Miracles and Statistics: The Casual Assumption of Independence
William Kruskal
77th ASA President, 1982

1. Introduction

Francis Henry Egerton, Earl of Bridgewater,[1] died in 1829: in his will he left 8,000 pounds for the writing and publication of works on "the Power, Wisdom, and Goodness of God, as manifested in Creation." Under eminent auspices, eight clerical and scientific notables each wrote and published a treatise. Charles Babbage, progenitor of the modern computer, felt that the eight treatises did not carry their argument far enough, and that at least one of them incorrectly argued that the pursuits of science are unfavorable to religion. In addition, Babbage had views about miracles, views in part statistical. So in 1837 he published what he called The Ninth Bridgewater Treatise: A Fragment. [3] (A second, revised edition appeared in 1838.)

I begin with Babbage's treatment of miracles in his own context, that of Hume's famous argument against testimony for miracles. Then I discuss miracles and testimony more generally, with special emphasis on the ubiquitous, critical, and often tacit assumption of independence among witnesses. I then discuss overcasual assumptions of independence in other contexts. If I have a moral, it is this: Do not multiply lightly.

Miracles are relevant to statistics in a number of ways. One analogy is that miracles are like apparent outliers that we study and worry over (e.g., see Anscombe [5]; Kruskal [6]). One might even say that miracles are the extreme outliers of nonscientific life. [7] It is widely argued of outliers that investigation of the mechanism for outlying may be far more important than the original study that led to the outlier; the discovery of penicillium is often given as an example. Just so, for those who take miracles as signs of revelation, the religious import of a miracle may be far more important than the daily life in which it occurred. In the same way, there are misleading apparent outliers and misleading false miracles, the results of poor, prejudiced, or careless observation. If identified as such, these outliers and miracles are worth little attention except for their connections with the psychological vagaries of human observation or with the effects on behavior of fortuitous coincidences.

2. Religion and Statistics

Religion and science generally are two basic realms of human effort that can hardly fail to interact and intersect. Gilkey expresses that strongly: "new religious forms appear and reappear out of and because of a scientific, technological culture in response, first, to the demand for a credible system of symbols giving structure, meaning, and direction to nature, history, society, and the self, and, second, to the particular sharp anxieties – and even terrors – of a technological age" ([9], p. 71). A current example of intersection is the debate about teaching evolution, especially in secondary schools, and Gilkey's article is about that debate. Again, the great physicist Frank wrote of how all scientific theories are related to "their fitness to support desirable attitudes in the domain of ethics, politics, and religion" ([10], p. 9). Some will be surprised at so normative a statement. I do not propose treating further the broad theme of religion and science, and I certainly will not exhibit aggressive incompetence by dipping deeply
into theology.

Now narrow our focus to connections between religion and the mathematical sciences, a part of science in general. There are many such connections without explicit statistical aspects; I think of a range from astronomical and calendar calculations to recent attempts (Brams [11]) toward applying game theory to biblical stories.

One also finds fatuous or superficial uses of mathematical terminology, for example, by Francis Hutcheson in 1724: "Benevolence is directly as the Moment of Good and inversely as their abilities, i.e., $B = M/A$" (Thompson [12], pp. 14-15). Again, David Hartley writes in 1749 that $W = F^2/L$, where $F =$ fear and $L =$ love of God, and $W =$ love of the world (see Stephen [13], Vol. 2, pp. 57-58). Such uses have been roundly criticized and also defended as useful metaphors. In the case of John Craig, whose 1699 calculation of the date of the Second Coming has been frequently derided, a recent brilliant interpretation by Stigler [14] provides a sympathetic recasting. In any case, the argument over apparently superficial use of mathematical expressions continues in bitter argument over membership in the U.S. National Academy of Sciences (see Koblitz [15]; Marshall [16]).

If we specialize further, from the mathematical sciences generally to probability and statistics, a number of developments come to mind. As Pearson said, De Moivre’s work on the central limit theorem was theological: "he was determining the frequency of irregularities from the Original Design of the Deity. Without grasping this side of the matter, it is impossible to understand the history of statistics from De Moivre through Derham and Sussmilch to Quetelet, culminating in the modern principle of the stability of statistical ratios" ([17], p. 404). In another place, Pearson carried the sequence of theologically motivated statisticians through Florence Nightingale, and added that "with a modified meaning we might almost add Francis Galton" ([18], p. 286). Indeed, several chapters of the lectures ([18]), especially those dealing with Thomas Bayes and Richard Price, are suffused with theological motivations and connections.

A different theme is the ancient religious fear of being counted, for example in a census. This appears in several parts of the Old Testament; see Madansky [19]. For general treatments of religion and statistics, see Portnoy [20] and Bartholomew [21, 22].

3. Prayer

I digress to discuss statistically shaped studies of the efficacy of prayer. These studies are interesting per se, and they bring us near the topic of miracles. For some argue that prayers – at any rate answered prayers – are in a sense miracles, if often small miracles. Yet Tyndall wrote – and many others have made the point – that no prayer "could call one shower from heaven or deflect towards us a single beam of the sun" without a "disturbance of natural law, quite as serious as the stoppage of an eclipse" ([23], p. 36).

Perhaps the first recorded empirical comparative trial of prayer was the competition between Elijah and the many priests of Baal (1 Kings 18), although the objectivity of its reporting may be questioned by some. Recall that the priests of Baal built an altar, put on it a sacrificed bullock, and called on Baal to start the fire. Nothing happened. Then Elijah had a bullock put on his altar,
poured water over it to make the task dramatically harder, and called on his deity. The response was overwhelmingly fiery. At least one scholar [24] has built on this example and proposed randomized trials as an approach to showing that miracles really happen – yet another relationship between religion and statistics.

Among relatively recent empirical studies of prayer, the earliest I know was by the great Francis Galton. He reasoned that the royal family and British clergymen probably generated more British prayers for good health than did others, so he compared longevity of members of the royal family and clergymen with the longevity of others. There are, of course, obvious problems of control and comparability, but Galton pushed ahead nonetheless, and came out with negative results that scandalized many contemporaries, perhaps as much for the method as the conclusions. His first publication on the topic was in 1872, and he included the material in his 1883 book (see [25, 26]). The book also had a more general discussion of what Galton called theocratic intervention and the possibility of ascertaining it. In the second edition, however, the theometric material was omitted, because – said Galton – it offended the sensibilities of a number of readers. I believe that he never returned to the topic in print.[27, 32]

The Galton [25] article and other contemporary exchanges were reprinted in The Prayer-Gauge Debate (Means [34]). (Galton's publication was slightly preceded by Tyndall [23] and Henry Thompson, but they only proposed studies where Galton carried one out.) Pearson's massive biography of Galton describes the study (see [33], Vol. 2, pp. 115-117). Haldane ([35], pp. 243-245) provides a slightly distorted account (wrongly substituting children of clergymen for their fathers), and Brush [36] gives a thorough discussion.[37] A recent medical evaluation of prayer by Joyce and Weldon [49] was double-blind and inconclusive. Another recent study ([50]) gave results more encouraging to prayer, but its description leaves me with questions. Of course there is a long literature on the curative effects of faith in various senses; for an overview and entry bibliography, see Frank [51].

4. Miracles

We turn now to miracles themselves. There is a huge literature on miracles, and I have been able only to dip into it selectively with a primary interest in evidence for miracles, especially the validity of testimonial evidence. Begin with David Hume's 1748 definition (see Hume [52]) because Hume was one of the first [53] to apply probabilistic ideas to miracles, and because Hume's famous argument against miracles is framed in terms of one probability outweighing another. Hume defined a miracle as "a transgression of a law of nature by a particular volition of the Deity, or by the interposition of some invisible agent" ([52], p. 115). See Pomeroy ([58], p. 6) for variant Hume characterizations of a miracle.

Aquinas had earlier said that a miracle is something done by "divine agency beyond the order commonly observed in nature." A more recent definition speaks of a "striking interposition of divine power by which the operations of the ordinary course of nature are overruled, suspended, or modified." See Flew ([59], p. 346) for a highly relevant treatment.

Every nuance of definitions like the above has been repeatedly examined and analyzed. For example, the idea that a miracle should be striking, remarkable, the Resurrection in bright color
with a full chorus singing the Messiah, has been both underscored and denied many times. Hume himself did not see the need for a striking appearance. He said:

The raising of a house or ship into the air is a visible miracle. The raising of a feather, when the wind wants ever so little of a force requisite for that purpose, is as real a miracle, though not so sensible with regard to us. ([52], p. 115)

Other authors might exclude the raising of a feather by insisting that an event must have clear religious significance to be counted a miracle; for example, see Swinburne ([60], pp. 7-10). For a discussion in the Old Testament context, see Eichrodt ([61], p. 163). Yet it is hard to know what religious significance might mean. Eliade insists that miracles may appear in the most ordinary parts of everyday life. He says, for example, that "Just as I believed in the unrecognizability of miracle, so I also believed in the necessity... of the camouflaging of the 'exceptional' in the banal, and of the trans-historic in historical events" ([62], p. 224). I have difficulty with the unrecognizable miracle, for how can that be a sign or evidence or explanation? For a fuller statement of this point see Smart ([63], pp. 48-49).

There are also viewpoints that simply take life or science or man as miracles in themselves. "The greatest miracle is the existence of the laws of nature." Or, "All of life is a miracle. When you look at an apple seed, you are really looking at a miniature fruit tree. This is as big a miracle as the news of the Resurrection." These are some quotations from scientists in The God of Science, a book of interviews with scientists about religion (Trinklein [64], pp. 80-81). [65] Prior says that "Faith... is an inward miracle of God's mercy" ([67], p. 8). To be more accurate, the statement is made by a Barthian Protestant in Prior's imaginary dialogue. Eichrodt stresses that the Old Testament takes miracles in the "widest possible terms. Even the course of Nature itself counts as a miracle" ([61], pp. 162-163).

An important problem with Hume's definition is that of explicating the idea of "law of nature." Another is to show that an alleged miracle is in fact willed by the deity, or by an appropriate "invisible agent" – presumably a benign one, say an angel, as opposed to a demon or devil (see Swinburne [68]; [60], chap. 5).

Yet I avoid such angelic temptations and hasten on to the central question of testimony about the happening of a miracle. For testimony, in particular allegedly independent testimony from two or more witnesses, forms a central topic of this exposition. [69, 76]

Hume's discussion of testimony is cynical, skeptical, and full of wit. He concludes that human testimony is so subject to bias, passion, self-interest, change over time, and so on, that "no testimony for any kind of miracle has ever amounted to a probability, much less to a proof; and that, even supposing it amounted to a proof, it would be opposed by another proof" from the weakness of testimony for remarkable events ([52], p. 127).[81] Hume goes on to say that experience, and only experience, informs us of the validity of human testimony and informs us of the constancy of the laws of nature. When "these two kinds of experience are contrary, we have nothing to do but subtract the one from the other. ... this subtraction, with regard to all popular religions, amounts to an entire annihilation; and therefore... no human testimony can... prove a miracle, and make it a just foundation for any such system of religion" ([52], p. 127).
One can try to formalize this argument in several ways, comparing two explanatory hypotheses about a reported miracle, say the changing of water into wine or the levitation of a holy anchorite. Hypothesis $H$ is that the miracle really happened, so the report is correct, putting aside questions of volition by a deity or other invisible agent. Hypothesis $\bar{H}$ is that the miracle did not happen and that the reporter (perhaps oneself) was mistaken, tricked, or even a liar. A fuller treatment might consider the possibility that the report was correct but misinterpreted. There are two sample points, $R$ (the miracle is reported) and $\bar{R}$ (it is not), to permit an analysis.

A likelihood approach might be tried, but it seems not rich enough to include Hume's argument. Turn instead to a simple, familiar Bayesian structure; after all, miracles usually have a prior improbability. (But keep in mind those everyday miracles stressed by some.) Suppose then that $\theta$ is the a priori probability for $H$, and $\bar{\theta} = 1 - \theta$ that for $\bar{H}$. Let $p$ be the probability of correct reporting: $p = \Pr(R \mid H) = \Pr(\bar{R} \mid \bar{H})$. Let $\bar{p} = 1 - p$. Then we have a standard $2 \times 2$ table:

$$
\begin{array}{c|cc|c}
 & R & \bar{R} & \\
\hline
H & p\theta & \bar{p}\theta & \theta \\
\bar{H} & p\bar{\theta} & \bar{p}\bar{\theta} & \bar{\theta} \\
\end{array}
$$

The familiar central calculation is to write the column-wise conditional probability as the unconditional one over the sum, $\Pr(H \mid R) = p\theta / (p\theta + \bar{p}\bar{\theta})$. Note that if $p = \bar{p} = \frac{1}{2}$, then $\Pr(H \mid R) = \theta = \Pr(H)$, a natural result. Even more naturally, if the witness is wholly trustworthy ($p = 1$), then $\Pr(H \mid R) = 1$.

This formulation may be criticized in many ways. For example, the propensity to lie or be duped is presumably very different if a highly likely event happens than if it does not; a bridge hand of 12 hearts is, I think, much more likely to be reported as 13, than one of 13 reported as 12. Thus the value of $p$ might well be taken as different in the two rows. More important, there is every reason to expect $p$ to be a function of $\theta$: the probability of reporting a bridge hand of 12 hearts as 13 is, I surmise, much greater than reporting a hand with 4 hearts as one of 5 (see Venn [87], p. 421). These are psychological and historical questions that cannot be neglected. There seems to be remarkably little discussion of reasonable numerical values for $\theta$ and $p$ in various settings. [88]

At a more fundamental level, we generally do not set up such models until after a surprising event has been reported or at least discussed. Thus there is inherent bias of selection in the very act of model formulation. Similarly, there is something arbitrary about setting up the two-point sample space. Coins do land on their edges sometimes, and more often witnesses have their own
in-between summaries: neither water nor wine, uncertainties about how dead the revived man was, and so forth. In some cases, it makes little sense to look at the probability of a simple event rather than at a tail probability of decreasingly probable events under one hypothesis.

Hume argued, as I interpret him, that

\[
\Pr(\text{incorrect observation} \mid R) = \Pr(\overline{H} \mid R) > \Pr(H \mid R)
\]

\[
= \Pr(\text{correct observation} \mid R),
\]

that is, that \( \Pr(H \mid R) < \frac{1}{2} \). A little calculation shows that this is equivalent to \( \theta < \overline{p} \); that is, the unconditional probability of the miracle is less than the probability of a lie (or mistake, or the like). He argues by anecdote, appeals to experience, and explanation of his cynical view of human nature. {That skeptical position was not new.[90] Indeed it long predates the Christian era; see Grant [91] for a detailed history of waves of skepticism and belief.} Hume might – but I think did not – have quoted the skeptical words of Shakespeare, who had Gloucester say "What means this noise? /Fellow, what miracle dost thou proclaim?" ([92], act 2, sc. 2). Still Hume put the skeptical case with special force, at a time of developing uncertainties about Christianity, and in terms – however early and rough – of probability.

5. Babbage’s Reply

Hume’s argument was met by a storm of protest, much too extensive to report here. {One account is by Brown [93], pp. 89-91.} But I do want to describe Babbage’s answer to Hume. Babbage said, in effect,

David Hume, you may be right for a single observer that \( 0 \) is less than \( p \), but suppose that there are several observers, and independent ones, who agree in reporting the miracle, that is, in \( R \). Isn't it obvious that, for any fixed \( \theta > 0 \) no matter how small, we can find an integer \( n \) such that agreement of \( n \) independent observers on \( R \) will produce a conditional probability for \( H \) that is arbitrarily close to 1?

Well of course that's right in an algebraic sense provided that \( p > p \), that is, that a correct observation is more probable than an incorrect one.

Indeed, the conditional probability of \( H \), given that all \( n \) independent observers agree in reporting that it happened, is
where we assume $0 < 0 < 1$. If $p > p$, then this probability as a function of $n$ is monotone increasing to the limit unity. Note that this assumes that $p$ is the same for all observers and not a function of $n$. It also assumes crucially that the observers behave; independently. At the opposite extreme, if the observers behave in a wholly dependent way and give the same answer, the conditional probability is the same as for a single observer.

A complication is that $p$ might decrease as $n$ increases, a phenomenon observable at magic shows. Under some circumstances, the larger the audience the greater the chance of misobservation.

Babbage went through calculations similar to those presented previously and concluded as follows:

\[
\frac{p^n \theta}{p^n \theta + \overline{p^n \theta}} = \frac{1}{1 + \left(\frac{\overline{p}}{p}\right)^n \overline{\theta}},
\]

where we assume $0 < 0 < 1$. If $p > p$, then this probability as a function of $n$ is monotone increasing to the limit unity. Note that this assumes that $p$ is the same for all observers and not a function of $n$. It also assumes crucially that the observers behave; independently. At the opposite extreme, if the observers behave in a wholly dependent way and give the same answer, the conditional probability is the same as for a single observer.

... if independent witnesses can be found, who speak truth more frequently than falsehood, it is ALWAYS possible to assign a number of independent witnesses, the improbability of the falsehood of whose concurring testimony shall be greater than that of the improbability of the miracle itself.

... the whole of this argument applies to independent witnesses. The possibility of the collusion, and the degree of credit to be assigned to witnesses under any given circumstances, depend on facts which have not yet been sufficiently collected to become the subject of mathematical inquiry. Some of those considerations which bear on this part of the subject, the reader will find treated of in the work of Dr. Conyers Middleton, entitled "A Free Inquiry Into the Miraculous Powers Which Are Supposed to Have Subsisted in the Christian Church, From the Earliest Ages Through Several Successive Centuries." ([3], pp. 202-203) [94]

Not surprisingly, others criticized Hume along Babbage-like lines.[95] Much earlier, Price ([99], p. 418) briefly made the point. Chalmers ([28], pp. 66-69, 130-135) gave a similar argument before Babbage. Chalmers excused his "mathematical style of reasoning" because it was "the best fitted to neutralize the mischievous influence superadded to the skepticism of Hume by the great name of Laplace" ([28], p. 66). Chalmers also scolded Laplace on pages 114-116 and 142-143; I shall return to Laplace.

Price and Chalmers are careful to require independence, although I do not see proper discussions of how strong that requirement is. (Indeed, Chalmers sometimes forgets and multiplies too casually, e.g., [28], p. 122.) In the same way the mathematician James R. Young carefully assumes independence – no collusion, in his language – without examining the strength of that assumption. Young's calculations are much like Babbage's; a first claim by Young ([100], pp. 366-367; [101], p. 80) is that Babbage ([3], pp. 198-199) made a small error in calculation that does affect his main conclusion. In fact, it seems to me that Young misread Babbage and
that the two agree.

Young's second claimed advance is more interesting. He points out that both Hume's and Babbage's treatments refer to a specified miracle; in fact, initial reports of miracles refer to previously unspecified miracles. "Suppose, for instance," says Young ([100]), "that a person testifies to the fact that he saw a dead man raised to life; and that ten other persons with every disposition to deceive, but without collusion, testify to the same thing. Now, supposing that these ten persons were limited within the very narrow range of only ten fabrications suitable to their purposes of deception: the probability that they would all fix upon the particular miracle mentioned is \(1/(10)^{10}\)"([100], pp. 363-365). This development is an early discussion of the familiar statistical problem of selection. It is also an attempt to get around the Hume-Babbage requirement \(p > \frac{1}{2}\). But of course the primary weakness in Young's argument is the difficulty in establishing the absence of collusion.

That some probability is very small says little or nothing by itself, or else every bridge hand would be astonishing. One natural statistical way of formalizing Young's approach is to consider the alternative that, if one of the witnesses in fact saw a dead man raised to life, he would (no matter how much a liar) report it with probability \(w > .1\). That leads to a likelihood ratio of \((w/1.1)^{10}\) for the sample point in which all report the revival – assuming independence.

The literature on miracles includes much on qualitative consistency among witnesses, for example, among the New Testament Gospels. A widely read nineteenth-century example is The Life of Jesus by Strauss ([102], e.g., see pp. 742-743). Ernest DeWitt Burton, president of the University of Chicago, devoted much of his scholarly energies to reconciling, and (as he put it) finding harmony among, the Gospels (e.g., see Stevens and Burton [103]). More recent related works are McArthur ([104]), which discusses possible dependencies among the Gospels, [105] and Perrin ([107]).

6. Other Fundamental Issues

In serious statistical thought about testimony, as about most other applications, fundamental problems may question the relevance of probability ideas at all, or point out ambiguities in framing a model. As Venn put it, "the individual presented himself, and the task was imposed upon us of selecting the requisite groups or series to which to refer him. . . . we have to select the conditions of frequency out of a plurality of more or less suitable ones" ([87], p. 395). In effect, Venn shows how \(p\) may well depend on the nature of the event in question, thus challenging our ability to set up a usable probability space at all. The point had been raised earlier, for example by Chalmers ([29], pp. 140-141) as he chides Hume for shifting reference populations.

So the reference set may be questioned, and there are further complications. For example, in surveys with several or many interviewers it may be sensible to regard them as randomly chosen, thus introducing dependence among the respondents interviewed by the same enumerator. In other circumstances, especially with one or two interviewers only, we may decide to regard any interviewer effect as fixed so that there is no stochastic dependence. There is a similar ambiguity in some discussions of covariance analysis where nearby plots of ground in agricultural experimentation are taken to have correlated fertilities on one page and considered as bearing
unknown fixed differential fertilities on the next. I think there are epistemological issues [108] here that remain to be sorted out.

7. France

Before coming to contemporary examples, I turn back to French thought. Starting with Condorcet and Laplace especially in the latter's famous *Philosophical Essay* – calculations something like those sketched earlier for testimony to miracles, decisions about law-court cases, and deliberations of legislatures. Poisson was a major contributor. In most of this work there was little reference to empirical evidence and great reliance on a priori simplifying assumptions: independence, equal probabilities over people, and probabilities of error not depending on circumstances. There was eventually a revulsion against what John Stuart Mill called the "opprobrium of mathematics" ([118], book 3, chap. 18, sec. 3, p. 353; [112], p. 538), an expression translated into French by Bertrand ([119], p. 327) as "la scan dale des Mathematiques." I hasten to add that there was also work of the era that tried to be more realistic and that dealt carefully with empirical evidence; in particular I cite Cournot ([120]) and Bienayme ([121]). For further discussion see Heyde and Seneta ([122], especially sec. 2.4, pp. 28-34). [123]

There was, in any case, a turning against direct probabilistic analyses of matters legal and legislative; only in recent years has such work – done far more carefully – had a rebirth. Yet even today egregious oversimplifications are made in legal settings. A popular example is the notorious Collins case, in which multiplication of alleged probabilities of characteristics was casually done; see Fairley and Mosteller ([126]) for citations and a tolerant exposition. An extensive discussion in the legal literature, with many references, is McCormick ([127], pp. 491-499). See also the exposition by Eggleston ([128]), in which the assumption of independence is carefully stated, but as a kind of incantation without due attention to its strength and the difficulty of confirming it. [129]

8. Recapitulation

I pause to recapitulate. Starting with questions about testimony for miracles, we have seen that one central issue is that of independence – or alleged independence – among witnesses. We have traced briefly the historical development of models for testimony generally, including testimony in judicial circumstances. Before turning to contemporary examples, one may well ask why the assumption of independence is so widespread.

One answer is ignorance. Many otherwise estimable people will write that, for example, "The combination of these two probabilities – that the core will dry out under accident conditions and that safeguard systems will fail to re-flood the core – is the product of the two individual probabilities, or 1 in 1 million" ([132], p. 72).

Far more important than simple ignorance, in my opinion, is seductive simplicity: It is so easy to multiply marginal probabilities, formulas simplify, and manipulation is relatively smooth, so the investigator neglects dependence, or hopes that it makes little difference. Sometimes that hope is realized, but more often dependence can make a tremendous difference.
There are intermediate positions in the constant tension between verisimilitude and simplicity. A grand theme in the development of stochastic models is the assumption of conditional independence, with the condition often turning on unobserved, hypothetical variables. Two examples are factor analysis and latent structure theory. In my view there is a central obligation to face squarely what we do and do not know, and to study robustness of conclusion against mistakes in a priori assumptions of independence, conditional or not.

9. Contemporary Examples

Over-easy acceptance of independence, and struggles with the complexities of not assuming it, continue about us. I give a few examples beyond those already mentioned, starting with accidents.

Estimates of accident probabilities in nuclear reactors are important, and much in the news. The difficulties of making those estimates, as with other low-probability accidents, turn on the relatively few serious accidents observed and on uncertainties in setting up physical models for events leading to accidents. The temptation to assume independence at various points must be strong; otherwise how can one get on with the job?

Most accident analysts understand, of course, that one cannot safely multiply marginal probabilities. For example, after explaining multiplication under independence, Bethe says that one cannot always do that: "There can be 'common mode' failures where one event triggers two or three failures of essential elements of the reactor; in that case the probability is the same as that of the triggering event, and one does not get any benefit from the multiplication of small probability numbers. The probability of such common-mode failures is of course the most difficult to estimate" ([136], p. 25). Exactly so. In fact many accident pathways are in between independence and complete dependence. One faces the difficult choice between simplifying assumptions—the simplest is independence and inserting guesses about many complex conditional probabilities where direct empirical evidence is nonexistent or tenuous. Or arbitrary averages of extreme possibilities may be taken; for a critical analysis of that approach, see Lewis ([137], p. 61).

It is not difficult to find a variety of statements at different levels of care. Deffeyes and MacGregor, for example, give the classically ingenuous form: "the probability that an ore deposit will be formed... is determined by multiplying the probabilities that each essential ingredient was present" ([138], p. 68). Swerdlow ([139], pp. 529-530) takes a contemporary historian of astronomy to task for naively multiplying probabilities.

Lorenz recognizes the need for independence, but passes lightly over judging it: "The improbability of coincidental similarity is proportional to the number of independent traits of similarity, and is, for \( n \) such characters, equal to \( 2^n - 1 \)" ([140], p. 230). He refers to striking resemblances in shape of, for example, a shark, a dolphin, and a torpedo. Other examples may readily be found in both ordinary science and exotic areas, for example, estimating the probability of extraterrestrial life or the probability of a world war next year. [141] Estimating
probabilities like the latter two stretches to the limit – perhaps beyond it – the appropriateness of probabilistic ideas.

How important dependence can be is vividly described by Machol ([147]) in connection with the sinking of the **SS Titanic**. Before the **Titanic** hit an iceberg and sank in 1912, she was declared unsinkable because of her double-bottom hull and watertight compartments. Unfortunately the iceberg cut a 300-foot gash on one side that flooded five adjacent compartments. In addition, the radio operator on a nearby ship was asleep, the **Titanic** was going too fast, and there were insufficient lifeboats.

It is ironic that Machol took the other side in a debate with the late J. Kiefer in an entirely different domain, the taxonomy and classification of mushrooms. Machol, together with R. Singer, had strongly suggested assigning mushroom species to genera in accord with pseudo-likelihood ratios formed by multiplying marginal probabilities of characteristics (like having amyloid spores or having an intermediate lamellae width). The procedure, which was called by Healy ([148], p. 169) in the medical diagnosis context "idiot's Bayes," is in fact the correct Bayes assignment procedure if all indicators are assumed independent. Healy points out that the idiot Bayes procedure sometimes works surprisingly well, but Kiefer ([149]) vigorously argues that responsibility for showing that it works or not is the proposer's rather than the reader's. Kiefer's analytical and critical article appeared in the mycological literature; it contains a full bibliography. Machol replied in 1980 [150].

In the earlier discussion of prayer, [37] I mentioned natural continued interest by the Bishops of the Church of England. A recent episcopal statement about religious belief (House of Bishops) deals in considerable part with miracles and testimony for them. It is explicit about dependence and independence, but its usage is, I think, somewhat vague. Thus, for example, it says that "if we bear in mind the dependence of Matthew and Luke on Mark, there are only two independent witnesses in the New Testament to the institution of the Eucharist, namely Mark and Paul" ([48], p. 28). Why Mark and Paul are taken as independent is not made clear. The same vagueness may be found in some historical literature, for example, Garraghan ([151], pp. 294-295 and elsewhere) as cited somewhat misleadingly by Burns ([56], p. 296). Both of these books contain much material relevant to our discussion.

We see then some of the traps in gratuitous or casual assumptions of independence. A grand irony is that independence seems rare in nature, and when we really want it, we go to great pains to achieve it, for example in the production of pseudorandom numbers for Selective Service lotteries (see Fienberg [152]), and in randomization for allocation of treatments in experiments. (These two cases are not, strictly speaking, aimed at full independence because of their without-replacement structure; the central point is that they are aimed at joint distributions of known form for the sample point. In fact those known forms are usually closely related to random sampling.)

In most real cases there is noticeable dependence between phenomena. We are surprised, for example, when social psychologists working with sentiments of contentment (see Bradburn [153], pp. 57-61) find little or no association between putatively polar traits of positive happiness and negative discontent. Similarly (Spence and Helmreich [154]) putatively polar
traits of stereotypic masculinity (self-confidence and competitiveness) versus stereotypic femininity (gentleness and concern for others) turn out to be almost vanishingly correlated given sex. {Cautionary comments: (a) weak correlation is not the same as independence; (b) the averages differ in standard stereotyped ways.}

An almost universal assumption in statistical models for repeated measurements of real-world quantities is that those measurements are independent, [155] yet we know that such independence is fragile. Berkson, Magath, and Hurn ([157]), for example, showed how laboratory blood-cell counts present surprising correlations. Repeated measurements by the same observer are bound to be affected by memory of prior observed values. To avoid that problem some measuring devices have arrangements for blind haphazard resetting of the zero point between successive observations. [158] Mahalanobis ([161]) suggested carrying this further by using different scales for separate observations, each with variant scale markings – perhaps even nonlinearly related! There is a related substantial literature on process control, in the sense of independent identically distributed observations. Much of that literature deals with shifts in expected value, but some discusses dependence. Page ([162]) gives an excellent overview.

So we have come from miracles, through testimony, to inflated or insidious claims of independence and how they can be damaging. I began with Charles Babbage and I end by quoting William Schwenk Gilbert, another wise Englishman.

In Act 2 of The Yeoman of the Guard Fairfax asks, "And thou didst see all this?" Point answers, Aye, with both eyes at once-this and that. The testimony of one eye is naught-he may lie. But when it is corroborated by the other, it is good evidence that none may gainsay. Here are both present in court, ready to swear to him. [163]

Notes and References

[Text set off by regular parentheses ( ) was provided by the author; text set off by script brackets { } was provided by the original editor; text set off by square brackets [ ] was provided by the current editors.]

[1] He was the eighth and last Earl, a renowned eccentric with family money, largely from the building and operation of English canals. Falk [2] gives a colorful account. The Earl's testamentary intentions were carried out by the President of the Royal Society, assisted by the Archbishop of Canterbury and the Bishop of London. See Babbage [3, p. xxi] and Barnes [4, p. 857].


[7] A philosophical discussion of the relationship between apparent miracles and "irregular or discrepant fact[s]" generally is given by McKinnon ([8], pp. 312-313).


[27] It may be an oversimplification to regard answered prayers as small miracles. For example, Chalmers has an ingenious distinction to explain the apparent inconsistency between the efficacy of prayer (taken for granted) and the observed uniformity of nature, save for the few "well accredited miracles of the Jewish and Christian dispensations" ([28], p. 237). Chalmers pointed out that we are unable to trace causative chains in nature back for more than a few steps, and that prayers are answered by unobservable divine intervention further back in the chain. Real miracles – those exceedingly rare events – arise from obvious, visible divine intervention. The same distinction had been made by Chalmers earlier ([29], pp. 149-150). He did not hesitate to repeat a favorite theme, and I note further that he wrote the first Bridgewater Treatise!

Romanes's book stresses the similarity between answered prayers and miracles (especially [30], pp. x, 174-187), treats constructively the dilemma that a proper miracle or an answered prayer seems to tear a hole in received natural law, and argues more or less quantitatively ([30], pp. 265-266) against Galton's empirical analysis. Part of that analysis shows that clergy live about two years longer than lawyers and doctors; Romanes says that only a fraction of prayers by clergy are truly sincere, say half, that only a fraction of those are for clergy, say half again, and that only half of the remaining prayers are for length of life as opposed to other boons. Multiplication gives an eightfold dilution, and multiplying again two years by eight provides an effective advantage of 16 years! Romanes, to be sure, recognizes the gaps in this counterargument and says that he gives it to show "that the statistical method is not a trustworthy instrument with which to gauge the physical efficacy of Prayer" ([30], p. 266). Romanes' thinking about religion fluctuated, probably under the influence of Charles Darwin. An account of perhaps questionable accuracy is given by Romanes's widow [31].
[28] Chalmers, Thomas (1842), On the Consistency Between the Efficacy of Prayer – and the Uniformity of Nature (Discourses on the Christian Revelation Viewed in Connection With Modern Astronomy To Which We Added Discourses Illustrative of the Connection Between Theology and General Science), in Works (Vol. 7), Glasgow: Collins, pp. 234-262. (The catalog of the British Museum gives a range of 1836-1842 for the Works; it also appears to say that the Discourses was first published in 1818.)


[32] The omission, according to Pearson, was at "the urgent request of the publishers" ([33], Vol. 3B, p. 448). In his preface to the second edition, Galton describes the omission ambiguously, but plainly says that he finds "nothing to recant" in the omitted chapters.


[37] There was a use of the prayer-gauge idea by Stead [38] in his article about a then-current scandal centering on Albert Edward, Prince of Wales, and alleged cheating at baccarat. Stead calculated that at least a thousand million prayers for the Prince of Wales had been made during his life, only to have him become the object of a virulent press attack over card cheating. Benson ([39], pp. 211-213) gives a lucid account of the incident, but he puts Stead with the attackers when in fact Stead is sympathetic toward the Prince – if not to the effectiveness of prayer. Stead explains the Prince's superficial life as the consequence of boredom and lack of serious responsibilities. Turner ([40], pp. 59-63) describes another controversy focused on the Prince of Wales and prayer – this one about serious illness and his recovery. Turner's article provides valuable further discussion about the argument over efficacy of prayer.

In recent years there have again been British commentaries on possible relationships between
prayer and healing, commentaries published under eminent and established auspices (see Archbishops' Commission [41]; British Medical Association [42]; Ramsey [43]). I have noted no reference in these publications to Chalmers' explanation, Galton's analysis, or the prayer-gauge debate otherwise. There are criticisms – mildly stated yet perhaps deep – of the British Medical Association report by the Archbishops' Commission (especially in the latter's chapter 3).

Debate about miracles continues within the Church of England. In particular, I cite the Church's Report of the Church of England Bishops [44]; it arose after York Minster was set afire by lightning soon after the consecration there of now-Bishop David Jenkins, who had been outspokenly skeptical about the fundamental Christian miracles. Two descriptive statements are by Thomas [45] and Kramer [46], pp. 80, 86-89. A strongly pro-miracle viewpoint is given by Berry [47] together with many further references. In the context of dependence among witnesses, I shall return later to a 1986 statement by the House of Bishops [48].

[38] Stead, W. T. (1891), The Prince of Wales, The Review of Reviews, 4, 23-34. {Ascription of authorship comes from Benson (1930, p. 211) and from internal evidence.}


[53] Not literally the first. Brown Grier (Northern Illinois University) has collected other examples, of which an unusually interesting one is the work of Hooper ([54], especially chap. 4; [55]) on transmission of human testimony. The centrality and persistence of Hume's argument nonetheless makes it a good point of departure. As Stephen says "The fact that the argument is being answered to this day proves that its efficacy is not exhausted. Every new assault is a tacit admission that previous assaults have not demolished the hostile works" ([13], vol. 1, chap. 6, sec. 1, p. 262).

Burns [56] treats predecessors of Hume; see pages 10, 59-60, 75 (about Annet), 77 (Woolston), 89 (Wollaston), and 90-95. {Annet's treatment ([57], especially p. 64) is striking, although it lacks even the rough probability idea that Hume uses. Annet stresses the shrewd point that, if a miracle worker really had the power to change nature marvelously, he might ipso facto, and much more simply, deceive the senses of witnesses.} Burns gives a systematic outline of Hume's several arguments about miracles; I discuss primarily the argument based on testimony and probability, called by Burns the a priori epistemological argument.


[55] Hooper, George (1699), A Calculation of the Credibility of Human Testimony, Philosophical Transactions of the Royal Society, 21, 359-365. (There has been disagreement about the authorship of this paper. It appeared in Hooper's Works, published by Oxford University Press in 1757 and 1855.)


[57] Annet, Peter (1744), The Resurrection of Jesus, London: M. Cooper


Philadelphia: Westminster. {Published in 1964 as Theologie des Alien Testament (5th ed.), Stuttgart: Ehrenfried Klotz.}


[65] Marty calls the Trinklein book "obscure and almost naive." Trinklein, says Marty, approached a "rather random sample of prestigious . . . scientists. . . . The opinions he heard were as varied as those . . . from 38 randomly selected patrons of any friendly neighborhood tavern" ([66], p. 31).


[69] Before going on, however, it is worth mentioning again the possibility of extraordinary events that are not miracles; this may be called, following Hume, the Indian Prince question. A South Indian Prince, says Hume, is told that water becomes hard and solid when the temperature goes down enough. The parochial Prince, continues Hume, ought to be skeptical and require strong testimony, but the alleged aqueous behavior is "not contrary to his experience. . . {but rather} not conformable to it" ([52], pp. 113-114). He simply had never observed water at low temperatures. The example goes back at least to John Locke in 1690 (see Locke [70], chap. 9, sec. 5) in terms of the King of Siam. Something like that is the story of meteorites, whose existence was at first denied for years (see Westrum [71]). There is an apocryphal Hume – like anecdote about Thomas Jefferson: When informed that two Yale professors had reported a meteorite fall in 1807, Jefferson is said to have remarked that "It is easier to believe that two Yankee professors would lie than that stones would fall from heaven" (Mason [72], p. 429). Whether ball lightning really exists has been questioned for years (Garfield [73]; Stewart [74], p. 403; Uman [75], p. 535). The Stewart discussion deals explicitly with supposedly independent observers. Of course these examples can be more than matched by many cases of false, tendentious, hallucinatory, and so forth observations of apparently extraordinary events. Most of these would not be called even potential miracles by most miracle enthusiasts, primarily because there is little religious relevance in ice, meteorites, ball lightning, and the like.

[70] Locke, John (1959), *An Essay Concerning Human Understanding*, ed. A. C. Fraser, New York: Dover. (First published in 1690; Fraser edition published in 1894.)
A basic divergence in the literature on miracles corresponds to the age-old debate about free will and divine intervention. To put it crudely: Did the creative divinity start the universe in a grand design and then simply sit back and watch it spin? Or does divinity tinker and interfere from time to time by suddenly introducing miracles?

Babbage ([3], pp. 32-49, 168-171) seems to take the first view; he regards miracles as preordained apparent shifts in natural law, but not real shifts – rather, parts of a more general law. His analog is to a computer program that gives the sequence of natural numbers for a thousand years and then suddenly shifts to another sequence in obedience to initial programming. Mill argues that the encompassing law-of-nature concept does not give a reasonable interpretation for miracles (Mill [77], pp. 473-478). Mill says that a miracle "must be produced by a direct volition, without the use of means; or at least, of any means which if simply repeated would produce it" ([77], pp. 224, 474).

Some thinkers seem to take both viewpoints at different places. An example is Maimonides; see Hartman ([78], pp. 157, 251) and Heller [79]. In Maimonides' ([80]) Guide of the Perplexed the general law viewpoint is taken on pages 480-484, but the more traditional interventionist position appears on pages 178, 498-499, and 528-529. For theologians who regard creation of the universe as timeless, or outside of time, the problem disappears, but fresh problems may then arise. For example, how can one reconcile a timeless universe and the specifically historical character of Christianity?


[76] A basic divergence in the literature on miracles corresponds to the age-old debate about free will and divine intervention. To put it crudely: Did the creative divinity start the universe in a grand design and then simply sit back and watch it spin? Or does divinity tinker and interfere from time to time by suddenly introducing miracles?

Babbage ([3], pp. 32-49, 168-171) seems to take the first view; he regards miracles as preordained apparent shifts in natural law, but not real shifts – rather, parts of a more general law. His analog is to a computer program that gives the sequence of natural numbers for a thousand years and then suddenly shifts to another sequence in obedience to initial programming. Mill argues that the encompassing law-of-nature concept does not give a reasonable interpretation for miracles (Mill [77], pp. 473-478). Mill says that a miracle "must be produced by a direct volition, without the use of means; or at least, of any means which if simply repeated would produce it" ([77], pp. 224, 474).

Some thinkers seem to take both viewpoints at different places. An example is Maimonides; see Hartman ([78], pp. 157, 251) and Heller [79]. In Maimonides' ([80]) Guide of the Perplexed the general law viewpoint is taken on pages 480-484, but the more traditional interventionist position appears on pages 178, 498-499, and 528-529. For theologians who regard creation of the universe as timeless, or outside of time, the problem disappears, but fresh problems may then arise. For example, how can one reconcile a timeless universe and the specifically historical character of Christianity?


Hume used the concept of probability in a rather rough way. For discussions of the state of probability at or before Hume see Burns [56], Van Leeuwen [82], and Peirce ([83], pp. 305-306). See also Sobel [84] and Owen [85] for relevant material, especially on Bayesian interpretations of Hume's statements. Patey [86] describes the development of probability concepts.


There is a sparse but extended literature presenting algebraic formulas for truth and testimony under various conditions. An excellent critical summary is by Zabell (in press) [89].


One theme of this essay has been a skepticism about testimony; indeed, statisticians – along with lawyers, historians, and others – by inclination and training listen with special caution to human testimony. Yet I must comment in passing that all of us take most testimony at face value most of the time. We could not otherwise carry on our lives.

Locke pointed out in 1690 the dangers in accepting opinions of others: "since there is much more falsehood and error among men, than truth and knowledge. . . men have reason to be Heathens in Japan, Mahometans in Turkey, Papists in Spain, Protestants in England, and Lutherans in Sweden" ([70], book 4, chap. 15, par. 6, p. 638). Yet Locke's editor, Alexander Campbell Fraser, writes in a footnote that we "cannot dispense with the authority of experts and, in many cases, dependence on men of higher intelligence and larger experience than our own is the most reasonable means we can use of the attainment of truth" ([70], p. 638).

However sensitive we are to the vagaries of testimony, we simply must accept most of it, at least provisionally. We have neither time nor energy to check more than a little of what we have
been told or read. That is certainly true of daily life. In scientific life, it is impossible for each of us to recapitulate more than a tiny fraction of other peoples' observations. Even in mathematics there is an analog: The most energetic and stringent mathematician cannot check and confirm more than a small fraction of statements and proofs in the mathematical literature. As Mill says, "Authority . . . is the evidence on which even the wisest receive all those truths of science, or facts in history or in life, of which they have not personally examined the proofs" ([77], p. 407). Thus profound, persistent skepticism would make life impossible. I find that paradoxical and puzzling.


[94] Hume in fact refers to multiple witnesses at several places in his essay (e.g., [52], pp. 112,116,121) but the references are brief and in passing. Swinburne ([60, 68]) picks up the multiple-witness point and wonders whether Hume has overstringent standards of evidence. Middleton's book was widely influential; it is discussed, for example, by Stephen ([13], vol. 1, especially pp. 222-230). I do not, however, find the question of independence addressed by Middleton.

[95] There were also restatements of Hume's argument, notably by Mill ([96], pp. 217-241; [77], pp. 470-489), who includes a bitter passage (pp. 236-237; 479-480) on the weakness of testimony for the biblical miracles. Venn ([87], chap. 17) took a dim view of probabilistic calculations about the veracity of joint testimony. A major reason for his skepticism was that "practically our main source of error and suspicion is the possible existence of some kind of collusion. Since we can seldom entirely get rid of the danger, and when it exists it can never be submitted to numerical calculation, it appears to me that combination of testimony . . . is yet more unfitted for consideration in Probability than even that of single testimony" ([87], p. 428).

Jean Jacques Rousseau earlier presented a skeptical approach to such calculations, presumably not in terms of explicit probability calculus. In "The Creed of a Savoyard Priest," he asked: "Who will venture to tell me how many eye-witnesses are required to make a miracle credible? What use are your miracles, performed in proof of your doctrine, if they themselves require so much proof?" ([97], p. 263).

A nineteenth-century eloquent supporter of the biblical miracles and their central importance for Christianity was Richard Trench. It is wryly ironic that his views about multiple witnesses should undercut the arguments put forward by other critics of Hume. Trench wrote that "however many they [the witnesses] may be, they are always but a few compared with the multitudes who attest a fact which excludes their fact, namely, the uninterrupted succession of a natural order in the world" ([98], p. 61). Trench's discernment is in any case questionable. He asserts, for example, that people are lighter when awake than when asleep, as first observed by Pliny and
verified by "every nurse that has carried a child" ([98], p. 232). Another curious Trench position is suspicion about variation among authoritative accounts of historical happenings, exactly the opposite view to that expressed by most defenders of the biblical miracles (that variation in accounts is natural, human, and increases credibility; [98], p. 233).


[105] There is a related literature about surprising degrees of agreement among the Fathers of the Church (see especially Daille [106]). Daille in effect discusses dependence among the views of the Fathers, and even draws analogies to contagion of illness. In the last chapter of his book, however, he seems to go the other way and argues that near unanimity of opinion strengthens the case (see [106], pp. 144-145, 170, 268-270, 349).


[108] Alleged circularity of argument is another basic question that arises in various related ways. Perhaps the most fundamental way is the validation of faith via miracles described in a Holy Book; the Book in turn is validated by the faith. As Matthew Tindal, an early English deist, wrote in 1730, "It's an odd jumble to prove the truth of a book by the truth of the doctrines it contains, and at the same time to conclude those doctrines to be true because contained in that book" (Stephen [13], p. 117). Yet, on the other side of the fence, one may find such boasts of circularity as the following: The New Testament "pretends to contain a divine revelation; and, in support of this high pretension, it gives a narrative of the very facts which. . . constitute the evidence that this pretension is founded in truth." That statement is by John Cook, professor of divinity at St. Mary's College, St. Andrews (109], pp. 351-352). Variants of Tindal's criticism come from the internal structure of the Bible. Frye ([110]) asks, "How do we know that the Gospel story is true? Because it confirms the prophecies of the Old Testament. But how do we know that the Old Testament prophecies are true? Because they are confirmed by the Gospel story. . . . The two testaments form a double mirror, each reflecting the other" [from the review by Kermode ([111], p. 32)].

A third circular argument is criticized by Mill. Mill says, as part of his recapitulation of Hume, that "the existence of God cannot possibly be proved by miracles, for unless a God is already recognized, the apparent miracle can always be accounted for on a more probable hypothesis. . . . Once admit a God, and the production by his direct volition of an effect. . . is no longer a purely arbitrary hypothesis. . . {and} must be reckoned with as a serious possibility" ([112], p. 477). In vulgar brevity, God implies miracles implies God.

Mill points this out again when he writes, "All, therefore, which Hume has made out. . . is that. . . no evidence can prove a miracle to anyone who did not previously believe the existence of a being or beings with supernatural power. . . . If we do not already believe in supernatural agencies, no miracle can prove to us their existence" ([112], vol. 7, chap. 25, sec. 2, p. 625).

Stephen makes a similar point in discussing Hume and Paley: "When Paley calmly says, if we believe in God, there is no difficulty in believing {in} miracles, Hume's answer is plain. If God is the cause of order, belief in him does not facilitate belief in miracles" ([13], vol. 1, chap. 6, sec. 34, p. 287; see also [13], chap. 8, p. 352).

Lewis, in his well-known book on miracles, makes a virtue of the circle, without mention of Stephen or Mill, when he writes "Theology says in effect, 'Admit God and with him the risk of a few miracles, and I in return will ratify your faith in uniformity as regards the overwhelming majority of events'" ([113], p. 106; p. 109 in 1947 ed.). Lewis' argument was not entirely new. The famous mathematical physicist G. G. Stokes had given a variant: "Admit the Existence of a God, of a personal God, and the possibility of a miracle follows at once" ([114], p. 24). For discussion see Mascall ([115], pp. 7-8, 180, 185). The quotation may remind us not to sneer at what appear obvious circularities of argument. Perhaps we are all forced willy-nilly into circularities that we dignify by calling them "coherences" or "interconsistencies." There is, however, this difference, that some of us are
willing to examine our own circularities, and some simply deny them without study.

Other discussions of circularity may be found. I cite as examples Lewis ([116], pp. 82-93), Chalmers ([29], p. 385), and Walker ([117], p. 162). The relevance of this subtopic is that it presents a methodological tangle different from those of dependence of testimony and reference set, yet alike in the frequency with which they are ignored.


The opprobrium statement seems to have appeared first in the second (1846) edition of Mill ([118]), although I do not find it in the 1848 American edition (see Mill [112], vol. 7, p. 538; vol. 8, app. F). Following the opprobrium statement, there is a brief, pointed discussion of oversimple probability ideas applied to tribunals. For more on Mill's distaste for formalism, including that of French mathematical-political thought, see Schabas ([124]). Eddington ([125], pp. 123-125) used the French legal example as the basis for his critique of casual independence.


Meier and Zabell ([130], especially pp. 501-502) give a fascinating story of 1867 expert testimony by two great American mathematicians, Benjamin and Charles S. Peirce (father and son), during a trial over a contested will. Apparently the independence assumption was used casually in a context where it could hardly hold.

Casual multiplication in a legal setting has a literary endorsement in Poe's "The Mystery of Marie Roget." Poe writes that "If the feet of Marie being small, those of the corpse were also small, the increase of probability that the body was that of Marie would not be . . . merely arithmetical, but . . . highly geometrical, or accumulative. Add to all this shoes such as she had been known to wear. . . . Give us, then, flowers in the hat corresponding to those worn by the missing girl, and we seek for nothing further. If only one flower, we seek for nothing further-what then if two or three, or more? Each successive one is multiple evidence-proof not added to proof, but multiplied by hundreds or thousands" ([131], pp. 285-286).


Poe, Edgar Allan (1904), The Mystery of Marie Roget, in *Works* (Vol. 1), New York: Collier, pp. 248-327. (First published in 1842.)

There is also a somewhat aloof mathematical viewpoint that simply avoids the verisimilitude questions emphasized here. Kac, for example, writes that "Statistical independence, once a shadowy partner of gamblers, experimental scientists and statisticians, has achieved the respectability that only an ancient discipline like number theory can bestow" ([134], p. 72).


Perhaps one reason for misunderstandings about independence is inadequate training in classrooms and lectures. In a modest probe, I looked at two issues of the Journal of the American Statistical Association and counted 11 reviews of introductory textbooks. I inspected the six of these books in our library and graded their treatments of independence tolerantly: no A's, one B, two C's, one D, and three F's, an unhappy record. Among the problems with treatment of independence in these books, I cite first difficulty in even finding the independence discussions: Indexes were absent or skimpy, and in some cases I gave up after a fruitless search of the index and table of contents together with rapid reading in likely chapters. When treatments of independence were given, they tended to be definitions plus a few simple combinatorial examples. With luck, there might be mention of independence for random variables and observations. In at most one of the books was there anything approaching responsible concern for what happens when there is dependence, how to recognize it, what analytic options there are, and so on.


An example of extraterrestrial calculation is the correspondence between Donaldson and Pollard ([142]). Lyttle ([143]) provides an example of the blithe assumption of independence to calculate the probability of accidental launch of a strategic missile. For a fine treatment of the dependence problem in estimating accident probabilities for liquified natural gas transportation see Fairley ([144], especially pp. 339-346). Related problems have arisen in examination of acoustic evidence about the assassination of John F. Kennedy (National Research Council [145],
p. 129). Portnoy and Petersen ([146]) criticize a study of biblical texts for casual assumptions of independence in statistical analyses. One could easily go on.


[155] An important exception is the position of De Finetti and his followers. {See De Finetti ([156]) for an intensely personal statement with key references.} The truly basic idea for De Finetti is not independence (plus identical distribution) but exchangeability, that is, invariance of the relevant joint distribution under permutations of variables. That option does not affect the
primary theme of this paper: to look critically at any simplifying assumption – independence, exchangeability, whatever – and to worry about consequences when the assumption fails.


[157] Berkson Joseph; Magath, Thomas B.; and Hurn, Margaret (1939), The Error of Estimate of the Blood Cell Count as Made With the Hemocytometer, American Journal of Physiology, 128, 309-323.

[158] Discussions of such a device for measuring blood pressure are given by Wright and Dore ([159]) and Labarthe, Hawkins, and Remington ([160], especially p. 553). In addition to decreasing or avoiding dependence among contiguous observations, devices of this kind may also decrease or avoid sources of bias and unwanted variation: integer preferences, observer expectations, and so on.


