

**Case comment—*United States v. Copeland*,
369 F. Supp. 2d 275 (E.D.N.Y. 2005): quantification of the
‘proof beyond reasonable doubt’ standard**

JAMES FRANKLIN[†]

*School of Mathematics and Statistics, University of New South Wales,
Sydney 2052, Australia*

[Received on 21 September 2006; accepted on 30 October 2006]

There are many reasons for objecting to quantifying the ‘proof beyond reasonable doubt’ standard of criminal law as a percentage probability. They are divided into ethical and policy reasons, on the one hand, and reasons arising from the nature of logical probabilities, on the other. It is argued that these reasons are substantial and suggest that the criminal standard of proof should not be given a precise number. But those reasons do not rule out a minimal imprecise number. ‘Well above 80%’ is suggested as a standard, implying that any attempt by a prosecutor or jury to take the ‘proof beyond reasonable doubt’ standard to be 80% or less should be ruled out as a matter of law.

Keywords: evidence standard; proof beyond reasonable doubt; quantification; logical probability.

Objections to the quantification of the criminal standard of proof come from two directions. From the direction of policy, ethics and psychology, the problems raised include the following:

- There may be different standards appropriate to different cases, e.g. a higher standard where the punishment is heavier.
- The jury is properly left to decide the standard in the light of the facts of the particular case.
- Since there is in fact considerable disagreement as to the correct numerical value of the standard, attempts to standardize it will create only confusion, evasions and a façade of uniformity where there is no true consensus.
- The majesty of the law and its powers of deterrence would be ill-served, if the law were forced to admit the truth about the number of false convictions it allows and the number of criminals it allows to go free.

Quite different objections arise from certain more conceptual problems about the nature of probability:

- Some probabilities may be inherently incapable of being given a precise number.
- Evidence suitable for conviction should be ‘substantial’ or ‘weighty’, and a numerical probability expresses only the balance between favourable and unfavourable reasons, not whether those reasons are substantial.

[†] Email: j.franklin@unsw.edu.au

- A numerical standard will tend to draw attention to evidence that is quantified and logically relevant but legally inadmissible, such as proportions in reference classes containing the defendant.

These conceptual problems have rarely been clearly distinguished from the more commonly discussed policy and ethical problems. They are therefore developed at greater length here. It is concluded, however, that though all these arguments have some force, it is still desirable to introduce some minimal quantification into the reasonable doubt standard. In particular, any probability less than 0.8 should be declared less than proof beyond reasonable doubt in all circumstances.

Although betting odds, biases of dice and relative frequencies in populations are inherently numerical, that is not obviously so with the logical probabilities that concern the relation of evidence to hypothesis. According to the classic exposition of Keynes' 'Treatise on Probability', the relation of evidence to hypothesis, in cases such as proof beyond reasonable doubt in law or the evaluation of scientific theories in the light of experimental evidence, is a logical matter, a kind of partial implication.¹ Certainly, there are cases where it is very natural to attach a precise number to that relation. For example, if the hypothesis is 'This swan is black' and the (sole relevant) evidence is '15% of swans are black', then it is natural to attach a precise numerical probability to the relation between the two, namely,

$$P(\text{This swan is black} | 15\% \text{ of swans are black}) = 0.15.$$

At the other extreme, it is unnatural to attach any number, precise or otherwise, to

$$P(\text{The moon is made of green cheese} | \text{Marilyn Monroe was murdered by the CIA}).$$

The 'evidence' has no logical relation to the hypothesis—there is no partial implication, hence no number expressing it. If one insists on numbers, one might admit that

$$0 < P(\text{The moon is made of green cheese} | \text{Marilyn Monroe was murdered by the CIA}) < 1$$

on the grounds that the conclusion is neither necessarily true nor necessarily false, but then one would have a maximally imprecise probability.

More typical cases of evidence evaluation, Keynes thought, lay between these two extremes of maximally precise and maximally imprecise probabilities.² Such cases of imprecision can arise even when the evidence is itself numerical. For example, it may be reasonable to conclude that

$$P(\text{This swan in Park Zoo is black} | 15\% \text{ of swans are black and } 30\% \text{ of fauna in Park Zoo are black})$$

lies between 0.15 and 0.3, but standard probability theory gives little guidance on where in that interval it should lie. We know even less how to combine the complex pieces of disparate kinds of evidence typical of a criminal case. Either there is no precise number even in principle to be attached to the probability of the defendant's guilt on the evidence, or the amount of background

¹ KEYNES, J. M. (1921) *Treatise on Probability*. London, Macmillan, c. 1.

² KEYNES, *Treatise*, c. 3.

knowledge needed to determine jurors' reasonable evaluations of testimony and argument is so large and amorphous that it is impossible to determine what the number is. In either case, forcing a single number on the relation of the evidence to the hypothesis of guilt falsifies the situation.

On the other hand, as Peter Tillers points out,³ quantification with imprecise numbers is still quantification, and so arguments that there is no precise number to be attached to a standard of proof do not carry over to arguments against imprecise but still numerical quantification. If there are reasons against choosing any one precise number such as 0.95 for the criminal standard of proof, that does not rule out an imprecise level such as 'considerably above 0.8' as a requirement for adherence to that standard.

A second problem arises from Keynes' problem of the 'weight of evidence'.⁴ A probability $P(h|e)$ expresses the balance of the reasons in evidence e for and against hypothesis h . But that balance may be a balance between few and light reasons or between many and solid reasons. The matter is most easily appreciated when $P(h|e)$ is a half, since it is easy to find cases where either few or many reasons for and against a conclusion balance. Keynes asks about the difference between

$P(\text{the next throw with this coin is a head}|\text{the coin appears symmetrical})$

and

$P(\text{the next throw with this coin is a head}|\text{the coin appears symmetrical and about 500 of the last 1000 throws of this coin have been heads}).$

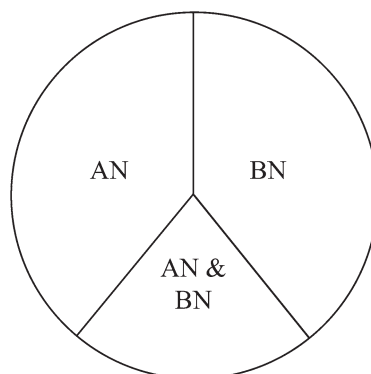
Both probabilities are $1/2$, but the second is based on the balance of much more evidence. It has greater 'weight'. The concept appears in a minimal way in the 'burden of production' of some appreciable amount of evidence that is required for a civil case to begin.⁵ But it is more evident in civil cases that involve decision on very little evidence. Since the civil standard of proof is $1/2$, or the 'preponderance of evidence' or 'balance of probabilities', realistic cases have arisen where decisions have had to be made on the balance of very small amounts of evidence, i.e. on probabilities of low weight. A celebrated instance is the Australian case *TNT Management Pty Ltd v. Brooks*. The plaintiff's husband was one of the two drivers killed in a head-on collision on a straight road. There were no witnesses and almost no further relevant evidence, and hence a symmetry in the evidence with respect to each driver. The legal situation required a decision as to whether, on the balance of probabilities, the other driver was negligent (irrespective of any possible negligence on the part of the plaintiff's husband). It was argued, using the following diagram, that on the balance of probabilities the other driver *was* negligent.⁶

³ TILLERS, P., Law, Probability and Risk, 5, 2006 (in press)

⁴ KEYNES, *Treatise*, c. 6; COHEN, L. J. (1985) Twelve questions about Keynes' concept of weight. *British Journal for the Philosophy of Science*, 37, 263; JAYNES, E. T. (2003) *Probability Theory: The Logic of Science* Cambridge, Cambridge University Press, c. 18; Similar ideas originally in PEIRCE, C. S. (1878) The probability of induction. *Popular Science Monthly*, 12, 705; A sixteenth century anticipation in FRANKLIN, J. (2001) *The Science of Conjecture: Evidence and Probability before Pascal*. Baltimore, Johns Hopkins University Press, 76–79.

⁵ FLEMING, J. & HAZARD, G. C. (1985) *Civil Procedure*, Little Brown 3rd edn. Boston, §7.7.

⁶ *T.N.T. Management v. Brooks* (1979) 23 A.L.R. 345, discussed in EGGLESTON, R. (1983) *Evidence, Proof and Probability*, 2nd edn. London, Weidenfeld and Nicolson p. 184 and in FRANKLIN, J. (2001) Resurrecting logical probability. *Erkenntnis*, 55, 277.



AN: Plaintiff's husband alone negligent

BN: Defendant's driver alone negligent

AN & BN: Both drivers negligent

Such cases are alarming, mainly because of the lack of robustness of the low-weight probabilities to new evidence. A small amount of further evidence would substantially change the probability. Therefore, the decision has some random element in it, arising from the randomness of the evidence that the court has available. Nevertheless, in civil cases it is arguable that this is the best one can do—a decision must be forced and the best one available on the evidence is the correct one, even if the evidence is scanty.

The use of probabilities of low weight in criminal cases is more worrying.⁷ A probability of guilt of 0.9 reached through balancing a small amount of evidence is different from a probability of 0.9 based on a mass of evidence, because the chance discovery of a new minor piece of evidence could well reduce the first to 0.7 but is unlikely to do so for the second. One might therefore be rationally less willing to condemn a defendant to a heavy sentence on a probability of 0.9 of low weight than on a probability of 0.9 of high weight. The purely qualitative language of 'beyond reasonable doubt' could be argued to mean a doubt that is both large enough in probability and of sufficient weight to rely on. Likewise, the 'merely fanciful' doubts that will not sway the jury from its conviction of guilt may be doubts that are not merely low in probability but low in weight, i.e. not based on any solid evidence but only on mere logical possibility or an ingenious imagination.

It is true that there is some necessary connection between high probabilities and high weight, in that it seems to be impossible to obtain extreme probabilities—those near 0 or 1—without some considerable weight. For example, if the probabilities are based purely on relative frequencies in some reference class, then a large reference class (and hence high weight) is needed to obtain a probability close to 0 or 1. Since

$$P(\text{this A is a B} | x\% \text{ of all As are Bs}) = x\%,$$

a probability of 0.9 can only be reached in this way with a population of at least 10, while a probability of 0.99 can only be reached with a population of at least 100. Nevertheless, it remains true

⁷ DAVIDSON, B. & PARGETTER, R. (1987) Guilt beyond a reasonable doubt. *Australasian Journal of Philosophy*, **65**, 182; Clarifications in DUNHAM, N. J. & BIRMINGHAM, R. L. (1989) On legal proof. *Australasian Journal of Philosophy*, **67**, 479.

that the probabilities considered by many to reach the standard of proof beyond reasonable doubt, those around 0.9 or just above, can be obtained with populations of only 10 or a dozen, a sample size which (in opinion polling, for example) is laughably inadequate as the foundation for any casual inference, let alone as sufficient for action in a serious matter like judicial punishment.

Again, these questions on the relation of high probability and high weight have no bearing on whether low numerical probabilities such as those less than 0.8 should be ruled out as failing to satisfy the standard of 'beyond reasonable doubt'. If a probability on the evidence is less than 0.8, then whether it is of high or low weight makes little difference. Conviction on that probability carries a high risk of condemning the innocent.

This brings us to the third conceptual problem with quantification of the reasonable doubt standard, the fact that the probabilities most usually and easily quantified, those arising from a proportion in a reference class like

$$P(\text{this A is a B} \mid x\% \text{ of all As are Bs}) = x\%,$$

are of very doubtful legal relevance at all. The evidence that 99% of drivers cut a corner is not allowed as evidence that a particular driver cut that corner on a particular occasion. It is neither allowed as sufficient evidence to prove that hypothesis nor allowed as partial evidence.⁸ While there have been debates on the admissibility of certain kinds of other 'similar fact' evidence such as evidence of prior convictions and of character, there has been no serious suggestion that simple membership of a reference class of high criminality (or high criminality of a certain sort) should be admissible as evidence.⁹ That is despite the fact that such evidence is logically relevant—indeed, in cases like drug trials, the evidence that a certain application of a drug will cure a patient with a certain disease often consists purely of the fact that 90% of patients with the disease are cured by the drug. It is also despite the fact that jurors are allowed, indeed encouraged, to evaluate evidence in the light of their general knowledge of human nature and of the common course of nature, knowledge which by and large arises from observation of relative frequencies (or from the testimony of others as to relative frequencies).

There is a special problem with frequencies in the reference class of which every defendant is a member, namely the class of defendants. Jurors have beliefs about this class, and widely differing beliefs. In the survey of Saunders on the opinions of 130 numerate adults on the level of probability equivalent to 'proof beyond reasonable doubt', all but two of the responses lay between 50% and 99.99999% inclusive. The two outliers both explained their opinion by their beliefs about defendants generally. One argued for 30%, on the grounds that defendants are generally guilty and so the standard for releasing accused murderers should be high; in his homeland (Nigeria), he thought, a presumption that a defendant was innocent would not be prudent. The other outlier, an African-American, believed defendants were often victims of police conspiracies and so demanded 100% certainty for conviction.¹⁰

⁸ Eggleston, pp. 59–60, 88–89.

⁹ Though possibly sometimes relevant to sentencing: TILLERS, P. (2005) If wishes were horses: discursive comments on attempts to prevent individuals from being unfairly burdened by their reference classes. *Law, Probability & Risk*, 4, 33; COLYVAN, M., REGAN, H. & FERSON, S. (2001) Is it a crime to belong to a reference class? *Journal of Political Philosophy*, 9, 166.

¹⁰ SAUNDERS, H. D. (2005) *Quantifying Reasonable Doubt: A Proposed Solution to an Equal Protection Problem*. ExpressO Preprint Series, paper 881 (<http://law.bepress.com/expresso/eps/881>).

Other frequencies in reference classes that could be but are not allowed to be considered as relevant include recidivism rates. Traditional justifications of high standards for proof beyond reasonable doubt along the lines of 'it is better to allow 10 (or 100) criminals to go free than to condemn one innocent' invite cost-benefit analyses that could only be conducted with close attention to relative frequencies. The cost of allowing criminals to go free depends on the chance of re-offending by those criminals. Recidivism rates are possibly sex-specific, race-specific and crime-specific and are certainly age-specific.¹¹ There is no prospect whatever that the presentation of such statistics will be permitted in court to influence a jury's setting of its standard of proof.

The reasons for the inadmissibility of all such reference-class evidence may be either psychological or ethical: either psychological claims as to its prejudicial effect on juries—i.e. claims that as a matter of psychological fact it tends to be overweighted and lead to wrong decisions—or ethical reasons on the injustice of condemning someone on the basis of the acts of others. In either case, it is to some degree problematic for the quantification of the standard of proof beyond reasonable doubt that the evidence that would most naturally lead to numbers is inadmissible, and therefore that admissible evidence is almost always qualitative. That takes us back to the first problem above, that it may be impossible in principle to assign a precise or even an imprecise number to some probabilities based on such evidence.

Again, none of this reasoning tends to rule out placing a floor of 0.8 on the criminal standard of proof. If the probability of guilt on the admissible evidence can be assigned a number and if that number (precise or otherwise) is not substantially greater than 0.8, then it carries a high risk of condemning the innocent.

Therefore none of the conceptual problems with probability constitute reasons against setting a minimum of 0.8 on the standard of proof beyond reasonable doubt. Nor do any of the ethical or psychological reasons mentioned at the beginning of the article. As these have been discussed at length, brief comments will suffice here. Undoubtedly, there are some reasons for insisting on an exceptionally high standard of proof in capital and other grave cases,¹² but the fact that some standards may be less stringent than others is no reason to relax standards at the bottom end. Jurors and the judges instructing them may need to be left with some discretion and flexibility,¹³ but 'some discretion' is not 'arbitrary discretion'; their discretion is already constrained by the requirement that 'proof beyond reasonable doubt' is a much higher probability than 'the preponderance of evidence', so some further restriction to eliminate the observed wide variation in reported standards should be acceptable in principle. It is true that attempts at constraint may be less than totally successful in their effects on real juries,¹⁴ but given the extreme present confusion revealed by surveys, any confusion resulting from a simple demand that the criminal standard should be more than 0.8 is likely to be much less than the current confusion, where a façade of uniform language hides an unacceptable large variation in numbers. In any case, the evidence suggests, as far as it goes, that jurors can understand quantified standards better than unquantified ones and act more uniformly as a result.¹⁵

¹¹ For example HANSON, R. K. (2002) Recidivism and age: follow-up data from 4,673 sexual offenders. *Journal of Interpersonal Violence*, **17**, 1046.

¹² Reasons for and against reviewed in LILLQUIST, E. (2005) Absolute certainty and the death penalty. *American Criminal Law Review*, **42**, 45.

¹³ Various reasons in STOFFELMAYR, E. & DIAMOND, S. S. (2000) The conflict between precision and flexibility in explaining "beyond a reasonable doubt." *Psychology, Public Policy, and Law*, **6**, 769.

¹⁴ Warnings in STRAWN, D. U. & BUCHANAN, R. W. (1976) Jury confusion: a threat to justice. *Judicature*, **59**, 478.

¹⁵ KAGEHIRO, D. K. (1990) Defining the standard of proof in jury instructions. *Psychological Science*, **1**, 194.

The public's respect for the law will probably not be diminished any further, by any unpleasant truths that may come to light about the numbers of false convictions or of acquitted criminals; in any case, such figures cannot be inferred from the standard of proof because the base rates in the class of defendants is unknown.

The main argument in favour of a quantitative constraint on the 'beyond reasonable doubt' standard is the gross divergence of opinions as to the numerical meaning of the standard. The survey results are clear.¹⁶ The consequences are also clear: there are many real juries in which a majority of jurors believe that a probability of 70% satisfies the standard of proof beyond reasonable doubt. That is a cause for alarm to all potential defendants, that is, to everyone. It is a travesty of the commitment to consistency on which the law prides itself in other areas.

A first step towards justice would be to rule out numerical probabilities that are clearly unreasonable. If a jury asks a judge for guidance on whether a 75% probability is sufficient, the law should be unambiguous in its answer. The answer should be 'no'.

An appropriate numerical standard to choose as an absolute minimum follows from Judge Weinstein's suggestion in 'Copeland III' (*United States v. Copeland*, 369 F. Supp. 2d (E.D.N.Y. 2005)) of 20% for a 'reasonable probability', and hence of 80% for its inverse or complement 'clear, unequivocal and convincing' evidence.¹⁷ The case was a civil one involving the serious consequence of deportation. Since proof beyond reasonable doubt is well above clear, unequivocal and convincing evidence, it follows that proof beyond reasonable doubt means 'well above a probability of 0.8'. Any suggestion from a jury that 0.8 or less is adequate can be ruled out, while the qualification 'well above' will avoid any suggestions that something just above 0.8 is in fact adequate, and will not obstruct any later attempts to quantify the standard more exactly.

¹⁶ 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; KAGEHIRO, D. K. & STANTON, W. C. (1985) Legal vs. quantified definitions of standards of proof. *Law and Human Behavior*, 9, 159; SAUNDERS, *op. cit.*; HOROWITZ, I. A. & KIRKPATRICK, L. C. (1996) A concept in search of a definition: the effects of reasonable doubt instructions on certainty of guilt standards and jury verdicts. *Law and Human Behavior*, 20, 655; HOROWITZ, I. A. (1997) Reasonable doubt instructions: commonsense justice and standard of proof. *Psychology, Public Policy and Law*, 3, 285.

¹⁷ As discussed in Tillers, previous article.