A Misuse of Bayes's Theorem

MICHAEL LEVIN *City College, City University of New York*

Abstract: A standard analysis of probabilistic reasoning in the legal and psychological literature implies that people commonly overestimate the reliabaility of witnesses. This paradoxical result arises from a misexplication of reliability. "Witness is right *n*% of the time" ordinarily means *P* (*p*/Witness says *p*) = *n*, not its converse—but the standard analysis takes reliability as the converse. When this misunderstanding is cleared away, the air of paradox dissipates.

Résumé: Une analyse courante dans les écrits en droit et en psychologie sur la raisonnement sur les probabilités implique que les gens surestiment d'habitude la sûreté des témoins. Ce résultat paradoxical se produit d'une explication fausse de la sûreté. "Témoin a raison en *n*% des cas" veut dire normalement *P* (*p*/Témoin dit *p*) = *n*, ne pas la converse—mais selon l'analyse courante, la sûreté est la converse. Quand on a défriché ce malentendu, l'air du paradoxe disperse.

Keywords: probabilistic reasoning, witness reliabilility, Tversky and Kahneman, Bayes's Theorem, reliability, conditional probability

To say someone is right *n*% of the time does not mean that he is right *m*% of the time, $m << n$, nor does it imply that his accuracy depends on the frequency of his guesses. Yet these seemingly evident truths are flouted by an analysis of "being right *n*% of the time" which has attained wide currency in the pedagogical and legal literature.[1]

In the paradigm case discussed in this literature, taken from Tversky and Kahneman (see n. 1), 15% of the taxicabs in a town are Blues and the rest are Greens. Walter Witness is said to be good at telling Greens from Blue; in fact, he is said to be right 80% of the time about when a cab is a Blue and when it is a Green. There is an accident involving a cab, and Witness maintains that the offending vehicle was Blue. According to common sense, the offending vehicle probably *was* Blue, and in fact the probability is .8. According to the analysis in question, the cab was probably *not* Blue.

That paradoxical result is reached as follows. Let *w* be "Witness said the cab was Blue," *h* be "The cab was Blue," $P(p)$ the background probability of *p*—where $p = h$, $P(h)$ is the "base rate"—and $P(p|q)$ the probability of *p* given *q*. Then

(1) $P(h) = .15$; 15% of the cabs are Blue;

(2) $P(\sim h) = .85$, since $P(\sim p) = 1-P(p)$;

(3) $P(w/h) = .8$; 80% of the time a cab is Blue, Witness says it is Blue;

(4) $P(w/\sim h) = .2$, assuming Witness always has an opinion.

© *Informal Logic* Vol. 19, No. 1 (1999): pp. 63-66.

Note also

(5) $P(\sim w/\sim h) = .8$

and

(6) $P(\sim w/h) = .2$,

where $\sim w \equiv$ "Witness says the cab was Green." (We again assume Witness is fully opinionated.)

The probability that the cab was Blue given that Witness says it was Blue, $P(h/w)$, can be derived from one case of Bayes's Theorem:

$$(7)\ P(h/w) = \frac{P(w/h) \times P(h)}{[P(w/h) \times P(h)] + [P(w/\sim h) \times P(\sim h)]}$$

Using (1) – (4), the term on the right becomes

$$\frac{.8 \times .15}{(.8 \times .15) + (.2 \times .85)} = .12/.29 = .41.$$

Amazingly, if Witness says the cab was Blue, the probability exceeds 58% that it was Green. This paradigm is expanded to the conclusion that in a wide variety of cases the declarations of reliable witnesses should be viewed with suspicion.

L.J. Cohen has also urged that there is something fishy about this argument,[2] but his diagnosis, or rather diagnoses, are somewhat obscure. At some points he seems to accuse the received analysis of conflating causal propensity with statistical frequency, or what is true of a concrete individual with what is true of a population in the long run; at other points he seems to accuse it of ignoring the need to narrow the reference class, or of insufficient attention to the principle of indifference. This confusion was mirrored in the variety of objections brought about by his critics.[3]

I suggest the culprit here is much more easily identified: it lies in initially explicating "the probability of Witness being right about the cab being Blue" as $P(w/h)$, and, generally, in taking the probability of someone's being right about a world-state to be the probability of his saying that that state obtains given that it does. As used in ordinary discourse, a phrase such as "the probability that Witness was right about the cab being Blue" is doubtless somewhat vague, and there may be no such thing as *the* idea it conveys. However, it seems to me that what is normally meant by it is simply $P(h/w)$, the probability that the Cab is blue conditional on Witness saying it is, and the probability of someone's being right about a world-state is the probability of that state obtaining given that he says it does. Overall, the probability that Witness is right about the color of the car, whatever color he says it is, is $[P(h/w) + P(\sim h/\sim w)]/ 2$, and, where $\{h_i\} i = 1, ..., n$ is a set of disjoint hypotheses about the color, the probability of Witness being right is $\Sigma_i P(h_i/w)/n$.

In other words, when Witness is said to be right 80% of the time, there is no need to calculate $P(h/w)$, by Bayes's Theorem or any other means, because $P(h/w)$ is what has been stipulated. If Witness says the errant cab was Blue, we have already been told how likely it is that it was Blue: .8. Using the

terminology of statistics, someone is right most of the time when his remarks are a *specific symptom* of truth, and the foregoing "proof" that the errant cab was Green errs in treating Witness's words as *sensitive* to truth.

Further evidence for this diagnosis is that estimates of chances of success are usually thought to be independent of the frequency of opportunities. To say that Mark Marksman's reliability with a rifle is .55, *i.e.* that he hits the bullseye a commendable 55% of the time, makes no reference to the frequency with which he or anyone else shoots at it. Likewise, "Witness is right about *h n*% of the time" should be invariant under changes in $P(h)$, and a term for evaluating his reliability should not—as the Bayesian quotient does—depend on $P(h)$. Of course, the reliability of an estimate of Marksman's reliability does depend on how many shots he has taken, but that is a different statistic. Marksman's and Witness's reliability may vary with circumstances, in which case a variable, *c*, ranging over circumstances must be introduced, and Witness's reliability be reckoned as $P(h/w\&c)$. But that too is independent of $P(h)$.

If the odds of Marksman's hitting the bullseye the next time he tries are calculated as the standard analysis calculates the odds of the cab Witness saw being Blue, a few reasonable assumptions show that he will almost surely miss. For if *m* is "Marksman shoots" and *s* is "A bullseye is scored," let "Marksman hits the bullseye 55% of the time" be interpreted as $P(m/s) = .55$, *i.e.*, that Marksman is the shooter 55 times out of every 100 times a bullseye is scored. Finally, suppose the rest of Marksman's regiment are such poor shots that the regimental average is .2. In other words, the background probability that a shot will hit the bullseye, or $P(s)$, is .2. The probability that Marksman will hit the bullseye the next time he shoots at it, explicated as $P(s/m)$, is then

$$\frac{P(m/s) \times P(s)}{[P(m/s) \times P(s)] + [P(m/\sim s) \times P(\sim s)]} = \frac{.55 \times .2}{(.55 \times .2) + (.45 \times .8)},$$

a feeble .23.

Marksman's chances of a bullseye, on this reckoning, will improve if the rest of his regiment improve their aim. But an accurate shot should not be inaccurate because his regiment is, nor, if his accuracy remains constant, should he improve by being transferred to a unit of sharpshooters. As ordinarily conceived, the probability of Marksman hitting the target his next time out is team-independent, and is already given by his average, .55.

"Reliability" should be explicated so as to preserve the apparent truism that someone equally reliable at two tasks—such as shooting for two different regiments, or identifying cabs of different colors—is equally likely to succeed at both. This principle is violated by the "Bayesian" analysis I have criticized. For let us assume, as does the received analysis, that Witness is precisely as reliable about Greens as about Blues, *i. e.*, (5) and (6). To evaluate the probability that the errant cab was Green if Witness says it was, switch *h* with $\sim h$ and *w* with $\sim w$ in (7); $P(\sim h/\sim w)$ is then $(.8 \times .85) \div [(.8 \times .85) + .2 \times .15)]$

= .95. That $P(\sim h/\sim w) >> P(h/w)$—the cab is more likely to have been Green if Witness says Green than to have been Blue if Witness says Blue—shows that, whatever we are discussing, it is not the probability that Witness is right.

What we *are* discussing, when Bayes's Theorem comes into play, is the cab's likely color when we do *not* know the probability that a cab is the color Witness says it is. Background information, including base rates, then becomes pertinent. If most cabs are Green, the cab Witness saw very likely was Green, all else equal. If in addition *most of the time Witness will say a cab is Green when it is, and say it is Blue when it is*, the cab he saw is almost certain to have been Green if he says Green—but less certain to have been Blue if he says Blue. Many situations, like this one, involve an indicator of unknown trustworthiness. We know the odds that a subject with clogged arteries will feel fatigue, and the odds that a subject with normal arteries will feel fatigue. What we would *like* to know is the specificity of fatigue, the probability that someone feeling fatigue has clogged arteries. In such cases we should not say we know how well fatigue predicts clogged arteries. Did we know that, further information would be superfluous. Indeed, knowing an idicator's trustworthiness and what the received analysis calls "trustworthiness" would us to solve for the base rate.

To repeat, I bear no ill-will toward Bayes's Theorem or base rates. I urge only that, when they are relevant, we use the descriptor "reliability" of the conditional probability we are calculating, not of the converse conditional probability, which generally has a much higher value. This "right ordering of names" (as Hobbes would have called it) disperses air of paradox, and the suggestion of stubborn human irrationality.[4]

Notes

[1] See: Paulos, J., *Innumeracy* (New York, 1988), p. 123; Epstein, R., *Forbidden Grounds* (Cambridge, MA,1992), pp. 40-41; Kaye, D., "The Law of Probability and the Law of the Land," *University of Chicago Law Review* 34 (1979); Koehler, J., and Shaviro, D., "Veridical Verdicts: Increasing Verdict Accuracy through the Use of Overtly Probabilistic Evidence and Methods," *Cornell University Law Review* 247 (1990); Tversky, A., and Kahneman, D. , "Causal Schemas in Judgments under Uncertainty," in Fishbein, ed., *Progress in Social Psychology* 117(1980).

[2] Cohen, L.J. and commentators, "Can Human Irrationality be Experimentally Demonstrated?" *The Behavioural and Brain Sciences* 4 (1981).

[3] See particularly the contributions of Diaconis and Freedman, Krantz, Mackie, Margalit and Bar-Hillel, Niiniluoto, Skyrms and Zabel in the *Behavioural and Brain Sciences* symposium cited in n. 2. Some of these critics, particularly Diaconis and Freeman and Mackie, mention the distinction I go on to stress.

[4] I wish to thank Jonathan Adler for helpful criticism and suggestions.

Michael Levin, City College and the Graduate Center, CUNY
138th St. and Convent Avenue, New York, NY 10031 U.S.A.