lehmann, Erich L. (1975), *Nonparametric Statistics: Methods Based on Ranks*, San Francisco: Holden-Day.

Lewis, Arthur H. (1963), *The Day They Shook the Plum Tree*, New York: Harcourt, Brace and World.

Mode, Elmer B. (1963), "Probability and Criminalistics," *Journal the American Statistical Association*, 58, 628–640.

Mosteller, Frederick, and Wallace, David L. (1964), *Inference and Disputed Authorship: The Federalist*, Reading, Mass.: Addison-Wesley Publishing Co.

Notable American Women, 1607–1950: A Biographical Dictionary (1971), Cambridge, Mass.: Belknap Press.

Osborn, Albert S. (1929), *Questioned Documents* (2nd ed.), Albany: Boyd Publishing Co.

Sparkes, Boyden, and Moore, Samuel T. (1931), *Hetty Green: A Woman Who Loved Money*, New York: Doubleday; reprinted (1948) as *Hetty Green: The Witch of Wall Street.*

Stigler, Stephen M. (1978), "Mathematical Statistics in the Early States," *Annals of Statistics*, 6, 239–265.

Tribe, Laurence H. (1971), "Trial by Mathematics: Precision and Ritual in the Legal Process," *Harvard Law Review*, 84, 1329–1393.

The Universal Cyclopedia (1900), New York: D. Appleton & Co.