PROBABILISTIC STANDARDS OF PROOF, THEIR COMPLEMENTS AND THE ERRORS THAT ARE EXPECTED TO FLOW FROM THEM

David Hamer[*]

ABSTRACT

Probability theory provides insights into the levels at which standards of proof are set. An analysis is provided of the rates of errors that flow from different standards, a necessary element of which is an assumption as to the level of certainty achieved in the flow of cases. Where one party is required to satisfy a standard of proof, a complementary standard is imposed on the other party.

I INTRODUCTION

Factual accuracy is the 'paramount'[1] goal of the civil or criminal trial. In this article I use probability theory to understand the role

---

[1] Marvin E Frankel, 'The Search for Truth: An Umpireal View' (1975) 123 *University of Pennsylvania Law Review* 1031, 1033, 1055. The goal of factual accuracy has also been described as 'foremost' (Adrian S Zuckerman, *Principles of Criminal Evidence* (1989) 7), 'fundamental' (Vern R Walker, 'Preponderance, Probability and Warranted Fact-finding'(1996) 62 *Brooklyn Law Review* 1075, 1081), 'principal' (Jonathan Koehler and Daniel N Shaviro, 'Veridical Verdicts: Increasing Verdict Accuracy through the use of Overtly Probabilistic Evidence and Methods' (1990) 75 *Cornell Law Review* 247, 250), 'overriding' (William Twining, *Theories of Evidence: Bentham and Wigmore* (1985) 117), 'necessary' (William Twining, 'Rationality and scepticism in judicial proof: some signposts' (1989) II *International Journal for the Semiotics of Law* 69, 72), 'primary' (ibid 70) and 'central' (Jack Weinstein, 'Some difficulties in devising rules for determining truth in judicial trials' (1966) 66 *Columbia Law Review* 223, 243).

played by the ordinary[2] civil and criminal standards of proof in the law's pursuit of factual accuracy.

Standards of proof inform the fact-finder of the basis on which a version of facts should be adopted. It might be argued that the standard best designed to achieve factual accuracy is obvious — demand absolute certainty. If the fact-finder is certain that the defendant is not liable, a finding should be made for the defendant. If the fact-finder is certain of the defendant's liability, a finding should be made for the plaintiff or prosecution. If the fact-finder is uncertain, no finding should be made.

This view is of course hopelessly idealistic. The material facts are not available for inspection by the fact-finder, but must be inferred from incomplete and inconclusive evidence.[3] Perhaps a fact-finder might occasionally feel totally certain about the material facts, but arguably she should not. The better view is that '[i]n the real world of human actions we can never be absolutely certain of anything'.[4] If certainty was a prerequisite to judgment, very few decisions would be made, if any. While the system would have the appearance of pursuing the goal of factual accuracy, other objectives would be severely compromised. There would be no peaceful resolution of civil disputes,[5] creating the danger that the parties may take matters into their own hands. Criminals would go unpunished; there would be no deterrence,

---

2   This article does not consider the question of intermediate or flexible standards. For a more general discussion see David Hamer, 'The Civil Standard of Proof Uncertainty: Probability, Belief and Justice' (1994) 16 *Sydney Law Review* 506; Mike Redmayne, 'Standards of Proof in Civil Litigation' (1999) 62 *Modern Law Review* 167; Charles R Williams, 'Burdens and Standards in Civil Litigation' (2003) 25 *Sydney Law Review* 165.

3   As has long been recognised: see W Twining, 'The Rationalist Tradition of Evidence Scholarship' (1982), in Enid Campbell and Louis Waller, *Well and Truly Tried: Essays on Evidence* (1982).

4   Barbara Shapiro, '"To a Moral Certainty": Theories of Knowledge and Anglo-American Juries 1600–1850' (1986) 38 *Hastings Law Journal* 153, 193; see also *Re Winship* 397 US 358 (1970), 370 (Harlan J).

5   See eg Charles Nesson, 'Reasonable Doubt and Permissive Inferences: The Value of Complexity' (1979) 92 *Harvard Law Review* 1187, 1194; John Thibaut, and Laurens Walker, 'A Theory of Procedure' (1978) 66 *California Law Review* 541.

rehabilitation or retribution.[6] And while limitless resources were being consumed, there would be an absence of closure for the parties and the community.[7]

The legal system is realistic. It does not demand the unattainable. Standards of proof, even the criminal standard, are recognised as inherently 'probabilistic'.[8] This is not to say that factual accuracy is sacrificed in favour of other more pragmatic ends. On the contrary, in this article I support the now 'conventional'[9] view that standards are set at levels of probability that minimise the costs of factual error. As explained in Parts II and III below, the ordinary civil standard of 50 per cent minimises the subjective expected rate of error while the criminal standard is set at a higher level to reduce the risk of the more costly error — the wrongful conviction. However, it must be recognised that the more the standard is increased, the greater the risk of the other error, the erroneous acquittal.

Clearly the level at which a standard is set will have an impact on the rates of the different kinds of errors. As far as the standard of proof is concerned, to reduce the risk of one kind of error is to increase the risk of the other. The trade-off is inescapable. But error rates are not solely determined by the standard of proof. Regard must also be had to the degree of certainty attained by the fact-finder in the flow of cases. All else being equal, the greater the fact-finder's certainty, the more possible it is to avoid error. The mathematical relationship between the probability distribution of the cases, the standard of proof and error rates is explored below in Part IV.

---

6   See eg Charles Nesson, 'The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts Thesis' (1985) 98 *Harvard Law Review* 1357, 1359; Jerome Frank, *Courts on Trial: Myth and Reality in American Justice* (1949) 101.

7   Weinstein, above n 1, 242.

8   Eg *Victor v Nebraska* 511 US 1 (1994) 14, 22 (O'Connor J), but contrast 25 (Ginsburg J, conc), 36 (Blackmun J, diss). See also *Re Winship* 397 US 358 (1970), 370 (Harlan J).

9   Dale Nance, 'Evidential Completeness and the Burden of Proof' (1998) 49 *Hastings Law Journal* 621, 622. See also Ron Allen et al, 'Probability and Proof in *State v Skipper*: An internet exchange' (1995) 35 *Jurimetrics Journal* 277 (22 August 1994) 309 (Friedman R).

Recognition of the probabilistic nature of the standard of proof — placing it at a certain position on the unit interval — leads immediately to an instructive corollary. There is, with regard to each issue at trial, not a single standard, but two complementary standards, one for each of the opposing parties. Talk of the criminal defendant's standard of proof may appear inconsistent with the presumption of innocence. However, such a concern should be addressed through careful instruction on burdens of proof. As outlined in Part V, an appreciation of the complementary corollary should lead to a better understanding and clearer instruction on the criminal standard, including those where the prosecution relies exclusively upon circumstantial evidence.

II THE ORDINARY CIVIL STANDARD: MINIMISING THE SUBJECTIVE EXPECTED RATE OF ERROR

The ordinary civil case is symmetrical.[10] The plaintiff and the defendant have an equal stake in the proceedings. The parties may, for example, be arguing as to who should bear the loss of an unsuccessful business dealing. The law seeks to treat the parties even-handedly. In the United States it is commonly held that the civil standard of proof requires the fact-finder to adopt the plaintiff's version of facts where they are established to a 'preponderance of probability' or a 'preponderance of evidence'.[11] In England and Australia, the predominant formula is that the plaintiff must establish her case on a 'balance of probabilities'.[12] Many authorities hold that these formulations can be interpreted as 'more probable than not' and as setting a standard of 50 per cent.[13] Putting to one side the burden of proof,

---

[10] Arguably, where this symmetry is not present and the defendant has more at stake than the plaintiff, the standard should increase or an intermediate standard should be imposed. See above, n 2.

[11] Eg *Grogan v Garner* 498 US 279 (1990), 286.

[12] Eg *Neat Holdings Pty Ltd v Karajan Holdings Pty Ltd* (1992) 110 ALR 449; *Re H* [1996] AC 563, 586.

[13] Eg *Re Winship* 397 US 358 (1970), 371–372 (Harlan J); *Bradshaw v McEwans Pty Ltd* (unreported, High Court of Australia, Dixon, Williams, Webb, Fullager and Kitto JJ, 27 April 1951), quoted in *Holloway*

which is generally borne by the plaintiff, the standard of proof is equally demanding on both parties. Each party will be concerned to demonstrate that her version of facts is more probable than the other party's competing version.

This standard minimises the risk of error in the instant case, and minimises the subjective expected rate of errors in civil litigation.[14] If a plaintiff proves her case to a probability of 60 per cent, a verdict will be rendered for the plaintiff and, in the fact-finder's view, this verdict will probably be correct. If ten plaintiffs proved their cases to a level of 60 per cent, all would succeed. Six of these verdicts would be expected to be factually correct, and four factually incorrect, though which were which would be unclear. Of course it would be preferable for all ten verdicts to be factually correct, but given the limitations of the evidence, this is not achievable. Expected accuracy would not be increased by increasing the standard of proof. If the standard were raised to, say, 65 per cent, the defendant would be successful in each of the ten cases. However, the fact-finder would then consider only four verdicts to be factually correct, and six to be factually incorrect. By increasing the standard of proof, the subjective expected rate of factually correct verdicts would be reduced. As these very simple examples demonstrate, there is a close relationship between the standard of proof and expected rates of the different kinds of errors, mediated by the probability distribution over the set of cases in question. The mathematical relationship between these three things will be examined in more detail below. First, however, I should clarify what is meant by subjective expected rate of errors.

The errors with which we are concerned in the present context are, of course, factual errors. The verdicts referred to in the examples above will all be legally correct in the sense that they were rendered in accordance with the appropriate standard of proof.[15] However, a legally correct verdict may nevertheless be

---

v McFeeters (1956) 94 CLR 470, 480–1 (Williams, Webb and Taylor JJ); Davies v Taylor [1974] AC 207, 219 (Lord Simon); Hamer, above n 1, 509.

[14] David Kaye 'Naked statistical evidence' (1980) 89 Yale Law Journal 601.

[15] Compare Zuckerman, above n 1, 122. Hay's recent economic analysis of burdens of proof appears inexplicably to focus entirely on legal error. He is concerned, inter alia, to minimize 'the costs of erroneous outcomes'. (Bruce L Hay, 'Allocating the Burden of Proof' (1997) 72 Indiana Law Journal

factually incorrect, in that the verdict is premised upon a version of the material facts that does not *correspond* with what really transpired.[16] The converse is also true. A legally incorrect verdict — one based, for example, on highly probative but inadmissible evidence — may be factually correct.[17] The same point can be made of inductive reasoning in general. Where evidence is incomplete or ambiguous, and it is not possible to eradicate all doubt, valid reasoning may unavoidably lead to factual error.[18]

651, 651.) However, he defines a correct or 'meritorious' outcome as one where the 'the evidence supports the claim' (657), and an 'erroneous outcome, as we have defined it, occurs when the party whom the evidence supports nonetheless loses' (662). His definition of error has no reference to any lack of correspondence between the material facts, as found, and what really occurred. His model appears to totally ignore the costs of error that flow from the unavailability and incompleteness of evidence.

[16] The correspondence theory of truth has been criticised for its 'naïve and outdated view that the outside world is able somehow to impress itself on our senses without change or remainder': Bernard Jackson, 'Semiotic Scepticism: A Response to Neil MacCormick' (1991) IV *International Journal for the Semiotics of Law* 175, 178; see also John Jackson 'Two Methods of Proof in Criminal Procedure' (1988) 51 *Modern Law Review* 549, 558. However, the charge of naivety does not appear warranted in the present context. An acknowledged difficulty in proving real world events provides the raison d'être of standards of proof. Perhaps it could be said that it would be naïve even to aspire to establish a correspondence. However, such an extreme scepticism about the possibility of knowing the outside world is as open to criticism as the most naïve realism. '[B]oth take for granted the relationship between evidence and reality'. 'Instead of dealing with the evidence as an open window, contemporary sceptics regard it as a wall, which by definition precludes any access to reality.' Carlo Ginzburg 'Checking the Evidence: The Judge and the Historian', in James Chandler, Arnold I Davidson, and Harry Harootunian (eds), *Questions of Evidence: Proof, Practice, and Persuasion across the Disciplines* (1994) 290, 294; Steven L Winter, 'The Cognitive Dimension of the *Agon* between Legal Power and Narrative Meaning' (1989) 87 *Michigan Law Review* 2225, 2244; William Twining, *Rethinking Evidence, Exploratory Essays* (1990) 96.

[17] See also Zuckerman, above n 1, 122; Andrew D Leipold, 'The Problem of the Innocent, Acquitted Defendant' (2000) 94 *Northwestern University Law Review* 1297, 1298 fn 9; Henry L Chambers, 'Reasonable Certainty and Reasonable Doubt' (1998) 81 *Marquette Law Review* 655, 658 fn 11.

[18] See also Detlof Von Winterfeldt, and Ward Edwards, *Decision Analysis and Behavioral Research* (1986) 2–3; Zuckerman, above n 1, 19; Willem A Wagenaar, Peter van Koppen, and Hans Crombag, *Anchored Narratives: The Psychology of Criminal Evidence* (1993) 14.

And conversely, invalid reasoning may lead to factually correct results.

The statement that a 50 per cent standard of proof will minimise the rate of error is qualified in two respects — it states an *expectation* which is based upon what is commonly termed a *subjective* probability assessment. Taking the subjectivity point first, the fact-finder's assessment of the probability of liability will generally lack the objective appearance of a probability based on statistical data or knowledge of the structure of a game of chance. Juridical fact-finders' assessments are rarely quantified, and if they were, it would be surprising to find different fact-finders arriving at precisely the same figure. For this reason, my use of probability theory in the analysis of human degrees of belief falls within what is generally termed the 'subjective theory' of probability.[19] It may be questioned whether the term 'subjective' is entirely appropriate in a couple of respects. First, some commentators argue that all probability assessments are subjective in the sense of being subject to a limited body of evidence. Probability, it is said, is not a feature of the world, but is an expression of partial belief based upon our limited knowledge of the world.[20]

More importantly in the present context, the term 'subjective' might be considered to carry the connotation of relativism.[21] Critics of the subjective theory have suggested that it exhibits 'great tolerance'[22] about different interpretations of evidence,

---

[19] Eg Henry Kyburg, and Howard Smokler, *Studies in Subjective Probability* (1964); Peter Ayton and George Wright (eds), *Subjective Probability* (1994).

[20] Eg Richard Wright, 'Causation, Responsibility, Probability, Naked Statistics and Proof: Pruning the Bramble Bush by Clarifying the Concepts' (1988) 73 *Iowa Law Review* 1001, 1041; Bernard Robertson, and G Anthony Vignaux, 'Probability—The Logic of the Law' (1993) 13 *Oxford Journal of Legal Studies* 457, 469–70.

[21] This leads Ayton and Wright to suggest that 'personalist' may be a better term than 'subjective' (Peter Ayton and George Wright, 'Subjective probability: What should we believe?', 165 fn 1, in Wright and Ayton (eds) above n 19). However, the term 'subjective probability' appears in the title of their article, and the book in which it appears.

[22] Eg, Wesley C Salmon, *The Foundations of Scientific Inference* (1967) 81; see also Kyburg and Smokler, above n 19, 7; James Logue, 'Weight of evidence, resiliency and second order probabilities', in Ellery Eells and

placing the correctness of probability assessments 'wholly outside the framework of rational controversy because it is not an objective issue'.[23] These criticisms are not necessarily fair. Rigorously applied, the subjective theory is capable of reconciling the divergent assessments of different individuals.

Suppose two jurors differ in their assessments of the probative value of evidence of flight from the police.[24] The standard measure of probative value within the subjective probability theory is the Bayesian likelihood ratio.[25] Juror X has had few and favourable experiences of the police and considers a guilty person would be far more likely to flee than an innocent person. He views the evidence as strongly incriminating and gives it a high likelihood ratio: $L_x(F) \gg 1$.[26] Juror Y has had far less pleasant experiences of police practices and considers flight an understandable response to police investigation irrespective of guilt. She views the evidence as having very little probative value on the question of guilt: $L_Y(F) \approx 1$.[27] While the jurors have interpreted the flight evidence quite differently, the inequality, $L_x(F) \neq L_Y(F)$, should not be considered as probabilistically

Tomasz Maruszewski (eds), *Probability and Rationality: Studies on L. Jonathan Cohen's Philosophy of Science* (1991) 155.

[23] Jonathan Cohen, 'Subjective Probability and the Paradox of the Gatecrasher' [1981] *Arizona State Law Journal* 627, 630; see also Patrick Suppes, 'The Qualitative Theory of Probability', in Wright and Ayton (eds), above n 19, 31; Salmon, above n 22, 81.

[24] Compare Richard E McGarvie, 'Equality, Justice and Confidence' (1996) 5 *Journal of Judicial Administration* 141, 145.

[25] Bayes' theorem provides a means of updating a prior probability assessment to take account of new evidence. In its odds-likelihood form, the posterior odds of a proposition, G, given the fresh evidence, F, is equal to the product of the prior odds of the proposition, and the likelihood ratio for the evidence: $O(G|F) = O(G) \times L(F)$.

$L(F)$ is defined as $P(F|G)/P(F|\neg G)$ where '$\neg$' is the negation symbol, meaning 'not'. Note that odds are a function of probability, $o(p) = p/(1 - p)$, and vice versa, $p(o) = o/(1 + o)$.

Bayes' theorem is so central to applications of the subjective theory, that it is alternatively known as the Bayesian theory: R D Friedman, 'A Presumption of Innocence, Not of Even Odds' (2000) 52 *Stanford Law Review* 873, 874–5; David Kaye, 'Introduction: What is Bayesianism?' in Peter Tillers and Eric Green (eds), *Probability and Inference in the Law of Evidence: The Limits and Uses of Bayesianism* (1988) 1, 9.

[26] The symbol '$\gg$' means 'much greater than'.

[27] The symbol '$\approx$' means 'is roughly equal to'.

incomprehensible.[28] To resolve the apparent inconsistency we need 'to be as clear as possible about the conditioning'.[29] If we label each juror's background evidence B_X and B_Y, then the inequality becomes perfectly understandable: $L(F|B_x) \neq L(F|B_Y)$.[30] The jurors assessments differ because they are, in a real sense, assessing different bodies of evidence. This is not to say that all differences of opinion can be reconciled in this fashion. But if they can not, then rather than tolerating inconsistency, the subjective theory can be a vehicle for interrogating it.[31]

The second qualification to the proposition that a 50 per cent standard will minimise errors is that this merely states an *expectation*. And, where legal trials are concerned, confirming that expectation will be virtually impossible. In this respect a contrast can be drawn with the more explicitly probabilistic activity of betting. Offered even odds, a bettor would be advised to select the more probable side of the bet, since this offers the greater expectation of winning. She would then have the opportunity to see whether she made the right choice. Betting is a predictive activity offering the bettor the opportunity to gather evidence that is generally conclusive. Indeed, the obtaining of such conclusive evidence is crucial to the activity of betting; if there was a common uncertainty as to which horses or lottery tickets were winners, gambling would be far less popular than it is. The courts, however, are usually engaged in postdiction — proof of past events.[32] As time passes the evidence tends to

---

[28] The symbol '$\neq$' means 'is not equal to'.

[29] Ian W Evett and Bruce S Weir, *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists* (1998) 8; see also Robertson and Vignaux above, n 29, 470.

[30] Stating the former likelihood ratio in full, $L(F|B_x) = P(F|G\&B_X)/P(F|\neg G\&B_X)$.

[31] Not that such a task would be easy. On the contrary, it is 'an open-ended problem, since there is no end to the variety of complicated information that might be contained in B; and therefore no end to the complicated mathematical problems of translating that information into numerical values': Edwin Jaynes, *Probability Theory: The Logic of Science* (Fragmentary Edition of March 1996) 213.

[32] Occasionally courts are involved in predictive proof which raise another set of problems: see David Hamer, '"Chance would be a fine thing": Proof of

decline in quality: memories fade, forensic traces deteriorate and documents are shredded.[33] After the trial it will become increasingly difficult to confirm the accuracy of the fact-finder's verdict.[34]

In isolated cases an external observer may be able to form a view with virtual certainty that a verdict is factually correct or incorrect. She may have reference to crucial additional evidence that was inadmissible or unavailable at trial, giving her greater certainty than was available to the juridical fact-finder. However, the twin objectives of efficiency and finality place an obstacle in the path of those that would prove the existence of errors. If a reopening of the first determination is allowed, why should the second determination not also be questioned?[35] The underlying problem — the absence of a 'criterion'[36] or 'gold standard'[37] for checking postdictions — is intractable.

And so it must be acknowledged that the proposition that the 50 per cent standard maximises factual accuracy is subject to qualification. '*If* the judge's estimates are good, so that we can

---

causation and quantum in an unpredictable world' (1999) 23 *Melbourne University Law Review* 557; Wright, above n 20.

[33] See also Jonathan Cohen, 'Subjective Probability and the Paradox of the Gatecrasher' [1981] *Arizona State Law Journal* 627, 632.

[34] Kaye suggests the distinction between postdictive trials and predictive bets is merely one of degree (David Kaye, 'The Laws of Probability and the Laws of the Land' (1979) 47 *University of Chicago Law Review* 34, 45). With enough resources, and 'sufficiently brutal methods', questions about the material facts could 'in theory, be settled'. Meanwhile, there may be 'some residual uncertainty' about the subject of a bet, such a horse race: 'The picture of the supposed winner may have been taken from an improper angle, it may have been doctored, or the laws of optics may have been suspended by a sinister force during the exposure. It is all a matter of degree.' Generally though, the distinction is quite clear-cut.

[35] *Jackson v Virginia* 443 US 307 (1979), 337 fn 12 (Stevens J conc) citing Paul M Bator, 'Finality in Criminal Law and Federal Habeas Corpus for State Prisoners (1963) 76 *Harvard Law Review* 441, 450–1, 450–451; see also Daniel Givelber, 'Meaningless Acquittals, Meaningful Convictions: Do We Reliably Acquit the Innocent?' (1997) 49 *Rutgers Law Review* 1317.

[36] Wagenaar, van Koppen and Crombag, above n 18, 182.

[37] Michale L DeKay, 'The Difference between Blackstone-Like Error Ratios and Probabilistic Standards of Proof' (1996) 21 *Law and Social Inquiry* 95, 125.

take them as accurate statements of the probability of X ...'[38] 'Assuming a positive correlation between the fact-finder's belief and the true state of affairs ...'[39] It is true that it is generally not possible to determine whether the expected accuracy benefits of the probabilistic standard of proof are achieved. Nevertheless, the 50 per cent standard is highly defensible in ordinary civil cases.[40] The alternative is to set the standard at a level that is subjectively expected to increase the rate of error.[41]

III THE CRIMINAL STANDARD: MINIMISING THE EXPECTED COST OF WRONGFUL CONVICTIONS

The symmetry of the ordinary civil case is wholly absent in the criminal trial. Whereas in the ordinary civil trial an error against either party is considered equally regrettable, no such equation holds in the criminal trial. As Deane J of the High Court of Australia observes:

> [T]he searing injustice and consequential social injury which is involved when the law turns upon itself and convicts an innocent person far outweigh the failure of justice and the consequential social injury involved when the processes of the law proclaim the innocence of a guilty one.[42]

As a result, the criminal standard of proof is almost universally recognised as being higher than the civil standard. While, on

---

[38] Kaye, above n 14, 604 (emphasis added).

[39] Koehler and Shaviro, above n 1, 250 (emphasis added).

[40] See also Redmayne, above n 2, 169.

[41] Specific challenges may be made to fact-finders' probability assessments on the basis of doubts about the validity of certain exclusion rules. Eg Givelber suggests that many of the doubts jurors have about guilt result from exclusion rules that protect the guilty rather than the innocent, undermining the assumption that 'one [would] have the greatest doubt about a proposition that is not true': Givelber, above n 35, 1368.

[42] *Van der Meer v The Queen* (1988) 82 ALR 10, 31 (Deane J). The classic statement in the US is provided by Harlan J in *Re Winship* 397 US 358 (1970), 370–372 (Harlan J conc); see also ibid 363–364 (Brennan J); *Jackson v Delaware L&W RR* 170 Atl 22 (1933), 23–24 discussed by Vaughan C Ball, 'The Moment of Truth: Probability Theory and Standards of Proof' (1961) 14 *Vanderbilt Law Review* 807, 815.

occasions, judges have questioned whether the distinction between the two standards is 'more a matter of words than of substance'[43] the better view is that 'it is a matter of critical substance'.[44] The most common formulation of the criminal standard, achieving the level of constitutional necessity in the United States, requires proof of guilt 'beyond reasonable doubt'.[45]

It has long been fashionable to describe by ratio precisely how much worse it is to convict the innocent than to acquit the guilty. Perhaps the most famous is Blackstone's maxim of 1765 that 'it is better that ten guilty persons escape than that one innocent should suffer'.[46] However, this was just the midpoint in a steadily rising trend. Other commentators recommended ratios of 5:1,[47] 20:1,[48] and even 99:1.[49] In 1829 Bentham warned that insufficient attention was being given to the ever increasing difficulty that this would pose for the prosecution of crime: 'we must be on our guard against those sentimental exaggerations which tend to give crime impunity, under the pretext of insuring the safety of innocence'.[50]

---

[43] *Edmunds v Edmunds and Ayscough* (1935) VLR 177, 183, quoted with apparent approval in *Briginshaw v Briginshaw* (1938) 60 CLR 336, 353 (Starke J); see also 368 (Dixon J); *R v Hepworth & Fearnley* [1955] 2 QB 600, 603 (Lord Goddard); *Khawaja v Secretary of State* [1984] AC 74, 112 (Lord Scarman); *US v Feinberg* 140 F2d 952 (1944) (Learned Hand J); *Larson v Jo Ann Cab Corp* 209 F2d 929 (1954).

[44] *Rejfek v McElroy* (1965) 112 CLR 517, 521; see also *Briginshaw v Briginshaw* (1938) 60 CLR 336, 344 (Latham CJ)); *Bater v Bater* [1951] P 35, 36 (Denning LJ); *Addington v Texas* 441 US 418 (1978), 424–425, 426–427.

[45] *Re Winship* 397 US 358 (1970); *Green v The Queen* (1971) 126 CLR 28; *R v Hepworth and Fearnley* [1955] 2 QB 600.

[46] William Blackstone, *Commentaries on the Laws of England* (1765) Vol 4, c 27, 352.

[47] Matthew Hale, *Pleas of the Crown* (1678), 289.

[48] John Fortescue, *De Laudibus Legum Angliae* (Amos' translation, 1825).

[49] *Schlup v Delo* 513 US 298 (1995), 325, quoting Starkie, *Evidence* (3rd ed, 1830), 751.

[50] Jeremy Bentham, *Principles of Judicial Procedure* (1829) 169, quoted in William S Laufer, 'The Rhetoric of Innocence' (1995) 70 *Washington Law Review* 329, 333 fn 17.

More recently, decision theorists have developed a method by which the relative weights attached to the competing objectives can be converted into probabilistic standards of proof — putting a figure to 'beyond reasonable doubt'. This method of utility modelling now provides the 'standard'[51] or 'conventional'[52] understanding of standards of proof.[53] The starting point is to arrive at a numerical measure of the utility of the two possible correct outcomes and the two possible mistaken outcomes: U_{CG} — convicting the guilty; U_{AI} — acquitting the innocent; U_{AG} — acquitting the guilty; U_{CI} — convicting the innocent. In reality, the utility of a verdict will have one of two values, depending on whether it is right or wrong. But reality is inaccessible — some degree of uncertainty appears unavoidable — and it is therefore necessary to talk in terms of the utility that is expected from a verdict given the fact-finder's probability of guilt, p. The expected utility of a conviction will rise from a minimum of U_{CI} where $p = 0$, to a maximum of U_{CG} where $p = 1$. For any p, the expected utility, E_C, of a conviction is:

$$EU_C = p \cdot U_{CG} + (1 - p) \cdot U_{CI} \qquad (1)$$

Similarly, the expected utility of an acquittal, EU_A, is:

$$EU_A = p \cdot U_{AG} + (1 - p) \cdot U_{AI} \qquad (2)$$

---

[51] Allen et al, above n 9, (22 August 1994) 309 (Friedman R). See eg, John Kaplan, 'Decision Theory and the Factfinding Process' (1968) 20 *Stanford Law Review* 1065, 1066–1083; Lawrence Tribe, 'Trial by Mathematics: Precision and Ritual in the Legal Process' (1971) 84 *Harvard Law Review* 1329, 1378–1386; Richard Lempert, 'Modelling Relevance' (1977) 75 *Michigan Law Review* 1021, 1032–1034; Andrew Ligertwood, 'The uncertainty of proof' (1976) 10 *Melbourne University Law Review* 367, 368–370; Richard Eggleston, *Evidence, Proof and Probability* (2nd ed, 1983) 114–5; Bernard Robertson and G Anthony Vignaux, *Interpreting Evidence: Evaluating Forensic Science in the Courtroom* (1995) 77–79; DeKay, above n 37; Redmayne, above n 2, 167–74.

[52] Nance, above n 9, 622.

[53] Strangely Wagenaar, van Koppen and Crombag, recently criticised the probabilistic theory of proof on the basis that it 'does not explain how … a [decision] criterion is chosen, nor even whether the criterion is constant or variable'. (Above n 18, 30.) Perhaps they were making the point that 'there is nothing in probability theory *per se* which could tell us where to put the critical levels … [This] obviously depends in some way on value judgments as well as on probabilities …': Jaynes, above n 31, 1301.

EC and EA are both linear functions of *p*. They are graphed in Figure 1 for one plausible interpretation of Blackstone's 10:1 ratio. The erroneous verdicts carry a negative utility, or harm, of $U_{AG} = -1$ and $U_{CI} = -10$. Blackstone made no explicit reference to the utility of the correct verdicts, and, for the time being, these can be assigned zero utility: $U_{CG} = 0$; $U_{AI} = 0$. Also graphed is the expected utility of the preferred verdict, which is simply the maximum of EU_C and EU_A. This V-shaped function has a critical point where EU_C and EU_A intersect. This critical point represents the standard of proof, which can be calculated as follows:

$$SP = \frac{U_{AI} - U_{CI}}{U_{CG} - U_{CI} + U_{AI} - U_{AG}} \quad (3)$$

For probability levels below the standard of proof, the preferred verdict is an acquittal. For probability levels above the standard of proof, the preferred verdict is a conviction. Where the probability is precisely equal to the standard of proof, the two verdicts will carry the same expected utility. It is one of the functions of the burden of persuasion to resolve such a tie.

On this interpretation of Blackstone's 10:1 ratio the criminal standard of proof would translate into a probabilistic standard of 91 per cent. However, as noted, there was some uncertainty about Blackstone's meaning as he made no reference to the utility of the correct outcomes.
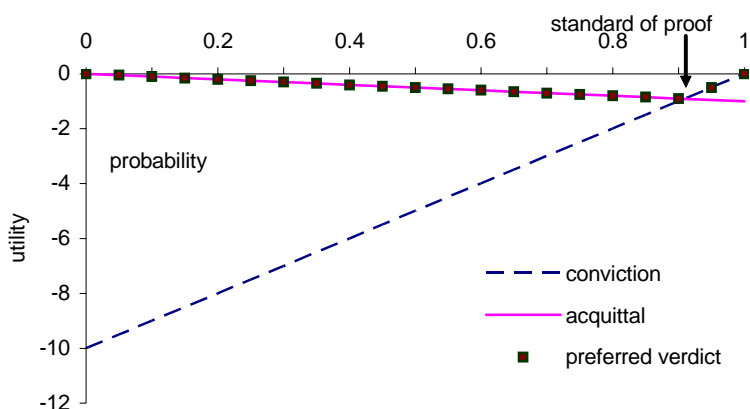
Figure 1 Expected utilities and standards of proof

It is arguable that Blackstone's ratio should be interpreted, not in terms of utilities, but in terms of the closely related notion of

costs.[54] The cost of convicting the innocent is defined as the difference between the utility of acquitting the innocent and the (negative) utility of convicting the innocent.

$$C_{CI} = U_{AI} - U_{CI} \qquad (4)$$

Thus defined, it takes account both of the benefit of the correct outcome and the harm of the erroneous outcome. The cost of acquitting the guilty is similarly defined.

$$C_{AG} = U_{CG} - U_{AG} \qquad (5)$$

Then the standard of proof can then be expressed more simply in terms of the two costs, rather than the four utilities.[55]

$$SP = \frac{C_{CI}}{C_{CI} + C_{AG}} \qquad (6)$$

Where the correct outcomes are assigned zero utility, as was done above for Blackstone's ratio, the disutility of error equals the cost of error, and so applying (6) to Blackstone would still result in a 91 per cent standard.

The other commentators mentioned above also spoke only in terms of the erroneous outcomes, and so they may be best interpreted as laying down cost ratios. Applying (6), the various commentators prescribe the probabilistic standards appearing in Table 1. The greater the cost of an erroneous conviction relative to a mistaken acquittal, the higher the criminal standard of proof. Note also that if the two errors are considered to produce carry an equal cost, as in ordinary civil cases, the V-shaped function is symmetrical about a standard of proof 50 per cent. The expected

---

[54] DeKay, above n 37, 99–100.

[55] As well as involving fewer terms, a further advantage of the cost analysis over the utility analysis is that it does not raise the awkward question of identifying the zero point. Costs are intervals, without direction, whereas utilities have a direction and length, relative to a zero point. In some circumstances the zero point can be taken as measuring the utility of the present situation: an improved situation has positive utility, while a worse situation has negative utility. Alternatively, one may set the best outcome as the zero point, and measure all the worse outcomes relative to that. DeKay suggests that '[f]or the purpose of determining the optimal threshold, it simply does not matter where zero falls'. (Ibid 117.) On this view it would not matter which convention was adopted. Nevertheless, for reasons outlined in the text, I would not advocate the abolition of the four utilities in favour of the two costs.

cost of error will be minimised simply by adopting the more probable version of events, the one that is least likely to be erroneous.

Table 1 Disutility ratios and standards of proof

Commentator	Cost ratio $C_{CI}:C_{AG}$	Standard of proof
Hale (1678)	5:1	0.83
Blackstone (1765)	10:1	0.91
Fortescue (1825)	20:1	0.95
Starkie (1865)	99:1	0.99

While it is simpler to talk in terms of the costs of the two possible errors rather than the utilities of the four possible outcomes, sight should not be lost of the benefits — and harms — of 'correct' outcomes. Laufer claims that '[l]egal scholarship assumes' that acquitting the innocent is the 'best' outcome, better than convicting the guilty.[56] However, this assumption, or perhaps Laufer's claim, is questionable. A conviction of a guilty defendant offers the greatest value in terms of deterrence, retribution and rehabilitation. The resources of the judicial system and any private interests will have been well spent. However, the acquittal of an innocent defendant, although a 'correct' outcome in one sense, signifies an incorrect prosecution. Resources have been expended in vain, and despite the acquittal, irreversible damage may have been caused to the defendant's reputation and relationships.[57] Meanwhile, assuming that the flaw in the prosecution concerned identification rather commission, the guilty party will still be at large. An appreciation of this would drive the standard of proof down, making it easier for the prosecution to obtain a conviction. In setting the standard of proof, attention should be paid to all the possible outcomes of a trial. The utility analysis can be reduced to a simpler cost analysis, but something may be lost in the reduction.

---

[56] Laufer, above n 50, 396 fn 309, citing only Terry Connolly, 'Decision Theory, Reasonable Doubt, and the Utility of Erroneous Acquittals' (1987) 11 *Law and Human Behavior* 101, 101.

[57] Leipold, above n 17, 1298.

IV STANDARDS, ERROR RATES AND PROBABILITY DISTRIBUTIONS

Having regard to the seriousness of a wrongful conviction relative to a mistaken acquittal, the criminal standard is a stringent one. Increasing the standard reduces the risk of a wrongful conviction, but increases the risk of a mistaken acquittal. As noted above, the 50 per cent civil standard minimises the expected overall error rate. That there is a relationship between standards of proof and expected error rates is clear. However, the precise nature of the relationship is not as straightforward as might be supposed.

Ligertwood is not entirely correct in suggesting that '[a] standard of criminal guilt of 0.9 means that of every 100 persons convicted, 10 will be wrongly convicted.'[58] Nor is Glanville Williams correct in stating: 'convicting in a criminal case on a probability of 0.95, high though that may seem at first sight, involves the probability of convicting one innocent person in 20'.[59] First, as noted above, it is only possible to talk of subjective expected error rates rather than actual error rates. And even then, the error rates that Ligertwood and Williams identify are the *maximum* expected error rates for the given standards of proof. Ligertwood and Williams assume that every case will be extremely difficult to decide, with the probability of guilt only just exceeding the standard of proof. Depending upon the degree of certainty obtained in the flow of cases the expected error rate may be much lower. If each case was proved to virtual certainty, the expected number of incorrect convictions would be close to zero. The expected error rate depends not only on the standard of proof, but also upon the probability distribution for the body of cases. The precise mathematical relationship can be described as follows.

Let the probability distribution be represented by the function $n(p)$, where n is the number of cases where guilt has been proved to a subjective probability level p. With standard of proof, SP, we can draw the conclusions appearing in Table 2.

---

[58] Andrew Ligertwood, *Australian Evidence* (3rd ed, 1998) 32.
[59] Glanville Williams, 'The Mathematics of Proof' [1979] *Crim LR* 297, 305–306.

**Table 2 Expected outcomes of a probability distribution of
contested cases**

1. Number of cases ending in an acquittal:

$$NA = \int_{0}^{SP} n(p)\, dp$$

2. Number of cases ending in a conviction:

$$NC = \int_{SP}^{1} n(p)\, dp$$

3. Expected number of mistaken or false acquittals:

$$FA = \int_{0}^{SP} n(p) \cdot p\, dp$$

4. Expected number of mistaken or false convictions:

$$FC = \int_{SP}^{1} n(p) \cdot (1-p)\, dp$$

5. Expected number of correct or true acquittals:

$$TA = NA - FA$$

6. Expected number of correct or true convictions:

$$TC = NC - FC$$

7. Total number of cases:

$$N = NA + NC$$

Note that this analysis assumes that the probability of guilt and
the standard of proof can be identified with exact precision, and
so there is a vanishingly small number of cases where the two are
equal. In reality there will be cases where the fact-finder

experiences a higher-order uncertainty as to whether the probability of guilt lies above or below the standard of proof, such cases being resolved against the party bearing the legal burden of proof.[60]

From these basic quantities, proportions are easily calculated: the prosecution's rate of success in achieving convictions is NC/N; the error rate for convictions is FC/NC; the overall error rate is (FC + FA)/N; and so on. Graphical representations of these quantities and proportions for several different probability distributions appear in the figures below.

The numbers of convictions and acquittals and the error rates are largely determined by the shape of the probability distribution. Of course, this ultimately is an empirical question without clear data. Statistics are available for the number of cases ending in convictions and acquittals,[61] but we do not have any figures indicating the fact-finders' raw probability assessments lying behind their verdicts. Furthermore, '[w]e naturally have no statistics about the number of cases in which the court has believed the wrong man'.[62] And, Frankel's claim notwithstanding, it cannot be said to be a '*statistical* fact ... that the preponderant majority of those brought to trial did substantially what they are charged with'.[63] It is, however, interesting to speculate as to the possible shape of the probability distribution of cases going to trial, and to calculate the expected rates of outcomes, correct and mistaken, that would result from the application of different standards of proof to such distributions.

---

[60] See eg Charles R Williams, above n 2, 169: 'The legal burden is decisive when the evidence is evenly balanced'.

[61] See eg, the figures published by the Australian Bureau of Statistics, the Australian Institute of Criminology, and the Bureau of Justice Statistics (US).

[62] Richard Eggleston, 'What is wrong with the adversarial system?' (1975) 49 *Australian Law Journal* 428, 431.

[63] Frankel, above n 1, 1037 (emphasis added). Wagenaar, van Koppen and Crombag, above n 18, 12, more accurately talk in terms of the 'generally accepted but hardly testable assumption that about 95 per cent of all defendants are actually guilty'.

Perhaps, in the absence of any data to the contrary, we should apply the principle of indifference.[64] If there is no reason to think that any particular probability assessment would be more common than any other, then we should assume a totally flat distribution with equal numbers of cases at each probability level (Figure 2); $n(p)$ equals a constant. The diagonal line represents the function $p.n(p)$ — the expected number of cases at each probability level in which the defendant is actually guilty. The intersection of this function with the standard of proof divides the probability distribution into four regions: true acquittals (TA), false acquittals (FA), true convictions (TC) and false convictions (FC).[65]

Figure 2(a) shows the impact of a Blackstonian standard of proof of 91 per cent on this distribution. Ninety-one per cent of the cases would result in acquittals, and 9 per cent in convictions. Forty-six per cent of the acquittals would be considered mistaken (FA), as would five per cent of the convictions (FC). By way of comparison Figure 2(b) shows the difference that would result from a standard of proof of 50 per cent. Half the cases would result in acquittals and half in convictions, and 25 per cent of each would be expected to be incorrect.

As expected, the result of increasing the standard has been to decrease the rate of mistaken convictions, but at the expense of mistaken acquittals, and overall accuracy. The overall error rate has increased from 25 per cent to 42 per cent.

The flat distribution might be warranted if we were totally ignorant as to its appearance. However, positive arguments can be put for alternative shapes. A realist, for example, might argue that the distribution would have more of a U-shape (Figure 3).
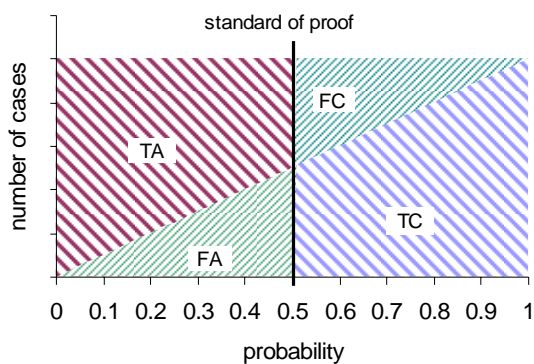
---

[64]   See eg Donald Gillies, *Philosophical Theories of Probability* (2000) 35.

[65]   I am grateful to Geoff Hamer for suggesting the addition of this function to the graph which greatly increases its explanatory power.
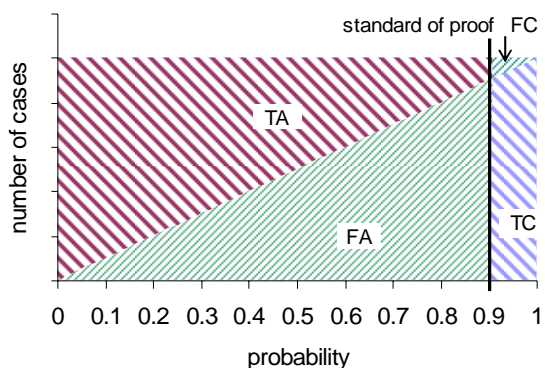
(a)

(b)

Figure 2 Flat distribution

The factual questions that arise out of criminal charges do have definite answers, and, given the competing parties' motivation to gather and present all available evidence favouring either side, the fact-finder should be able to achieve some degree of certainty in the defendant's guilt or innocence in the majority of cases. Conceding that absolute certainty is unachievable, the distribution would tail down sharply at either end. Figure 3(a) shows that with a 91 per cent standard of proof, 88 per cent of cases would end in an acquittal and 12 per cent in a conviction. Forty-four per cent of the acquittals would be expected to be incorrect (FA), as would five per cent of the convictions (FC). The expected overall error rate would be 39 per cent.
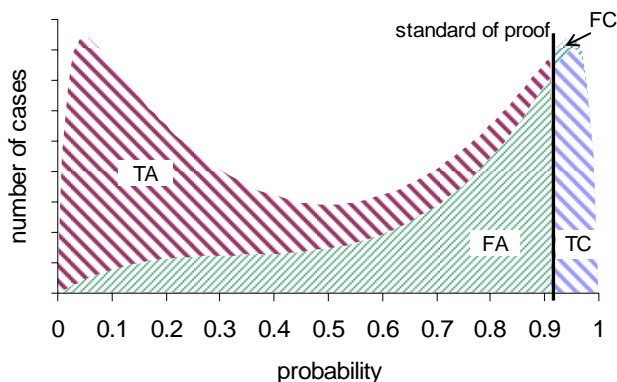
(a)
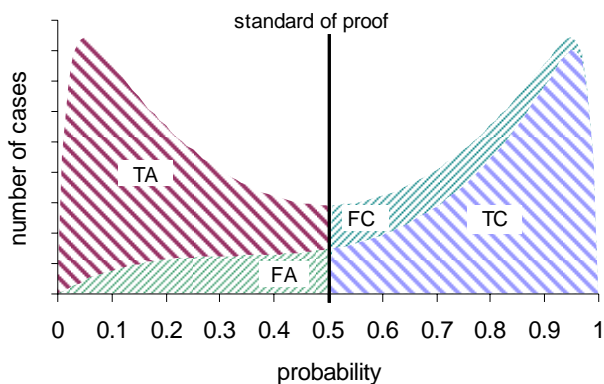
(b)

Figure 3 Realist distribution

As illustrated in Figure 3(b), with a 50 per cent standard, the symmetry of the distribution would result in half acquittals and half convictions. Twenty-one per cent of the acquittals would be expected to be mistaken (FA) and 20 per cent of the convictions (FC). The effect of the increased standard of proof is as previously noted — the expected rate of mistaken convictions is reduced, but with an increase in the expected rate of mistaken acquittals, and an increase in the expected overall error rate. Given the greater certainty in the U-shaped distribution, the expected error rates are lower than those for the flat distribution.

Whatever the merits of its realist foundations in other settings, it is arguable that the U-shaped distribution displays a lack of appreciations of the realities of litigation. The vast bulk of the

cases where the available evidence offers virtual certainty would never get to court. If the probability of guilt clearly exceeded the standard of proof, the defendant would plead guilty to avoid the public ordeal of trial and in the hope of obtaining a lighter penalty. If the probability of guilt fell far short of the standard, the prosecution would not waste resources in pursuing the charges. It is only the 'hard cases', those where the probability of guilt is near the standard with each party seeing some chance of success, that would go to trial to be decided by the application of the standard of proof.

The force of this argument depends, in part, on the extent to which the parties have a grasp on all of the evidence and can predict how it will appear to the fact-finder. It is undermined by 'the unpredictable hazards of the forensic process'.[66] It is also arguable that the parties' determination to litigate would be influenced to some degree by their perception of the underlying reality — the criminal defendant's belief in her innocence, the plaintiff's certainty that the defendant is responsible for her injuries — regardless of the evidence.[67] And if a party thinks she has little or nothing to gain by settling before trial — consider for example a criminal trial in which the prosecution was likely to insist upon the maximum penalty — cases may tend to reach trial despite decisive evidence in the other party's favour. Whereas the hard-cases reasoning suggests that the distribution would be concentrated about the standard of proof, these factors would tend spread at least some the distribution across the width of the unit interval. Applying the hard-cases reasoning to the Blackstonian ratio, we might expect the cases to form a normal distribution centred on 91 per cent, as in Figure 4(a).

---

[66] *Wilsher v Essex Area Health Authority* [1988] AC 1074, 1092 (Lord Bridge); see also Redmayne, above n 2, 182; Barbara D Underwood, 'The Thumb on the Scales of Justice: Burdens of Persuasion in Criminal Cases' (1977) 86 *Yale LJ* 1299, 1335. Nevertheless, the US Supreme Court in *Santosky v Kramer* 455 US 745 (1982) 757, assumed that the strength of the evidence appropriately impacted upon the parties' litigious behaviour, suggesting that 'the standard of proof necessarily must be calibrated in advance'. Compare George M Dery III, 'The Atrophying of the Reasonable Doubt Standard: The United States Supreme Court's Missed Opportunity in *Victor v Nebraska* And Its Implications in the Courtroom' (1995) 99 *Dickinson Law Review* 613, 614; Hay, above n 15, 657.

[67] See also Givelber, above n 35, 1367.

The distribution is symmetrical about the standard of proof, and so the cases would be split evenly between acquittals and convictions. However, the vast bulk of the distribution lies towards the certain guilt end of the unit interval. The result is that 89 per cent of the acquittals would be expected to be mistaken (FA), but only 8 per cent of the convictions (FC). Overall, the expected error rate is 53 per cent. If the standard of proof were decreased, the hard-cases reasoning suggests the distribution would also shift down the probability scale. In Figure 4(b) we see the corresponding normal distribution centred on 0.50. The symmetry of the distribution means that half the cases will result in acquittals and half in convictions, 48 per cent of each will be expected to be erroneous. Once again, the higher standard reduces the expected rate of wrongful convictions, but increases the expected rate of erroneous acquittals, and the overall error rate.
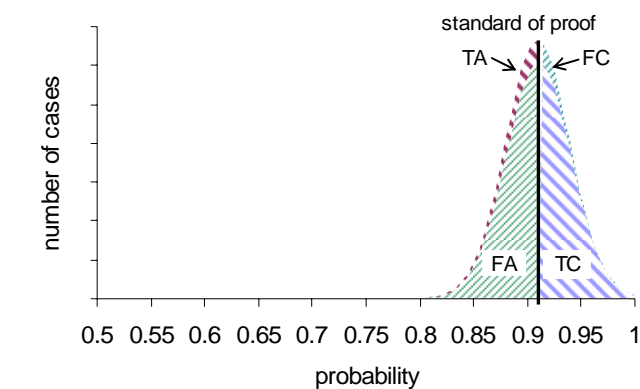
(a)
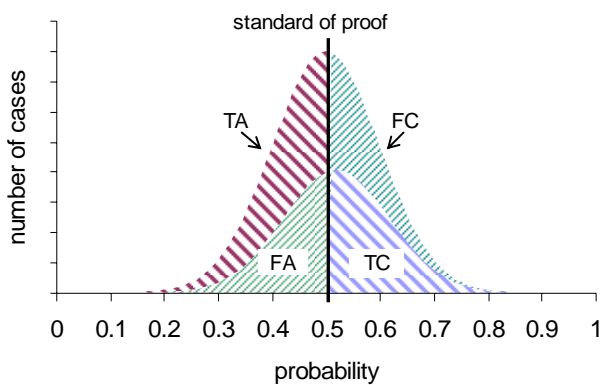
(b)

Figure 4 'Hard cases' distributions

Comparing the three distributions, it is unsurprising that the greatest expected accuracy is offered by the 'realist' U-shaped distribution in which relatively high certainty is attained in most cases. The highest error rates are expected in the 'hard-case' normal distribution which assumes that the realist's easy cases never get to court. Common to all three distributions is the notion that to reduce the risk of one error-type is to increase the risk of the other error-type. The trade-off appears unavoidable.[68]

It appears inconsistent with this latter observation for commentators to advance prescriptions that putatively reduce both error types. Newman argues that the courts' 'unwillingness to apply the "reasonable doubt" standard rigorously has resulted not only in some unjust convictions, it has also precipitated some unwarranted acquittals', conceding that this appears to be a 'paradoxical effect'.[69] Givelber similarly suggests that '[t]o a surprising degree, the risks [of each error-type] are unrelated'.[70] However, the paradoxical nature of these comments can be resolved by recognising that they stem from recommendations for reform that extend beyond the criminal standard of proof. Newman recommends 'broaden[ing] the categories of evidence that juries are permitted to hear … so that the standard might serve as a *more precise divider of the guilty from the innocent*'.[71] Givelber also expresses concern about the exclusion of relevant evidence[72] and other of the 'current processes [that] compromise the criminal justice system's ability to *identify the innocent as well as the guilty*'.[73] In effect they both suggest reforms to the system that will provide fact-finders with a greater quantity of relevant evidence and more coherent guidance in its use, enabling them to get closer to the real facts, and reduce the number of hard cases.[74] It is arguable, however, that the impact of such reforms, if successful, would be in the shadow of the trial court, with an increase in warranted guilty pleas and the dropping

---

[68] See also Underwood, above n 66, 1331.

[69] Jon O Newman, 'Beyond "Reasonable Doubt"' (1993) 68 *New York University Law Review* 979, 1000; see also at 980.

[70] Givelber, above n 35, 1321.

[71] Newman, above n 69, 1000 (emphasis added).

[72] Givelber, above n 35, 1378, 1394.

[73] Ibid 1396 (emphasis added).

[74] Underwood, above n 66, 1333.

of unjustified charges. The staple of the court's reduced diet would be likely to remain one of hard cases.

V THE COMPLEMENTARY STANDARDS OF THE PROSECUTION AND THE CRIMINAL DEFENDANT

In Parts II and III probability theory was employed to explore the rationale behind the contrasting civil and criminal standards of proof, and Part IV examined the expected error rates that the contrasting standards would produce for different probability distributions. This Part concerns another aspect of probabilistic standards of proof. While it is common to refer to the standard of proof that must be met by a single party, usually the plaintiff or prosecution, probability theory implies that a standard of proof is also applicable to the defendant, the two standards being complementary, in combination making up the unit probability interval.

In discussing the ordinary civil standard above, I noted that effectively both parties are presented with the same standard of proof. The plaintiff will seek to show that it is more probable than not that the defendant is liable. Likewise, it is in the defendant's interest to establish that it is more probable than not that she is not liable. Each standard is 0.5, and these are complementary, summing to unity. Less obviously, in criminal trials, there are also two complementary criminal standards, one for the prosecution and one for the defendant.[75]

---

[75] See also John T McNaughton, 'Burden of Production of Evidence: A Function of a Burden of Persuasion' (1955) 68 *Harvard Law Review* 1382, 1382–1383; Robert C Power, 'Reasonable and Other Doubts: The Problem of Jury Instructions' (1999) 67 *Tennessee Law Review* 45, 74. R Allen, 'Presumptions in Civil Actions Reconsidered' (1981) 66 *Iowa Law Review* 843, 327 fn 22, suggests that the defendant's standard is the 'reciprocal' of the prosecution's standard. In its everyday meaning this may appear an appropriate term, but, mathematically, the term is incorrect. The reciprocal of $^1/_2$ is 2; the *product* of a number and its reciprocal is one. 'Complement' is a better term. The complement of $^1/_2$ is $^1/_2$; the *sum* of a number and its complement is one. The two complementary standards, added together, make up the unit probability interval. This is an instance of the addition law, one of the fundamental axioms of probability theory. See Gillies, above n 64, 59.

The two complementary criminal standards are most apparent if expressed probabilistically. As discussed above, the Blackstonian 10:1 ratio suggests that the prosecution needs to prove the defendant's guilt to a probability of greater than 0.91 to obtain a conviction. At the same time the defendant will need to prove that her innocence has a probability greater than 0.09 per cent, to secure an acquittal. The two standards, 0.91 per cent and 0.09 per cent are complementary summing to unity. In general, if the prosecution needs to prove guilt to a standard of S to obtain a conviction, then the defendant the defendant's complementary standard is $1 - S$. (See Figure 5.)
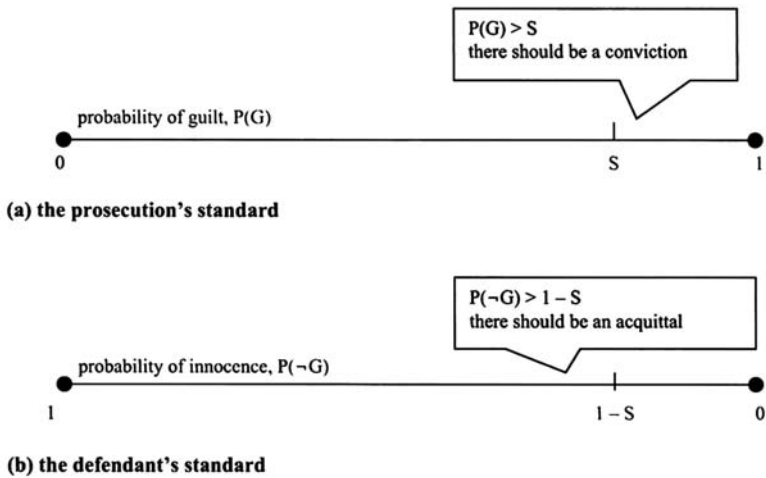

(a) the prosecution's standard

(b) the defendant's standard

Figure 5 Complementary standards

A *Inadvertent Interchange*

The two complementary criminal standards are readily identified in numerical formulations. The two are also discernible in verbal formulations, however, these are less clearly identified and trial judges have, on occasions, got the two confused.[76] The prosecution must establish guilt beyond a reasonable doubt to obtain a conviction. It is in the defendant's interest to ensure there is a reasonable doubt about her guilt in order to secure an acquittal. However, in *Agnew v US*[77] the trial judge instructed the

---

76 See also Power, above n 75, 77–78.
77 165 US 36 (1897).

jury that there is a rebuttable presumption 'that every man intends the legitimate consequences of his own acts'[78] and that the evidence in rebuttal 'must be sufficiently strong to satisfy you beyond a reasonable doubt that there was no such guilty intent'.[79] Surprisingly, the Supreme Court did not take the error seriously, holding that although this 'was open to objection for want of accuracy, we are unable to perceive that this could have tended to prejudice the defendant when the charge is considered as a whole.'[80]

There are a number of alternative definitions of the criminal standard beyond the classic 'beyond reasonable doubt' formulation. These also imply two complementary standards which trial judges, on occasions, have got the wrong way around. In *Dunn v Perrin*[81] the trial judge instructed the jury that a 'reasonable doubt' was 'such a strong and abiding conviction as still remains after careful consideration of all the facts and arguments'. The Court of Appeal for the First Circuit commented that this was 'exact inverse of what it should have been'.[82] The trial judge was actually describing proof *beyond* reasonable doubt, not a reasonable doubt. The defendant's appeal was upheld.

Another popular formulation suggests that proof beyond reasonable doubt is 'proof of such a convincing character that you would be willing to rely and act upon it without hesitation in the most important of your own affairs'.[83] In *Holland v US*[84] the trial judge defined 'reasonable doubt' as 'the kind of doubt … which you folks in the more serious and important affairs of your own lives might be *willing* to act upon'.[85] Again, the trial judge got the two standards mixed up. '[T]he charge should have been in terms of the kind of doubt that would make a person *hesitate*

---

[78]   Ibid 53.

[79]   Ibid 49.

[80]   Ibid 50.

[81]   570 F2d 21 (1st Cir, 1978).

[82]   Ibid 24.

[83]   *US v Williams* 20 F3d 125 (5th Cir, 1994), 129.

[84]   348 US 121 (1954).

[85]   Ibid 140 (emphasis added).

to act'.[86] However, the Supreme Court again brushed the error aside, suggesting that this 'would seem to create confusion rather than misapprehension … [and] that, taken as a whole, the instructions correctly conveyed the concept of reasonable doubt to the jury.' It is regrettable that an interchanging of the prosecution's standard and the defendant's complementary standard was once more taken so lightly. These decisions may need to be reconsidered in the light of the Supreme Court's increased awareness of proof beyond reasonable doubt as a constitutionally mandated standard.[87]

B *Circumstantial Direction*

Another way of expressing the defendant's complement of the beyond-reasonable-doubt standard is as follows: 'if there is any reasonable hypothesis consistent with the [defendant's] innocence'[88] there should be an acquittal. The High Court of Australia has described this complementary proposition as an 'amplification'[89] of the classic beyond-reasonable-doubt formulation, holding that it will not always be necessary for the trial judge include this in her directions to the jury. In *Shepherd* Dawson J suggested that it can be 'helpful' where the

---

[86] Ibid (emphasis added); see also *Holt v US* 218 US 245 (1910), 254 ('hesitate to act') and *Hopt v Utah* 120 US 430 (1887), 439 ('willing to act').

[87] Eg *Re Winship* 397 US 358 (1970). But see George M Dery III, 'The Atrophying of the Reasonable Doubt Standard: The United States Supreme Court's Missed Opportunity in *Victor v Nebraska* And its Implications in the Courtroom' (1995) 99 *Dickinson Law Review* 613, discussing *Victor v Nebraska* 114 S Ct 1239 (1994).

[88] *Peacock v The King* (1911) 13 CLR 619, 630 (Griffith CJ); see also ibid 634; *Hodges Case* (1838) 2 Lewin 228; 168 ER 1136; *Martin v Osborne* (1936) 55 CLR 367, 375, 378 (Dixon J), 381–382 (Evatt J); *Thomas v The Queen* (1960) 102 CLR 584, 605–606; *Plomp v The Queen* (1963) 110 CLR 234, 243 (Dixon CJ), 252 (Menzies J); *Barca v The Queen* (1975) 133 CLR 82, 104 (Gibbs, Stephen and Mason JJ), 109 (Murphy J); *Knight v The Queen* (1992) 175 CLR 495, 502 (Mason CJ, Dawson and Toohey JJ), 511 (Brennan and Gaudron JJ).

[89] *Grant v The Queen* (1975) 11 CLR 503, 505 (Barwick CJ, McTiernan, Mason, Jacobs and Murphy JJ agreeing); *Knight v The Queen* (1992) 175 CLR 495, 502 (Mason CJ, Dawson and Toohey JJ).

prosecution relies heavily on circumstantial evidence,[90] but 'may be confusing rather than helpful' where the prosecution relies primarily on direct evidence.[91]

The rationale of providing the 'reasonable hypothesis' direction in circumstantial cases is that, while circumstantial evidence can be just as strong as direct evidence, it invites a different form of reasoning. 'To say the two types of evidence are equal in weight … is not to say that they are equal in all respects.'[92] Whereas direct evidence 'expressly narrate[s]'[93] the ultimate fact with respect to which it was adduced, circumstantial evidence narrates a fact more or less remote from the ultimate fact. The fact-finder is required to traverse the gap between the narrated fact and the ultimate fact by an interpretative act, and this carries a certain risk. There is greater scope for 'conjecture and speculation'.[94] In the seminal case on the circumstantial evidence direction Baron Alderson warned that'[t]he mind [can be too] apt to take a pleasure in adapting circumstances to one another, and even straining them a little, if need be, to force them to form parts of one connected whole'.[95] More recently, Onion J of the Texas Court of Criminal Appeals noted the view that 'people have a

---

[90] *Shepherd v R* (1990) 170 CLR 573, 578 (Dawson J); see also *Grant v The Queen* (1975) 11 ALR 503, 504 (Barwick CJ); *Peacock v The King* (1912) 13 CLR 619, 630 (Griffith CJ).

[91] *Shepherd v R* (1990) 170 CLR 573, 578 (Dawson J).

[92] *Hankins v State* 646 SW2d 191 (1983), 215 (Onion J, diss), quoting from *State v Lasley* 583 SW2d 511 (1979).

[93] Jeremy Bentham, *Rationale of Judicial Evidence* (Hunt and Clarke 1827), 3:7–8; discussed in Alexander Welsh, *Strong Representations: Narrative and Circumstantial Evidence in England* (1992), 36. See also *Shepherd v R* (1990) 170 CLR 573, 579; William Wills, *Wills on Circumstantial Evidence* (6th ed, 1912) 412.

[94] *Hankins v State* 646 SW 2d 191 (1983), 207, 217 (Onion J, diss).

[95] Baron Alderson in *Hodges Case* (1838) 2 Lewin 228; 168 ER 1136, as discussed in Wills, above n 93, 48. See also *Pfennig v The Queen* (1995) 182 CLR 461, 536 (McHugh J); Nancy Pennington and Reid Hastie, 'Explaining the evidence: Tests of the story model for jury decision-making' (1992) 62 *Journal of Personality and Social Psychology* 189, 252; W Lance Bennet, and Martha Feldman, *Reconstructing Reality in the Courtroom: Justice and Judgment in American Culture* (1981) 66–68; Wagenaar, van Koppen, and Crombag, above n 18, 141–3, 211, 218; Paul Bergman and Alan Moore, 'Mistrial by likelihood ratio: Bayesian analysis meets the F word' (1991) 13 *Cardozo Law Review* 589, 602.

psychological propensity to weave theories from circumstantial evidence and then to defend their theories because of vanity or pride'.[96] For this reason, it is 'necessary to warn the jury against the danger of being misled by a train of circumstantial evidence'.[97] It may be helpful to direct the jury that it is not enough that the defendant's guilt be a 'rational inference … it should be "the only rational inference that the circumstances would enable them to draw"'.[98] 'In the ordered processes of logical thought … regard must be had to every hypothesis which can compete with the hypothesis suggested by experience and knowledge of human affairs.'[99] The story of guilt may be attractive, but the jury must also consider whether stories of innocence are also open. If any such story appears reasonable the defendant should be acquitted.[100]

A number of United States courts have misunderstood the logic of the 'reasonable hypothesis' direction, resulting in a relaxation of the prosecution's standard of proof. The confusion appears to have begun with the Supreme Court in *Holland v US*, although properly understood, the decision is unproblematic.[101] One of the defendants' grounds of appeal was that the trial judge had failed to provide a 'reasonable hypothesis' direction even though the prosecution case was entirely circumstantial. In a unanimous judgment delivered by Clark J the appeal was dismissed. The court held that 'where the jury is properly instructed on the

---

[96] *Hankins v State* 646 SW2d 191 (1983), 204–205.

[97] Baron Alderson in *Hodges Case* (1838) 2 Lewin 228; 168 ER 1136, as discussed in Wills, above n 93, 48.

[98] *Plomp v The Queen* (1963) 110 CLR 234, 252, quoted in *Barca v The Queen* (1975) 133 CLR 82, 104 (Gibbs, Stephen and Mason JJ).

[99] *Martin v Osborne* (1936) 55 CLR 367, 381 (Evatt J).

[100] The potential weaknesses of circumstantial evidence should not diminish the fact that the directness of eye-witness evidence carries its own dangers. (See eg Elizabeth Loftus, *Eyewitness Testimony* (revised ed, 1986); *R v Turnbull* [1977] QB 224; *Davies v The King* (1937) 57 CLR 170, 182; *Hampton v State* 285 NW2d 868 (1979), 872.) Indeed, an accumulation of circumstantial evidence may appear far stronger than a single eyewitness: '[C]ircumstances cannot lie or be mistaken, whereas witnesses can'. (TC Brennan, 'Circumstantial Evidence' (1930) 4 *Australian Law Journal* 106, 106; see also Welsh, above n 93, 16.) The reasoning behind this is that 'a narrative cannot be shaped from circumstantial evidence without considerably more work than goes into most lies'. (Welsh, above n 93, 17.)

[101] 348 US 121 (1954).

standards for reasonable doubt, such an additional instruction on circumstantial is confusing and incorrect'.[102] It appears clear in the context of the case that the court was questioning the necessity for the direction, and not its logic. The court recognised that 'the cogency of [the prosecution's] proof depends upon its effective negation of reasonable explanations by the [defendant] inconsistent with guilt';[103] if the defendants were able to offer a reasonable explanation — a reasonable hypothesis consistent with innocence — they would be entitled to an acquittal.

In *United States v La Rose*[104] the United States Court of Appeals for the Sixth Circuit provided a logically correct statement of the law: 'in a circumstantial evidence case the inferences to be drawn from the evidence must not only be consistent with guilt but inconsistent with every reasonable hypothesis of innocence.'[105] However, in *United States v Scott*[106] the court declined to follow *La Rose* on the ground that it had failed to consider *Holland* and was 'an aberration'.[107] In *United States v Stone*[108] the court held 'that circumstantial evidence alone can sustain a guilty verdict and that to do so, circumstantial evidence need not remove every

---

[102] Ibid 139–140 (emphasis added).

[103] Ibid 135. The court may have had concerns particular to the circumstances of that case. The defendant was convicted on charges of tax evasion, the prosecution pointing to an apparent increase in the defendant's net worth over a given period far exceeding declared income. The taxpayer responded by offering a number of 'leads' – explanations as to how the money was in fact acquired prior to the relevant period. 'Were the Government required to run down all such leads it would face grave investigative difficulties; still its failure to do so might jeopardize the position of the taxpayer.' (Ibid 127; see also at 137–139.) Much of the reasoning in the case sought to resolve the conflict between these competing concerns. A 'reasonable hypothesis' direction may have been considered to tip the balance too far in the defendant's favour. It is interesting to note that in *Jackson v Virginia* 443 US 307 (1979), 318 fn 9, the Supreme Court suggested that the theory 'rejected' in *Holland* was 'that the prosecution was under an affirmative duty to rule out every hypothesis except that of guilt'; the court did not suggest that it was wrong to require the exclusion of every 'reasonable' hypothesis.

[104] 459 F2d 361 (1972).

[105] Ibid 363, quoting from *Fitzpatrick v US* 410 F2d 513(1969), 516.

[106] 578 F2d 1186 (1978).

[107] Ibid 1192.

[108] 748 F2d 361 (1984).

reasonable hypothesis except that of guilt',[109] following *Scott* and expressly overruling *La Rose*.[110] This statement is logically inconsistent with the Constitutional requirement that guilt be proved beyond reasonable doubt.[111] If there is a reasonable hypothesis consistent with innocence, then there should be a reasonable doubt about the defendant's guilt.

The Sixth Circuit is not alone in its unconstitutional confusion. In *Hankins v State*[112] a majority of the Court of Criminal Appeal of Texas held that the 'reasonable hypothesis' principle is in 'conflict'[113] with, and sets a 'more rigorous standard'[114] than the requirement of proof beyond reasonable doubt. To an extent, this view may have been the product of an awkward formulation of the principle that was then conventional in Texas: the circumstantial evidence 'must exclude, *to a moral certainty*, every other *reasonable hypothesis*, except the defendant's guilt; and unless they do so, *beyond a reasonable doubt*, you will find the defendant not guilty.'[115] As already outlined, a 'reasonable hypothesis' of innocence is basis for a reasonable doubt, and the expression 'moral certainty' is another version of the criminal standard.[116] The Texas formulation is therefore, at best, doubly tautological; at worst it sets an exponentially higher standard than the proper criminal standard. The majority of the Court of Criminal Appeal in *Hankins* was right to express disapproval of this instruction,[117] But, as Onions J in dissent recognised, the 'reasonable hypothesis' principle 'is only another application of

---

[109]   Ibid 362.

[110]   Ibid 363.

[111]   *Re Winship* 397 US 358 (1970), 364; see also *Jackson v Virginia* 443 US 307 (1979).

[112]   646 SW2d 191 (1983).

[113]   Ibid 199.

[114]   Ibid 198, quoting from *State v LeClair* 425 A2d 182 (1981).

[115]   *Hankins v State* 646 SW2d 191 (1983), 207 (emphasis added), quoting the State Bar of Texas, *Texas Criminal Pattern Jury Charges*, §0.01 (1975).

[116]   See eg Shapiro, above n 4.

[117]   *Hankins v State* 646 SW2d 191 (1983), 200; see also *Charles v State* 955 SW2d 400 (1997), 405.

the doctrine of reasonable doubt'.[118] The decisions in *Hankins*, *Stone* and *Scott* 'have redefined the reasonable doubt standards in circumstantial evidence cases'.[119]

In fact it is arguable that any weakness with the 'reasonable hypothesis' standard is in favour of the prosecution rather than the defendant. It might lead a juror to think that, to avoid prosecution, it must be possible to mould the evidence into a single concrete story of innocence. An assortment of weaknesses in the prosecution case — for example, evidence only of a weak motive, possible bias in the eyewitness, effective cross-examination of the prosecution's expert forensic witness, equivocal evidence of post-offence conduct — would not suffice; 'a generalized unease or scepticism about the prosecution's evidence is not a valid basis to resist entreaties to vote for conviction'.[120] The 'reasonable hypothesis' principle should not be interpreted as meaning 'a doubt is reasonable only if the juror can articulate to himself or herself some *particular* reason for it.'[121]

C *Concerns Over Burden Reversal*

The present discussion may raise the objection that standards are not imposed on *both* parties. Only one party is required to satisfy the standard of proof, generally the plaintiff or prosecution, although exceptionally the defendant must prove an affirmative defence to a certain standard.[122] In criminal cases in particular, to

---

[118] *Hankins v State* 646 SW2d 191 (1983), 206 (Onion J, diss), quoting from *Hunt v State* 7 Tex Ct App 212 (1879), 237–238. See also *Hankins v State* 646 SW2d 191 (1983), 215–217 (Onion J, diss).

[119] Chambers, above n 17, 688.

[120] Newman, above n 69, 983. Newman was here addressing the definition of a reasonable doubt as 'a doubt based on a reason' (see also Thomas Mulrine, 'Reasonable Doubt: How in the World is it Defined?' (1997) 12 *American University Journal of International Law and Policy* 195, 224). However, the comments have clear application to the 'reasonable hypothesis' principle.

[121] Newman, above n 69, 983 (emphasis added).

[122] Generally the defendant bears at most an evidential burden, requiring her to point to sufficient evidence to raise certain defences. The prosecution will then be required to negate the defence beyond reasonable doubt. The only common law exception to this is the insanity defence. The defendant bears a legal burden, and must prove the facts constituting the defence on the

talk in terms of the defendant's standard of proof appears inconsistent with the presumption of innocence. But this objection is misplaced. It conflates the principles governing standards of proof with those governing the closely related topic of burdens of proof. It could be argued that the present discussion of standards of proof is incomplete due to its failure to address burdens of proof. But such an argument should not be permitted to divert attention from the fact that the probabilistic nature of standards of proof makes the existence of the complementary standards an inescapable fact. The confusions explored above indicate that it is better to acknowledge and understand the defendant's standard than to pretend it does not exist.

Solan suggests that even the classic 'beyond reasonable doubt' formulation may have a burden-shifting effect.[123] He prefers a formulation which would require the jury to be 'firmly convinced' in the defendant's guilt in order to convict. This, he suggests 'focuses more on the government's burden',[124] whereas the 'reasonable doubt' instruction 'focus[es] the jury on the defendant's ability to come up with alternative explanations'.[125] The classic formulation, he suggests, 'encourages jurors to take the government's case as a given and then challenges them to find alternatives'.[126]

But, as Solan acknowledges, even his preferred formulation has its inevitable complement.[127] A model instruction advanced by

---

balance of probabilities. However, the general rule has been greatly weakened by numerous statutory defences which also impose the legal burden on the defendant. Such legislation raises concerns about the erosion of the presumption of innocence; however, the courts' ability to override the legislation on this ground varies from jurisdiction to jurisdiction. See eg *Woolmington v Director of Public Prosecutions* [1935] AC 462; *Patterson v New York* 432 US 197 (1977); *R v Hunt* [1987] 1 All ER 1; *R v Lambert* [2002] AC 545.

[123] Other aspects of the law of criminal proof raise far more genuine concerns about burden reversal. See previous n.

[124] Lawrence M Solan, 'Refocusing the Burden of Proof in Criminal Cases: Some Doubt about Reasonable Doubt' (1999) 78 *Texas Law Review* 105, 107.

[125] Ibid 129.

[126] Ibid 134.

[127] Ibid 144.

the Federal Judicial Center requires the jury to be 'firmly convinced' of the defendant's guilt in order to convict, adding that the jury should acquit if they 'think there is a real possibility that he is not guilty'.[128] One could perceive equal burden-shifting risks with this instruction.[129] The only version of the criminal standard that does not pose burden-shifting risks is the one with no complement — a standard requiring the prosecution to prove guilt to absolute certainty, a probability of one hundred per cent. But where empirical proof is involved some doubt appears inevitable. To require absolute certainty would be to give crime impunity.

As stated at the outset, factual accuracy is the paramount goal of the trial. But it would be a perversion of this goal to seek to project a 'myth of certainty'[130] on the ground that 'some things are best left unsaid',[131] or out of a concern that to acknowledge the risk of error would jeopardise 'the continued confidence in the present system of criminal trial'.[132] The standing of the judicial system would suffer far more from the debunking of the myth when inevitable miscarriages of justice occasionally came to light. The system is likely to maintain far greater health by the acknowledgement, rather than the concealment, of the fundamental fact that in administering criminal justice, 'some risk of convicting the innocent must be run'.[133]

---

[128] Federal Judicial Center, *Pattern Criminal Jury Instructions* (1988) No 21, at 28.

[129] *US v Porter* 821 F 2d 968 (4th Cir, 1987). At 973 the court entertained the defendant's argument that this shifted the burden: 'Implying the evidence must show a real possibility of innocence to justify acquittal trenches on the principle that a defendant is presumed to be innocent.' Ultimately, however, any such risk was overcome in other instructions. Contrast, *US v Conway* 73 F 3d 975 (10th Cir, 1995) 981. See generally Power, above n 75, 84–85.

[130] Tribe, above n 51, 1373.

[131] Kaye, above n 34, 40. See also Tribe, above n 51, 1372; Nesson, above n 5, 1225; Ligertwood, above n 58, 79.

[132] Ligertwood, above n 51, 387.

[133] Glanville Williams, *The Proof of Guilt: A Study of the English Criminal Trial* (1963) 190; compare Daniel N Shaviro, 'Statistical-Probability Evidence and the Appearance of Justice' (1989) 103 *Harvard Law Review* 530, 544.

VI CONCLUSION

This article has examined the role played by standards of proof in the pursuit of factual accuracy at trial. Recognising that absolute certainty is generally unattainable and that some risk of error is unavoidable, standards of proof are set at probability levels that minimise the cost of expected errors. In the ordinary civil trial, an error against either party is considered to carry equal cost. A standard of 50 per cent directs the fact-finder to the verdict that is more probably correct, and an erroneous verdict in each case is not probable. Each party will be concerned to demonstrate that her version of facts is the more probable. In criminal trials a wrongful conviction is considered to carry a far greater cost than a mistaken acquittal. An elevated standard reduces the risk of the more serious error, at the same time inevitably increasing the risk of the lesser mistake. According to the most common formulation, the prosecution must show guilt is beyond reasonable doubt, and the defendant will seek to present a reasonable explanation of the evidence consistent with innocence. It is a corollary of the probabilistic nature of juridical proof that the opposing parties are both presented with complementary standards.

The level at which a standard is set is a key determinant of the type of error that may occur. The other determining factor is the probability of guilt or liability that the fact-finder arrives at in each case. If absolute certainty were attainable, errors would be avoidable regardless of the standard of proof. The closer that a fact-finder is able to get to certainty, the greater the opportunity of avoiding error. In this article I have provided a clear statement of the mathematical relationship between the standard, the probability distribution and error rates. A realist may argue that evidence will generally be forthcoming enabling high levels of certainty and low error rates. However, many of these easy cases will be resolved before reaching court. The trial court is likely to be presented with the most difficult cases with a maximal risk of costly error.