

Chapter 1: Probability: Classical and Bayesian

Probability in mathematical statistics is classically defined in terms of the outcomes of conceptual experiments, such as tossing ideal coins and throwing ideal dice. In such experiments the probability of an event, such as tossing heads with a coin, is defined as its relative frequency in long-run trials. Since the long-run relative frequency of heads in tosses of a fair coin is one-half, we say that the probability of heads on a single toss is one-half. Or, to take a more complicated example, if we tossed a coin 50 times and repeated the series many times, we would tend to see 30 or more heads in 50 tosses only about 10% of the time; so we say that the probability of such a result is one-tenth. We refer to this relative frequency interpretation as classical probability. Calculations of classical probability generally are made assuming the underlying conditions by which the experiment is conducted, in the above examples with a fair coin and fair tosses.

This is not to say that the ratio of heads in a reasonably large number of tosses invariably equals the probability of heads on a single toss. Contrary to what some people think, a run of heads does not make tails more likely to balance out the results. Nature is not so obliging. All she gives us is a fuzziest determinism, which we call the law of large numbers. It was originally formulated by Jacob Bernoulli (1654-1705), the “bilious and melancholy” elder brother of the famous Bernoulli clan of Swiss mathematicians, who was the first to publish mathematical formulas for computing the probabilities of outcomes in trials like coin tosses. The law of large numbers is a formal statement, proved mathematically, of the vague notion that, as Bernoulli biliously put it, “Even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is in wandering from one’s goal.”¹

To understand the formal content of the commonplace intuition, think of the difference between the ratio of successes in a series of trials and the probability of success on a single trial

¹ Jacob Bernoulli, *Ars Conjectandi* [The Art of Conjecturing], 225 (1713), quoted in Stephen M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900*, at 65 (1986).

as the error of estimating the probability from the series. Bernoulli proved that the probability that the error exceeds any given arbitrary amount can be made as small as one chooses by increasing sufficiently the number of trials. This result represented a fateful first step in the process of measuring the uncertainty of what has been learned from nature by observation. Its message is obvious: the more data the better.

What has classical probability to do with the law? The concept of probability as relative frequency is the one used by most experts who testify to scientific matters in judicial proceedings. When a scientific expert witness testifies that in a study of smokers and non-smokers the rate of colon cancer among smokers is higher than the rate among non-smokers and that the difference is statistically significant at the 5% level, he is making a statement about long-range relative frequency. What he means is that *if* smoking did not cause colon cancer and *if* repeated samples of smokers and non-smokers were drawn from the population to test that hypothesis, a difference in colon cancer rates at least as large as that observed would appear less than 5% of the time. The concept of statistical significance, which plays a fundamental role in science, thus rests on probability as relative frequency in repeated sampling.

Notice that the expert in the above example is addressing the probability of the data (rates of colon cancer in smokers and non-smokers) given an hypothesis about the cause of cancer (smoking does not cause colon cancer). However, in most legal settings, the ultimate issue is the inverse conditional of that, i.e., the probability of the cause (smoking does not cause colon cancer) given the data. Probabilities of causes given data are called *inverse probabilities* and in general are not the same as probabilities of data given causes. In an example attributed to Keynes, if the Archbishop of Canterbury were playing poker, the probability that the Archbishop would deal himself a straight flush given honest play on his part is not the same as the probability of honest play on his part given that he has dealt himself a straight flush. The first is 36 in 2,598,960; the second most people would put at close to 1 (he is, after all, an archbishop).

One might object that since plaintiff has the burden of proof in a law suit, the question in the legal setting is not whether smoking does not cause cancer, but whether it does. This is true, but does not affect the point being made here. The probability that, given the data, smoking causes colon cancer is equal to one minus the probability that it doesn't, and neither will in

general be equal to the probability of the data, assuming that smoking doesn't cause colon cancer. Or to vary our earlier example, the probability that the Archbishop was dishonest when he dealt himself a straight flush is not equal to the probability that if he were honest he would deal himself a straight flush.

The inverse mode of probabilistic reasoning is usually traced to Thomas Bayes, an English Nonconformist minister from Tunbridge Wells, who was also an amateur mathematician. When Bayes died in 1761 he left his papers to another minister, Richard Price. Although Bayes evidently did not know Price very well there was a good reason for the bequest: Price was a prominent writer on mathematical subjects and Bayes had a mathematical insight to deliver to posterity that he had withheld during his lifetime.

Among Bayes's papers Price found a curious and difficult essay that he later entitled, "Toward solving a problem in the doctrine of chances." The problem the essay addressed was succinctly stated: "Given the number of times in which an unknown event has happened and [has] failed: Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named." Price added to the essay, read it to the Royal Society of London in 1763, and published it in *Philosophical Transactions* in 1764. Despite this exposure and the independent exploration of inverse probability by Laplace in 1773, for over a century Bayes's essay remained obscure. In fact it was not until the twentieth century that the epochal nature of his work was widely recognized. Today, Bayes is seen as the father of a controversial branch of modern statistics eponymously known as Bayesian inference and the probabilities of causes he described are called Bayesian or inverse probabilities.

Legal probabilities are mostly Bayesian (i.e., inverse). The more-likely-than-not standard of probability for civil cases and beyond-a-reasonable-doubt standard for criminal cases import Bayesian probabilities because they express the probabilities of past events given the evidence, rather than the probabilities of the evidence, given past events. Similarly, the definition of "relevant evidence" in Rule 401 of the Federal Rules of Evidence is "evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence." This definition imports Bayesian probability because it assumes that relevant facts have probabilities attached to

them. By contrast, the traditional scientific definition of relevant evidence, using classical probability, would be any “evidence that is more likely to appear if any fact of consequence to the determination of the action existed than if it didn’t.”

The fact that classical and Bayesian probabilities are different has caused some confusion in the law. For example, in an old case, *People v. Risley*,² a lawyer was accused of removing a document from the court file and inserting a typed phrase that helped his case. Eleven defects in the typewritten letters of the phrase were similar to those produced by defendant’s machine. The prosecution called a professor of mathematics to testify to the chances of a random typewriter producing the defects found in the added words. The expert assumed that each defect had a certain probability of appearing and multiplied these probabilities together to come up with a probability of one in four billion, which he described as “the probability of these defects being reproduced by the work of a typewriting machine, other than the machine of the defendant....” The lawyer was convicted. On appeal, the New York Court of Appeals reversed, expressing the view that probabilistic evidence relates only to future events, not the past. “The fact to be established in this case was not the probability of a future event, but whether an occurrence asserted by the People to have happened had actually taken place.”³

There are two problems with this objection. First, the expert did not compute the probability that defendant’s machine did not type the insert, the occurrence asserted by the People to have taken place. Although his statement is somewhat ambiguous, he could reasonably be understood to refer to the probability that there would have been matching defects *if* another machine had been used. Second, even if the expert had computed the probability that the insert had been typed on defendant’s machine, the law, as we have seen, *does* treat past events as having probabilities.⁴ If probabilities of past events are properly used to define the certainty needed for the final verdict, there would seem to be no reason why they are not properly used for

² 214 N.Y. 75 (1915).

³ *Id.* at 85, 108 N.E. at 203.

⁴ To be sure, the court’s rejection of the testimony was correct because there were other, valid objections to it. See p. 4 -11 *infra*.

subsidiary issues leading up to the final verdict. As we shall see, the objection is not to such probabilities *per se*, but to the expert's competence to calculate them.

A similar confusion arose in a notorious case in Atlanta, Georgia. After a series of murders of young black males, one Wayne Williams was arrested and charged with two of the murders. Critical evidence against him included certain unusual trilobal fibers found on the bodies. These fibers matched those in a carpet in Williams' home. A prosecution expert testified that he estimated that only 82 out of 638,992 occupied homes in Atlanta, or about 1 in 8000, had carpeting with that fiber. This type of statistic has been called "population frequency" evidence. Based on this testimony, the prosecutor argued in summation that "there would be only one chance in eight thousand that there would be another house in Atlanta that would have the same kind of carpeting as the Williams home." On appeal, the Georgia Court of Appeals rejected a challenge to this argument, holding that the prosecution was not precluded from "suggesting inferences to be drawn from the probabilistic evidence."

Taken literally, the prosecutor's statement is nonsense because his own expert derived the frequency of such carpets by estimating that 82 Atlanta homes had them. To give the prosecutor the benefit of the doubt, he probably meant that there was 1 chance in 8,000 that the fibers came from a home other than the defendant's. The 1-in-8,000 figure, however, is not that, but the probability of the particular kind of fiber, given that it came from an Atlanta home picked at random.

Mistakes of this sort are known as the fallacy of the inverted conditional. That they should occur is not surprising. It is not obvious how classical probability based, for example, on population frequency evidence, bears on the probability of defendant's criminal or civil responsibility, whereas inverse Bayesian probability purports to address the issue directly. In classical terms we are given the probability of seeing the incriminating trace if defendant did not leave it, but what we really want to know is the probability that he did leave it. Or to revert to our expert on smoking and cancer, he testifies to the probability of observing the study data given that smoking does not cause colon cancer, when we are after the probability that smoking *does* cause colon cancer. In a litigation, the temptation to restate things in Bayesian terms is very strong. The Minnesota Supreme Court was so impressed by the risk of this kind of mistake by

jurors that it ruled out population frequency evidence, even when correctly stated.⁵ The court apprehended a “real danger that the jury will use the evidence as a measure of the defendant’s guilt or innocence.”⁶ The court evidently feared that if, for example, the population frequency of a certain incriminating trace is 1 in a 1,000 the jury might interpret this figure as meaning that the probability of defendant’s innocence was 1 in a 1,000. And, as we have seen, it is not only jurors who can make such mistakes. This particular misinterpretation, which arises from inverting the conditional, is sometimes called the prosecutor’s fallacy.⁷

Is the prosecutor’s fallacy in fact prejudicial to the accused? Recent studies with simulated cases before juries of law students show higher rates of conviction when prosecutors are allowed to misinterpret population frequency statistics as probabilities of guilt than when the correct statement is made. The effect was most pronounced when population frequencies were as high as one in a thousand, but some effect also appeared for frequencies as low as one in a billion. The correctness of the interpretation does seem to matter.

The defendant also has his fallacy, albeit of a different sort. This is the argument that the evidence does no more than put defendant in a group consisting of all those who have the trace in question, so that the probability that defendant left the trace is only one over the size of the group. If this were correct, then only a show of uniqueness (which is perhaps possible for DNA evidence, but all but impossible as a general matter) would permit us to identify a defendant from a trace. This is not a fallacy of the inverted conditional, but is fallacious because, as we shall see, it ignores the other evidence in the case.

To examine the prosecutor’s and defendant’s fallacies a little more closely we ask a more

⁵ State v. Kim, 398 N.W. 2d 544 (1987); State v. Boyd, 331 N.W. 480 (1983).

⁶ State v. Boyd, 331 N.W. 2d 480, 483 (1983).

⁷ The court’s opinion seems unduly restrictive of valuable evidence and its rulings were overturned by statute with respect to genetic or blood test results. The statute provides: “In a civil or criminal trial or hearing, statistical population frequency evidence, based on genetic or blood test results, is admissible to demonstrate the fraction of the population that would have the same combination of genetic markers as was found in a specific human biological specimen. ‘Genetic marker’ means the various blood types or DNA types that an individual may possess.” Minn. Stat. Sec. 634.26 (1992).

general question: If it is wrong to interpret the probabilities of evidence given assumed causes as probabilities of causes given assumed evidence, what is the relation between the two probabilities? Specifically, what is the probative significance of scientific probabilities of the kind generated by statistical evidence to the probabilities of causes implied by legal standards? The answer is given by what is now called Bayes's theorem, which Bayes derived for a special case using a conceptual model involving billiard balls. We do not give his answer here. Instead, to explain what his result implies for law, we use a doctored example from a law professor's chestnut: the case of the unidentified bus.⁸ It is at once more general and mathematically more tractable than the problem that Bayes addressed.

The facts are simply stated. On a rainy night, a driver is forced into collision with a parked car by an unidentified bus. Of the two companies that run buses on the street, Company A owns 85% of the buses and Company B owns 15%. Which company was responsible? That icon of the law, an eyewitness, testifies that it was a Company B bus. A psychologist testifies without dispute that eyewitnesses in such circumstances tend to be no more than 80% accurate. Our problem is to find the probabilities associated with the cause of the accident (Company A or B) given the case-specific evidence (the eyewitness report) and the background evidence (the market shares of Companies A and B). To be specific, let us ask for the probability that it was a Company B bus, assuming that the guilty bus had to belong to either Company A or Company B.

The Bayesian result of relevance to this problem is most simply stated in terms of odds:⁹ the posterior odds that it was a Company B bus are equal to the prior odds that it was a Company B bus times the likelihood ratio that it was such a bus. Thus,

$$\textit{posterior odds} = \textit{prior odds} \times \textit{likelihood ratio} .$$

In this formula the *posterior odds* are the odds that the cause of the accident was a

⁸ The chestnut is from *Smith v. Rapid Transit, Inc.*, 307 Mass. 469, 58 N.E.2d 754 (1945). The doctoring is from D. Kahneman, A. Tversky & P. Slovic (eds.), *Judgment Under Uncertainty: Heuristics and Biases* 1 (1982).

⁹ The odds on an event are defined as the probability that the event will occur, p , divided by the probability that it will not, $1-p$. Conversely, the probability of an event is equal to the odds on the event divided by one plus the odds.

Company B bus, given (or posterior to) the background and case-specific evidence. These are Bayesian odds. The *prior odds* are the odds that the cause of the accident was a Company B bus prior to considering the case-specific evidence. These are also Bayesian. If one assumes in our bus problem that, in the absence of other evidence bearing on routes and schedules, the prior probabilities are proportional to the sizes of the respective bus fleets, the probability that it was a Company A bus is 0.85 and a Company B bus is 0.15. Hence the prior odds that the bus was from Company A are $0.85/0.15 = 5.67$ and from Company B are $0.15/0.85 = 0.1765$. Since probabilities greater than 50% and odds greater than 1 meet the more-likely-than-not standard for civil cases, plaintiff should have enough for a verdict against Company A, *if* we allow the sufficiency of statistical evidence. That is a big “if.” We return to this point later.

To solve the problem that he set for himself, Bayes made the restrictive assumption that the prior probabilities for the causes were equal. Prior probability distributions that assign equal or nearly equal probabilities to the possible causes are now known as diffuse or flat priors because they do not favor one possibility over another. The prior in our example is informative and not diffuse or flat because it assigns much greater probability to one possible cause than to the other.

The third element in Bayes’s theorem is the *likelihood ratio* for the event given the evidence. The likelihood ratio for an event is defined as the probability of the evidence if the event occurred divided by the probability of the evidence if the event did not occur. These are classical probabilities because they are probabilities of data given causes. In our bus example, the likelihood ratio that it was a Company B bus given the eyewitness identification is the probability of the witness reporting that it was a Company B bus if in fact it was, divided by the probability of such a report if it were not. On the facts stated, the numerator is 0.80, since the witness is 80% likely to report a Company B bus if it was such a bus. The denominator is 0.20, since 20% of the time the witness would mistakenly report a Company B bus when it was a Company A bus. The ratio of the two is $0.80/0.20 = 4$. We are four times as likely to receive a report that it was a Company B bus if it was in fact such a bus than if it was not.

The likelihood ratio is an important statistical measure of the weight of evidence. It is intuitively reasonable. The bloody knife found in the suspect’s home is potent evidence because

we think we were far more likely to find such evidence if the suspect committed the crime than if he didn't. In general, large values of the likelihood ratio imply that the evidence is strong; small values the opposite; a ratio of 1 means that the evidence has no probative value.

Putting together the prior odds and the likelihood ratio, the posterior odds that it was a Company B bus given the evidence are $0.1765 \times 4.00 = 0.706$. The probability that it was a Company B bus is $0.706/(1 + 0.706) = 0.4138$. Thus, despite eyewitness identification, the probability of a Company B bus is less than 50%. If there were a second eyewitness with the same testimony and the same accuracy, the posterior odds with respect to the first witness could be used as the prior odds with respect to the second witness and Bayes's theorem applied again. In that case the new posterior odds would be $0.706 \times 4.00 = 2.824$ and the new posterior probability would be $2.824/3.824 = 0.739$.

Some people object to these results. They argue that if the eyewitness is right 80% of the time and she says it was a Company B bus, why isn't the probability 80% that it was a Company B bus? Yet we find that, despite the eyewitness, the preponderance of probability is against the witness's testimony. The matter is perhaps even more counter-intuitive when there are two eyewitnesses. Most people would think that two eyewitnesses establish a proposition beyond a reasonable doubt, yet we conclude that the probability of their being correct is only about 74%, even when there is no contrary testimony. Surely Bayes's theorem is off the mark here.

There are two things wrong with this argument. The first is that it confuses two conditional probabilities: the probability that the witness would so testify conditional on the fact that it was a Company B bus (which is indeed 80%) and the probability that it was a Company B bus conditional on the fact that the witness has so testified (which is not necessarily 80%; remember the Archbishop playing poker).

The second and related objection is that the 80% figure ignores the effect of the statistical background, i.e., the fact that there are many more Company A buses than Company B buses. For every 100 buses that come along only 15 will be Company B buses but 17 (0.20×85) will be Company A buses that are wrongly identified by the first witness. Because of the fact that there are many more Company A buses, the witness has a greater chance of wrongly identifying a Company A bus as a Company B bus than of correctly identifying a Company B bus. The

probabilities generated by Bayes's theorem reflect that fact. In this context its application corrects for the tendency to undervalue the evidentiary force of the statistical background in appraising the case-specific evidence.

The correction for statistical background supplied by Bayes's theorem becomes increasingly important when the events recorded in the background are rare. In that case even highly accurate particular evidence may become surprisingly inaccurate. Screening devices are of this character. For example, the Federal Aviation Administration is said to use a statistically based hijacker profile program to help identify persons who might attempt to hijack a plane using a nonmetallic weapon. The accuracy of such screening tests is usually measured by their sensitivity and specificity. *Sensitivity* is the probability that the test will register a positive result if the person has the condition for which the test is given. *Specificity* is the probability the test will be negative if the person doesn't have the condition. Assume that the test has a sensitivity of 90% (i.e., 90% of all hijackers are detected) and a specificity of 99.95% (i.e., 99.95% of all non-hijackers are correctly identified). This seems and is very accurate. But if the rate of hijackers is 1 in 25,000 passengers, Bayes's theorem tells us that this seemingly accurate instrument makes many false accusations.

The odds that a passenger being identified as a hijacker by the test is actually a hijacker are equal to the prior odds of a person being a hijacker times the likelihood ratio associated with the test. Our assumption is that the prior odds that a passenger is a hijacker are $(1/25,000) / (24,999/25,000) = 1/24,999$. The likelihood ratio for the test is the probability of a positive identification if the person is a hijacker (0.90) divided by the probability of such an identification if the person is not $(1 - 0.9995 = 0.0005)$. The ratio is thus $0.90/0.0005 = 1,800$. The test is powerful evidence, but hijackers are so rare that it becomes quite inaccurate. The posterior odds of a correct identification are only $(1/24,999) \times 1,800 = 0.072$. The posterior probability of a correct identification is only $0.072/1.072 = 0.067$; there is only a 6.7% chance that a person identified as a hijacker by this accurate test is really a hijacker. This result has an obvious bearing on whether the test affords either probable cause to justify an arrest or even reasonable

suspicion to justify a brief investigative detention.¹⁰

The failure to distinguish posterior odds and likelihood ratios has led to some confusion in the Supreme Court's jurisprudence on the constitutionality of stopping passengers to search for drugs based on a profile.¹¹ In particular, debates have swirled around the question whether a profile can be a basis for "reasonable suspicion" that the person is a drug courier, which would justify a brief investigative detention, a lower standard than "probable cause" needed for an arrest. Since drug couriers are rare among passengers even profiles with impressive operating characteristics are likely to generate small posterior odds that the person is a courier. For example, imagine a profile which is quite accurate: 70% of drug couriers would fit the profile, but only 1 normal passenger in a 1,000 would do so. The likelihood ratio for the profile is $0.70/.001 = 700$, which means that the odds that a person is a courier are increased 700-fold by matching the profile. But if the rate of couriers is 1 in 10,000 passengers, the odds on the person matching the profile is a courier are only $(1/9,999) \times 700 = 0.07$, or a probability of 6.5%. This result is not fanciful: A study by the U.S. Customs Service found that the hit rate for such stops was about 4%.

These facts lead us to the legal issue that the Supreme Court has not faced in these cases: whether a profile that significantly increases the probability that the person is a criminal is a sufficient basis for reasonable suspicion, even if that increased probability remains low because couriers are so rare. By greatly increasing the probability that a person who fits the profile is a courier, a valid profile clearly can provide a rational basis for investigation. If "reasonable suspicion" requires no more than a "rational basis" for suspicion, then a profile could be sufficient. On the other hand, if "reasonable suspicion" requires a reasonable probability of crime, a profile is unlikely to be sufficient. Which standard applies is presently unclear.

We have been considering the effects of the likelihood ratios of screening tests on their predictive values. We now turn things around and consider the implications of predictive values

¹⁰ *Cf.* *United States v. Lopez*, 328 F. Supp. 1077 (E.D.N.Y. 1971).

¹¹ For a discussion of the Supreme Court cases, *see* Michael O. Finkelstein & Bruce Levin, *On the Probative Value of Evidence from a Screening Search*, 43 *Jurimetrics J.* 265, 270-276 (2003).

for likelihood ratios. In *United States v. Scheffer*,¹² the Supreme Court considered a challenge to the constitutionality of Military Rule of Evidence 707, which made polygraph evidence inadmissible in military courts-martial, because it denied the accused the right to introduce an exonerating test. Justice Thomas, writing for the majority, upheld the rule, arguing that the reliability of the polygraph was uncertain. In support of that, Thomas pointed out that views of polygraph accuracy ranged from 87% to little better than the coin toss (50%). However, in taking the position that polygraphs were not sufficiently reliable to be admitted in evidence, the government had to deal with the interesting fact that the Defense Department itself used polygraph tests to screen employees in a counterintelligence program. If good enough for counterintelligence, why not for court?

The negative predictive values of the test were very strong in the counterintelligence context: if an employee passed the test the probabilities were very high that he was not a security risk meriting further investigation. But this high value is probably not applicable to polygraphs given in criminal cases because it depends on the operating characteristics of the test *and* on the prior odds of guilt; and the prior odds on guilt would probably be much larger in a criminal case than in a screening program. Nevertheless, admissibility should not turn on the test's predictive values because the strength of other evidence of guilt is no justification for keeping out strong exonerating test results. Probative strength for these purposes has to be measured by the likelihood ratio associated with a negative test result, which measures how much the predictive values change with the results. Information about the likelihood ratio can be gleaned from the results of screening, assuming that the likelihood ratio remains the same across contexts. (This is a conservative assumption because it is probable that test is more accurate in the criminal context than in the more amorphous arena of security screening.)

In 1997, about the time of *Scheffer*, the Defense Department used the polygraph to screen 7,440 employees; the polygraph identified 176 employees as possible security risks, to be further investigated, and passed 7,418 employees. Of the 176 employees at the time the data were reported to Congress 6 had lost clearance and the cases of 16 were pending. Assume for the

¹² 523 U.S. 303 (1998).

moment that all pending cases were resolved unfavorably to the individuals, so there would be 22 security risks among the 176 identified by the polygraph screen. How many security risks were not identified by the polygraph? If the polygraph has only a 50% chance of identifying a security risk (Justice Thomas's coin toss) there would be another 22 risks that were not so identified; the negative predictive value of the test would be $7418/7440 = 0.997$, and the negative predictive odds would be $0.997/0.003 = 332$. Since there are assumed to be 44 security risks out of 7616 employees, the prior probability that an employee is not a security risk is $7572/7616 = 0.994$, and the prior negative odds are $0.994/0.006 = 166$. Using Bayes's theorem, the likelihood ratio associated with a negative polygraph test is therefore about $332/166 = 2.0$. If the polygraph were 87% accurate, a similar calculation shows that the likelihood ratio for an exonerating test would rise to about 8. Thus the odds on innocence are increased between two and eight times by an exonerating polygraph.¹³ An incriminating polygraph has low predictive power in the counterintelligence arena—between 3.4% and 12.5%—but an even stronger likelihood ratio than the exonerating polygraph. The odds on guilt are increased by between 25 and 43 times for a defendant who fails a polygraph test. Either way, the polygraph qualifies as highly relevant evidence, and it illustrates that a test used for screening may supply important evidence for criminal cases.

Classical statisticians have two basic objections to Bayesian analysis. The first is philosophical. They point out that for events above the atomic level, which is the arena for legal disputes, the state of nature is not probabilistic; only the inferences to be drawn from our evidence are uncertain. From a classical statistician's point of view, Bayesians misplace the uncertainty by treating the state of nature itself, rather than our measurement of it, as having a probability distribution. This was the point of view of the *Risley* court. However, the classicists are not completely consistent in this and sometimes compute Bayesian probabilities when there are data on which to base prior probabilities.

The second objection is more practical. In most real-life situations, prior probabilities

¹³ These results are not much changed if it is assumed that none of the pending cases resulted in adverse action against individuals. In that scenario, the likelihood ratio is between 1.95 and 6.8, respectively.

that are the starting point for Bayesian calculations cannot be based on objective quantitative data of the type at least theoretically available in our example, but can only reflect the strength of belief in the proposition asserted. Such probabilities are called subjective or personal probabilities. They are defined not in terms of relative frequency, but as the odds we would require to bet on the proposition. While the late Professor Savage has shown that subjective or personal probabilities satisfying a few seemingly fundamental axioms can be manipulated with the same methods of calculation used for probabilities associated with idealized coins or dice, and thus are validly used as starting points in Bayesian calculations, there is a sharp dispute in the statistical community over the acceptability of numerical measures of persuasion in scientific calculations.¹⁴ After all, subjective prior probabilities may vary from person to person without a rational basis. No other branch of science depends for its calculations on such overtly personal and subjective determinants.

To these objections the Bayesians reply that since we have to make Bayesian judgments in any event, at some point we must introduce the very subjectivity that we reject in the classical theory. This answer seems particularly apt in law. We require proof that makes us believe in the existence of past events to certain levels of probability; this attachment of probabilities to competing states of nature is fundamentally Bayesian in its conception. And since we have to make these judgments in any event, the Bayesians argue that it is better to bring them within the formal theory because we can correct for biases that are far more important than the variability of subjective judgments with which we started.

To make this point specific, let me report on my personal experiments with the theory. I give each class of my law students the following hypothetical and ask them to report their personal probabilities.¹⁵ A woman is found in a ditch in an urban area. She has been stabbed with a knife found at the scene. The chief suspect is her boy friend. They were together the day before and were known to have quarreled. He has been violent on other occasions. Based on

¹⁴ Whether people conform their personal probabilities to the assumptions has also been questioned. [INSERT ELLSBERG].

¹⁵ Taken from Michael O. Finkelstein & William Fairley, "A Bayesian Approach to Identification Evidence," 83 Harv. L. Rev. 489 (1970).

these facts, I ask the students for their estimate of the probability that he killed her. For the most part the students give me estimates between 25% and 75%. Now I tell them that there is a partial palm print on the knife, so configured that it is clear that it was left there by someone who used to stab rather than cut. The palm print matches the boy friend's palm, but it also appears in one person in a thousand in the general population. I ask again: what is the probability that he killed her? Usually (not always!) The estimate goes up, but generally not over 90%.

Bayes's theorem teaches us that the adjustments made by the students are too small. For the person whose prior was 0.25, his or her posterior probability should be 0.997.¹⁶ The upper range of the range, At On the other hand, if there is no evidence of responsibility apart from the trace, the accused is no more likely to have left it than anyone else in the relevant population who shared the trace. If N is the size of the relevant population (excluding the accused) then there are $N/1000$ people in the population who share the trace, plus the accused. The probability the

accused left the print is $\frac{1}{(N/1,000) + 1} = \frac{1,000}{N + 1,000}$. If we apply Bayes's theorem,

assuming there is no other evidence against the accused other than the trace, the prior probability of guilt is $1/(N + 1)$, and the prior odds on guilt are simply $1/N$. Since the likelihood ratio is still 1,000, the posterior odds are $1,000/N$, and the posterior probability

is $\frac{1,000/N}{(1,000/N) + 1} = \frac{1,000}{N + 1,000}$, which is as the accused argued. So while the prosecutor's

fallacy assumes a 50% prior probability that the accused left the print, the defendant's fallacy assumes a prior probability no greater than that of a randomly selected person.

Thus the admissibility and sufficiency of population frequency data may, as a theoretical matter, turn on the strength of the other evidence in the case. One must say as a theoretical

¹⁶ For such a person the prior odds are $0.25/0.75 = 1/3$. The likelihood ratio is 1 (we are certain that the palm print would be like the defendant's if he left it) divided by $1/1000$ (the rate of such palm prints in the population if he didn't leave it, or 1,000). Applying Bayes's theorem, $1/3 \times 1,000 = 333.33$ (the posterior odds). The posterior probability of guilt is thus $333.33/334.33 = 0.997$.

matter because as a practical matter it is hard to imagine a real criminal case in which the other evidence would not suffice to permit the jury also to hear the statistical evidence. The prosecutor's fallacy is likely to be closer to the truth than the defendant's fallacy.

In civil cases, the story is more complicated.

In some early cases in which there were no data, a few judges muddied the waters by sounding the theme that quantitative probability is not evidence. For example, in *Day v. Boston & Maine R.R.*,¹⁷ a case involving the need to determine the cause of death in a railway accident, but in which there was really no evidence of how the accident had occurred, Judge Emery gratuitously commented that mathematical probability was akin to pure speculation:

Quantitative probability – is only the greater chance. It is not proof, nor even probative evidence of the proposition to be proved. That in one throw of the dice there is a quantitative probability, or greater chance, that a less number of spots than sixes will fall uppermost [the probability is 35/36] is no evidence whatever that in a given throw such was the actual result . . . The slightest real evidence that sixes did in fact fall uppermost would outweigh all the probability otherwise.¹⁸

This theme was picked up in *Sargent v. Massachusetts Accident Company*,¹⁹ in which the court had to decide whether the deceased, canoeing in the wilderness, met his end by accident (which would have been within the insurance policy) or by other causes such as starvation (which would not). Of course, there were no data. Citing *Day*, Justice Lummus elaborated, in much-quoted language, the theme of mathematical probability as insufficient proof:

It has been held not enough that mathematically the chances somewhat favor a proposition to be proved; for example, the fact that colored automobiles made in the

¹⁷ 96 Me. 207, 52 A. 771 (1902).

¹⁸ *Id.* at 217-218, 52 A. at 774.

¹⁹ 307 Mass. 246, 29 N.E.2d 825 (1940)

current year out-number black ones would not warrant a finding that an un-described automobile of the current year is colored and not black, nor would the fact that only a minority of men die of cancer warrant a finding that a particular man did not die of cancer. The weight or preponderance of the evidence is its power to convince the tribunal which has the determination of the fact of the actual truth of the proposition to be proved. After the evidence has been weighed, that proposition is proved by a preponderance of the evidence if it is made to appear more likely or probable in the sense that actual belief in its truth, derived from the evidence, exists in the mind or minds of the tribunal notwithstanding any doubts that may linger there.²⁰

It was this language that was cited by the court in *Smith v. Rapid Transit, Inc.*, a case in which there were data.

Since the *Sargent* dictum cited in *Smith* originated in cases in which there were case-specific facts that were only suggestive, but not compelling, the notion that probability is not evidence seems to spring from the false idea that the uncertainty associated with mathematical probability is no better than ungrounded speculation to fill in gaps in proof. There is of course a world of difference between the two. As Bayes's theorem shows us, the former, when supported, would justify adjusting our view of the probative force of particular evidence, while the latter would not. Nor is it correct, as we have seen, that the slightest real evidence (presumably an eyewitness would qualify) should outweigh all the probability otherwise.

Yet the view that "bare" statistics are insufficient persists.²¹ In a recent case, *Krim v. pcOrder.com, Inc.*,²² the issue was whether the plaintiffs had standing to sue for fraud in the sale of securities under the federal securities laws. Since the Securities Act of 1933 was designed to

²⁰ 307 Mass. at 250-251, 29 N.E.2d at 827.

²¹ See, e.g., *Guenther v. Armstrong Rubber Co.*, 406 F.2d 1315 (3rd Cir. 1969) (holding that although 75% to 80% of tires marketed by Sears were made by defendant manufacturer, plaintiff would have lost on a directed verdict even if he had been injured by a tire bought at Sears).

²² 2005 U.S. App. LEXIS 3471 (5th Cir. 2005).

protect only purchasers in public offerings, the court held that at least one share had to be traced to those issued in two public offerings by the company for the plaintiffs to have standing to sue. There were approximately 2.5 million shares issued in the public offerings and registered in street name, Cede & Co., a nominee of the Depository Trust Company. When two of the plaintiffs bought their shares from this pool there were intermingled with it shares held by insiders that were also offered for sale. Despite the fact that the pool of shares consisted of 99.85% public offering shares when one plaintiff bought 3000 shares and 91% public offering shares when another bought his shares, there was no way of determining whether in fact they had bought the public offering shares or the insider shares. Assuming random selection from the pool, plaintiffs' expert computed the probability that at least one share would have been from the public offering segment of the pool and the court did not dispute that this probability was near 100%. The court acknowledged that the burden of proof of standing was by a preponderance of the evidence and that all evidence is probabilistic, but nevertheless ruled that probabilities were not enough, citing *Sargent and Smith*. In particular, the court concluded that allowing the probabilistic argument would contravene the statute by extending it to all purchasers of shares from the pool, since all would have the same argument. It also rejected the "fungible mass" argument by which each purchaser would be deemed to have purchased his proportionate share of public offering shares; allowing that argument, the court held, would undermine the requirement that plaintiff trace his shares to the public offering. When the statute was passed, the practice of holding shares in street name was not widespread and shares could in many cases be traced. Since that was no longer true, the remedy, the court concluded, was to amend the statute.

Most commentators think that *Smith* was rightly decided. Their reasons vary. For example, Judge Richard Posner, discussing *Smith*, argues that plaintiff needs an incentive to ferret out case-specific evidence, in addition to the statistics, so he must lose unless he shows that it was infeasible for him to get more particular evidence.²³ Unless he makes that showing, his

²³ *Howard v. Wal-Mart Stores, Inc.*, 160 F.3d 358, 359-360 (7th Cir. 1998). See also Richard A. Posner, "An Economic Approach to the Law of Evidence," 51 *Stan. L. Rev.* 1477, 1508-1509 (1999). For a rare dissent, see Daniel Shaviro, "Statistical Probability Evidence and

failure to produce more evidence must be due either to the fact that the other evidence was adverse or that he failed to make an investigation—and in either case he should lose. The *Krim* court cited Judge Posner’s decision without noting the distinction that in *Krim* there was no other evidence possible of which source the shares came from, and so the decision should fall within Judge Posner’s caveat holding statistics sufficient in such cases. Also, the hypothetical Judge Posner discussed was a bare statistical preponderance of 51%, far from the overwhelming probability in *Kim*. Another objection is that if statistics were sufficient, Company B (the Company with fewer buses) would enjoy immunity from suit, all the errors being made against Company A, a significant economic advantage. But this assumes there are many such cases. If accidents involving unidentified buses become numerous enough to be a statistical phenomenon, the solution is an application of enterprise liability, the ultimate statistical justice, in which each company bears its proportionate share of liability in each case.

Those who find “bare” statistical evidence intrinsically insufficient must explain why it somehow should become magically sufficient if there is even a smidgeon of case-specific evidence to support it.

Fortunately the utterances in the early cases are only dicta that have not been uniformly solidified into bad rules of evidence or law. The actual rules are much sounder from a Bayesian point of view. The learned Justice Lummus may disparage statistical evidence by writing that the fact that only a minority of men die of cancer would not warrant a finding that a particular man did not die of cancer, and yet in an earlier case appraise the force of statistical proof by soundly holding that “the fact that a great majority of men are sane, and the probability that any particular man is sane, may be deemed by the jury to outweigh, in evidential value, testimony that he is insane. . . [I]t is not the presumption of sanity that may be weighed as evidence, but rather the rational probability on which the presumption rests.”²⁴

When the event is unusual, but not so rare as to justify a presumption against it, we may

the Appearance of Justice,” 103 Harv. L. Rev. 530 (1989)(criticizing arguments against the sufficiency of statistics).

²⁴ Commonwealth v. Clark, 229 Mass. 409, 415, 198 N.E. 641, ___ (1935) (Lummus, J.).

fairly require especially persuasive proof to overcome the weight of statistical or background evidence. That, arguably, is the reason for the rule requiring “clear and convincing” evidence in civil cases in which the claimed event is unusual, as in cases involving the impeachment of an instrument that is regular on its face. By this standard we stipulate the need for strong evidence to overcome the negative background. When the statistics are very strong and negative to the proposition asserted, particular proof may be precluded altogether. In the Agent Orange, Bendectin, and silicone-breast-implant litigations some courts refused to permit experts to testify that defendants’ products caused plaintiffs’ harms, since the overwhelming weight of epidemiological evidence showed no causal relation.²⁵ It is entirely consistent with Bayes’s theorem to conclude that weak case-specific evidence is insufficient when the background evidence creates very strong prior probabilities that negate it.

It is one thing to allow Bayes’s theorem to help us understand the force of statistical evidence, it is another to make explicit use of it in the courtroom. The latter possibility has provoked considerable academic and some judicial debate. In paternity cases, where there is no jury, some courts have permitted blood typing experts to testify to the posterior probability of paternity by incorporating a 50% prior probability of paternity based on the non-blood-type evidence in the case. But in one case a judge rejected the testimony and recomputed the posterior probability because he disagreed with the expert’s assumed prior.²⁶

The better rule is that such evidence should not be admitted. The issue was explored in a criminal case, from New Jersey, in which a black prison guard was prosecuted for having sexual intercourse with an inmate, which is a crime under state law. She became pregnant and the child’s blood type matched that of the guard’s. At the trial an expert testified that 1.9% of black males had that blood type and so the exclusion rate was 98 to 99%. Using a 50% prior she

²⁵ In re Agent Orange Liability Lit., 597 F. Supp. 740, 787–795 (E.D.N.Y. 1984) and 611 F. Supp. 1223, 1231–1234 (Weinstein, J. 1985), *aff’d in relevant part*, 818 F.2d 145, 171–173 (2d Cir. 1987)(INSERT DESCRIPTION); *Daubert v. Merrell Dow Pharmaceuticals*, 43 F.3d 1311 (9th Cir. 1995)(Bendectin and birth defects); *Hall v. Baxter Healthcare Corp.*, 1996 U.S. Dist. Lexis 18960 at p. 15 (D. Ore. 1996)(silicone breast implants and connective tissue disease).

²⁶ *Hawkins v. Jones*, Doc. Nos. P-2287/86K and P-3480/86K (N.Y. Family Ct. Jan. 9, 1987).

testified that there was a 96.5% probability that the accused was the father. On appeal, the intermediate appellate court reversed the conviction, quoting from an opinion by the Wisconsin Supreme Court: “It is antithetical to our system of criminal justice to allow the state, through the use of statistical evidence which assumes that the defendant committed the crime, to prove that the defendant committed the crime.”²⁷ This is clearly incorrect because the prior in Bayes’s theorem does not assume that the accused committed the crime, but only that there was a probability that he did so. The court was right, however, in refusing to let the expert testify based on his prior because he had no expertise in picking a prior and because testimony based on his prior would not be relevant for jurors who had different priors. On final appeal, the Supreme Court of New Jersey affirmed the intermediate court’s reversal of the conviction, suggesting that the expert should have given the jurors posterior probabilities for a range of priors so they could match their own quantified prior views with the statistical evidence.²⁸ If the expert cannot use his own prior, is the New Jersey Supreme Court right that she can give jurors a formula and tell them to insert their prior, or give them illustrative results for a range of priors? In addition to New Jersey, one other state supreme court has suggested that an expert may give illustrative results for a range of priors.²⁹ At least one court has held otherwise.³⁰

Another, perhaps better way of proceeding, is to allow the expert to describe the effect of the likelihood ratio associated with the test without making any assumption about the prior probabilities. Thus the expert could testify that we are a thousand times more likely to see a matching print if defendant left it than if someone else did. Or, alternatively, the odds that defendant left the print are increased a thousand fold by finding a match. In both formulations the expert makes no assumption about prior probabilities, but only testifies to the *change* in

²⁷ “Paternity Test at Issue in New Jersey Sex-Assault Case,” N.Y. times, November 28, 1999 at B1.

²⁸ State v. Spann, 130 N.J. 484, 617 A.2d 247 (Sup. Ct. N.J. 1993).

²⁹ Plemel v. Walter, 303 Ore. 262, 735 P.2d 1209 (1987).

³⁰ Connecticut v. Skipper, 228 Conn. 610, 637 A.2d 1104 (1994).

probabilities or odds associated with the test. One expert panel has endorsed this approach.³¹

However this issue may be resolved, as a practical matter the proper scope of Bayes's theorem in the courtroom is probably not all that important. Some simulation studies suggest that jurors simply ignore expert testimony making explicit use of Bayes's theorem.³² What is important is the larger teaching that is itself often ignored: a matching trace does not have to be unique or nearly unique to the defendant to make a powerful case when combined with other evidence.

Summary

The probability of an event is generally interpreted either as the relative frequency of the event in a long series of trials or as the degree of belief (as in "the odds" on the event. Classical statisticians calculate such probabilities for observed data given assumed states of nature. Bayesian statisticians, and in some cases classical statisticians, calculate "inverse" probabilities of states of nature given the observed data. These two conditional probabilities usually are not the same. The legal standards of beyond-a-reasonable-doubt in criminal cases and more-likely-than-not in civil cases are Bayesian because they invoke the probabilities of past events given the Bayes's theorem teaches us that the adjustments made by the students are too small. For the person whose prior was 0.25, his or her posterior probability should be 0.997.³³ At the upper end of the range, the prior is 0.75 and the posterior probability should be 0.9997. The difference between these posterior probabilities is not significant for decision purposes; we convict in either case. The point is that underestimation of the force of statistical evidence (the 1 in a 1,000

³¹ Committee on Scientific Assessment of Bullet Lead Elemental Composition Comparison, "Forensic Analysis: Weighing Bullet Lead Evidence," 96, 97, 112 (The National Academies Press, 2004).

³² See, e.g., David L. Faigman & A.J. Baglioni, Jr., "Bayes' Theorem in the Trial Process," 12 Law & Hum. Rel. 1 (1988).

³³ For such a person the prior odds are $0.25/0.75 = 1/3$. The likelihood ratio is 1 (we are certain that the palm print would like defendant's if he left it) divided by 1/1000 (the rate of such prints in the population, if he didn't leave it), or 1,000. Applying Bayes's theorem, $1/3 \times 1,000 = 333.33$ (the posterior odds). The posterior probability of guilt is thus $333.33/334.33 = 0.997$.

statistic) when it is informally integrated with other evidence is a source of systematic bias that is far more important than the variation due to subjectivity in estimating the prior probability of guilt, even if we assume that the subjective variation is entirely due to error.

Besides correcting for bias, Bayes's theorem helps shed some light on the value of traces used for identification. Take the prosecutor's and defendant's fallacies previously discussed. We see from Bayes's theorem that these two arguments depend on the assumed prior probabilities. If the prior probability that the accused was the source of the trace is 50%, then the prior odds are 1 (.50/.50) and the posterior odds given the match are equal to the likelihood ratio, which is the reciprocal of the population frequency. Thus if the population frequency of an observed trace is 1 in a 1,000, the posterior odds that the accused left it are 1,000 and the posterior probability of the accused having left the print is $1000/1001 = 0.999$, which is as the prosecutor asserted. The evidence is so powerful that it could not be excluded on the ground that the jury might overvalue it.

On the other hand, if there is no evidence of responsibility apart from the trace, the accused is no more likely to have left it than anyone else in the relevant population who shared the trace. If N is the size of the relevant population (excluding the accused) then there are $N/1000$ people in the population who share the trace, plus the accused. The probability the accused left the print is $\frac{1}{(N/1,000) + 1} = \frac{1,000}{N + 1,000}$. If we apply Bayes's theorem, assuming there is no other evidence against the accused other than the trace, the prior probability of guilt is $1/(N + 1)$, and the prior odds on guilt are simply $1/N$. Since the likelihood ratio is still 1,000, the posterior odds are $1,000/N$, and the posterior probability is $\frac{1,000/N}{(1,000/N) + 1} = \frac{1,000}{N + 1,000}$, which is as the accused argued. So while the prosecutor's fallacy assumes a 50% prior probability that the accused left the print, the defendant's fallacy assumes a prior probability no greater than that of a randomly selected person.

Thus the admissibility and sufficiency of population frequency data may, as a theoretical matter, turn on the strength of the other evidence in the case. One must say as a theoretical matter because as a practical matter it is hard to imagine a real criminal case in which the other evidence would not suffice to permit the jury also to hear the statistical evidence. The prosecutor's fallacy is likely to be closer to the truth than the defendant's fallacy.

In civil cases, the story is more complicated.

In some early cases in which there were no data, a few judges muddied the waters by

sounding the theme that quantitative probability is not evidence. For example, in *Day v. Boston & Maine R.R.*,³⁴ a case involving the need to determine the cause of death in a railway accident, but in which there was really no evidence of how the accident had occurred, Judge Emery gratuitously commented that mathematical probability was akin to pure speculation:

Quantitative probability – is only the greater chance. It is not proof, nor even probative evidence of the proposition to be proved. That in one throw of the dice there is a quantitative probability, or greater chance, that a less number of spots than sixes will fall uppermost [the probability is 35/36] is no evidence whatever that in a given throw such was the actual result . . . The slightest real evidence that sixes did in fact fall uppermost would outweigh all the probability otherwise.³⁵

This theme was picked up in *Sargent v. Massachusetts Accident Company*,³⁶ in which the court had to decide whether the deceased, canoeing in the wilderness, met his end by accident (which would have been within the insurance policy) or by other causes such as starvation (which would not). Of course, there were no data. Citing *Day*, Justice Lummus elaborated, in much-quoted language, the theme of mathematical probability as insufficient proof:

It has been held not enough that mathematically the chances somewhat favor a proposition to be proved; for example, the fact that colored automobiles made in the current year out-number black ones would not warrant a finding that an un-described automobile of the current year is colored and not black, nor would the fact that only a minority of men die of cancer warrant a finding that a particular man did not die of cancer. The weight or preponderance of the evidence is its power to convince the tribunal which has the determination of the fact of the actual truth of the proposition to be proved. After the evidence has been weighed, that proposition is proved by a preponderance of the evidence if it is made to appear more likely or probable in the sense that actual belief in its truth, derived from the evidence, exists in the mind or minds of the tribunal notwithstanding any doubts that may linger there.³⁷

It was this language that was cited by the court in *Smith v. Rapid Transit, Inc.*, a case in which there *were* data.

Since the *Sargent* dictum cited in *Smith* originated in cases in which there were case-

³⁴ 96 Me. 207, 52 A. 771 (1902).

³⁵ *Id.* at 217-218, 52 A. at 774.

³⁶ 307 Mass. 246, 29 N.E.2d 825 (1940)

³⁷ 307 Mass. at 250-251, 29 N.E.2d at 827.

specific facts that were only suggestive, but not compelling, the notion that probability is not evidence seems to spring from the false idea that the uncertainty associated with mathematical probability is no better than ungrounded speculation to fill in gaps in proof. There is of course a world of difference between the two. As Bayes's theorem shows us, the former, when supported, would justify adjusting our view of the probative force of particular evidence, while the latter would not. Nor is it correct, as we have seen, that the slightest real evidence (presumably an eyewitness would qualify) should outweigh all the probability otherwise.

Yet the view that "bare" statistics are insufficient persists.³⁸ In a recent case, *Krim v. pcOrder.com, Inc.*,³⁹ the issue was whether the plaintiffs had standing to sue for fraud in the sale of securities under the federal securities laws. Since the Securities Act of 1933 was designed to protect only purchasers in public offerings, the court held that at least one share had to be traced to those issued in two public offerings by the company for the plaintiffs to have standing to sue. There were approximately 2.5 million shares issued in the public offerings and registered in street name, Cede & Co., a nominee of the Depository Trust Company. When two of the plaintiffs bought their shares from this pool there were intermingled with it shares held by insiders that were also offered for sale. Despite the fact that the pool of shares consisted of 99.85% public offering shares when one plaintiff bought 3000 shares and 91% public offering shares when another bought his shares, there was no way of determining whether in fact they had bought the public offering shares or the insider shares. Assuming random selection from the pool, plaintiffs' expert computed the probability that at least one share would have been from the public offering segment of the pool and the court did not dispute that this probability was near 100%. The court acknowledged that the burden of proof of standing was by a preponderance of the evidence and that all evidence is probabilistic, but nevertheless ruled that probabilities were not enough, citing *Sargent and Smith*. In particular, the court concluded that allowing the probabilistic argument would contravene the statute by extending it to all purchasers of shares

³⁸ See, e.g., *Guenther v. Armstrong Rubber Co.*, 406 F.2d 1315 (3rd Cir. 1969) (holding that although 75% to 80% of tires marketed by Sears were made by defendant manufacturer, plaintiff would have lost on a directed verdict even if he had been injured by a tire bought at Sears).

³⁹ 2005 U.S. App. LEXIS 3471 (5th Cir. 2005).

from the pool, since all would have the same argument. It also rejected the “fungible mass” argument by which each purchaser would be deemed to have purchased his proportionate share of public offering shares; allowing that argument, the court held, would undermine the requirement that plaintiff trace his shares to the public offering. When the statute was passed, the practice of holding shares in street name was not widespread and shares could in many cases be traced. Since that was no longer true, the remedy, the court concluded, was to amend the statute.

Most commentators think that *Smith* was rightly decided. Their reasons vary. For example, Judge Richard Posner, discussing *Smith*, argues that plaintiff needs an incentive to ferret out case-specific evidence, in addition to the statistics, so he must lose unless he shows that it was infeasible for him to get more particular evidence.⁴⁰ Unless he makes that showing, his failure to produce more evidence must be due either to the fact that the other evidence was adverse or that he failed to make an investigation—and in either case he should lose. The *Krim* court cited Judge Posner’s decision without noting the distinction that in *Krim* there was no other evidence possible of which source the shares came from, and so the decision should fall within Judge Posner’s caveat holding statistics sufficient in such cases. Also, the hypothetical Judge Posner discussed was a bare statistical preponderance of 51%, far from the overwhelming probability in *Kim*. Another objection is that if statistics were sufficient, Company B (the Company with fewer buses) would enjoy immunity from suit, all the errors being made against Company A, a significant economic advantage. But this assumes there are many such cases. If accidents involving unidentified buses become numerous enough to be a statistical phenomenon, the solution is an application of enterprise liability, the ultimate statistical justice, in which each company bears its proportionate share of liability in each case.

Those who find “bare” statistical evidence intrinsically insufficient must explain why it somehow should become magically sufficient if there is even a smidgeon of case-specific evidence to support it.

⁴⁰ *Howard v. Wal-Mart Stores, Inc.*, 160 F.3d 358, 359-360 (7th Cir. 1998). See also Richard A. Posner, “An Economic Approach to the Law of Evidence,” 51 *Stan. L. Rev.* 1477, 1508–1509 (1999). For a rare dissent, see Daniel Shaviro, “Statistical Probability Evidence and the Appearance of Justice,” 103 *Harv. L. Rev.* 530 (1989)(criticizing arguments against the sufficiency of statistics).

Fortunately the utterances in the early cases are only dicta that have not been uniformly solidified into bad rules of evidence or law. The actual rules are much sounder from a Bayesian point of view. The learned Justice Lummus may disparage statistical evidence by writing that the fact that only a minority of men die of cancer would not warrant a finding that a particular man did not die of cancer, and yet in an earlier case appraise the force of statistical proof by soundly holding that “the fact that a great majority of men are sane, and the probability that any particular man is sane, may be deemed by the jury to outweigh, in evidential value, testimony that he is insane. . . [I]t is not the presumption of sanity that may be weighed as evidence, but rather the rational probability on which the presumption rests.”⁴¹

When the event is unusual, but not so rare as to justify a presumption against it, we may fairly require especially persuasive proof to overcome the weight of statistical or background evidence. That, arguably, is the reason for the rule requiring “clear and convincing” evidence in civil cases in which the claimed event is unusual, as in cases involving the impeachment of an instrument that is regular on its face. By this standard we stipulate the need for strong evidence to overcome the negative background. When the statistics are very strong and negative to the proposition asserted, particular proof may be precluded altogether. In the Agent Orange, Bendectin, and silicone-breast-implant litigations some courts refused to permit experts to testify that defendants’ products caused plaintiffs’ harms, since the overwhelming weight of epidemiological evidence showed no causal relation.⁴² It is entirely consistent with Bayes’s theorem to conclude that weak case-specific evidence is insufficient when the background evidence creates very strong prior probabilities that negate it.

It is one thing to allow Bayes’s theorem to help us understand the force of statistical evidence, it is another to make explicit use of it in the courtroom. The latter possibility has provoked considerable academic and some judicial debate. In paternity cases, where there is no jury, some courts have permitted blood typing experts to testify to the posterior probability of

⁴¹ Commonwealth v. Clark, 229 Mass. 409, 415, 198 N.E. 641, ___ (1935) (Lummus, J.).

⁴² In re Agent Orange Liability Lit., 597 F. Supp. 740, 787–795 (E.D.N.Y. 1984) and 611 F. Supp. 1223, 1231–1234 (Weinstein, J. 1985), *aff’d in relevant part*, 818 F.2d 145, 171–173 (2d Cir. 1987)(INSERT DESCRIPTION); Daubert v. Merrell Dow Pharmaceuticals, 43 F.3d 1311 (9th Cir. 1995)(Bendectin and birth defects); Hall v. Baxter Healthcare Corp., 1996 U.S. Dist. Lexis 18960 at p. 15 (D. Ore. 1996)(silicone breast implants and connective tissue disease).

paternity by incorporating a 50% prior probability of paternity based on the non-blood-type evidence in the case. But in one case a judge rejected the testimony and recomputed the posterior probability because he disagreed with the expert's assumed prior.⁴³

The better rule is that such evidence should not be admitted. The issue was explored in a criminal case, from New Jersey, in which a black prison guard was prosecuted for having sexual intercourse with an inmate, which is a crime under state law. She became pregnant and the child's blood type matched that of the guard's. At the trial an expert testified that 1.9% of black males had that blood type and so the exclusion rate was 98 to 99%. Using a 50% prior she testified that there was a 96.5% probability that the accused was the father. On appeal, the intermediate appellate court reversed the conviction, quoting from an opinion by the Wisconsin Supreme Court: "It is antithetical to our system of criminal justice to allow the state, through the use of statistical evidence which assumes that the defendant committed the crime, to prove that the defendant committed the crime."⁴⁴ This is clearly incorrect because the prior in Bayes's theorem does not assume that the accused committed the crime, but only that there was a probability that he did so. The court was right, however, in refusing to let the expert testify based on his prior because he had no expertise in picking a prior and because testimony based on his prior would not be relevant for jurors who had different priors. On final appeal, the Supreme Court of New Jersey affirmed the intermediate court's reversal of the conviction, suggesting that the expert should have given the jurors posterior probabilities for a range of priors so they could match their own quantified prior views with the statistical evidence.⁴⁵

If the expert cannot use his own prior, is the New Jersey Supreme Court right that she can give jurors a formula and tell them to insert their prior, or give them illustrative results for a range of priors? In addition to New Jersey, one other state supreme court has suggested that an expert

⁴³ Hawkins v. Jones, Doc. Nos. P-2287/86K and P-3480/86K (N.Y. Family Ct. Jan. 9, 1987).

⁴⁴ "Paternity Test at Issue in New Jersey Sex-Assault Case," N.Y. times, November 28, 1999 at B1.

⁴⁵ State v. Spann, 130 N.J. 484, 617 A.2d 247 (Sup. Ct. N.J. 1993).

may give illustrative results for a range of priors.⁴⁶ At least one court has held otherwise.⁴⁷

Another, perhaps better way of proceeding, is to allow the expert to describe the effect of the likelihood ratio associated with the test without making any assumption about the prior probabilities. Thus the expert could testify that we are a thousand times more likely to see a matching print if defendant left it than if someone else did. Or, alternatively, the odds that defendant left the print are increased a thousand fold by finding a match. In both formulations the expert makes no assumption about prior probabilities, but only testifies to the *change* in probabilities or odds associated with the test. One expert panel has endorsed this approach.⁴⁸

However this issue may be resolved, as a practical matter the proper scope of Bayes's theorem in the courtroom is probably not all that important. Some simulation studies suggest that jurors simply ignore expert testimony making explicit use of Bayes's theorem.⁴⁹ What is important is the larger teaching that is itself often ignored: a matching trace does not have to be unique or nearly unique to the defendant to make a powerful case when combined with other evidence.

Sum mary

The probability of an event is generally interpreted either as the relative frequency of the event in a long series of trials or as the degree of belief (as in "the odds" on the event. Classical statisticians calculate such probabilities for observed data given assumed states of nature. Bayesian statisticians, and in some cases classical statisticians, calculate "inverse" probabilities of states of nature given the observed data. These two conditional probabilities usually are not

⁴⁶ Plemel v. Walter, 303 Ore. 262, 735 P.2d 1209 (1987).

⁴⁷ Connecticut v. Skipper, 228 Conn. 610, 637 A.2d 1104 (1994).

⁴⁸ Committee on Scientific Assessment of Bullet Lead Elemental Composition Comparison, "Forensic Analysis: Weighing Bullet Lead Evidence," 96, 97, 112 (The National Academies Press, 2004).

⁴⁹ See, e.g., David L. Faigman & A.J. Baglioni, Jr., "Bayes' Theorem in the Trial Process," 12 Law & Hum. Rel. 1 (1988).

the same. The legal standards of beyond-a-reasonable-doubt in criminal cases and more-likely-than-not in civil cases are Bayesian because they invoke the probabilities of past events given the evidence. In real-life situations, calculations of Bayesian probabilities generally must begin with prior probabilities that do not express relative frequencies, but rather are personal or subjective, reflecting degrees of belief. There is a dispute in the statistical community whether the use of personal probabilities is valid. Credibility beliefs can be combined with relative frequency evidence using Bayes's theorem, which states that the posterior odds on an event are equal to the prior odds times the likelihood ratio for the event. There is a legal debate over whether an expert may make explicit use of Bayes's theorem by giving fact finders posterior probabilities for a range of priors. Less controversial would be to limit the expert to describing the change in probabilities or posterior odds associated with the test, without making any assumption about prior probabilities. However that is resolved, Bayes's theorem teaches us that identification evidence need not point to the defendant uniquely to make a powerful case when combined with other evidence.

evidence. In real-life situations, calculations of Bayesian probabilities generally must begin with prior probabilities that do not express relative frequencies, but rather are personal or subjective, reflecting degrees of belief. There is a dispute in the statistical community whether the use of personal probabilities is valid. Credibility beliefs can be combined with relative frequency evidence using Bayes's theorem, which states that the posterior odds on an event are equal to the prior odds times the likelihood ratio for the event. There is a legal debate over whether an expert may make explicit use of Bayes's theorem by giving fact finders posterior probabilities for a range of priors. Less controversial would be to limit the expert to describing the change in probabilities or posterior odds associated with the test, without making any assumption about prior probabilities. However that is resolved, Bayes's theorem teaches us that identification evidence need not point to the defendant uniquely to make a powerful case when combined with other evidence.