*Economic Perspectives—Volume 1, Number 1—Summer 1987—Pages 155–160*

# Evaluating Uncertain Evidence With Sir Thomas Bayes: A Note For Teachers

Steven C. Salop

Consider the following fact-finding problem: On the night of March 1, 1986, in Lorain, Ohio, John Doe was struck by a speeding taxi as he crossed the street. The taxi was driving the wrong way down a one-way street and did not stop. An eyewitness thought that the taxi was blue. Doe has sued the Blue Cab Company for his medical expenses in a tort claim.

Lorain has only two taxi companies, Blue Cab and Green Cab. Green Cab is the dominant firm with 85 percent of the taxis registered in the town. No "counterfeit" taxis have ever been seen in Lorain. Drivers are employees of the cab companies and the parties have stipulated that Blue Cab company would be liable for injuries suffered in such an accident, if the owner of the offending taxi could be identified. At the trial, the eyewitness' live testimony was buttressed with evidence regarding the ability of the eyewitness to make correct color identifications at night. According to uncontroverted evidence, the eyewitness was perfectly reliable in distinguishing taxis from private automobiles and trucks and was 80 percent reliable in identifying the color of taxis. That is, he was able to identify the correct color of taxis 80 percent of the time, under conditions approximating those of the night of the accident.

The case is being heard by a judge. Suppose the legal standard is "preponderance of the evidence." Is the eyewitness reliable enough to carry the plaintiff's burden? If you were the judge, what would you decide?

1. Suppose that "preponderance of the evidence" is interpreted to mean a probability of 50 percent? Do you think that it is more likely than not that the taxi was blue rather than green?

■ *Steven C. Salop is Professor of Economics, Georgetown University Law Center, Washington, DC.*

2. Alternatively, suppose that "preponderance of the evidence" means a likelihood of 75 percent? Would the reliability of the evidence meet this standard?

When asked these questions, most people answer affirmatively to both. They say that the likelihood that the cab is blue, given that the witness thinks that it is blue, is 80 percent. Therefore, the probability that the cab is blue exceeds even the 75 percent threshold. This conclusion is correct, but only if the fact finder relies *exclusively* on the eyewitness. Yet ignoring other reliable information would be irrational.

What about the additional information that 85 percent of the taxis in the town are green? Why should this information be ignored? When it is taken into account, the likelihood that the offending taxi is blue falls sharply—below 50 percent, in fact. How much weight to place on imperfect evidence is a problem of statistical interference. The correct statistical methodology of combining such evidence is called Bayes Theorem, after Sir Thomas Bayes, who devised the method.

## Statistical Inference with Bayes Theorem

Because the eyewitness is known to be fallible, additional evidence could improve the fact finder's accuracy. Thus, the fraction of green cabs is relevant evidence. Indeed, as demonstrated below, it is actually more reliable than the eyewitness.

This counterintuitive result can be understood as follows. Suppose the motor vehicle registration evidence was proffered by the defendant through live testimony by the Registrar of the local department of motor vehicles. The Registrar could testify that, on the basis of his knowledge of taxi registrations, the offending taxi was probably green. In effect, this testimony has an error rate of only 15 percent, because 85 percent of the taxis are green.

Viewed in this light, the Registrar's testimony is more reliable than that of the eyewitness. The eyewitness, who testified that the taxi was blue, has a reliability rate of 80 percent. The Registrar, who testified that the taxi was green, has a reliability rate of 85 percent. The weight of the evidence tips the scale towards the cab being green. In statistical terms, the probability that the taxi was green exceeds one-half and the probability the taxi was blue is less than one-half.

The following matrix can inform the analysis of using these two sources of evidence. An entry in the matrix states how many taxis of a certain color would be perceived as a certain color by the witness, given the known accuracy of the witness. The first entry (blue, blue) of 12, for example, is how many of the blue cabs would be perceived as blue by the eyewitness in a repetition of the situation. There are, in total, 15 blue cabs and 85 green cabs in town, as proved by the Registrar. Because the eyewitness is 80 percent accurate, out of the 15 blue cabs he would have correctly perceived 12 of them as blue (that is, 80 percent of 15) and erroneously perceived 3 as green (that is, 20 percent of 15). Of the 85 green cabs, he would have correctly perceived 68 as green (80 percent of 85) and erroneously perceived 17 as blue (20 percent of 85). These figures are entered in the table.

| | | Color perceived by eyewitness | | |
|---|---|---|---|---|
| | | *Blue* | *Green* | *Total* |
| **Color in fact** | Blue | 12 | 3 | 15 |
| | Green | 17 | 68 | 85 |
| | **Total** | 29 | 71 | 100 |

Thus, the eyewitness would perceive 29 blue taxis, of which 12 are truly blue and 17 are truly green. Stated in a slightly different way, in the 29 situations where the eyewitness says the taxi is blue, the taxi is blue in fact only 12 times and green in fact 17 times. Therefore, when the eyewitness testifies that a particular taxi is blue, the likelihood that the taxi is blue in fact is only 12/29, or 41 percent, once the Registrar's testimony also is taken into account.[1]

This example might suggest that the plaintiff should have sued the Green Cab company, on the strength of the Registrar's testimony. Under classical legal rule of evidence, however, the Registrar's testimony would be insufficient to hold Green Cab liable. Indeed, a joint suit against both taxi firms also would fail. This outcome may seem peculiar, since it was undisputed that the hit-and-run vehicle was a taxi. But tort law is concerned with providing incentives for being careful as well compensating victims. A rule holding Green Cab solely liable, or sharing the damages between the two taxi firms in proportion to their market shares, might reduce their incentives to drive carefully. Stated another way, the fact that 85 percent of the taxis are green does not prove that the fraction of all the *carelessly driven* taxis in town that are owned by Green Cab also is 85 percent. (Of course, if color identification is impossible, then allowing the cab companies to escape liability also would reduce their incentives.)

This area of the law is changing. In a recent class action involving birth defects resulting from pregnant women taking the drug DES, a court upheld a claim by a class of plaintiffs to sue jointly a number of DES producers and collect damages in proportion to their market shares. See *Sindell v. Abbott Labs*, 607 P. 2d 924 (1980). Similarly, in the case of *Summers v. Tice*, 199 P. 2d 1 (1948), a shooting victim was able to establish that he had been injured by shots negligently fired by one of two hunters, but he could not tell which one. His joint suit against both hunters succeeded. But, he proved both hunters were negligent.

[1]Another example will help illustrate the technique. Suppose the eyewitness was 90 percent accurate. Then, ignoring the fact that the number of cabs must be an integer, of the 15 blue cabs, the eyewitness would have perceived 13.5 as blue. Of the 85 green cabs, he would have perceived 8.5 blue cabs. Thus, the probability that a cab is blue in fact, given that he testifies that it is blue, is 13.5/22, or 63 percent. Because the perceptions of the eyewitness are more reliable in this example, more weight is placed on his testimony.

## The Fallibility of Human Judgment

The taxi problem focuses on the visual fallibility of the eyewitness; however, the statistical fallibility of fact finders is perhaps a more significant issue. Most human beings, at least those untrained in statistics, overestimate the likelihood that the taxi is blue. Most people would state that the probability that the taxi is blue is 80 percent. They make several errors.

Most people do not understand Bayes Theorem, the statistical method of weighing evidence illustrated here. Given two sources of evidence, most would not weight them correctly. Mistakes might be a less serious problem, if human errors were unbiased. If some people overweighted the eyewitness and some people overweighted the Registrar, perhaps the errors would cancel out in the aggregate. But people are not unbiased.

People tend to underweight or altogether ignore fundamental background data like the statistical information represented by the motor vehicles data and tend to overweight anecdotal evidence of the sort represented by the eyewitness identification or descriptive information about a particular situation. They also tend to overweight the most recent information received at the expense of earlier information.[2] Note that in the statement of the problem, the eyewitness testimony came after the background information about the sample. If the order of the question was reversed, more people would answer green.

This bias presents a serious problem for the judicial system. First, it suggests that statistically sophisticated judges should be more willing to reverse the factual conclusions of juries. These experiments show that individuals are poor at weighing evidence. Second, it should go without saying that judges should take the time to become well trained in statistics. Third, until the judicial system adjusts, litigators should take an interest in learning how to frame evidence to tilt the bias in the preferred direction.

Take the example of a choice among Asian flu vaccines. Experiments demonstrate that the majority choice between vaccines depends on whether the outcomes are framed in terms of lives saved or lives lost. People tend to be *risk averse* with respect to gains, but *risk loving* with respect to losses. For example, consider the following pair of problems.[3]

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease

[2]Of course, these statistical errors are not the only ones people make. In a famous case involving a robbery carried out by a white woman and a bearded black man who escaped in a yellow sports car, the California prosecutor argued that the defendants, who met that description, must have been guilty because the probability that a randomly selected couple would have those characteristics was but one chance in 12 million. Not only was the probability calculation erroneously based on a false assumption of independent events, but it ignored the fact that in a population of 24 million couples, the likelihood that the police arrested the right one was only one-half, hardly "beyond a reasonable doubt." *People v. Collins*, 438 P. 2d 33 (1968), 68 at Appendix. See Tribe (1971) for a discussion of this case and a number of fundamental criticisms of the use of Bayes law in litigation.

[3]See the article by Machina in this journal for a detailed discussion of this problem, other related problems of how people perceive uncertain situations, and their implications for economics.

have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

> If program A is adopted, 200 people will be saved. If program B is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved.
>
> If Program C is adopted, 400 people will die. If Program D is adopted, there is a 1/3 probability that nobody will die, and a 2/3 probability that 600 people will die.

Some people are given a choice between programs A and B, while others are asked to choose between programs C and D. Of course, program A and C have identical outcomes, as do programs B and D. Yet, 72 percent chose program A over program B, while only 22 percent chose C over program D. In other words, the majority choice between vaccines depends on whether the outcomes are framed in terms of lives saved or lives lost.

This outcome is a litigator's dream. By affecting the fact finder's frame of reference, the litigator can control the outcome. For example, suppose the family of one of the victims of the Asian disease brought a tort action against the drug company that produced the vaccine. Suppose a drug company took the cautious approach, selecting the safer vaccine that saved 200 lives for sure; that is, suppose it willfully and wantonly selected the drug that led to 400 people dying for sure.

A clever plaintiff's attorney would present the jury with the choice framed in terms of people dying (i.e. program C vs. program D), to induce it to conclude that the all-or-nothing approach (program D) would have been more prudent. A skilled drug company defense attorney would frame the choice as one of lives saved (program A vs. program B). Where both attorneys are skilled litigators, their efforts might cancel each other out. Given both frames of reference, each juror would be forced to choose one or the other. But where only one litigator is sophisticated in these tools, that attorney gains a significant advantage.

■ *I would like to thank Donald Elliott, Thomas Krattenmaker, Elizabeth Loftus, Peter Schuck, and the editors for helpful comments.*

## References

**Tribe, Lawrence,** "Trial by Mathematics: Precision and Ritual in the Legal Process," *Harvard Law Review*, April 1971, *84*, 1329–93.

**This article has been cited by:**


1. Lubomír Cingl. 2018. Social learning under acute stress. *PLOS ONE* **13**:8, e0202335. [Crossref]
2. Chris Yung, Roberto B. Pinheiro. 2012. CEOs in Family Firms: Does Junior Know What He's Doing?. *SSRN Electronic Journal* **1**. . [Crossref]
3. Roberto B. Pinheiro, Chris Yung. 2011. CEOs in Family Firms: Does Junior Know What He's Doing?. *SSRN Electronic Journal* **122**. . [Crossref]
4. Ernst Fehr, Jean-Robert Tyran. 2005. Individual Irrationality and Aggregate Outcomes. *Journal of Economic Perspectives* **19**:4, 43-66. [Abstract] [View PDF article] [PDF with links]
5. Ernst Fehr, Jean-Robert Tyran. 2005. Individual Irrationality and Aggregate Outcomes. *SSRN Electronic Journal* **71**. . [Crossref]
6. Richard A. Posner,. 1999. The Law and Economics of the Economic Expert Witness. *Journal of Economic Perspectives* **13**:2, 91-100. [Citation] [View PDF article] [PDF with links]
7. Charles A. Holt,, Lisa R. Anderson. 1996. Classroom Games: Understanding Bayes' Rule. *Journal of Economic Perspectives* **10**:2, 179-187. [Abstract] [View PDF article] [PDF with links]
8. Philip Moon, Andrew Martin. 1996. The search for consistency in economic search. *Journal of Economic Behavior & Organization* **29**:2, 311-321. [Crossref]