Law, Probability and Risk (2006) **5**, 167–173 doi:10.1093/lpr/mgl016
Advance Access publication on January 15, 2007

Comment on the meaning of 'proof beyond a reasonable doubt'

JACK B. WEINSTEIN

Senior Judge United States District Court, Eastern District of New York
United States Courthouse, 225 Cadman Plaza East, Brooklyn, NY 11201

AND

IAN DEWSBURY

Student Law Clerk to Judge Weinstein

[Received on 30 October 2006; revised on 8 November 2006;
accepted on 10 November 2006]

In view of the variety of predispositions among jurors regarding the meaning of proof 'beyond a reasonable doubt', some quantitative definition of the probability of guilt required for conviction in addition to descriptive phrasing— in particular stressing the importance of avoiding conviction of the innocent—is desirable. Some flexibility should be left to the discretion of the jurors. Whether any articulation of a standard—in words or percentages—will have an appreciable effect on the rate of improper convictions is not clear.

Keywords: burden of proof; reasonable doubt; quantification; jury instructions.

1. Introduction

Prof. Tillers's and Jonathan Gottfried's analysis supporting a quantified jury instruction on the meaning of proof beyond a reasonable doubt is persuasive. That it will be adopted in practice seems doubtful, for some of the reasons suggested by Prof. Franklin and in our Comment.

In the usual civil case, the law does not favour either party, except that it slightly prefers the status quo. 'More probable than not', with the burden of getting over equipoise, satisfies society's need for stability and is relatively easy for the jury to understand. If, for some policy reason, the law wants to make it harder or easier for the proponent to win, it puts its thumb on the scale with a variety of burden formulations.[1]

In a criminal case, the law tilts in favour of defendants; it prefers that some guilty go free rather than that some innocents be convicted. The questions are (1) how high the required minimum probability should be set and (2) how should the test be articulated. Quantification is one way of stating the standard. Requiring a combined descriptive explanation and an explicit percentage would likely work well for the thousands of jurors we have observed.

[1] See, e.g. *United States* v. *Fatico*, 458 F. Supp. 388 (E.D.N.Y. 1978); *Vargas* v. *Keane*, 86 F.3d 1273 (2nd Cir. 1996) and *United States* v. *Copeland*, 369 F. Supp.2d 275 (E.D.N.Y. 2005), cited by Prof. Tillers. See also WEINSTEIN, J. B., MANSFIELD, J. H., ABRAMS, N. & BERGER, M. A. (1997) *Evidence*, 9th edn. The Foundation Press, pp. 1087–1158; *ibid.* at 1156–1158 (instructing juries in criminal cases).

2. Variability among jurors

Based on their genetic differences and divergences in experience, variations among jurors—a cross-section of the community under our practice—are huge. This affects the jury's impression of such matters as what the rules of law are and ought to be, evaluation of credibility of witnesses and the inferences to be drawn from the evidence.

For example, in theory, the 'presumption of innocence' should mean that, before hearing any evidence, the jury should start with the assumption that it is close to 0% likely that the defendant (out of all the people in the universe) committed the crime charged. In practice, most trier's initial anchoring assumption is probably considerably higher than this. Jurors will assume, to some extent, depending on their individual beliefs about how the criminal justice system works (and despite directions not to do so), that if there were no substantial evidence of guilt: (a) the prosecutors and investigators would not have gone forward with the case, (b) the grand jury would not have indicted and (c) the trial judge would not have started up the machinery of selecting a jury and holding a trial. Some jurors also probably have been swayed towards or away from a preliminary hypothesis of guilt or innocence by the jury selection process and opening statements of counsel.

It is not unlikely that many jurors begin hearing evidence believing that it is roughly 50% probable that the defendant committed the offense charged, with the state having the responsibility to drive up this probability. Some few will believe that a defendant they identify with almost certainly did not commit any crime and is in the dock because of a trumped-up case. In our heterogeneous society, prejudice and life experiences also cause enormous differences in evaluation of what constitutes beyond a reasonable doubt. There have been only a few empirical studies of the effect on the jury of varying the instruction on the weight of the burden in a criminal case—none of them entirely satisfactory.[2]

3. Policy on height of burden

As jury instructions are commonly given, the precise meaning of the standard—i.e. how high it is—is left to the jury. There is a difference in peoples' views of the degree that the law should abhor the conviction of the innocent and let the guilty go free. Should society be willing to risk 10 guilty defendants going free rather than one innocent person convicted? Or is the proper ratio 100 to one? Should we be willing to accept lower risks in a 'spitting on the sidewalk' case than in a capital homicide case? In a bomb-terrorism case, should the risks be inverted with a preference for

[2] See, e.g. the Jury Project at the London School of Economics reported in L.S.E. JURY PROJECT (1973) Juries and the rules of evidence. *Criminal Law Review* **1973**, 208. The subject is briefly discussed in UNDERWOOD, B. D. (1977) The thumb on the scales of justice: burdens of persuasion in criminal cases. *Yale Law Journal*, **86**, 1299, 1309–1310. She notes that '[t]he result of [the study is] inconclusive, but [it suggests that] the instruction can affect the outcome of a case'. See also STOFFELMAYOR, E. & DIAMOND, S. S. (2000) The conflict between precision and flexibility in explaining "Beyond a Reasonable Doubt". *Psychology, Public Policy and Law*, **6**, 769; MCCAULIFF, C. M. A. (1982) Burdens of proof: degrees of belief, quantification of evidence, or constitutional guarantees? *Vanderbilt Law Review*, **35**, 1293, 1324–1326 (author's survey showed variations from 50% to 100% with most common response at 90% and next most common at 95%); SIMON, R. J. & MAHAN, L (1971) Quantifying burdens of proof: a view from the bench, the jury, and the classroom. *Law and Society Review*, **5**, 319, 328. See also authorities cited in *supra* n. 1. Consider also the variation among judges described in the *Fatico case, supra.*

convicting 10 innocents rather than letting one guilty go free?[3] The public may believe that not all crimes warrant the same burden at all times.

In most cases, a large part of the evidence presented by the prosecution is redundant. Its cumulative nature gives weight to the simple computation of probability of guilt calculated on the evidence alone, making inapplicable Prof. Franklin's warning about computed probabilities based upon the occasional relative light weight of the evidence. Our standard charge that the jury can take into account 'lack of evidence' points the jury towards considering the 'thinness' of the prosecutor's case.[4] But there are cases, as Prof. Franklin points out, where there is little corroborating evidence, as in a single stranger's eyewitness identification, or a prosecution based almost wholly on the defendant's confession.[5] And, in fact, these are the kinds of cases where the courts are most often wrong—i.e. a trial is most likely to result in the conviction of an innocent defendant.[6] There are also a substantial number of cases involving inept investigations or even deliberate fingering of the wrong person, fairly conclusively established long after the event where DNA evidence is available.

Whether any articulation of the highest practicable standards for reasonable doubt will appreciably change the ratio of wrongful convictions to improper acquittals remains a matter of some doubt.[7] The remedy lies more in reform of the investigative, prosecutorial and defense phases than in the formulation of probability standards for the jury. Nevertheless, a high burden will tend to force the prosecution to do a more thorough job of investigation.

We personally favour burden of proof in the realm of 95+% probability of guilt. Yet, if all our jurors had been given this quantitative definition of the standard, we doubt that the result would have changed in more than in the order of 1% of our jury-tried cases—and the increased acquittals would not necessarily be in cases where the defendant was innocent.[8]

The jury is composed of members of the community who have different views (even though not stated) of the acceptable risks of wrongful convictions and acquittals. By putting a specific

[3] See, e.g. the case of an apparent general lack of strong disapproval of the police killing an innocent suspect in London after the public transport bombings. In New York City, particularly if the person shot were from a minority group, the outcry would be enormous. Cf. HOROWITZ, I. A. (1997) Reasonable doubt instructions. *Psychology, Public Policy and Law*, **3**, 285 (possibility that rather than the well-established 'leniency bias' leading to acquittal in criminal trials, jurors may exhibit a conviction bias in certain kinds of cases; discusses various standards approved by courts).

[4] See Sixth Circuit Pattern Criminal Jury Instructions 1.03. But cf. SAND, L. B., SEIFFERT, J. S., LOUGHLIN, W. P. & REISS, S. A. (2002) *Modern Federal Jury Instructions*. Lexis Nexis, pp. 4–18 (considering the pattern instruction's language to the effect that reasonable doubt 'may arise from the evidence, the lack of evidence or the nature of the evidence' but preferring the 'simpler language that the jury should carefully weigh all of the evidence').

[5] There may also be a problem when the prosecution's case is built on multiple contingent or circumstantial inferences, each of which must be true to establish the defendant's guilt. This is due to peoples' systematic overestimation of the likelihood of conjunctive, as opposed to disjunctive events. See PIATTELLI-PALMARINI, M. (1991) Probability blindness: neither rational nor capricious. *Bostonia*, **March/April**, 28–35.

[6] See, e.g. BORCHARD, E. M. (1932) *Convicting the Innocent: Error of Criminal Justice*. Yale University Press. See also DWYER, J., NEUFELD, P. & SCHECK, B. (2003) *Actual Innocence: When Justice Goes Wrong And How To Make It Right*; New American Library. EGGERS, D. & VOLLEN, L. (2005) *Surviving Justice: America's Wrongfully Convicted and Exonerated*; McSweeney's Books. GRISHAM, J. (2006) *The Innocent Man: A True Story*; Doubleday Publishing. WEINSTEIN, J. B. (2006) *Avoiding Convicting the Innocent and Acquitting the Guilty*; Unpublished, text of speech given to the Great Neck Lawyers' Association on November 14, 2006.

[7] This empirical question would be nearly impossible to answer. Collecting information about the number of wrongly *acquitted* defendants, in particular, would be difficult, if not impossible. This calls into question Prof. Franklin's observation that quantification of the standard of proof would mean that the law would be 'forced to admit the truth about the number of false convictions it allows and the number of criminals it allows to go free'.

[8] See KALVEN, H. & ZEISEL, H. (1996) The American Jury, p. 62 (summary view of judge–jury disagreements on guilt).

percentage to the jury, the court, rather than the jury, would be making this policy choice. Our traditional practice is to leave this policy decision largely to the cross-section of the community represented by the members of the jury. In criminal cases, the law does not permit the direction of a guilty verdict by the judge, so the jury has more leeway in protecting the defendant—or failing to do so. While the court may direct for the defendant, the power is little exercised unless a point of law is decisive against conviction.[9] Unlike the English practice, the reluctance of American judges to sum up the evidence gives us less control.[10]

The jury decides the critical issue of credibility almost without limit—providing wide power to nullify.[11] While the cases say a jury may not nullify in criminal trials, it can and sometimes does so, misinterpreting the evidence and burden of proof standard deliberately—usually to acquit, but sometimes to convict improperly. The policy arguments for and against the desirability of allowing jurors to rely on their own conception of whether a law is just apply somewhat to the desirability of allowing jurors to rely on their own conception of whether a given case warrants a lower or higher standard of proof.[12] The court is not allowed to say to the jury, 'You may nullify', but, that is the partial effect of leaving the burden of proof in a criminal case unquantified.

4. Articulation of a standard

None of this is to say that we might not do better in defining what beyond a reasonable doubt means. The many formulations in use are somewhat helpful, but there is wide variation in their interpretation. Some evidence suggests that to the extent jurors misunderstand their instructions on the point, they err on the side of *lowering* the burden of proof, effectively failing to differentiate between proof beyond a reasonable doubt and lesser civil standards.[13]

At least some judges of the Supreme Court of the United States have favoured the following instruction which is based on suggestions from the United States Judicial Center:[14]

> [T]he government has the burden of proving the defendant guilty beyond a reasonable doubt. Some of you may have served as jurors in civil cases, where you were told that it is only necessary to prove that a fact is more likely true than not true. In criminal cases, the government's proof must be more powerful than that. It must be beyond a reasonable doubt.
>
> Proof beyond a reasonable doubt is proof that leaves you firmly convinced of the defendant's guilt. There are very few things in this world that we know with absolute

[9] See, e.g. *United States* v. *Eppolito and Caracappa*, 436 F.Supp. 2d 532 (E.D.N.Y. 2006) (statute of limitations and general discussion).

[10] See WEINSTEIN, J. B. (1988) The power and duty of federal judges to marshall and comment on the evidence in jury trials and some suggestions on charging juries. *Federal Rules Decisions*, **118**, 161.

[11] See, e.g. HOROWITZ, I. A., KERR, N. L. & MEDIRMEIR, K. E. (2001) Jury nullification: legal and psychological perspectives. *Brooklyn Law Review*, **66**, 1207; WEINSTEIN, J. B. (1998) The many dimensions of jury nullification. *Judicature*, **81**, 168; WEINSTEIN, J. B. (1993) Comments on Jury Nullification, Proceedings of the Fifty-third Judicial Conference of the District of Columbia Circuit, *Federal Rules Decisions*, **145**, 149; WEINSTEIN, J. B. (1993) Considering "Jury Nullification," when may and should a jury reject the law to do justice. *American Criminal Law Review*, **30**, 239.

[12] See authorities in footnotes 1, 2, *supra*. See also the different probabilities assigned to 14 different formulations explaining the meaning of reasonable doubt in *Vargas* v. *Keane*, 86 F.3d 1273, 1281 ff. (2nd Cir. 1996) (Jack B. Weinstein concurring).

[13] Stoffelmayor and Diamond at 774–776, *supra* n. 2.

[14] *Victor* v. *Nebraska*, 511 U.S. 1, 27, 114 S.Ct. 1239, 1253 (1994) (Justice Ginsberg concurring).

> certainty, and in criminal cases the law does not require proof that overcomes every possible doubt. If, based on your consideration of the evidence, you are firmly convinced that the defendant is guilty of the crime charged, you must find him guilty. If on the other hand, you think there is a real possibility that he is not guilty, you must give him the benefit of the doubt and find him not guilty."

Other federal courts have required the following shorter instruction:[15]

'It is not required that the government prove guilt beyond all possible doubt. The test is one of reasonable doubt. A reasonable doubt is a doubt based upon reason and common sense—the kind of doubt that would make a reasonable person hesitate to act. Proof beyond a reasonable doubt must, therefore, be *proof of such a convincing character that a reasonable person would not hesitate to rely and act upon it in the most important of his own affairs*. The jury will remember that a defendant is never to be convicted on mere suspicion or conjecture.'

In the last case we tried, we added to this last charge the following with the consent of all counsel: 'The burden on the government is extremely high because the law abhors the conviction of a person who is not guilty'.[16] We were influenced in part by the Tillers–Franklin debate. The jury convicted all four defendants, though we almost entertained a 'reasonable doubt' about one of them.

Were the parties to agree, we would like to add to our charge: 'In my opinion, a probability of guilt of no less than 95% should be necessary to support a conviction'.[17] Aside from the question of appellate courts' antipathy towards such an instruction, as properly noted by Prof. Tillers, it would probably be opposed by prosecutors *and* defense counsel. Prosecutors would object on the grounds that this quantification requires too high a probability. Some defense counsel would object because 'fuzziness' in the burden of proof would allow jurors to take more account of the case-specific heuristics, biases and other reasons why *their* clients should be found not guilty.

While the modest research to date indicates that clarity in our charge to the jury is desirable, it is not obvious that a quantified instruction will lead to more precise communication between bench and jury. Comprehensive empirical research would be helpful. It is interesting to note, as Prof. Franklin does, that some of the easiest information to quantify, such as recidivism rates based on past commission of crimes are precisely the kind of information that the rules of evidence, not to mention the Constitutional guarantees of due process and equal protection, prohibit the jury from considering.

5. Factors the jury will probably take into account

As already suggested, of the many factors which may influence a particular juror to modify the reasonable doubt standard, some may be seen as permissible or even desirable, while others are clearly inappropriate. A brief catalogue of some of the possible factors illustrates the point:

- Jurors may evaluate the same evidence differently and adjust the burden of proof based on personal experience. Some of these variations are valid and may be relied on, some are valid (in the

[15] *United States* v. *Savulj*, 700 F.2d 51, 69 (2nd Cir. 1983) (emphasis supplied). That we sometimes hesitate more about which tie we should select than who should be our spouse takes some of the edge off this charge; intuition is often wrong, but almost always swift. This charge was criticized by Justice Ginsberg in *Victor*, who points to other circuits which want *no* definition. See also, e.g. SAND, L. B., SEIFFERT, J. S., LOUGHLIN, W. P. & REISS, S. A. (2002) *Modern Federal Jury Instructions*, Instr. 4-2, with an extensive comment.

[16] *United States* v. *Noira*, CR 06-135 (E.D.N.Y. 2006).

[17] See the slight modification of the formulation in part VI, *infra*.

sense of logically relevant) but may not be relied on (such as information about the defendant's criminal history) and some may have some experiential basis, but are not permissible (such as racial stereotypes).

- Jurors may alter the probability required for conviction based on their degree of confidence in the police, the prosecutor, the court and the justice system as a whole. Feelings about such matters can vary significantly from one community to another.[18]
- Jurors may require a lower standard of proof based on their perception of the danger of the moment. In a 'rash' of sex crimes or robberies publicized by the media or post 9/11 in a case where the defendant is charged with attempting to commit terrorist acts, the jury may feel that the risk of setting a guilty defendant free weighs heavily.
- The jury may elect to convict a defendant for the crime charged on a lower standard of proof based on evidence of defendant's past crimes or problematic life style, which has been admitted only for a limited purpose, on the theory that the defendant probably should be in prison anyway.
- Jurors may be influenced by their beliefs about the future dangerousness of the defendant. While this may or may not be relevant to sentence, depending on one's theory of punishment,[19] it should not be relevant to a determination of guilt or innocence with respect to the crime charged.

Even if we could agree in principle on which of these factors are permissible as a basis for requiring a higher or lower standard of proof in a given case, a jury instruction explaining this in any detail would be unwieldy, perhaps to the point of uselessness, and would unnecessarily put ideas in jurors' minds, akin to telling young children not to put beans up their noses.

6. Taking the wisdom of Tillers, Gottfried and Franklin into account

To reach the problems posed by the Tillers–Gottfried–Franklin debate directly, the jury would be provided with a charge something like the following:

> The burden of proof on the government is beyond a reasonable doubt. Why do we have such a burden? The law prefers to see that guilty persons go free rather than an innocent be convicted.
>
> What is the height of the burden? That will vary. It may depend on your evaluation of the nature of the crime charged, the dangers of letting the guilty go free, the great unfairness in convicting the innocent, and other factors you find appropriate. But you may not change the balance to convict on a lesser burden because of the particular gender, age, religion, place of origin, or other personal characteristics of the defendant.
>
> In general we can all agree that convicting the innocent is a great harm that should be avoided. That is why we have a presumption of innocence about which I've talked to

[18] For example, African–American jurors from some inner city communities, who have witnessed police abuse, may hold the state to a higher burden when the evidence is based on a police investigation resulting in the arrest of an African–American defendant. By contrast, stereotypes on the part of white jurors in the same case may lead to a weakening of the presumption of innocence or a tendency to multiply the probative force of the evidence by a substantial factor. Other strong biases may lead to hung juries but seldom, in our diverse and sophisticated New York jurisdiction, force the jury into an improper verdict. Cf. KALVEN JR., H. & ZEISEL, H. (1966) *The American Jury*. Little Brown and Company, pp. 453–463 (hung juries).

[19] See WEINSTEIN, J. B. & WIMMER, C. (2006) *Sentencing in the United States, passim* (forthcoming).

> you already. It is also one of the reasons why we require proof beyond a reasonable doubt. A reasonable doubt may arise from the evidence, the lack of evidence, or the nature of the evidence. Were I the trier of fact, I would require a probability of guilt of no less than 95%. But it is for you to decide how high the burden should be as long as it is much higher than the highest civil standard—"clear and convincing evidence," which is itself much higher than "more probable than not."

Such an instruction would seem to comport (though it needs refinement) with Tillers's and Gottfried's discussion of *Winship* and other cases. It has the advantage of drawing the jurors' attention to the high standard of proof in criminal cases, as well as the law's policy of avoiding conviction of the innocent. It also points the jurors towards the need to consider absence of proof that would be expected were the investigation thorough. For a particular lay juror, who may not be used to thinking in terms of precise numerical probabilities, a nonquantitative formulation *plus* a numerical formulation such as 'in the high 90's' seems to make the most sense.

7. Conclusion

On the whole, Tillers and Gottfried are right that focusing on quantification would reduce the adverse effects of stereotyping, heuristic bias and other factors that should not influence the jury to adjust the standard of proof. Franklin is also right in suggesting that—except for gross quantification, such as his suggested percentage of 'greatly above 80%'—more precision is not practicable.[20] Finally, we think we are correct in suggesting that variations in the community with respect to what risks should be borne of convicting the innocent and how much pressure should be put on the police to investigate fully and fairly provide good reason for requiring a 'very high' probability of guilt for conviction—a probability in the high 90's—as well as a clear articulation of the standard that stresses our abhorrence of convicting the innocent.[21]

A verbal articulation plus a quantitative statement is justifiable, and on this point Tillers, Gottfried, Franklin and we seem to agree. Our jurors are now subjected to state-sponsored gambling and widespread health calculations based on pharmaceutical studies, and many of them have taken statistics course and have college degrees, so they are well habituated to thinking in terms of numerical probabilities.[22]

[20] Some research supports the proposition that any distinction finer than the difference between 'preponderance of the evidence' and beyond reasonable doubt would be cognitively meaningless. See Piattelli-Palmarini, *supra* n. 5 ('[W]e seem to understand only four degrees of probability for an event: very likely, somewhat likely (more likely to happen than not), somewhat unlikely (more likely not to happen), and very unlikely. Inside those four compartments, all is gray. No difference makes any difference. A 6 percent probability appears to us already sufficiently "very unlikely," that the significantly inferior probability of 1 percent is just "the same."').

[21] Cf. NEWMAN, D. J. (1966) *Conviction: The Determination of Guilt or Innocence Without Trial*. Little Brown and Company, pp. 188–196 (acquittal of the guilty to control police enforcement methods); *ibid.* at 134–151 (judicial discretion to acquit the guilty).

[22] But see, e.g. the heuristics problems in VISCUSI, W. K. (1983) *Risk By Choice*. Harvard University Press, pp. 76–78 (imperfections in Workers' Understanding of Risks); Piattelli–Palmarini, *supra* n. 5. Cf. BREYER, S. (1993) *Breaking the Vicious Circle: Toward Effective Risk Regulation*. Harvard University Press, pp. 33–39 (surveying 'several examples of thinking that impede rational understanding' and quoting Oliver Wendell Holmes to the effect that 'most people think dramatically, not quantitatively').