Law, Probability and Risk (2006) **5**, 135–157 doi:10.1093/lpr/mgl015
Advance Access publication on March 3, 2007

Case comment—*United States* v. *Copeland*, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond A Reasonable Doubt Is Unquantifiable?

PETER TILLERS†
Cardozo School of Law, Yeshiva University, New York, USA

AND

JONATHAN GOTTFRIED‡
Attorney, Friedman, Kaplan Seiler & Adelman LLP, New York, USA

[Received on 21 August 2006; revised on 27 October 2006; accepted on 30 October 2006]

There is a well-settled maxim that the standard of persuasion in criminal trials—proof beyond a reasonable doubt—is unquantifiable. However, the usual reasons given for the unquantifiability of reasonable doubt are unsatisfactory; and a recent case, *United States* v. *Copeland*, serves as a reminder that strong considerations favour quantification of at least some standards of persuasion. This comment attempts to bring greater clarity to the question of the advantages and disadvantages of some form of quantification of the reasonable doubt standard.

Keywords: evidence; inference; proof; standard of persuasion; proof beyond a reasonable doubt; mathematics in trials; proof in criminal trials; trial by mathematics; proof and mathematics.

1. Judicial hostility towards 'quantification' of reasonable doubt

The U.S. constitutional guarantee of due process[1] permits an accused to be convicted of a crime after trial, only if the evidence presented at the trial proves the accused's guilt beyond a reasonable doubt in the eyes of the trier of fact.[2] Occasionally, an actor in a criminal trial—typically a prosecutor, but sometimes a trial judge—will use numbers of one kind or another in an attempt to explain or clarify the reasonable doubt standard in some fashion. Appellate courts have condemned such attempts at quantification of reasonable doubt whenever they have encountered them. For example, in one case, a court condemned a prosecutor's use of a bar graph that displayed, in percentages, the prosecutor's view of the numerical equivalents of various levels of proof.[3] In several cases, courts

† Email: peter@tillers.net
‡ Email: jgottfried@fklaw.com

[1] U.S. CONST. amends. V & XIV.
[2] *In re Winship*, 397 U.S. 358, 364 (1970).
[3] *People* v. *Ibarra*, No. H021123, 2001 WL 1330296 (Cal. Ct. App. 26 October 2001). In Ibarra, the prosecutor's numerical characterizations of different levels of persuasions were curious: 'She displayed to the jury a bar graph on a standard-sized sheet of paper. The graph showed different levels from 100 percent certainty to beyond a reasonable doubt at the lowest level, 60 percent. The graph included beyond a shadow of a doubt at 70 percent, beyond all doubt at 80 percent, and absolute certainty at 90 percent'. *Ibid.* at *10. Perhaps, 100% amounts to utterly certain proof beyond any shadow of a hint of any suspicion of any conceivable doubt? *Compare* TILLERS, P. "Intellectual history, probability, and the law of

have disapproved the use of football field[4] analogies that are designed to suggest numerical values for proof beyond a reasonable doubt.[5] In one case, the prosecutor told a juror that proof beyond a reasonable doubt is ‘kind of like 75 percent’ and like being ‘[s]omewhere [between] [the] 75 and 90’ yard line on a 100-yard-long ‘football field’.[6] In another case, the U.S. Court of Appeals for the Ninth Circuit disapproved the prosecutor’s off-hand comment to prospective jurors during voir dire that proof beyond a reasonable doubt is ‘like getting to the 97 yard line’.[7] Baseball analogies have also been used and occasionally have also been condemned.[8] Prosecutors and trial judges often use jigsaw puzzle analogies to explain the assessment of circumstantial evidence under the reasonable doubt standard. Appellate courts usually do not frown on such analogies. But puzzle analogies that refer to specific numbers of missing puzzle pieces, it seems, can raise the hackles. In one of those cases, the trial judge instructed the jury to compare circumstantial proof beyond a reasonable doubt to ‘a 1,000 piece puzzle with sixty pieces missing’.[9] Although the trial judge’s analogy was held to be harmless error, the Massachusetts Supreme Judicial Court said, ‘Given the close connection between circumstantial evidence instructions and reasonable doubt instructions, it is best for judges to avoid examples that have numeric or quantifiable implications’.[10] In another case, a prosecutor invoked an Eiffel Tower jigsaw puzzle.[11] In other cases, courts have condemned the submission of statistical evidence on the ground (sometimes mistaken, sometimes not) that the statistical evidence as presented amounted to quantification of the reasonable doubt standard.[12] It is possible to find the occasional case in which a court hints at its willingness to embrace at least some kind of

evidence,” 91 *Michigan Law Review* 1465, 1476–1480 (1993) (criticizing suggestion that proof can achieve a level of certainty beyond certainty).

[4] In American football, six points are earned (six points for a touchdown) when the football is carried or thrown (under certain circumstances, in the right direction, and in a permissible fashion) across the end of a 100-yard-long football field (beyond the ‘touchdown line’).

[5] See, e.g. *State* v. *DelVecchio*, 191 Conn. 412, 417–18, 464 A.2d 813, 817–18 (1983).

[6] *State* v. *Casey*, No. 19940, 2004 WL 405738, at *6 (Ohio App. 5 March 2004).

[7] *Petrocelli* v. *Angelone*, 248 F.3d 877, 888 (9th Cir. 2000).

[8] See, e.g. *People* v. *Hughes*, Nos. C036658, C036888, 2002 WL 725133, at *7-*8 (Cal. Ct. App. 24 April 2002).

[9] *Commonwealth* v. *Rosa*, 422 Mass. 18, 27, 661 N.E.2d 56, 62 (1996).

[10] *Ibid.*, 422 Mass. at 28, 661 N.E.2d at 63.

[11] *United States ex rel. Bartlett* v. *Briley*, No. 04 C 4204, 2005 WL 217029, at *6 (N.D. Ill. 25 January 2005), *affirmed*, 453 F.3d 796 (7th Cir. 2006).

[12] See, e.g. *Wilson* v. *Maryland*, 370 Md. 191, 214, 803 A.2d 1034, 1047 (2002) (evidence about frequency of sudden infant death syndrome (SIDS) deaths was admitted at trial; ‘[t]he State’s Attorney . . . improperly used the statistics to argue that there was only a minuscule possibility that the defendant was innocent. The State’s Attorney was well aware that the statistical evidence could not be used to calculate the probability of petitioner’s innocence. . . . The courts that have considered this issue have concluded that it is impermissible to assign a number to the probability of guilt or innocence.’); but cf. *State* v. *Pankow*, 144 Wis.2d 23, 38–40, 422 N.W.2d 913, 918–19 (Wis. App. 1988) (admission of ‘mathematical probability evidence’ about frequency of SIDS deaths did not constitute an impermissible attempt to quantify the reasonable doubt standard); *see also State* v. *Bailey*, 677 N.W.2d 380, 403 (Minn. 2004) (‘We have cautioned the state to avoid any attempt to equate DNA probability statistics with proof beyond a reasonable doubt’.) (quoting *State* v. *Carlson*, 267 N.W.2d 170, 176 (Minn. 1978)) (citing TRIBE, L. “Trial by mathematics: precision and ritual in the legal process,” 84 *Harvard Law Review* 1329) (1971) (‘Testimony expressing opinions or conclusions in terms of statistical probabilities can make the uncertain seem all but proven, and suggest, by quantification, satisfaction of the requirement that guilt be established “beyond a reasonable doubt.” ’)

quantification of the reasonable doubt standard.[13] But such cases are rare, and in general, the proposition that mathematical quantification of the reasonable doubt standard is impermissible appears to be established beyond a reasonable doubt.[14]

It could be argued that judicial hostility to quantification of reasonable doubt is only a transitory state of affairs. Many American judges now accept that mathematical and quantitative methods can shed light on many legal problems. All American judges now either accept or must accept that the results generated by mathematical and quantitative methods are often admissible at trial. Perhaps, the ever-increasing use of mathematical and quantitative methods in litigation both foreshadows and reflects a transformation in judicial attitudes towards the hard sciences. Perhaps, such a change in the intellectual culture of the judiciary will create fertile judicial soil for the eventual 'mathematization' of the reasonable doubt standard. Perhaps so, but solid evidence that mathematization of the reasonable doubt standard will come to pass is hard to find.

Consider Judge Weinstein, a leading authority on the American law of evidence. He has long advocated more extensive forensic use of statistical methods. If any reputable judge were to advocate quantification of the reasonable doubt standard, one might expect that Weinstein would be the one to do so. A search of Weinstein's judicial record does show that Weinstein has written two opinions that discuss quantification of the reasonable doubt standard. See *United States* v. *Fatico*, 458 F. Supp. 388, 409–11 (E.D.N.Y. 1978) and *Vargas* v. *Keane*, 86 F.3d 1273, 1281–84 (2nd Cir. 1996) (Weinstein, concurring, sitting 'by designation'—i.e. temporarily—on the United States Court of Appeals for the Second Circuit). However, neither of these opinions directly embraces quantification of the reasonable doubt standard. In these opinions, Weinstein reports the results of two surveys that were conducted under his auspices. In both surveys—one survey canvassed some federal trial

[13] See, e.g. *In re As.H.*, 851 A.2d 456 (D.C. 2004). The only issue was the identity of the perpetrator. The only evidence was the testimony of an eyewitness, an eyewitness who testified that 'her level of certainty . . . [was] "seven or eight" on a scale of one to ten'. *Ibid.* at 457. The District of Columbia Court of Appeals held: 'It is, of course, difficult (if not impossible) to place a meaningful numerical value on reasonable doubt . . . . Here, the doubt of the sole identifying witness in a night-time robbery by strangers to her stood at two or three out of ten, or 20-30%. We conclude, at least on this record, that this level of uncertainty constituted reasonable doubt as a matter of law.' *Ibid.* at 461–62 (citations omitted).

Judge Farrell, dissenting, wrote 'that the entire effort to quantify the standard of proof beyond a reasonable doubt is a search for fool's gold'. *Ibid.* at 463. He added, 'Even in popular usage, the "scale of one to ten" as an indicator of belief is notoriously imprecise. People who in any ultimate sense—and unascertainable—sense probably share the same level of conviction may translate that very differently into numbers. . . Treating "one to ten" as a decisive indicator of the sufficiency of identification evidence thus elevates to a legal standard a popular measure that makes no claim at all to precision.' *Ibid.* at 464.

Judge Farrell presents us with an instance in which a judge complains about quantification on the ground that a particular numerical measure of probative value is too 'imprecise'. The more common complaint is that mathematical measures of the probative value of evidence are too 'precise'. See text *infra* at 2, p. 144 — (discussing 'The myth of excessive mathematical precision').

[14] In addition to the cases already cited, see, e.g. *People v. Collins*, 68 Cal.2d 319, 331-32, 66 Cal. Rptr. 497, 504-05 (1968); *State v. Spann*, 130 N. J. 484, 497, 501, 617 A.2d 247, 253-55 (1993); *Evans v. State*, 117 Nev. 609, 631-32, 28 P.3d 498, 514 (2001); *State v. Rizzo*, 266 Conn. 171, 221, 833 A.2d 363, 399 (2003).

The idea of viewing the reasonable doubt standard through a mathematical lens does not seem absurd to some academics. During the last 35 years, many students of the law of evidence have reached the conclusion that mathematical methods can shed light on problems of evidence and inference in legal settings. See, e.g. KAYE, D., "The laws of probability and the law of the land," 47 *University of Chicago Law Review* 34 (1979); SCHUM, D., THE EVIDENTIAL FOUNDATIONS OF PROBABILISTIC REASONING (John Wylie & Sons 1994). Many of these 'new evidence scholars' – see LEMPERT, R., "The new evidence scholarship: analyzing the process of proof," 66 *Boston University Law Review* 439 (1986) – have used mathematical concepts and numbers to discuss the reasonable doubt standard. See, e.g. MARTIN, A. & SCHUM, D., "Quantifying burdens of proof: a likelihood ratio approach," 27 *Jurimetrics Journal* 383 (1987); HAMER, D., "Probabilistic standards of proof, their complements, and the errors that are expected to flow from them," 1 *University of New England Law Journal* 71 (2004).

judges; the other, jurors—the respondents were asked to use numbers (in the form of percentages) to express their judgement of the amount of certainty that is required for conviction of crime under the reasonable doubt standard.[15] To the extent that the actual language of Fatico and Vargas is a true barometer of Weinstein's motivations for conducting those surveys, Weinstein was investigating how judges and jurors view the reasonable doubt standard, not because he was working towards the day when the fuzzy legal concept of reasonable doubt would be replaced with a crisp mathematical formulation, but because he wanted to 'precisiate'[16] the reasonable doubt standard—to explain, precisely, how the fuzzy concept of reasonable doubt works in trials. If even math-friendly judges such as Jack Weinstein do not endorse the use of numbers in criminal trials to clarify or reformulate the reasonable doubt standard, the prospects for mathematical quantification at trial of the reasonable doubt standard would seem to be virtually nonexistent.

But what are we to make of another decision by the very same Jack B. Weinstein: *United States* v. *Copeland*, 369 F. Supp. 2d 275 (E.D.N.Y. 2005)? In Copeland, Weinstein used a numerical probability (expressed as a percentage) to quantify a standard of persuasion ('reasonable probability'). Is Copeland compatible with the prevailing rule that reasonable doubt cannot be quantified in trials? If 'substantial probability' can and should be quantified, why cannot and why should not 'reasonable doubt' be quantified? Does *Copeland* amount to a collateral attack on the rule prohibiting quantification of the reasonable doubt standard?

2. *United States* v. *Copeland* (*Copeland III*)

In 1978, Richard Copeland, a Jamaican citizen, was adopted by his grandmother, a naturalized citizen of the United States. When Copeland moved to the United States in 1982 at the age of 12, he became a lawful permanent alien of the United States. But from 1988 to 1995, Copeland committed serious crimes on four separate occasions. He was convicted of committing one crime on each of those four occasions. These four convictions, however, greatly understate the gravity of Copeland's criminal misbehaviour on those four occasions. (We have plea bargaining to thank for this.)

Until 1995, Copeland had managed to avoid extensive imprisonment; he had spent only 3 days in prison. In October 1995, however, Copeland's fortunes changed; he was sentenced to a term of imprisonment for 18–54 months. After he was imprisoned, the U.S. Immigration and Naturalization Service (INS) (as it was then called) began a deportation proceeding against him. Two deportation hearings were held in 1996 before an INS immigration judge. These hearings resulted in a deportation order against Copeland late in that same year. After further proceedings before the INS, Copeland was deported in 1998.

Copeland illegally re-entered the United States in 1999. He was later indicted for illegally re-entering the United States after having been removed from the United States. Copeland moved to dismiss this indictment on the ground that the 1996 deportation order that ultimately resulted in his removal (by deportation) was fundamentally unfair and therefore unlawful. In *United States* v. *Copeland*, 228 F. Supp. 2d 267 (E.D.N.Y. 2002) (*Copeland I*), Judge Weinstein accepted Copeland's argument.

[15] In Vargas, the respondents—jurors—were asked to use quantitative mathematical language—the language of percentages—to describe how sure they had to be of the guilt of the defendant under different nonnumerical formulations of the reasonable doubt standard.

[16] This word is Lotfi Zadeh's neologism. See ZADEH, L., "Toward a generalized theory of uncertainty (GTU)—an outline,". 172 *Information Sciences* 1 (2005).

Under the law applicable in 1996, Copeland was eligible for discretionary relief from a deportation order. However, the immigration judge who conducted the 1996 deportation hearings erroneously concluded that Copeland was not eligible for discretionary relief from deportation and he so advised Copeland. Judge Weinstein wrote:

> The deportation order was fundamentally unfair . . . because the Immigration Judge not only failed to advise the defendant of the existence of discretionary relief, but affirmatively misled him by indicating that he was ineligible for such relief. This misadvice improperly, as it turned out, discouraged the defendant from seeking discretionary relief. As a result, no immigration judge ever determined whether the defendant was worthy of [discretionary] relief [from deportation]. Fundamental unfairness was prejudicial to the defendant because there is a reasonable likelihood he would have been granted [discretionary] relief.

Copeland I at 271–272.

Judge Weinstein dismissed the indictment against Richard Copeland. The government appealed to the U.S. Court of Appeals for the Second Circuit. In *United States* V. *Copeland*, 376 F.3d 61 (2d Cir. 2004) (*Copeland II*), the Second Circuit reversed. Although the Second Circuit did not reject Weinstein's finding that the INS immigration judge had violated due process, it did rule that Weinstein's dismissal of the indictment of Richard Copeland was improper because the indictment could be dismissed, only if the due process violation committed by the immigration judge required was 'prejudicial'—only if the immigration judge's violation was not harmless error—and Judge Weinstein had failed to consider certain evidence relevant to the question of whether the immigration judge's violation of Copeland's due process rights was or was not harmless. In addition, the Second Circuit further said:

> . . . Copeland must show that he likely would have been granted [discretionary] relief if he had obtained a hearing. Have not decided what level of proof is required for a showing that an alien likely would not have been removed, but we have flirted with two possible standards: a 'reasonable likelihood' and a 'plausible showing'.
>
> In our view . . . the appropriate test for prejudice is the one used to decide ineffective assistance of counsel claims, namely, prejudice is shown where 'there is a reasonable probability that, but for counsel's unprofessional errors, the result of the proceeding would have been different', *Strickland v. Washington* 466 U.S. 668, 694, 80 L. Ed. 2d 674, 104 S. Ct. 2052 (1984).

Copeland II, 376 F.3d at 73 (citation omitted).

The Second Circuit remanded the case to Weinstein, instructing him to determine whether or not there was a reasonable probability that Richard Copeland was prejudiced by the immigration judge's error. Judge Weinstein responded in *United States* v. *Copeland*, 369 F. Supp. 2d 275 (E.D.N.Y. 2005) (*Copeland III*). He effectively asked himself, 'How much of a probability is a "reasonable probability"?' Weinstein gave a startlingly but disarmingly forthright answer to his own question: in this case, he would equate the Second Circuit's requirement of a 'reasonable probability' with a probability of 0.2—or, as Weinstein chose to put it, 20%. Weinstein explained his choice of the number 20% thus:

> Typically, courts have not quantified burdens of proof. *See* J. Maguire, J. Chadbourn., J. Mansfield, Et Al., CASES AND MATERIALS ON EVIDENCE 871-73 (6th ed. 1973) (collecting literature on quantification of burdens of proof); Richard H. Gaskins, BURDENS OF PROOF IN MODERN DISCOURSE 20 (1992) (discussing implications of burdens of proof); Terence Anderson & William Twining, ANALYSIS OF EVIDENCE 338 (1991) (correlating verbal and mathematical measures of certainty and doubt).
>
> The term 'reasonable probability' should be quantified to the extent possible, given the difficulty of assessing what another adjudicator would have done when applying complex and subjective criteria.. . . Agreement on quantification, while not a solution, does help move beyond the mere anecdotal to at least a rough consensus in application.
>
> Quantification requires analysis in terms of probability.
>
> While 'reasonable probability,' the term of art selected by the Court of Appeals, seems deliberately designed to be fuzzy in concept and articulation, it is suggested that a probability of 20%—the approximate inverse of 'clear, unequivocal and convincing evidence'—represents a sensible and enforceable standard, considering that deportation often has such serious consequences for the deportee and his or her family.
>
> * * *
>
> When, as here, a relatively 'simple fact'—what happened in the real world of defendant's life—is combined with what an unknown administrative judge would have done in evaluating the evidence supporting that finding of 'fact', and analyzing the 'fact' in the context of a 'legal rule', the problem of determining how the judge would have decided the 'law-fact' issue is complex. It is compounded by many factors—among them the egocentricity of the judge. At most a band of probabilities is all that we can expect. Since the defendant's constitutional rights have been violated he is entitled, it is submitted, to the most favorable band border—here, it is proposed, 20%. An attempt to quantify in order to provide some uniformity in application of the rule is justified even though it must be conceded that the percentage chosen is based on public policy favoring enforcement of constitutional rights and somewhat arbitrary.
>
> The present illegal reentry case involves a potentially unconstitutional deportation of great consequence to a deportee-defendant. . . . Unlike a standard deportation case, which would be nominally civil, this case also involves a criminal prosecution. Requiring a petitioner to meet a burden greater than 20% to establish a 'reasonable probability' that he would have been granted section 212(c) [discretionary] relief would therefore seem unfair and unreasonable.

Copeland III at 286–88 (some citations and internal quotation marks omitted).

Is Judge Weinstein's quantification of the 'reasonable probability' standard (together with his quantification of the complementary 'clear and convincing evidence' standard[17]) compatible or incompatible

[17] In *Copeland III*, Weinstein quantified the 'clear and convincing evidence' standard as well as the 'reasonable probability' standard. He said, not unreasonably, that the 'inverse' of the 'reasonable probability' requirement is the 'clear and convincing

with the customary legal maxim that the standard of 'proof beyond a reasonable doubt' is not quantifiable? Wags have noted that lawyers are trained to distinguish the indistinguishable. Viewed from a technical legal perspective, Weinstein's decision in *Copeland III* is not directly inconsistent with the customary legal rule barring quantification of reasonable doubt.[18] Broadly viewed, however, *Copeland III* raises serious questions about the soundness of the well-entrenched legal maxim that the reasonable doubt standard is unquantifiable. If standards such as 'reasonable probability' and 'clear and convincing evidence' can be quantified, why can't reasonable doubt be quantified? What is it, precisely, that allegedly makes it impossible or inadvisable to quantify reasonable doubt?

The objectives of this case comment are limited. We make no attempt to answer the question whether it is advisable or inadvisable to quantify proof beyond a reasonable doubt. But in the remainder of this case comment, we do try to clarify how the question of quantification should be debated. We try to show that some of the usual objections to quantification of the reasonable doubt standard are without merit. We further try to show that there is a very important sense in which the reasonable doubt standard involves 'quantification' even if no numbers are used. Finally, without taking a position on the ultimate question of whether some sorts of numbers should be used to characterize the standard of persuasion in criminal trials, we describe several considerations that provide a promising foundation for arguments in favour of numerical quantification of the reasonable doubt standard, and we describe (in a general way) the kinds of rejoinders that can and cannot be advanced against such arguments.

3. Myths about quantification of reasonable doubt

The myth of 'trial by mathematics (or statistics)'

The language of some judicial opinions suggests that some judges believe that quantification of the reasonable doubt standard entails the vice of trial by *statistics*.[19] Now trial by statistics—whatever it is—might or might not be a bad thing. But it is important to understand that 'trial by mathematics' does not necessarily entail 'trial by statistics'. Assume that the phrase trial by mathematics refers to trials in which decisions at trial are governed by the (use of the methods of) probability calculus. Assume further that a judicial trial becomes a trial by mathematics, if the law quantifies burdens of persuasion in criminal trials by informing triers of fact that they may find a defendant guilty of crime (or find facts essential to criminal guilt) if and only if they believe that the probability of criminal guilt (or of each fact essential to guilt) exceeds some specified numerical probability. This sort of trial by mathematics—if it *be* trial by mathematics—does not necessarily involve statistics.

evidence requirement'. A probabilist would say that Weinstein viewed the clear and convincing evidence requirement as the complement of the reasonable probability requirement. This is why Weinstein concluded that the clear and convincing evidence requirement amounts to a probability requirement of 0.8 (80%). Having concluded that a reasonable probability amounts to 0.2 and that a clear and convincing evidence requirement is the probability of the negation of the matter that has a reasonable probability, Weinstein had no choice but to conclude that a clear and convincing probability must be 0.8; the probability of an event and the probability of the negation of that event must sum to one (1).

[18] This is because (i) the reasonable doubt standard—which governs only decisions at trial by triers of fact about the ultimate question of criminal guilt or innocence—was not directly in question in *Copeland III*, (ii) in *Copeland III*, Weinstein had nothing directly to say about the feasibility or advisability of quantifying the reasonable doubt standard and (iii) there is at a minimum, a plausible argument that the two questions—quantification of reasonable doubt and quantification of other standards of proof and persuasion—are fundamentally different.

[19] See, e.g. *Rosa*, 422 Mass. at 28, 661 N.E.2d at 63 ('We are aware that to attempt to quantify proof beyond a reasonable doubt changes the nature of the legal concept of "beyond a reasonable doubt," which seeks "abiding conviction" or "moral certainty" rather than statistical probability').

Probabilities are not the same thing as statistically grounded probabilities. Yes, modern statistical analysis does involve the probability calculus. But, as the word ‘statistics’ implies, statistical analysis involves and requires systematic collection of data or observations, data and observations that can be summarized in the form of *statistics*. It is possible to talk—and talk coherently—about odds or probabilities without systematically gathering data, compiling statistics or analysing systematically gathered collections of data. In short, although it is not possible to do statistics without doing probability, it is possible to do probability without doing statistics. Hence, any uneasiness about the use of statistical methods in criminal trials does not explain the judiciary’s uneasiness about quantification of the reasonable doubt standard.

The language of some judicial opinions suggests that some judges believe that quantification of the reasonable doubt standard would constitute an invidious type of mathematization of criminal trials simply because a quantified standard of persuasion would require the trier of fact to assess the trier’s own degree of uncertainty about criminal guilt or innocence by thinking and deliberating in terms of numbers in the form of probabilities or odds. Is there something inherently invidious and harmful about this sort of mathematization of the reasonable doubt standard?

A scholarly debate about the virtues and vices of mathematical analysis of evidence has raged for more than three decades. The outcome of that debate remains unclear: it is unclear whether the proponents or the opponents of mathematical analysis of evidence and inference will ultimately prevail. (Tillers confesses that he’s betting on the advocates of heuristic mathematical analysis.) But one thing about that long-running and often acrimonious debate is relatively clear: most of that debate is immaterial to the question of quantification of the reasonable doubt standard. Scholarly arguments about mathematical analysis of evidence and inference largely have to do with the logic or structure of *argument* about and from evidence—i.e. the logic or structure of factual or evidential *inference* or *evidential argument*. Like other forms of inference, evidential inference involves at least one step—a step, e.g. from an evidential premise to a factual conclusion. (That a step is required is the reason why we call the step ‘inference’.) Disagreements about mathematical analysis of evidence and inference mainly involve disagreements about how inferences are or should be drawn. The sorts of quantitatively phrased standards of persuasion under discussion here do not implicate controversies about the structure of evidential inference because the type of quantification under discussion here specifies only how much uncertainty is acceptable at the end of the day—after the trier has used whatever logic it chooses to use to draw inferences from and about the available evidence. Quantified standards of persuasion of this sort appear to say nothing about the kind of logic or reasoning the trier should use to reach its final (uncertain) conclusion.

Quantification of bottom-line inferences does contemplate that the trier of fact will measure and express its uncertainty by using the language of probabilities and odds. But it is hard to see why a trier’s use of the language of probabilities and odds to describe the extent of its uncertainty about its ultimate factual conclusions compels the trier to use any particular method for drawing inferences from and about evidence, let alone a method of inferential analysis that is rooted in the standard probability calculus. For example, it seems possible for a person to state what he or she thinks the chances of rain are without having to do a mathematical analysis of the evidence that bears directly or indirectly on the hypothesis of rain or no rain. (One would not say that a person who says, ‘The odds of rain are two to one’ has done or must do a mathematical analysis of the chances of rain.) Similarly, it seems possible—not illogical or incoherent—for a trier of fact such as a juror to formulate an opinion of the odds of a defendant’s guilt, even though the trier has not done a mathematical analysis of the evidence of guilt or innocence.

If numerical quantification of a standard of persuasion does not require that mathematics or numbers be used to analyse evidential inference, not much is left of the claim that quantification of a standard of persuasion amounts to trial by mathematics. It must be granted, of course, that quantification of the reasonable doubt standard in terms of odds, probabilities or chances—e.g. quantification by a rule providing that proof beyond a reasonable doubt exists and a juror can vote to convict only if the juror believes that there is more than a 95% chance that the defendant is guilty (or, alternatively, by a rule which provides that a juror can vote to convict only if the juror believes that the chances are greater than 95% that every fact essential to conviction, such as the fact that defendant caused the victim's death, is true)—such a quantification of the burden of persuasion in criminal trials would require a trier such as a juror to use numbers when interrogating itself about the sufficiency and strength of the evidence against an accused. It might even be necessary to grant that even this sort of limited quantification of the reasonable doubt standard does in a sense entail trial by mathematics—precisely because in trials in which odds or probabilities are used to characterize the government's burden of persuasion, triers of fact would be required to use *some* sort of numbers—and, hence, mathematics—when making decisions bearing on guilt or innocence. But so what? Numbers are not inherently evil things. The use of numbers to express the degree of a person's uncertainty about a factual possibility does not require the use of higher mathematics—or even intermediate mathematics. Arithmetic will do.

Unless the objection to trial by mathematics has become entirely talismanic and formulaic, a plausible objection to quantification of the reasonable doubt standard must have some thrust or basis beyond the mere fact that overt quantification ordinarily involves the use of numbers. For example, a nonridiculous argument against the use of odds and probabilities to characterize reasonable doubt in trials could rest on the premise that the inevitability of uncertainty about matters of fact should be kept hidden from jurors, on the premise that judges should not decide how much uncertainty about criminal guilt is acceptable, or on the premise that juries ought to be empowered to decide (within the broad limits imposed by the reasonable doubt standard) how much uncertainty is acceptable or unacceptable in the criminal trial in which they happen to be sitting as jurors. See discussion at pp. 21–22, below. But it borders on the ridiculous to argue that the use of numerical probabilities or odds in criminal trials is harmful simply because the use of numbers (or quantitative expressions) in criminal trials is harmful.

The curious myth of 'mathematical certainty'

Occasionally, it is said that mathematical analysis of evidence or mathematical accounts of inference is unacceptable because mathematical analysis aims for a kind of certainty—mathematical certainty—that is unattainable in ordinary affairs or in inferential deliberation.[20] Is it conceivable that this sort of argument would be made about quantification of the reasonable doubt standard—that quantification of the reasonable doubt standard would somehow convert the standard into one that requires mathematical certainty of guilt? We hope not. But if the argument were to be made, it would be so preposterous that it might be difficult to know what to say about it.

[20] Cf. *People* v. *Harbold*, 124 Ill. App. 3d 363, 383, 464 N.E.2d 734, **7**49 (Ill. App. Ct. 1984) ('Testimony expressing opinions or conclusions in terms of statistical "probabilities" can make the uncertain seem all but proven, and suggest, by quantification, satisfaction of the requirement that guilt be established "beyond a reasonable doubt." ' (quoting *State* v. *Carlson* 267 N.W.2d 170, 176 (Minn. 1978)) (emphasis added).

The objection that mathematical analysis of evidence and inference entails a (spurious) mathematical certainty about evidence and inference fundamentally misconceives the entire point of using probability theory to analyse factual proof. Although it is possible that a particular argument rooted in probability theory will demonstrate that a particular factual proposition is certain, the general purpose of using probability theory is not to eliminate factual uncertainty. Probability theory assumes the existence of uncertainty. (One might say that probability theory takes uncertainty *seriously*.) The entire point of using probability theory is to talk coherently about uncertainty—not to eliminate uncertainty.

The myth of excessive mathematical precision

Courts often suggest that quantification of the reasonable doubt standard entails *precise* quantification of the standard—and that such precise quantification would be a bad thing because a quantitatively precise formulation of the burden of persuasion in criminal trials would be *excessively* precise. Consider, e.g. the following comment by the U.S. Supreme Court about the meaning of ‘probable cause’, a comment that seems to presuppose that quantification of probable cause is necessarily *precise* quantification:

> The long-prevailing standard of probable cause protects ‘citizens from rash and unreasonable interferences with privacy and from unfounded charges of crime’, while giving ‘fair leeway for enforcing the law in the community’s protection’. On many occasions, we have reiterated that the probable-cause standard is a practical, nontechnical conception that deals with the factual and practical considerations of everyday life on which reasonable and prudent men, not legal technicians, act. Probable cause is a fluid concept—turning on the assessment of probabilities in particular factual contexts—not readily, or even usefully, reduced to a neat set of legal rules.
>
> The probable-cause standard is incapable of precise definition or quantification into percentages because it deals with probabilities and depends on the totality of the circumstances. We have stated, however, that the substance of all the definitions of probable cause is a reasonable ground for belief of guilt, and that the belief of guilt must be particularized with respect to the person to be searched or seized.

Maryland v. *Pringle*, 540 U.S. 366, 370–371 (2003) (citations and internal quotation marks omitted).

The objection to quantification of standards of persuasion on the ground that quantified standards are precise may seem to require no explanation. The notion of ‘precise quantification’, however, has various connotations, and each of these connotations seems to have different wellsprings.

In some instances, the thesis (or suspicion) that quantification of matters such as probable cause and reasonable doubt necessarily produces an excessive and spurious degree of precision about uncertainty may be rooted in the following two related assumptions:

(i) Any quantification of the reasonable doubt standard in terms of probabilities would have to use relatively discrete rather than relatively coarse probabilities—such as probabilities that run to three or even to five or more decimal places, e.g. the probability 0.953 or the probability 0.95312 and

(ii) The degree of doubt and uncertainty about matters such as criminal guilt is necessarily, relatively imprecise; it is always comparatively coarse.

The objection to quantification of standards of persuasion is not well grounded if it rests only on these two propositions. It is very probably true that triers' uncertainty about many types of facts that are legally essential to a finding of criminal guilt—about possible facts such as 'intent to kill'—is ordinarily, relatively coarse. However, nothing in mathematical logic or in probability theory dictates that mathematical measures of uncertainty must be highly granular. Today there is an entire family of mathematical theories of uncertainty that are dedicated to the study of 'imprecise probabilities'.[21] Even before the advent of nonstandard mathematical approaches to uncertainty, it was well known that probabilities can be imprecise. For example, students of probability and statistics have been familiar with the following matters for quite a long time:

- Mathematical probabilities can be stated with varying degrees of discreteness, or granularity. For example, a probability can be stated with the discreteness of 0.23167 (five decimal places) or 0.23 (two decimal places) or 0.2 (one decimal place). From a mathematical perspective, the selection of any one of these numerical probabilities rather than some other is a matter of choice.
- Point probabilities can be both relatively coarse and also approximate—e.g. $P(X) \approx 0.3$ is a permissible expression, one that translates as 'the probability of X is approximately 0.3'.
- It is mathematically possible to express uncertainty about the location of a point probability—e.g. $P(Y) = 0.5 \pm 0.1$.
- Probabilities can be expressed in the form of intervals (e.g. 'a probability between 0.2 and 0.4, inclusive'), rather than as point probabilities.[22]
- The terminal points of probability intervals can be approximate—e.g. 'the probability of rain is between approximately 20% and approximately 30%'.
- It is possible to quantify uncertainty and make valid mathematical arguments about uncertainty without using any (cardinal) numbers, whatever, let alone very discrete (cardinal) numbers. There is a mathematics of inequalities, and this sort of math can be used to quantify uncertainties without assigning cardinal numerical values to uncertainty. For example, consider
    - the expression $[P(R) > P(\sim R)]$, which can be translated as 'the probability of R is greater than the probability of not-R'', which might be interpreted as 'the probability rain is greater than the probability of not-rain';
    - the expression $[P(R) \gg P(\sim R)]$, which can be translated as 'the probability of R is much greater than the probability of not-R'; and
    - the argument [if $\{P(A) \gg P(Q)\}$ & $\{P(Q) > P(J)\} \rightarrow P(A) \gg P(J)$], which can be translated as 'if the probability of A is much greater than the probability of Q AND

[21] See, e.g. Berkeley Initiative in Soft Computing (BISC), Electrical Engineering and Computer Sciences Department, University of California at Berkeley. Available at http://www-bisc.cs.berkeley.edu/. Accessed 26 September 2006; JIN, Y. *A Definition of Soft Computing*. Available at http://www.soft-computing.de/def.html. Accessed 26 September 2006. ('Soft computing differs from conventional (hard) computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth and approximation. In effect, the role model for soft computing is the human mind.')

See also WALLEY, P., STATISTICAL REASONING WITH IMPRECISE PROBABILITIES (1991).

[22] *Copeland III* alludes to probability intervals: 'At most a band of probabilities [for the legal standard of "reasonable probability"] is all that we can expect'. 369 F. Supp. 2d at 288.

if the probability of Q is greater than the probability of J, THEN the probability of A is much greater than the probability of J'. This argument can be given an interpretation such as the following: 'if the probability of an Ace is much greater than the probability of a Queen AND if the probability of a Queen is greater than the probability of a Jack, then the probability of an Ace is much greater than the probability of a Jack'.

In sum, the notion that there is something in the nature of mathematics that dictates that a quantified standard of persuasion must have a single discrete cardinal numerical point value is a myth.

The objection to 'precise quantification' of burdens of persuasion sometimes may have a basis entirely different from the (erroneous) notion that mathematical probabilities must be granular. Consider again the passage by the U.S. Supreme Court quoted above. In part of that passage, the Court emphasized that precision about probable cause is bad because probable cause 'depends on the totality of the circumstances'. *Pringle*, 540 U.S. at 371. The evil hinted at by this part of the Court's language is not any excessive granularity of probability judgements, but the *invariability* of the degree of probability that, the Court suggests, would be required for a finding of 'probable cause' were the probable cause requirement quantified.

The notion that mere use of the language of mathematical probability to describe the relationship between uncertainty and probable cause requires that 'probable cause' be assigned some invariant numerical ('mathematical') probability is almost silly beyond words. The choice of a numerical probability to express the meaning of probable cause in a variety of situations could only be a social, political, legal or normative choice; such a choice could not rest on mathematical considerations. The Court would have spoken more honestly—or, in any event, more plainly—if it had simply said that constitutional law does not supply a trans-situational rule that describes how large a probability is necessary for probable cause'. The only sensible way to view the Supreme Court's caveat that the meaning of probable cause is 'fluid' and depends on the 'totality of the circumstances' is to view it as an objection to the use of a *uniform*, or *invariant*, probability to describe the requirements of a standard such as probable cause.[23]

The U.S. Supreme Court's authority and influence alone practically guarantee that other courts will sometimes use the word 'precision' to lodge an objection to quantification of some standard of persuasion (such as probable cause), not on the ground that quantification necessarily involves granular probabilities, but on the ground that the probabilities required for some burdens of proof (such as probable cause) must be allowed to vary from case to case (or, in the jargon of judges and lawyers, such burdens should be 'case specific'). Such a thesis is not intrinsically absurd. But it has nothing to do with the supposed *precision* of numerical probabilities.

[23] Unlike the jurisprudence of the probable cause standard, the official jurisprudence of the reasonable doubt standard (in the U.S.) generally does not openly admit that the level of probability required to satisfy the reasonable doubt standard is 'fact specific': the orthodox gospel is that the demands of the reasonable doubt standard remain the same from case to case and are 'not' affected by the 'totality of the circumstances' in each case**.** See, e.g. *Pringle*, 540 U.S. at 371 ('[T]he quanta of proof appropriate in ordinary judicial proceedings are inapplicable to the decision to issue a warrant. Finely tuned standards such as proof beyond a reasonable doubt or by a preponderance of the evidence, useful in formal trials, have no place in the probable-cause decision.') (quotations and citations omitted). However, capital punishment cases effectively fall into a special category and are treated differently**.** See, e.g. *Snipes* v. *State,* 733 So.2d 1000, 1007 (Fla. 1999**)** ('A proportionality review involves consideration of the totality of the circumstances of a case and comparison of that case with other death penalty cases. When we compare the totality of the circumstances of this case to other similar cases, we conclude that a sentence of death is inappropriate.'); *State* v. *Cohen,* 604 A.2d 846, 849 (Del. 1992) (discussing 'totality of the circumstances analysis' in death-penalty case).

The myth of an absolute disjunction between qualitative and quantitative judgements

Courts frequently declare that the reasonable doubt standard requires the trier of fact to make qualitative rather than quantitative judgements.[24] To make sense of this proposition—to make it amount to more than the tautology that a verbal formulation of the reasonable doubt standard is not a numerical formulation—it is necessary to understand it as an assertion that judgements about states of the world are either qualitative or quantitative, but not both. If this is the kind of notion that is at work here, it is hard to understand.

Perhaps, the thesis of a disjunction between quantitative and qualitative judgements rests on the premise that numbers somehow speak for themselves—and that, thus, no qualitative human thinking is required when numbers are involved in an argument or assessment. There are any number of difficulties with this idea. The first is that numbers often do not come into existence 'on their own'. That is the case here, where numbers are not being used to tally—to enumerate—the number of entities (such as automobiles) in some domain (such as some street or city). Furthermore, even after numbers have appeared or have been made to appear, they must usually be interpreted by human actors and often arguments about the significance of the available numbers for the thesis in question must be constructed and assessed. Such activities seem to involve 'qualitative' mental processes as well as quantitative ones.

Perhaps, the thesis of a disjunction between quantitative and qualitative judgements about evidence involves the notion that mathematical procedures for the assessment of evidence amount to mechanical recipes—'algorithms'—that automatically—or, in any event, in a machine-like fashion—determine the probative value of evidence. But debates about the advantages and disadvantages of 'algorithmic' methods of analysing evidence are beside the point here: Algorithmic reasoning would not be required by a quantified legal standard of persuasion that merely specifies the level of certitude that must exist in the mind of a trier of fact if the trier is to take some action such as casting a vote in favour of verdict of guilty in a criminal case. A quantified legal rule of this sort assumes that the trier somehow reaches a conclusion about his level of certitude. It does not describe the type of reasoning that the trier should use to reach a conclusion about his or her level of certitude or incertitude. See discussion at pp. 7–9.

The myth of the unquantifiability of degrees of belief

More than half a century ago, the dean of all scholars of the Anglo-American law of evidence—John Henry Wigmore—wrote:

> The truth is that no one has yet invented or discovered a mode of measurement for the intensity of human belief. Hence there can be yet no successful method of communicating intelligibly to a jury a sound method of self-analysis for one's belief. If this truth be appreciated, courts will cease to treat any particular form of words as necessary or decisive in the law for that purpose; for the law cannot expect to do what logic and psychology have not yet done.

[24] See, e.g. *McCullough* v. *State*, 99 Nev. 72, 75–76, 657 P.2d 1157, 1159 (1983).

9 John H. Wigmore, EVIDENCE IN TRIALS AT COMMON LAW Section 2497 (3d ed. 1940)

Wigmore's language is sweeping. The sentiment it expresses is practically hypermodern. Just as Kenneth Arrow argued that interpersonal comparisons of preferences are impossible,[25] Wigmore seemed to suggest that interpersonal comparisons of the strength of credal states—interpersonal comparisons of the strength of beliefs about states of the world—are impossible. Indeed, Wigmore seemed to go yet further: he seemed to assert that 'intrapersonal' comparisons of the strength of credal states are also impossible—that individuals cannot compare the degree of their own uncertainty about the truth or falsity of different propositions about the world. In short, Wigmore seemed to suggest that in the end, we just feel that this or that proposition is true or false and that we cannot tell others or even ourselves just how strongly we feel that this or that proposition is in fact true or false.[26]

If it is true that both intrapersonal and interpersonal comparisons of degrees of persuasion or degrees of uncertainties are impossible, it seems to follow that *all* legal rules mandating a certain level of certitude on the part of the trier of fact in specified situations are both meaningless and useless. In particular, if the degree of human certainty about facts (or, if you prefer, the intensity of human beliefs about facts) is neither measurable nor communicable, and if standards of persuasion are legal rules that require triers of fact to have a specified degree of certitude about specified types of facts (or, if you prefer, a specified degree of intensity in their beliefs about specified types of facts) before triers of fact in litigation report that they have made certain factual findings, it seems to follow that

- the reasonable doubt standard itself (and not just differences in how the reasonable doubt standard is formulated or explained) is meaningless and pointless;
- there is no useful or meaningful difference between a legal requirement that certain facts must be established beyond a reasonable doubt and a legal requirement that certain facts in certain kinds of cases must be established only by a preponderance of the evidence (or, as the English say, by a balance of the probabilities);
- a trier of fact can *never* determine whether or not the requirement of proof beyond a reasonable—or any other standard of persuasion—has been satisfied; and
- it would be best if legal standards of persuasion ceased to exist.

But American law on standards of persuasion does not bear traces of such hyper-skepticism. A legally mandated standard of persuasion for criminal trials—the reasonable doubt standard—does exist. Furthermore, American law mandates the use of various other standards of persuasion for other kinds of cases and situations. This is some evidence that American lawmakers (i.e. legislators and judges) generally believe or assume that there is indeed a difference, e.g. between proof beyond a reasonable doubt and proof by a preponderance of the evidence, between proof beyond a reasonable

[25] ARROW, K., SOCIAL CHOICE AND INDIVIDUAL VALUES (John Wiley) (1951).

[26] Although the quoted passage from Section 2497 of Wigmore's treatise has occasionally been cited by courts to support the thesis of the unquantifiability of the reasonable doubt standard, Wigmore's focus in Section 2497 was not on the inability of *numbers* to capture differences in the intensity of credal states. The quotation speaks only of the inability of *words* to capture and express such differences. The reasonable doubt standard is the topic of Section 2497 and the point Wigmore was driving at in Section 2497 was that variations in verbal formulations of the *reasonable doubt standard* serve no purpose.

doubt and proof by clear and convincing evidence and between proof beyond a reasonable doubt and a showing of probable cause.[27]

That's the way things stand. But do legal standards of persuasion amount to a shell game? Do they amount to a kind of verbal sound and fury signifying nothing?

The thesis that the strength of human credal states is not knowable or communicable cannot be comprehensively evaluated in a paper such as this; this comment would have to become a treatise. However, it should be noted that it is not self-evident that Wigmore's radical thesis about credal states is true.[28] It should also be noted that many people in many walks of life do not assume that Wigmore's skeptical thesis is true; many people in many walks of life assume exactly the opposite. Decision theory, e.g. rests on the assumption that in some instances decision makers are able to grade not only the utilities but also the probabilities of the possible outcomes of human choices. Versions of probability theory that emphasize bet-making as an important key to judgements about uncertainty seem to presuppose that meaningful interpersonal comparisons of probability judgements are possible. In game theory, some noncooperative games involve adversaries who, it is posited, can communicate their personal probability judgements to other players. (In game theory, it is sometimes also posited that players can even infer—at least to some degree—the probability judgements of adversaries who would prefer to keep their probability judgements to themselves.) Finally, ordinary people also often believe that the strength of factual beliefs can be graded and communicated. For example, ordinary people in their ordinary lives frequently talk about the chances that some thing has happened, is happening or will happen, and they frequently assume that their stated estimates of such chances—e.g. 'I think the chances of war in the Middle East next year are better than even', 'I'm quite sure that John won't show up at the party tonight'—do sometimes communicate meaningful information to other people. Ordinary people sometimes even bet their lives and their fortunes on their estimate of the probabilities of a wide variety of events—e.g. 'I'm pretty sure my boss hates me, so I'm going quit my job' or 'I'm not going to put the hay in the barn now because I very much doubt there will be a thunderstorm here today'—and they sometimes bet their lives and fortunes on the assumption that other people who have heard their probability estimates understand those estimates to some substantial degree.

The immediate impetus for Wigmore's expression of skepticism about the ability of people to determine and describe the degree of their uncertainty was not Wigmore's wish to demonstrate the futility of using *numbers* to quantify standards of persuasion: the immediate impetus for Wigmore's skeptical outburst was instead his desire to demonstrate the futility of using *words* to explain the reasonable doubt standard.[29] Of course, had Wigmore been asked, he would also have condemned the use of numbers to describe the meaning of the reasonable doubt standard. But the point remains

[27] But cf. *United States* v. *Dominguez Benitez*, 542 U.S. 74, 86–87 (2004) (Scalia, J., concurring) (suggesting that a meaningful distinction can be made only between the reasonable doubt standard and the more-likely-than-not standard).

[28] The popularity of Wigmore's implausible thesis may be due to the clever way in which he phrased his thesis. Wigmore links personal judgements about uncertainty to the *intensity* of beliefs about facts. This way of talking suggests that human judgements about uncertain factual matters are nothing more than a reflection of human *feelings* about the truth or falsity of propositions about factual matters. If it is true that probability judgements are nothing more than manifestations of the intensity or strength of human feelings, interpersonal comparisons of probability judgements may then indeed be either impossible or difficult to make. But in fact there may be a difference between the intensity of a person's feeling about the truth of a factual proposition, on the one hand, and a person's belief or personal judgement about the magnitude of the chance that some factual proposition is true or false.

[29] See *supra* n.26.

that Wigmore's critique cuts at least as much against *verbalization* as against *quantification* of the reasonable doubt standard. We emphasize this point because it suggests an important insight about the true nature of debates about quantification of standards of persuasion such as the reasonable doubt standard.

The true question is not whether a standard such as the reasonable doubt standard should be quantified or not quantified. The question of quantification is tied up with the more general question of the advantages and disadvantages of using both words and numbers to describe a standard of persuasion such as the reasonable doubt standard. When the question of quantification is framed in this way, we can more readily appreciate that words as well as numbers can be used and are used to grade—quantify!—degrees of certainty or uncertainty. The debate about quantification is not really about quantification. If we reject (as we should) the radical thesis that uncertainty is not subject to any discernible gradations, *the debate about quantification is really about the kind of language that should be used to grade and quantify uncertainty* and to communicate to triers of fact in legal proceedings, society's judgement about the kind and amount of factual uncertainty that society views as acceptable or unacceptable in criminal trials.

The myth of the (allegedly) necessary—but (allegedly) spurious—objectivity of quantifications of reasonable doubt

This myth is a noxious but hardy weed. It first erupted—in modern legal memory—in 1971, when Laurence Tribe made his renowned attack on trial by mathematics.[30] Mathematical analysis of evidence, he argued, can perhaps do a nice job of handling 'hard variables', but quantitative analysis (in the form of probability theory) either cannot quantify soft variables or does a lousy job of quantifying them.[31] Although Tribe does not define hard variables, he intimates that they amount to readily enumerable—readily countable—phenomena.[32]

The notion that soft variables cannot be quantified is a myth. For example, I can and do make uncertain judgements about how my neighbour will feel next time I see her—and, if asked, I can and will tell you what I think are the chances that I am right. There are, of course, more systematic critiques of the notion of the unquantifiability of soft variables, but it is not useful or possible to review them here. Suffice it to say that there is an entire family of approaches to probability—they can be called personal or subjective probability—that rest on the assumption that ALL (or almost all) probability judgements are expressions of personal and subjective probability estimates. Proponents of this sort of interpretation of the probability calculus maintain (roughly) that there are no 'objective'

[30] Tribe, *supra* n. 12.

[31] *Ibid.* at 1361–1366.

[32] *Ibid.* at 1361 ('The syndrome [the dwarfing of soft variables] is a familiar one: If you can't count it, it doesn't exist'.)

Tribe's attack on trial by mathematics, however, was multilayered. In addition to suggesting that hard variables are quantifiable and soft variables are not, he argued that hard variables would tend to swamp soft variables—they would be considered more readily by triers of fact—partly because, Tribe implies, the illusion of proof approaching certainty is thereby more easily attainable. See, e.g. *ibid*. at 1365 ('At the outset some way of integrating the mathematical evidence with the non-mathematical was sought [by Tribe?], so that the jury would not be confronted with an impressive number that it could not intelligently combine with the rest of the evidence, and to which it would therefore be tempted to assign disproportionate weight. . . . Yet, on closer inspection, [the] method [of Finkelstein and Fairley] too left a number . . . which the jury must again be asked to balance against such fuzzy imponderables as the risk of frameup or of misobservation, if indeed it is not induced to ignore those imponderables altogether'.) However, Tribe also attacks trial by probabilistic mathematics on the ground that it would more plainly expose the inability of proof in criminal trials to demonstrate a defendant's guilt conclusively. *Ibid.* at 1372–1375 (under heading 'The Quantification of Sacrifice').

probability estimates or 'hard' variables that disclose their relevant statistical probabilistic properties entirely without the corrosive intervention of subjective human judgement.[33]

Given the withering scholarly criticism that has been directed at the myth that probability theory deals with objective or hard facts and therefore cannot regulate uncertain (or inconclusive) reasoning about nonobjective phenomena, one might think that even judges would now refrain from asserting that quantitative methods cannot deal with 'soft variables'. But it is not so—at least not universally so: the wrong-headed notion that mathematical measures of the strength of evidence can measure only the strength of evidence (or judgements about the strength of evidence) about 'objective' phenomena has resurfaced in judicial discussions of the reasonable doubt standard.[34]

4. Genuine issues

In 1970, Justice John Harlan articulated a profoundly important insight about the reasonable doubt standard. He did so in his much-cited and much-quoted concurring opinion in *In re Winship*, 397 U.S. 358, 368 (1970). He saw that although the reasonable doubt standard is about probabilities, the reasonable doubt standard is not *only* about probabilities. Yes, analysis of the reasonable doubt standard must begin, he said, with the premise that factual proof is always or almost always less than conclusive: yes, we must begin by acknowledging that factual proof is necessarily probabilistic. But this insight, he suggested, is just the beginning of wisdom about the standard of persuasion in criminal trials. The level of probability that society requires for a criminal conviction—and the amount of uncertainty that it finds tolerable—depends, he effectively said, on society's values and how they are prioritized (or, in the parlance of decision theory, 'weighted').[35]

He wrote:

> [E]ven though the labels used for alternative standards of proof are vague and not a very sure guide to decisionmaking, the choice of the standard for a particular variety of adjudication does, I think, reflect a very fundamental assessment of the comparative social costs of erroneous factual determinations.
>
> ***
>
> The standard of proof influences the relative frequency of these two types of erroneous outcomes [acquittal of the guilty and conviction of the innocent]. If, for example, the

[33] See HÁJEK, A. *Interpretations of Probability, The Stanford Encyclopedia of Philosophy* (E. N. Zalta ed.). Available at http://plato.stanford.edu/archives/sum2003/entries/probability-interpret/. Accessed 28 July 2006.

[34] Even Judge Richard Posner, a founder of the law and economics movement, fell victim to the 'soft variable fallacy'. In *United States* v. *Hall*, 854 F.2d 1036, 1043 (7th Cir. 1988) (concurring opinion) he said, 'Numerical estimates of probability are helpful in investments, gambling, scientific research and many other activities but are not likely to be helpful in the setting of jury deliberations. No objective probability of a defendant's guilt can be estimated other than in the rare case that turns entirely on evidence whose accuracy can be rigorously expressed in statistical terms (e.g., fingerprints and paternity tests). In other cases the jury's subjective estimate would float free of check and context. It is one thing to tell jurors to set aside unreasonable doubts, another to tell them to determine whether the probability that the defendant is guilty is more than 75, or 95, or 99 percent.' *Ibid*. at 1044.

[35] Harlan was certainly not the first jurist or thinker to have these insights about the standard of persuasion in criminal trials; the notion of a tradeoff between the chances of convicting the innocent and the chances of acquitting the guilty, accompanied by the suggestion that the trade-off should be strongly tilted in favour of protecting the innocent from conviction, has an ancient lineage. In *United States* v. *Fatico*, 458 F. 388, 410–11 (E.D.N.Y. 1978), Judge Weinstein notes that the idea of a tilt in favour of the accused appears in the work of the medieval Jewish philosopher Maimonides and also in Justinian's Digest.

> standard of proof for a criminal trial were a preponderance of the evidence rather than proof beyond a reasonable doubt, there would be a smaller risk of factual errors that result in freeing guilty persons, but a far greater risk of factual errors that result in convicting the innocent. Because the standard of proof affects the comparative frequency of these two types of erroneous outcomes, the choice of the standard to be applied in a particular kind of litigation should, in a rational world, reflect an assessment of the comparative social disutility of each.
>
> ***
>
> In a criminal case we do not view the social disutility of convicting an innocent man as equivalent to the disutility of acquitting someone who is guilty.
>
> ***
>
> I view the requirement of proof beyond a reasonable doubt in a criminal case as bottomed on a fundamental value determination of our society that it is far worse to convict an innocent man than to let a guilty man go free.

Ibid. at 369–72 (footnote omitted)

Harlan's analysis of the reasonable doubt standard was not perfect. For example, Harlan emphasized how society quantifies and balances certain harms; he emphasized, in particular, the harm resulting from an improper acquittal and the harm resulting from a false conviction.[36] Harlan's use of a calculus that takes into account only the costs of errors (but not the benefits of correct results) might be questioned.[37] Furthermore, even if Harlan's focus on harm is justifiable, it is also possible to question the sort of calculus that Harlan hinted should be used to evaluate the comparative harm inflicted by alternative erroneous outcomes.[38] So, no, Justice Harlan surely did not get everything right. But it is hard to quarrel with Harlan's general thesis that the degree of certainty that society requires for conviction of crime is influenced by—or, in any event, *should* be influenced by—a comparative evaluation of the amount of good and evil that will come society's way as the degree of persuasion and probability required for conviction of crime vary and as, thus, the frequency of various outcomes in the criminal process also changes. Harlan quite sensibly suggested that in a rational world, society will calibrate proof requirements so that outcomes fall out in a way that maximizes social welfare.

An important point to be extracted from Harlan's analysis is that the *existing* standard—the existing *verbal* standard—of 'proof beyond a reasonable' doubt 'is' a 'calibration' of how much certainty

[36] Harlan's approach, which emphasized comparative harms, anticipated Richard Lempert's use of 'regret matrices' to analyse standards of persuasion in criminal trials. See LEMPERT, R., "Modeling relevance," 75 *Michigan Law Review* 1021 (1977).

[37] For example, some current 'expected utility analysis' takes into account *four* possible outcomes: acquittal of the guilty, conviction of the innocent, conviction of the guilty and acquittal of the innocent. See Hamer, *supra* n.14.

[38] There are several reasons why the level of probability required by a standard of persuasion does not translate directly into rates of error. One often-neglected reason is that the probabilities established by proof in particular cases can exceed the minimum probability that some legal standard of persuasion requires. See generally David Hamer, *ibid.*; cf. KAYE, D., "The error of equal error rates," 1 *Law, Probability and Risk* 3 (2002) (more-probable-than-not standard does not necessarily equalize the rates of (1) erroneous verdicts for plaintiffs and (2) erroneous verdicts for defendants); "Two theories of the civil burden of persuasion," 2 *Law, Probability and Risk* 9 (2003) (more-probable-than-not standard of persuasion minimizes expected losses due to erroneous jury verdicts, and it does not necessarily equalize the allocation of errors among plaintiffs and defendants).

about criminal guilt our society must demand, if social welfare is to be maximized. This is true even if the requirement of proof beyond a reasonable doubt is a fuzzy requirement (as it almost surely is). For even if the existing verbal standard of proof beyond a reasonable doubt standard is a fuzzy standard, it is not a *completely* fuzzy standard. The existing—entirely verbal—version of the standard of persuasion in criminal trials does quantify *to some extent* how much uncertainty is acceptable and how much is not. For example, practically everyone seems to agree that the reasonable doubt standard requires a greater probability (i.e. less uncertainty) than does the preponderance of the evidence standard, which, it is generally thought, requires that the trier of fact believe that the factual propositions in issue are more probably true than false. Furthermore, today almost everyone (but not everyone) agrees that proof beyond a reasonable doubt does not require that the trier of fact feels utterly certain that the factual propositions essential to criminal guilt are true. So it should be agreed that the reasonable doubt standard is not completely meaningless and that it does convey some information about society's expectations to the trier of fact. But, as already noted, the reasonable doubt standard is nevertheless a very fuzzy standard: although the reasonable doubt standard is not open to any interpretation, it is open to a wide range of interpretations. The question is whether society can better achieve its objectives by quantifying the reasonable doubt standard in some way.

But the mere fact of the fuzziness of the reasonable doubt standard does not justify the conclusion that quantifying reasonable doubt is the right thing to do. There are several reasons why this is so. First, *even numbers can be fuzzy*.[39] Second, *there are degrees of fuzziness*. Third, *fuzziness is not always a bad thing*; *sometimes fuzziness is good*. The true question is not whether the law should use a fuzzy rule or a nonfuzzy rule to describe the prosecution's burden of persuasion in criminal trials; in the legal domain all possible rules are fuzzy—to some degree. In the legal domain, the real question is instead *how much* fuzziness is optimal. Another important question is *what sort* of legal fuzziness works best. So the issue of quantification of reasonable doubt should take the following general form: is it good that the reasonable doubt standard is as fuzzy—and as unfuzzy—as it is; and is it good that the reasonable doubt standard is fuzzy in the way that it is now fuzzy? Once the problem of 'quantification' is framed in this way—and once, therefore, it is apparent that the central problem is not *whether* the reasonable doubt standard should be quantified but rather *how granularly* and *in what fashion* the reasonable doubt standard should be quantified—it should be apparent that there are serious arguments for as well as against quantification of the reasonable doubt standard. More precisely stated, it should be apparent that there are serious arguments for and against *more* discrete quantification and for and against a *different kind* of quantification.

Several considerations speak in favour of *numerical* formulations of the standard of persuasion in criminal trials. As Weinstein suggested in *Copeland III*, one possible powerful argument for some sort of numerical quantification is the importance of the *uniformity* of legal standards. Weinstein correctly suggested that differential legal standards are particularly disturbing when they appear in the criminal justice system. *If* numerical quantification of the burden of persuasion in criminal trials can reduce differences in the operational meaning of the standard of persuasion in criminal trials, a powerful argument in favour of numerical quantification is at hand.

Other considerations speak in favour of some sort of numerical quantification of the reasonable doubt standard. Justice Harlan's concurring opinion in *Winship* is an eloquent reminder that the level of persuasion that a society chooses to impose in criminal trials (and in criminal litigation generally) reflects or should reflect the way a society chooses to balance its priorities. Two of the items

[39] Perhaps *all* numbers are fuzzy. But if so, some numbers are probably *relatively* unfuzzy.

on society's scale (though very probably not the only items) must be the value that society places on preventing guilty people from being acquitted and the value that society places on preventing conviction of the innocent. However, if the reasonable doubt standard is very fuzzy—and it is—the manner in which society has chosen to balance its priorities or—the scale that it has chosen to use to balance its priorities—is also fuzzy. This fuzziness may significantly contribute to several evils:

1. A very fuzzy formulation of the standard of persuasion in criminal trials may contribute to lack of transparency about society's preferences about matters such as the importance of preventing the conviction of innocent people.
2. A related evil that may be aggravated by the great fuzziness of the reasonable doubt standard is societal hypocrisy or the appearance of societal hypocrisy: the extraordinary fuzziness of the reasonable doubt standard perhaps increases the chances that the public will believe that the probability requirements actually applied in criminal trials are substantially different from the standard of persuasion that society professes to demand in criminal trials, and that the public will think that this perceived disparity is deliberate rather than accidental.
3. The very fuzzy standard of persuasion now applied in criminal trials may increase the frequency, with which a grossly improper balance is struck between values such as the security of the community and protecting the liberty of innocent people; the extreme fuzziness of the reasonable doubt standard may significantly increase the frequency of the use of probability requirements that are widely at variance with societal preferences and with our society's sense of justice.

There is some evidence that the present standard of persuasion in criminal trials may in fact significantly contribute to this last evil. Several informal surveys of judges and jurors—two of them conducted under Judge Weinstein's auspices—suggest that a disturbingly large fraction of trial judges and jurors believe that the existing reasonable doubt standard imposes a probability requirement of only 0.6 (60%) to 0.75 (75%) in criminal trials. A requirement that the trier's certitude of guilt must be a minimum of only 0.6 or 0.75 may not strike an appropriate balance between considerations such as the importance of freedom from crime, the importance of convicting the guilty and the importance of not convicting the innocent. We readily grant that we do not *know*—and we suspect many or most readers also do not know—whether or not the balance presupposed by such a 'liberal' interpretation of the reasonable doubt standard is appropriate or inappropriate; many variables have to be considered if we are to make reasonably educated guesses about the real-world upshot of trials, in which some triers use a probability benchmark as 'low' as 0.6 or 0.75. For example, pretrial processes have to be taken into account if one thinks that the harm of 'false convictions' should be framed in terms of the fraction of innocent people who are likely to be adjudged guilty in the 'criminal process'—rather than just the fraction of innocent people who are likely to be adjudged guilty in criminal *trials*. Even if the latter fraction is very large, the former fraction might be very small. Another variable that might have to be considered is small group behaviour: in jury trials, it is possible, e.g. that the effect of outlying opinions among jurors about the meaning of reasonable doubt will be largely washed out by the process of collective jury deliberation. Nonetheless, it is also possible that some triers' 'liberal' interpretation of the reasonable doubt standard (e.g. the belief that a probability of guilt of 0.6 or 0.75 suffices) may produce error rates that are wildly at variance with our society's values.

Suppose authorized lawmakers reach the conclusion that the amount of subjective uncertainty about guilt roughly corresponds to the relative frequency of false convictions. Further suppose that authorized lawmakers reach the conclusion that the reasonable doubt standard leads many juries (in jury trials) and trial judges (in bench trials) to measure guilt and innocence by probability requirements that are widely at variance with our society's values. Suppose that these conclusions or beliefs are plausible. Under these circumstances, a strong *prima facie* argument can be made that

(i) it is wrong for society to give the trier of fact a legal standard that allows the trierwhether wittingly or unwittingly to strike a balance that is wildly at variance with the values of society at large;

(ii) it would be far better for authorized lawmakers—e.g. legislators and judges—to evaluate society's preferences, to decide how those preferences are best balanced and to craft a relatively unfuzzy probability requirement that in their judgement—in the view of society's authorized representatives—best accommodates society's values, priorities, concerns and capacities and

(iii) society can most effectively communicate to triers of fact society's preferred subjective probability requirements and it can best enforce such socially chosen probability requirements by using *numbers* to designate the maximum amount of subjective uncertainty that is legally compatible with criminal punishment.

We do not claim that the preceding argument for numerical quantification of the standard of persuasion in criminal trials is ironclad. It is in fact not ironclad. But we are not trying to establish that the case for numerical quantification is invincible. We are instead trying to establish three principal points:

1. There are powerful arguments in favour of numerical quantification of the burden of persuasion in criminal trials.
2. Many of the usual arguments against numerical quantification have little or no substance.
3. Only an appreciation of the first two points makes it possible to evaluate objections to quantification that may actually have substance.

Perhaps two general examples will suffice to illustrate the third point.

There are two broad classes of potentially valid objections to numerical quantification of the reasonable doubt standard: (i) feasibility objections and (ii) objections rooted in political theory. Feasibility objections involve principally the question of intelligibility. A numerically quantified burden of persuasion for criminal trials may be 'in principle' less fuzzy than the existing nonnumerical standard of persuasion. But how a numerical standard works 'in principle' is largely immaterial. What matters is how such quantified standards actually work. So one critical question is what sorts of numbers or numerical expressions are intelligible (or can be made intelligible) to triers of fact such as jurors. Psychologists and other scholars have done empirical research on this sort of question. And further research about this should be done; the intelligibility of different kinds of numbers to actors such as judges and jurors is a proper question to consider, in relation to proposals for numerical quantification of the standard of persuasion in criminal trials. But our discussion here suggests that it is not permissible for judges or lawmakers simply to assume that jurors and trial judges are incapable of making sense of any numbers.

Another class of rationally defensible objections to numerical quantification has a political or 'ideological' character. We cannot enumerate (let alone evaluate) all such objections. But we can say that a clear appreciation of how 'quantification' actually works—and how it does *not* work—makes it possible to make more clear-headed judgements about normatively grounded objections to numerical quantification of the standard of persuasion in criminal trials. For example, it is entirely possible to make an argument—it is possible to make a *strong* argument—that reasonable doubt should not be quantified because (i) fuzzy standards of persuasion in criminal trials are a good thing because the fuzziness of such standards allows juries to decide how much probability should be required for a criminal conviction and (ii) it is good for juries to be empowered to make such legislative or quasi-legislative choices. This kind of argument might be further rooted in some sort of a quasi-populist political theory that maintains that juries are appropriate representatives of the community at large and that juries, as such representatives, should be given the power to make some law in the trials in which they participate. For purposes of this case comment, it does not matter whether this sort of normative argument about the role of juries is good or bad. What matters here is only the proposition that clarity about matters such as probability and quantification makes it possible to use the proper terms and concepts to evaluate such normative or political arguments about the structure of proof in criminal trials, and that clarity about mathematical matters prevents political or normative arguments about forensic proof from being obscured by wrong-headed notions about how numbers can lead jurors and others astray.[40]

5. Conclusion

In *United States* v. *Hall*, 854 F.2d 1036 (7th Cir. 1988) Judge Richard Posner, concurring, wrote:

> When . . . judges and juries are asked to translate the requisite confidence into percentage terms or betting odds, they sometimes come up with ridiculously low figures-in one survey, as low as 76 percent, see *United States v. Fatico,* 458 F. Supp. 388, 410 (E.D.N.Y. 1978); in another, as low as 50 percent, *see* McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?,* 35 Vand. L. Rev. 1293, 1325 (1982) (tab. 2). The higher of these two figures implies that, in the absence of screening by the prosecutor's office, of every 100 defendants who were convicted 24 (on average) might well be innocent.

Ibid. at 1044

Judge Posner may be right that it is ridiculous for triers of fact—trial judges and jurors—to come up with widely varying probabilities to describe the degree of certitude that they (these judges and jurors) think the reasonable doubt standard requires for conviction. But if the choice of such widely varying numerical probabilities is an ailment, an important question remains: is the ailment a disease or is it a symptom? It is unclear that outlawing the use of numbers by trial judges and jurors is the right medicine for the tendency of jurors and judges to select widely divergent numerical probabilities to describe the requirements imposed by the reasonable doubt standard. To outlaw such erratic

[40] It is of course true that numbers can lead jurors astray. The real task is to figure out the actual reasons why some types of numbers sometimes do so—and why some other types of numbers sometimes do not do so.

quantification of the reasonable doubt standard by outlawing its quantification might be akin to trying to cure a disease by excising the symptoms of the disease. It is possible that Judge Posner has gotten cause and effect mixed up: it is possible that the possibly ridiculous numbers that some triers of fact choose when they are asked to quantify the reasonable doubt standard *reflect* a serious problem rather than *cause* one. It is possible that different triers of fact choose very different probabilities to describe the reasonable doubt standard, not because triers of fact are confused, but because the reasonable doubt standard as it now exists is *in reality* extraordinarily fuzzy.[41] In short, dear Brutus, the fault may lie not in the numbers, but in our law: perhaps judges and jurors use strange numbers to describe the burden of persuasion in criminal trials primarily because of the strange way that our law now describes the burden of proof and persuasion in criminal trials.

Acknowledgements

We are grateful for the comments and suggestions of Professors George Fisher, James Franklin and Alex Stein, who bear no responsibility for this paper's flaws and limitations.

[41] There is an old saying that an incoherent question may be correctly answered by any answer, including an incoherent one. It is equally true that a completely fuzzy question may be answered in any way one pleases. Less dramatically, a relatively fuzzy question may be answered in a variety of ways—because of the fuzziness of the question. Hence, it is no surprise that different people use different numbers to describe a fuzzy legal requirement. Why would anyone expect respondents to do otherwise?