Trial and Error

Harry D. Saunders
Joshua G. Genser

Available at: http://works.bepress.com/harry_saunders/30/
TRIAL AND ERROR

Jurors interpret the standard of "reasonable doubt" in different ways. The result may be an intolerable inconsistency in the law

BY HARRY D. SAUNDERS AND JOSHUA G. GENSER

Death, taxes... and jury duty. Few things in life that are certain are also distasteful. So it was that, finding himself at the courthouse in Contra Costa County, California, one Monday in July several years ago, one of us (Saunders) was feeling rather put out. At the time, his small business was vying with a larger company for a lucrative contract, and the competition was promoting itself on the grounds that "bigger is better"; its larger staff could absorb the impact if a key employee fell sick—or, say, happened to be called for jury duty.

So Saunders sat there, a computer on his lap, his briefcase open on the floor, resolved to do his civic duty, yet annoyed that it should exact such a sacrifice. But by the time he had spent two days in the courthouse, watching the tedious yet strangely hypnotic process of assembling a panel for trial, he was hooked. The defendant, a man in his twenties accused of raping two young girls, looked on impassively as his lawyer and the prosecutor questioned a group of strangers, some of whom would decide his fate. As Saunders followed the proceedings, he could not help evaluating them through the lens of his professional experience. His business, Decision Processes, Inc., specializes in teaching corporate executives how to make reasoned decisions on the basis of the information available. One of the fundamental techniques his students learn is how to quantify the uncertainty they face. Here in the courthouse, however, the measure of uncertainty was conveyed by a single phrase, "reasonable doubt"—a standard that seemed exceedingly vague.

"How could I, or any juror, know whether my opinion about the defendant's guilt had surpassed the benchmark of reasonable doubt?" Saunders wondered. Did justice require that a juror had to be at least 99.9 percent certain the man had raped those girls? Or would it be enough to be just 75 percent sure? Would 51 percent suffice? It seemed wrong that in the jury room, of all places—where people are called on to make the most important decision of all, about the fate of another human being—those decisions were arrived at so inconsistently.

Saunders told the judge he was finding it hard to interpret the standard of reasonable doubt. The judge replied dismissively: "Don't worry. You'll receive clear instructions when the time comes." But Saunders was not appeased. When the opportunity arose, he asked to speak with the judge privately. In chambers, before the judge and the attorneys for both sides, he explained his dilemma. Until then, he had had the distinct impression that the prosecuting attorney, a sharp-edged, intense woman in her late thirties, was well disposed toward impaneling him. But as he talked about the science of decision theory, and the doubts it instilled in him about whether a verdict could be arrived at with adequate clarity and precision, both of the lawyers in the room turned icy. "This guy is trouble," they were clearly thinking. "He could create chaos in a jury deliberation."

Needless to say, Saunders was passed over as a juror for the rape trial. His company went on to win its bid, and he gained a fail-safe method of avoiding jury duty for the rest of his life. Still, he wasn't satisfied. Something seemed desperately wrong with the entire legal system.

If you don't think the vagueness of "reasonable doubt" is a problem for the criminal justice system, imagine that you yourself must serve as a juror in the following hypothetical murder case. A secretary at a one-man law firm—we'll call her Sarah Neill—is at her desk one afternoon, when she hears an unusual noise coming from the office of her boss, Rob Gunderson. Gunderson has put Neill under strict orders never to disturb him, so she lets it go. Unbeknownst to her, a man has entered Gunderson's office and stabbed him...
to death with a ceremonial knife that Gunderson had collected during his travels abroad.

The next morning, Neill discovers Gunderson's body in his office and calls the police. The knife is still there, wiped clean, and there is no other evidence. Later that day Neill is arrested and charged with the murder. The brutality of the crime leads the district attorney to call for the death penalty. The state's case is simple. There is no sign of forced entry. Gunderson's office is accessible only through two locked doors: one next to Neill's desk, the other out of her sight. Each door is secured by a keypad lock. Gunderson, a cautious and obsessive man, had required Neill to change the code for the locks daily. Hence, besides Gunderson, Neill was presumably the only person who knew the code. As for motive, Neill had recently argued with Gunderson over her salary.

The case is unusual because the likelihood of Neill's guilt can be precisely determined. The keypads guarding Gunderson's office have ten digits each, and investigators learn that the entry code had always been a single digit. The electronic log shows that the code had been entered only once on the afternoon of the murder. If someone other than Neill, someone who did not know the entry code, had tried to enter, he or she would have had a one-in-ten chance of hitting the right number on the first try.

Neill's attorney soon discovers that a security camera had filmed a man entering the hallway near Gunderson's office at around the estimated time of the murder. The man was identified as a vagrant who held a grudge against Gunderson. Unfortunately for Neill's case, however, the man had been found dead on the street two days before the video surfaced.

The facts of the case are clear-cut. There are no witnesses, and Neill and her attorney decide that her negative feelings about Gunderson make it too risky for her to testify. Consequently, both the defense and the prosecution argue that there is a 10 percent chance the vagrant killed Gunderson, and a 90 percent chance that Neill did. Thus when the summations are over and the deliberations begin, the jury need not hash out the differing opinions of twelve people about the likelihood of Neill's guilt. No, in this case the verdict comes down to a single question: Does 90 percent certainty constitute proof of guilt beyond a reasonable doubt?

The members of a jury—though they may not know it—have a double task. First, each of them must determine how likely they think it is that the accused person is guilty. Because it is impossible for jurors to know for sure, they must make a reasoned judgment of probability based on their own interpretation of the evidence and arguments presented at the trial.

The second task for the jurors—the one they will probably accomplish unconsciously—is to decide what, for them, represents guilt beyond a reasonable doubt. (The

Olivia Parker, Wheel of Uncertainty. 1995

September/October 1999 • THE SCIENCES 19
problem is unique to criminal proceedings, by the way. Civil cases are judged by a different standard, "preponderance of the evidence.") When jurors have accomplished those two steps, the decision to be made becomes a matter of the simplest mathematics. If the probability of guilt a juror has arrived at is greater than the reasonable-doubt threshold, the juror must vote to convict. Otherwise, the juror must vote to acquit.

Now let's return to the case of Sarah Neill. Suppose the jury convicts Neill and she is sentenced to death. Clearly, that panel's threshold for reasonable doubt was less than 90 percent certainty. Another jury, meanwhile, possessed of the same belief in Neill's guilt but demanding, say, 96 percent certainty to meet the test of reasonable doubt, would have let her go free. Can juries in the real world differ so markedly in how they would act, given the same beliefs about the guilt or innocence of a defendant? The results of our research strongly suggest that the answer is yes. And if so, judicial standards in the United States are unevenly applied and inherently unfair—and justice is indeed blind.

In a series of experiments we have conducted with a few dozen managers of large U.S. corporations, such as General Motors, Chevron and Hewlett-Packard, we have discovered that people hold an amazingly wide range of attitudes about what constitutes reasonable doubt. We gave our subjects the standard California jury instruction, which is similar to that of most states. It reads, in part:

Reasonable doubt is defined as follows: It is not a mere possible doubt; because everything relating to human affairs is open to some possible or imaginary doubt. It is that state of the case which, after the entire comparison and consideration of all the evidence, leaves the minds of the jurors in that condition that they cannot say they feel an abiding conviction of the truth of the charge.

For all its long legal history and carefully constructed wording, the jury instruction on reasonable doubt merely replaces one vague phrase with several other phrases of equal vagueness. Many people, when they read or hear the jury instruction, seem to develop a strong sense of what it means. The problem is, their sense is rarely in accord with the equally steadfast opinions of others.

Having read the formal definition, our subjects were asked: What is the probability of guilt, expressed as a percentage, that for you would represent guilt beyond a reasonable doubt? Of the sixty subjects we interviewed, 35 percent said beyond a reasonable doubt meant greater than 90 percent probability of guilt, and most of those—18 percent of the overall population—insisted on 99 percent likelihood or greater. Another 35 percent put the threshold between 80 percent and 90 percent; 18 percent put it between 70 percent and 80 percent; and the remaining 12 percent thought it meant between 50 percent and 70 percent. One subject said the standard could be as low as 51 percent: in other words, just a smidgen better than a fifty-fifty chance of guilt was enough to convict. The average was around 87 percent—thus an "average" juror drawn from the pool would have voted to convict Sarah Neill, though we, as privileged observers, know she is innocent.

Granted, our subjects were midlevel executives—not your standard jury pool. But if a relatively sophisticated and homogeneous population of businesspeople cannot agree on what is meant by reasonable doubt, how can jurors randomly selected from the entire adult population be expected to apply a consistent standard?

Why not make the process more scientific? It would be straightforward to set the standard of reasonable doubt at some constant percentage. Some people would argue that finding and applying a numerical interpretation for the venerable words of the criminal code would be too daunting a task. Yet the reality is that, implicit in the current system, jurors must each find a way to measure the meaning of the standard for themselves.

Assigning a precise number to one's sense of certainty might seem nearly impossible. If I decide not to go home for Thanksgiving, am I 43 percent sure that I won't regret it, even though my stepmother will be angry? Or 44 percent sure? Or 42 percent? How could one ever be so
precise about something as ineffable as a multilayered decision involving unknown factors ("Will the traffic on the way be intolerable? Will the cranberry sauce be full of walnuts, which I hate, the way it was last year? Will my favorite cousin be there?"). as well as complex interactions and personalities? Can people really be taught to quantify their opinions?

The answer is yes. By presenting people with a series of hypothetical "bets," decision analysts can extract from anyone a numerical assessment of confidence. Moreover, once the bets have been completed, people agree that the numbers they arrive at reflect all the nuances they want to consider, including knowledge, intuition and even feelings about the likelihood of any given event.

Here's how it works: The decision analyst shows the subject a probability wheel, invented by Carl S. Spetzler of Navigator Consulting, Inc. in Menlo Park, California. The wheel is similar to a pie chart containing two adjustable colors. It can show all red, or it can show mostly red with a skinny pie-slice of blue, or any other proportion of red to blue. On the back is a scale that tells what percentage of the wheel is showing red at any given time. Attached to the center of the wheel is a pointer, which can be spun much like the pointer in a board game.

Suppose the decision analyst asks the subject: How confident are you that your company's stock will double its price in the next five years? To help the subject quantify her answer, the analyst tells her she will play a game that involves making a large (though fictitious) bet. The player can choose to place her bet in one of two ways: she can put her money on the probability wheel or on a "clairvoyant." The clairvoyant, for the purposes of the game, is a hypothetical person who can look into the future and know with absolute certainty how things will turn out.

The rewards are as follows: If the player chooses to bet on the wheel and the pointer ends up on red, she will win $100 million. If the pointer ends up on blue, she will win nothing. Alternatively, if the player chooses to bet on the clairvoyant, and the clairvoyant reveals that the stock will double, the player wins $100 million. If the clairvoyant reveals that the stock will not double, the player wins nothing. The question for the player, at each stage in the game, is whether to bet on the wheel or on the clairvoyant, so as to maximize her potential earnings.

A quick glance at the wheel makes it clear to the player that the odds are overwhelming for the spinner to end up on red. If, by contrast, she thinks there is only about a 60 percent chance that the clairvoyant's answer will be yes, she will quickly decide she has a better chance of winning her $100 million by betting on the wheel.

Next, the analyst adjusts the wheel to the opposite extreme: 5 percent red and 95 percent blue. This time the player perceives the small likelihood that the spinner will end up on red; she opts for the clairvoyant, which now seems a safer bet. Thus a series of increasingly fine adjustments is made, much like the adjustments that take place in an ophthalmologist's office. The patient being fitted for eyeglasses is asked to read an eye chart through lenses of differing strengths—stronger, then less strong. The decision analyst continues to adjust the amount of red showing on the wheel, each time asking the player to choose whether she prefers to bet her money on the wheel or on the clairvoyant. Each time the player chooses the wheel, the analyst decreases the amount of red and asks again. Each time the player chooses the clairvoyant, the analyst increases the amount of red and asks again. When the amount of red showing leaves the subject just as willing to bet on the wheel as on the clairvoyant—when, in other words, the subject is indifferent about her choice of bets—then her degree of confidence has been quantified. The number can be read off the back of the wheel. It takes about forty-five minutes to teach people the probability-wheel technique; conceivably, jurors could receive such instruction while waiting to be impaneled.

If jurors were to use the probability-wheel technique, presumably each person would come up with a different number. That is entirely appropriate. One would expect that different people judging a case would come to different conclusions about the likelihood of the defendant's guilt or innocence. In fact, it could be argued that such differing assessments are the very reason the law calls for multiperson juries.

Now suppose an objective standard for reasonable doubt had been instituted, after much debate, in the U.S. justice system—set it randomly, for argument's sake, at 87 percent. What, then, would be the next step? Would the jury foreman simply take out a pocket calculator and average the twelve numbers, entering a guilty verdict if the average were above 87 percent, and not guilty if it fell at or below 87 percent? Would the sacred angst of jury deliberations be reduced to a mere mathematical formula?

That is not at all what we are suggesting. Jury decisions in criminal cases must be unanimous. A guilty verdict should be reached only if, after discussion, debate and the assorted adjustments of opinion that result, each of the twelve jurors reached a personal determination of guilt that exceeded the 87 percent threshold. Introducing clarity and objectivity into the deliberations, however, would aid jury mem-
bers in coming to a fair and reasoned decision that accurately reflects the will of the group.

Probability wheels are an offshoot of decision theory, a discipline that was popularized in the early 1960s by Howard Raiffa, now a professor emeritus at Harvard Business School. Decision theory attempts to introduce logic and the mathematical description of probabilities into decisions that must be made in the real world. The question of probability is an age-old one; Pascal and Fermat grappled with it in the seventeenth century. But until the 1960s, no one had carried probability into the realm of professional worrying: decision making.

Among statisticians, there are those who endorse so-called classical statistics and those who follow what is known as Bayesian statistics—named for the eighteenth-century English minister and part-time mathematician Thomas Bayes. In classical statistics—which derives in part from work done in the nineteenth century by the German mathematician Karl Friedrich Gauss—the probability of a particular event's outcome is determined by how that event has played out in the past. Thus the probability that a coin will show heads—50 percent—is based on previous experiences with coin tosses. The Bayesian method, by contrast, takes a more flexible view: it allows a decision maker to draw not only on past experiences with the phenomenon in question, but also on personal beliefs and judgments, as well as on expert opinion. In other words, Bayesian statistics views probability as a statement about available knowledge, rather than as a statement about the laws of nature.

For example, suppose you observe a magician tossing an unusual-looking coin three times, and each time it turns up heads. In the classical world, you are left with a dilemma regarding the probability that the coin will turn up heads in a future toss. The coin may be a heavily biased one, and its source (the magician) gives ample cause for suspicion. But the only usable data about coin tosses suggest that the inherent probability of heads is 50 percent. The Bayesian approach, by contrast, is more adaptable to novel circumstances, in that it enables you to combine your prior knowledge of coin tosses with your observation of this unusual coin and your assumptions about its owner.

In the real world, when one-of-a-kind decisions must be made, the Bayesian approach has its advantages. Suppose, for instance, that you are an executive contemplating a $100 million acquisition of a unique business. With classical statistics, such a problem would require the introduction of a series of complex models—to make up for the fact that no experimental data can be collected because the contemplated action is unique. (You cannot undertake the business acquisition fifty times and record how many times it was successful?) Bayesian statistics, on the other hand, does not require hard data derived from comparable events, and so it lends itself more readily to one-of-a-kind problems. The countless arenas. The analyses of complex decisions are embedded, for instance, in the choice of vehicles produced by the major auto makers (which face large uncertainties about the actions of competitors and consumers), and in the fare structures of airlines (which face large uncertainties about the number of seats they will sell at various price intervals in advance of a flight).

We are suggesting that the Bayesian method is just as applicable to the courtroom.

Would you have voted to convict Sarah Neil? Should you have? A national debate to establish an objective standard of reasonable doubt would be a healthy exercise.

Agreement on an objective standard would bring welcome discipline to the growing number of cases being argued on the basis of probabilities. Jurors are increasingly being presented with probabilistic analyses of the results of tests of DNA and blood, as well as testimony about the probability that a fingerprint, a fiber, or the rifling marks on a bullet from a crime scene match their counterparts as collected from the accused. Yet in the face of so many quantifiable probabilities, jurors are given no guidance on how high those probabilities need to be to meet the ambiguous legal standards.

The path to an objective standard will, however, be rocky. Could a nation that cannot agree about universal health care or campaign-finance reform ever pin a number on reasonable doubt? Who would bring the diverse opinions together and propose a hard figure? And how would it be implemented? State legislatures would be an obvious route, because they write most criminal laws. But lawmakers would be loath to tackle the issue: they would be called soft on crime if the proposed standard were too high, and vilified by civil libertarians if it were too low. Meanwhile, if they did go through with it, the existence of different standards in different states—84 percent in Texas, say, and 96 percent in New York—would violate the very notion of fairness that an objective standard is intended to achieve.

The scary thing is that, right now, just about any guilty verdict could be appealed on the grounds that the standard of reasonable doubt is unclear. Such a move would not only force the issue of an objective standard, but could throw the nation's entire judicial system, past and present, into disarray. Virtually every criminal case ever tried could be
subject to review and invalidation, paving the way for widespread retrials. Such a situation would bring chaos—but it might be necessary in order to avoid injustice. Currently, some juries are undoubtedly using a low standard of proof that invites them to convict, whereas others are using a high standard that biases them toward acquittals. Although the process of setting an objective standard would undoubtedly be messy, what price could be too high to pay to protect future generations from wrongful verdicts?

There is, however, another option. Perhaps instead of creating a universal standard, the legal system should formally accept the principle of no universal standard. That might not be such a bad outcome if everyone understood and agreed to it, and it certainly appears the less painful choice. The legal system would simply ask jurors to assume the role of determining for themselves the standard of proof. Each juror, as we have shown, would end up interpreting the reasonable-doubt standard differently, and a range of opinions would be accepted as appropriate.

The consequences of such a policy would be far-reaching. For one thing, the law would not apply equally to all: defendants could not assume that prosecutors would be held to the same standard of proof in their particular case as they would in other similar cases. Attorneys would begin selecting jurors based on their views about reasonable doubt, and they would be free to argue what standard of proof jurors ought to apply.

Even the choice of no objective standard, however, requires a clear-headed national discussion, and it speaks to the need for educating judges, lawyers and the general public in probability and statistics. Whether a universal standard is adopted or dispensed with, some conscious choice must be made to rectify the ambiguity that currently reigns in the jury room. Every time a jury sits down to deliberate, those twelve good men and women must make tough decisions—about truth and deceit, guilt and innocence. The task would be hard enough if there were a clear standard; now, however, it is unnecessarily difficult, even confusing. It's time to stop asking jurors to make their judgments in the dark.

Harry D. Saunders is managing director of Decision Processes, Inc., in Danville, California, where he helps executives learn to quantify their uncertainties in order to become better decision makers. Joshua G. Gensler, an economist and attorney, heads The Gensler Group, a legal consulting firm in Emeryville, California.

Andy Warhol, Statue of Liberty, 1963
PEER REVIEW
Continued from Page 3

antitobacco work of the era is not. It
would be comforting to believe that sup-
pression, secrecy, abusive experimenta-
tion and the like are never to be found
in political democracies, but life is not
so simple. Recognizing that Nazi med-
icine was more complex than people
have generally realized could help shake
up our complacent sense that their world
was such a totally different one from ours.

OBJECTION!
I was amused but somewhat appalled
by Harry D. Saunders and Joshua G.
Genser’s attempt to quantify the unquantifiable in their essay “Trial and
Error” [September/October] about jury
deliberations and the standard of proof.
I am a member of that most dreaded of
human subspecies, the personal-injury
trial lawyer, but I also have an academ-
ical background in the sciences, engi-
nineering and philosophy.

A number of years ago I became a stu-
dent of Bayesian analysis while involved
in a complex case combining medical
malpractice and drug liability. The bio-
statistician Richard Peto of the Univer-
sity of Oxford had conducted a so-called
meta-analysis of a great many clinical tri-
als of a then-new drug, which was high-
ly relevant to my case. Unfortunately, I
did not think I could make the mem-
bers of a jury understand Peto’s meta-
analysis any more than I could make
them understand Bayesian analysis—and
neither did Peto. Game theory, classical
statistics and probability wheels are all
well and good, given infinite time and
infinite resources to train juries to ana-
lyze a particular set of facts. For resolv-
ing real problems in the real world of the
courts, however, they are useless.

If the point of Messrs. Saunders and
Genser’s article is that the standards of
proof enunciated by our courts—
“beyond a reasonable doubt” and “by a
preponderance of the evidence”—are
imprecise, then yes, they are correct.
The Anglo-American legal system has
struggled with definitions of those terms
for more than 200 years, and they remain
a subject of debate among legal schol-
ars. But even if more agreement exis-
ted, a “universal standard” for reasonable
doubt would be unthinkable, because it
would erode the autonomy of juries.

RANDALL C. JACKSON
San Antonio, Texas

Thomas Bayes and Karl Friedrich Gauss
both must be turning over in their graves
at the statistical absurdity of the hypo-
thetical trial described by Harry Saun-
ders and Joshua Genser. Surely the fact
that the vagrant in their anecdote had
only one chance in ten of hitting the right
number on the door lock does not estab-
lish that there is only a 10 percent chance
of his being the murderer. What if that
door had been the only means of access
to the victim? From the fact that the vic-
tim is dead, one can infer that the mur-
derer almost certainly hit the right lock
number on the first try.

As a politician who has used polls
extensively, I also find the poll taken by
Messrs. Saunders and Genser to be quite
suspect. They asked their subjects: “What
is the probability of guilt, expressed as a
percentage, that for you would represent
guilt beyond a reasonable doubt?” But
what if the poll had focused instead on
the widely accepted definition of rea-
sonable doubt cited in their article, ask-
ing: “What degree of doubt about a
defendant’s guilt, expressed as a percent-
age, would cause you to acknowledge
that you cannot say you feel ‘an abiding
conviction of the truth of the charge’?”
I think it likely that the results would be
significantly smaller than 100 percent
minus the percentages reported in Messrs.
Saunders and Genser’s poll.

It would be nice to have a simple
mathematical criterion for reasonable
doubt—comparable to the standard for
“preponderance of evidence” that ap-
plies in most civil cases (greater than 50
percent). But the patent flaws in “Trial
and Error” show that setting up such a
criterion may be very difficult indeed.

BILL GREEN
New York, New York
[Editor’s Note: The letter writer is
chairman of the Board of Governors of
the New York Academy of Sciences
and a former U.S. Congressman from
New York State.]

Harry Saunders and Joshua Genser raise
intriguing questions, but current deci-
sion theory leads to quite different
answers than theirs. The authors note
correctly that utility theory makes every
decision dependent on estimates of prob-
ability, but decisions also depend on esti-
mates of “cost” or “value.” A juror who
believes that convicting an innocent
person is twice as bad an error as acquis-
ting a guilty person will not be willing
to convict until, after weighing the evi-
dence, her estimate of the probability
that the defendant is guilty exceeds two-
thirds. Only at that point will her con-
cerns about the meaning of the term
“reasonable doubt” come into play.
Thus a better understanding of the situation is that each juror has a threshold probability above which his or her utility calculations make it preferable to convict. Each juror can then read “reasonable doubt” to mean that the choice has to be clearer than just barely above the threshold. Thus, if the authors want a debate on the meaning of reasonable doubt, they must consider how that doubt will act when regarded as a function of the specific cutoff probability that each juror requires.

Paul B. Kantor
Rutgers University
New Brunswick, New Jersey

The essay by Harry Saunders and Joshua Gensler gives the false impression that the authors were the first to discover the connection between probability calculus (particularly the Bayesian approach) and the problem of “objective” judicial decision making. In fact, that relation is age-old. Applying the calculus of probabilities to important societal problems such as jury behavior was a favorite subject for the French school of probabilists during the late eighteenth and early nineteenth centuries. For example, much relevant work, using Bayesian assumptions, was done by the French mathematicians Pierre-Simon Laplace and especially Siméon-Denis Poisson. Modern American scholars have elaborated on those models; for instance, Poisson’s ideas were extended in the 1970s by the statisticians Alan E. Gelfand of the University of Connecticut in Storrs and Herbert Solomon of Stanford University. Indeed, the analysis of judicial decisions has a long history in the United States and does not need to be reinvented.

I also take issue with the techniques employed by Messrs. Saunders and Gensler. Instead of guessing the present level of “intolerable inconsistency in the law,” they might want to determine it on the basis of the appropriate statistical data, which are available from studies of juries in the academic literature or from the U.S. Department of Justice.

Incidentally, as an “expert witness,” I fully support Messrs. Saunders and Gensler’s suggestion that judges and, more particularly, lawyers should learn more about probability, statistics, and modern decision theory.

George E. Bozoki
Cliffside Park, N.J.

Harry D. Saunders and Joshua G. Gensler reply: Our central point is that the standard of proof in criminal trials is

---

Basic Books

One Renegade Cell
The Quest for the Origins of Cancer
Robert A. Weinberg
“Offers a breathtaking picture, both wonderful and frightening, of the fantastic intricacy of aberrant cellular functioning.”
NEW YORK TIMES
176 pages • $13.00 paper

The Pattern on the Stone
The Simple Ideas That Make Computers Work
Daniel Hillis
“Danny Hillis’s ability to be both original and clear is astounding. Even where I thought I knew the material he was talking about, he was always teaching me something, either by his clear command of the subject or by his original perspective on it.”
KEVIN KELLEY, Wired
176 pages • $12.00 paper

Symbiotic Planet
A New Look at Evolution
Lynn Margulis

Once Upon a Number
A Mathematician Bridges Stories and Statistics
John Allen Paulos
“Paulos’s insightful and amusing observations on how the truths discovered through mathematics should be applied to our everyday lives will appeal to an audience beyond math and science enthusiasts.”

Philosophy in the Flesh
The Embodied Mind and Its Challenge to Western Thought
George Lakoff and Mark Johnson
“This book will be an instant academic bestseller.”

DATABASE

November/December 1999 • THE SCIENCES 47
so vague that defendants are potentially subject to wildly different likelihoods of being convicted, even when the evidence and the jurors’ determinations of the probability of guilt are identical. Such a problem, we believe, warrants a public debate, one that would either set an objective standard for reasonable doubt or declare that jurors can interpret it as they will. Randall Jackson finds an objective standard unthinkable because it would require juries to forfeit their autonomy. But arguably, judging the likelihood of a defendant’s guilt is a big enough burden

The fact that the vagrant had one chance in ten of opening the locked door does not imply that there is only a 10 percent chance of his being the murderer.

already, and the last thing juries need is the added responsibility and confusion of defending the standard of proof.

Bill Green is correct in finding fault with the probabilities in our contrived example—we somehow managed to ignore our own Bayesian sermonizing! The point of the example, however, remains: Even if juries agree completely about the likelihood of guilt, they may disagree about whether it surpasses the “beyond a reasonable doubt” threshold. It is difficult (as our error demonstrates) to concoct a realistic case in which the probability of guilt is indisputable. That very difficulty, in fact, may be one reason

interpretations than our results suggest. Paul Kantor argues that jurors consider the utilities of outcomes in determining their subjective standards. But he should note that considering outcomes is against the law: jurors are prohibited from considering the consequences for the defendant of a verdict and instructed only to judge the weight of the evidence. Surely, however, Kantor is right that a sensible standard of reasonable doubt should reflect the difficulty of wrongful verdicts. The conundrum that poses for juries would be resolved if a standard were set universally rather than by individual juries.

Finally, George Bozoki suggests that we avail ourselves of statistical data based on actual convictions and acquittals. But such statistics can never reveal the underlying interpretation that jurors attached to “reasonable doubt”—unless the statistics were also to provide objective probabilities of guilt (which, as our critics have helped point out, are pretty much impossible to come by). No, juries’ interpretation of reasonable doubt must be addressed directly, as we have done. And despite Mr. Bozoki’s assertion that we neglected to give credit where it was due, our research indicates no precedent for our work among the authors he cites.

The Sciences welcomes correspondence from readers. Please send electronic mail to <sciences@nyai.org> and regular mail to 655 Madison Ave., 16th floor, New York, N.Y. 10021-8043. All letters should include a daytime telephone number and a complete street address, and all letters are subject to editing.