Law, Probability and Risk (2012) **11**, 135–158 doi:10.1093/lpr/mgs004

Probability, anti-resilience, and the weight of expectation

David Hamer†
Faculty of Law, Sydney University, Australia

[Received on 27 September 2011; revised on 13 November 2011; accepted on 19 March 2012]

The probabilistic representation of the proof is often criticized for failing adequately, to reflect the quantity or weight of evidence. There are cases (such as the naked statistical evidence hypotheticals) where a slight body of evidence appears to provide a strong measure of probabilistic support; however, a common intuition is that the evidence would lack sufficient weight to constitute legal proof. Further, if the probability measure has no relation to the weight of evidence, it would appear to express scepticism about the value of evidence. Why bother considering fresh evidence, increasing the weight of evidence, if it provides no epistemic benefit? Probability theory can avoid the spectre of scepticism. Fresh evidence should be considered because it offers the promise of increased certainty. In this article, I prove the relationship between quantity of evidence and these expected increases, and provide a computer model of the relationship. As the weight of evidence increases, greater certainty can be expected. Equivalently, on numerous runs of the model, greater certainty is achieved on average. But in particular cases, certainty may remain the same or decrease. The average and expected increase will be more accentuated where more decisive evidence is available, but the result still holds for situations where the evidence is merely probative. Notwithstanding that certainty can be expected to increase, the probability measure expected from considering fresh evidence is, by definition, equal to the prior probability measure. The model also throws doubt on the notion that an increase in the weight of evidence will bring greater stability or weight in the probability assessment. Indeed, the opposite appears to be the case. The more detailed evidence shifts cases from the general to the particular, and this is reflected in sharper movements in probability assessments.

Keywords: evidence; inference; probability; resilience; standard of proof; weight.

1. Introduction

This article revisits one of the main arguments against the use of mathematical probability theory to understand juridical proof. Probability theory appears to be a rational way of measuring the strength of evidence. To decide contrary to the probabilistic *strength* of evidence appears to be courting factual error. But in another respect, the probabilistic measure seems a poor basis for rational decision-making. It may say something about the evidence to hand, but it says nothing about the completeness or *weight* of that evidence. A slight body of evidence may give rise to a reasonable degree of probabilistic certainty,[1] but what of the fragility of that assessment, given the many items of further evidence that have not yet been considered? Furthermore, if the probabilistic measure has no

† Email: david.hamer@sydney.edu.au

[1] At one point in time the expression ‘probabilistic certainty’ may have appeared an oxymoron, like ‘relatively unique’. Many medieval philosophers viewed probability as the absence of certainty; the two were ‘opposed categories’ like true and false: Lorraine Daston, *Classical Probability in the Enlightenment* (Princeton University Press, Princeton, 1988) 37. However, Bernoulli and his successors put the two on the same scale. ‘[S]eventeenth century mathematicians [came] to view the

connection with the weight or completeness of a body of evidence it would express scepticism about the value of evidence and the feasibility of proof. If fresh evidence promises no probabilistic benefit, why bother considering it? On this view, proof at trial requires a body of evidence that not only delivers a high enough probability assessment, but one that is also sufficiently complete or weighty.

As I explore below, probability theory has answers to these concerns. Probabilistically, it is not possible to take account of unavailable evidence for the simple reason that it is unavailable and its content is unknown. Weight advocates express concern that the missing evidence, should it become available, may contradict the indications from existing evidence. But it appears equally likely that the further evidence would confirm these indications. The two possibilities—contradiction and confirmation—cancel each other out. But this is not to say that, probabilistically, fresh evidence is not worth considering. The probability calculus does provide a reason for increasing the weight of evidence—fresh evidence can be expected to lead to an increase in certainty. In view of this, there is no need for the juridical standard of proof to incorporate a weight requirement.

The relationship between probabilistic strength and the weight of evidence is explored further through use of a computer model. Admittedly, the model is simplistic, however, it still provides some interesting insights. Of particular interest, the model throws doubt on claims that increasing the weight of evidence results in assessments with greater resilience. Some commentators draw an analogy with the narrowing of confidence intervals as sample size increases. However, this analogy is false. As will be shown, there is actually reason to expect the opposite. To increase sample size is an exercise in generalization. The details of specific cases are averaged out and the scope for variation decreases. However, in juridical proof, to increase the weight of evidence is to add detail, and individualize the present case, increasing the scope for variation.

2. Illustrations of the conflict

Before seeking to model the relationship between the probabilistic strength and weight of a body of evidence, a few problem cases will be provided that illustrate the tension. Much of the discussion of weight has focused on hypothetical cases. In the Blue Bus case, for example, the plaintiff was negligently run over by an unidentified bus. The plaintiff sues the Blue Bus Co on the basis they own 80% of the buses in town.[2] It is more probable than not—80% probable—that the bus was blue. Arguably, this satisfies the civil standard of proof requiring a preponderance of probability,[3] or proof on the balance of probabilities.[4] However, most commentators consider that the plaintiff would fail.[5] And many commentators attribute this to the slightness of the plaintiff's evidence. Both in evidence law scholarship and the broader philosophy of proof many hold the view that proof is not purely

relationship of probability to certainty as that of parts to the whole': Ibid, 38. Probability was developed as a method of 'quantifying, or at least ordering, *degrees* of certainty': ibid (emphasis in original).

[2] Laurence H Tribe, 'Trial by Mathematics: Precision and Ritual in the Legal Process' (1971) 84 Harvard Law Review 1329, 1340–41, 1346–50. Another often discussed hypothetical is the Gatecrasher Case: LJ Cohen, *The Probable and the Provable* (Clarendon Press, Oxford 1977) 75.

[3] Eg *Grogan v Garner* 498 US 279 (1991).

[4] Eg *Davies v Taylor* [1974] AC 207, 219 (Lord Simon).

[5] But see James Brook 'The Use of Statistical Evidence of Identification in Civil Litigation: Well-worn Hypotheticals, Real Cases, and Controversy' (1985) St Louis University Law Journal 293, 299.

probabilistic, but is dependent upon the 'weight',[6] 'completeness',[7] or 'comprehensiveness'[8] of the evidence.

Hypothetical cases carry the risk of being unrealistic, generating questionable intuitive responses. Why is there such little evidence in the Blue Bus Case? Is the plaintiff withholding evidence? Perhaps the absence of evidence is itself evidence against the plaintiff, supporting a spoliation inference, in which case a finding for the defendant may be reconciled with the probabilistic measure.[9] On the other hand, much of the further evidence that might be expected—bus routes, drivers, damaged buses—would appear to be in the defendant's domain.[10] Why should an adverse inference be raised against the plaintiff rather than the defendant? The Blue Bus Case appears to display the conflict between probabilistic strength and weight, but the potential presence of a spoliation inference confuses matters. Moreover, given their artificiality, hypotheticals like the Blue Bus Case may say little about the practical significance of the conflict.

But the conflict is not purely hypothetical. There are real cases where slight evidence appears to generate high probabilities and in circumstances where there is no scope for a spoliation inference. The toxic tort is one such class of case, illustrated by the recent decision of the UK Supreme Court in *Sienkiewicz v Greif (UK) Ltd*.[11] At issue was the causation of mesothelioma. The only known cause of mesothelioma is exposure to asbestos fibres, but was it the defendant's negligent exposure, or some other exposure? Given the limited knowledge as to the aetiology of the disease and the long latency period, direct evidence of causation is unavailable. The court has little to go on other than epidemiological evidence of the relative risk of disease from the different exposures. This gives rise to the question whether the civil standard of proof translates to a statistical requirement that the negligent exposure more than doubles the risk from the non-negligent exposures.[12] In *Sienkiewicz*, however, the court expressed considerable resistance to exclusively statistical proof, and to a large extent, this appeared to reflect concerns about its slight weight.[13]

A number of the judgements drew a sharp distinction between statistical proof and proof in the individual case.[14] While this distinction has a moral flavour in the value it places on the individual,[15] it also has an epistemic aspect. The instant case is given greater individuality by increasing the weight of

[6] JM Keynes, *A Treatise on Probability* (MacMillan and Co, London 1921) 71; LJ Cohen, '12 Questions about Keynes's Concept of Weight' (1985) 37 British Journal of Philosophy of Science 263; Barbara Davidson and Robert Pargetter, 'Weight' (1986) 49 Philosophical Studies 219; Dale Nance, 'The Weights of Evidence' [2008] *Episteme* 267, 268; Alex Stein, *Foundations of Evidence Law* (OUP, Oxford 2005) 47–48.

[7] Richard Friedman, 'Assessing Evidence' (1996) 94 Michigan Law Review 1810, 1819; Nance (n 6) 268.

[8] Hock Lai Ho, *A Philosophy of Evidence Law – Justice in the Search for Truth* (OUP, Oxford 2008) 166.

[9] Eg David Kaye, 'The Paradox of the Gatecrasher and Other Stories' [1979] Ariz State LJ 101.

[10] David Hamer, 'The Civil Standard of Proof Uncertainty: Probability, Belief and Justice' (1994) 16 Sydney Law Review 506, 516.

[11] [2011] UKSC 10 (9 March 2011).

[12] As the Court noted, this question has been considered elsewhere, eg *Sienkiewicz* (n 11) [85] (Lord Phillips), discussing *Merrell Dow Pharmaceuticals Inc v Havner* 953 SW 2d 706 (1997).

[13] In addition to this theoretical concern, the court raised practical concerns about the significance and reliability of the statistics, the availability and reliability of evidence of the plaintiff's situation, and whether the statistics cover the plaintiff's situation: eg [98] (Lord Phillips).

[14] *Sienkiewicz* (n 11) [154], [163] (Lord Rodger); [170] (Lady Hale); [190] (Lord Mance); [205] (Lord Kerr).

[15] Eg JJ Thomson, 'Liability and Individualized Evidence' (1986) 49 Law and Contemporary Problems 199; D T Wasserman, 'The Morality of Statistical Proof and the Risk of Mistaken Liability' (1991) 13 Cardozo Law Review 935; Amit Pundik, 'Statistical Evidence and Individual Litigants: A Reconsideration of Wasserman's Argument from Autonomy' (2008) International Journal of Evidence & Proof 303.

evidence relating to the case, distinguishing it from other similar cases. Lord Rodger indicated that 'since, by its very nature, the statistical evidence does not deal with the individual case, *something more will be required* before the court will be able to reach a conclusion, on the balance of probability, as to what happened in that case.'[16] The Court discussed a variation on the Blue Bus Case involving cabs,[17] and the further evidence that might be expected—eyewitness evidence of a blue cab in the vicinity, physical evidence of damage to a blue cab, the relative accident rates of the two cab firms.[18] In the absence of such evidence, a finding against the Blue Cab Co. on the basis that it has more cabs appears unwarranted. And in *Sienkiewicz*, a conclusion on the causation of the plaintiff's mesothelioma requires more than just epidemiological evidence.[19]

Concern has been expressed about the low weight of evidence in another class of case, far removed from road accidents and toxic torts—sexual assault cases where the complainant has delayed complaint for many years. Because of the delay, the evidence available to the court may be very slight. In the intervening years, records will have been lost, memories will have faded, and the opportunity for medical or forensic examination will have passed, leaving the court with little more than the complainant's allegations, and the defendant's denials. Often, as with *Sienkiewicz*, the lack of evidence is understandable—it is now recognized that there are many reasons for child victims of sexual assault delaying their complaint, such as confusion, fear and humiliation—and this will generally not raise an adverse inference.[20] While the extant evidence is not statistical, it is of low weight, and raises similar issues to the Blue Bus Case, and toxic tort cases.

It will often be difficult for the complainant to win the battle of credibility decisively enough to satisfy the criminal standard of proof,[21] however, the lack of weight of the prosecution case adds a further obstacle. In the leading Australian decision, *Longman v R*,[22] Deane J observed that 'the evidence of the complainant reads convincingly. It is not surprising that the jury accepted her as an honest witness. The same could not be said of the evidence of the [defendant]. . . It is not surprising that the jury plainly rejected the [defendant] as a witness . . .'[23] But can a conviction rest upon such a slight weight of evidence? '[T]he fact remains that the only evidence of the applicant's guilt . . . was the oral evidence of the complainant.'[24] There is a marked concern about a prosecution proceeding on the basis of such slight evidence. Depending upon the jurisdiction, the prosecution may be blocked by a statute

[16] Above (n 11) [163] (emphasis added).

[17] Above (n 11) [95] (Lord Phillips); [171] (Lady Hale); [217] (Lord Dyson); drawing on earlier discussions in *Herskovits v Group Health Cooperative of Puget Sound* 664 P 2d 474 (1983) 491 (Brachtenbach J) and *Hotson v East Berkshire Area Health Authority* [1987] AC 750, 789 (Lord Mackay).

[18] Above (n 11) [95]–[96] (Lord Phillips).

[19] Ultimately *Sienkiewicz* was decided in favour of the plaintiffs by the application of a principle from *Fairchild v Glenhaven Funeral Services Ltd* [2003] 1 AC 32 which, in mesothelioma cases, deems contribution to risk to be causal contribution, even where the risk is insufficient to establish causation on the balance of probabilities. The *Sienkiewicz* court, however, also expressed misgivings about this departure from orthodoxy: Above (n 11) [186] (Lord Brown); see also at [167] (Lady Hale).

[20] Historically, the complainant's delay did call into question his or her credibility: Eg *R v DD* [2000] 2 SCR 275 [60]; *R v Johnston* (1998) 45 NSWLR 362, 367. The credibility issue can be clearly separated from the weight/forensic disadvantage issue: David Hamer, 'Trying Delays: Forensic Disadvantage in Child Sexual Assault Trials' [2010] Criminal Law Review 671, 672.

[21] David Hamer, 'Delayed Complaint, Lost Evidence and Fair Trial: Epistemic and Non-epistemic Concerns' in Paul Roberts and Jill Hunter (eds), *Criminal Evidence and Human Rights*, Ch. 9, p. 215 (Hart, Oxford, 2012).

[22] (1989) 168 CLR 79.

[23] Ibid 98.

[24] Ibid 99.

of limitations,[25] a prohibition order,[26] or as an abuse of process.[27] More frequently in Australia and England, the prosecution will be allowed to proceed, but the jury is directed to have regard to the fact that much evidence is missing. In Australia, at common law, the jury is told 'that, as the evidence of the complainant could not be adequately tested after the passage of [so many] years, it would be dangerous to convict on that evidence alone unless the jury, scrutinising the evidence with great care, considering the circumstances relevant to its evaluation and paying heed to the warning, were satisfied of its truth and accuracy.'[28] This has been described as 'a not too subtle encouragement by the trial judge to acquit'.[29]

3. Bringing missing evidence to account?

The concern about the lack of weight of evidence is often expressed in terms of the impact the missing evidence would have if it became available. Richard Friedman suggested some time ago that 'incompleteness of evidence with respect to a proposition in itself tends to prevent an observer from assigning a high probability to the proposition because the observer should consider the possibility that the missing evidence would, if produced, tend to negate the proposition'.[30] Other commentators have more recently expressed concern about the 'unrealized potential of the missing evidence to produce a different factual conclusion'.[31] In the delayed complaint cases courts appear to assume that the missing evidence would contradict the complainant. The loss of evidence therefore presents the 'twin possibilities both of serious prejudice to the defence, and positive benefit to the prosecution'.[32] Delay creates 'significant disadvantage',[33] 'great disadvantage,'[34] and 'serious disadvantages'[35] for the defendant.

But this pro-defendant orthodoxy is puzzling. As pointed out by Wood CJ at CL, 'the impact of the delay is double edged, since it is just as likely to occasion practical difficulty for the prosecution'.[36] The supposed missing alibi evidence may in fact be missing opportunity evidence. Had forensic or medical examinations been carried out at the time, they may have produced inculpatory not exculpatory results.[37] More generally, if the evidence is missing it cannot be known which way it points. It appears equally possible that the missing evidence would confirm the current factual conclusion as contradict it. The competing possibilities cancel each other out. There is no warrant for the assumption that the missing evidence will point one way rather than the other.

[25] In many US jurisdictions: eg Brian L Porto, 'New Hampshire's New Statute of Limitations for Child Sexual Assault: Is It Constitutional and Is It Good Public Policy?' (1991) 26 New England Law Review 141.

[26] In Ireland: *PL v DPP* [2004] 4 IR 494.

[27] Eg *Joynson* [2008] EWCA Crim 3049; *R v Littler* [2001] NSWCCA 173.

[28] *Longman* (1989) 168 CLR 79, 91 (Brennan, Dawson and Toohey JJ). For the recommended English direction see: Judicial Studies Board, *Crown Court Bench Book: Directing the Jury* (March 2010), 34.

[29] *R v BWT* (2002) 54 NSWLR 241 [35] (Wood CJ at CL); see also at [118] (Sully J). The direction has recently been toned down by legislation (eg *Evidence Act 1995* (NSW) s 165B(4)), but remains strongly pro-defendant.

[30] Richard Friedman 'Assessing Evidence' (1996) 94 Michigan Law Review 1810, 1819, fn 23.

[31] Stein (n 6) 120; see also at 97–100, 228; Ho (n 8) 278.

[32] *Turner* (unreported, EWCA Crim Div, 27 March 2000) [19].

[33] *Crampton v R* (2000) 206 CLR 161 [45] (Gaudron, Gummow and Callinan JJ); [140] (Hayne J).

[34] Ibid [131] (Kirby J).

[35] Ibid [132] (Kirby J).

[36] *BWT* (n 29) [23] (Wood CJ at CL); see also *R v Inston* (2009) 103 SASR 265 [112] (Vanstone J).

[37] Wood CJ at CL recognised that his views were unorthodox: *BWT* (n 29) [13]; see also *Percival* EWCA Crim Div, 97/6746/X4, June 19, 1998. Other possible explanations such as the presumption of innocence and the defendant's right to present a defence are examined and found wanting in Hamer (n 20) and Hamer (n 21).

The probability calculus captures this logic. The content of the missing evidence is not known. It could point one way or the other. Therefore the probability assessment expected from the missing evidence is an average of these possibilities. If this calculation is made it is found that the expected probability is precisely equal to the current probability. This 'expected probability' result is proven in Appendix A.

4. Bias and scepticism

The discussion in the previous sections illustrates some of the situations in which the weight notion plays a role, and also highlights its potential conflict with the probabilistic measure of evidential strength. From the probabilistic perspective, the weight notion appears illogical. Where evidence is of low weight courts appear reluctant to make a positive finding, even if the evidence is sufficient probabilistically. But to refuse to make a positive finding on the basis that evidence lacks weight is, in effect, to make a finding against the party bearing the burden of proof, in most cases the plaintiff or prosecution. Similarly, to discount a probability assessment due to a low weight of evidence will often result in a finding against the plaintiff or prosecution. These responses to the weight concern give too much force to the burden of proof. A decision contrary to the probabilities is a decision that is likely to be wrong. This can be expected to increase error costs, with a bias favouring the defendant.[38] Deserving plaintiffs will be denied compensation. Guilty defendants will go free. The weight requirement appears to undermine factual accuracy, which is widely regarded as the 'paramount'[39] goal of juridical proof.

However, the probabilistic critique as characterized above is potentially problematic. The probabilistic perspective is that, once standards of proof are established,[40] a determination should be made on the basis of the probabilistic strength of the evidence, regardless of the fact that the evidence lacks weight. But does this suggest that, from the probabilistic perspective, the weight of evidence is epistemically unimportant? This would express an unwelcome scepticism about the enterprise of fact-finding. If a larger weight of evidence is not epistemically preferable to a lower weight of evidence, why bother increasing the weight of the evidence?[41] Why bother considering any evidence?

To avoid the charge of scepticism, the probabilistic measure must, in some way, be positively correlated with the weight of evidence. Without this, the probabilistic measure would fail to provide any motivation for a fact-finder to consider fresh evidence. And this would throw doubt upon the

[38] Eg J Kaplan, 'Decision Theory and the Factfinding Process' (1968) 20 Stanford Law Review 1065; David Hamer 'Probabilistic Standards of Proof, Their Complements, and the Errors That are Expected to Flow from Them' (2004) 1 University of New England Law Journal 71. See also Nance (n 6) 274–277.

[39] Marvin E Frankel, 'The Search for Truth: An Umpireal View' (1975) 123 University of Pennsylvania Law Review 1031, 1033, 1055. As I have previously noted (Hamer (n 38) 71 fn 1) the goal of factual accuracy has also been described as 'foremost' (Adrian S Zuckerman, *Principles of Criminal Evidence* (OUP, Oxford 1989) 7), 'fundamental' (Vern R Walker, 'Preponderance, Probability and Warranted Factfinding' (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).

[40] Standards of proof may be set at levels of probability to minimise the expected error costs: Kaplan (n 38); Hamer (n 38). Of course, there may be a range of objections to this explicitly probabilistic approach to proof. I put these to one side in order to focus on the weight issue.

[41] Eg IJ Good, *Good Thinking: The Foundations of Probability and Its Applications* (University of Minnesota Press, Minneapolis, 1983) 178.

credentials of the probabilistic measure. The weight-strength correlation is clearly not simple and direct. As the Blue Bus Case, toxic tort and delayed complaint cases illustrate, it is possible for a low weight of evidence to generate a high level of probabilistic strength. Conversely, it is easy to imagine large bodies of evidence that fail to provide probabilistic certainty. But this does not rule out the possibility of a less immediate correlation.

5. The pursuit of objective certainty

Broadly speaking, there are two different responses to the sceptical challenge weight poses for the probabilistic perspective. The motivation for considering fresh evidence may be either internal to the probability calculus or external to it. Externally, it might be argued, for example, that our experience of the world is that as we gather further evidence we generally get closer to the truth, and reduce the risk of factual error. While this is not a necessary consequence of the probability calculus, it is an experience that the probability measure reflects. We may not feel confident of immediately gaining certainty. It is not like the question of whether a card dealt face down is the Jack of Hearts (which can be answered simply by turning the card). But generally, as the weight of evidence increases, our probabilistic certainty increases. 'The truth will out.'

It is an empirical question whether, generally, as we accumulate a greater weight of evidence we experience greater certainty. It is a perception that evidence commentators sometimes make reference to. Ron Allen and Michael Pardo, for example, refer to the 'strongly held belief that personalizing data – evidence that includes observational data about the particular individual– reduces ambiguity [and] the probability of an error'.[42] Alex Stein suggests that the effect of increasing weight is to 'exclude[] from consideration middle-range probabilities, far removed from both certainty . . . and impossibility';[43] '[t]he estimate would come close to certainty'.[44]

This attitude to proof may motivate a fact-finder to consider further evidence. Furthermore, the degree of motivation will depend upon the degree of prior uncertainty. According to ET Jaynes, where a probability is close to one (or zero), one is entitled to 'conclude that [one's hypothesis] is very likely to be true (false) and act accordingly'.[45] However, where the probability is close the midpoint of the unit interval, this should be taken as a 'warning . . . that the available evidence is not sufficient to justify any very confident conclusion, and we need to get more and better evidence'.[46] The road to certainty is paved with evidence.

Of course, those advancing this relationship between the weight of evidence and degree of certainty do not suggest it is simple, direct or necessary. Allen and Pardo note, 'it is theoretically possible to have more personalizing and better data and to simultaneously increase the probability of an error – but generally things do not seem to work that way'.[47] Stein describes the postulate of 'a linear progression relationship between the amount of information that fact-finders have and the accuracy of their decision' as 'unwarranted', 'fallacious', and 'plainly wrong'.[48] Indeed, it is difficult to entirely ward off the

[42] Ronald J Allen and Michael S Pardo, 'The Problematic Value of Mathematical Models of Evidence' (2007) 36 Journal of Legal Studies 107, 134.

[43] Stein (n 6) 90.

[44] Ibid 89.

[45] ET Jaynes, *Probability Theory: The Logic of Science* (Fragmentary Edition of March 1996) 404. See http://thiqaruni.org/mathpdf9/%2886%29.pdf.

[46] Ibid.

[47] Allen and Pardo (n 42), 134.

[48] Stein (n 6) 122–3.

spectre of scepticism. Jaynes notes, ‘uncovering new facts, can lead us to feel either more certain or less certain about our conclusions, depending on what we have learned. New facts may support our previous conclusions, or may refute them’.[49] Stein suggests that ‘new information . . . merely substitute[s] the risk of error that existed before with a new risk of error [with] no guarantee that the new risk of error will be smaller than the old one’.[50] Pardo suggests that there is no ‘guarantee that smaller or more detailed classes [roughly, increasing the weight of evidence] will take one closer to the truth’.[51]

Legal authorities highlight a potential restriction operating on the pursuit of probabilistic certainty. It may be limited to past events whereas courts, particularly in torts cases, are often interested in future and hypothetical events.[52] In *Sienkiewicz*, for example, the causation issue raised the question whether the plaintiff *would have* developed mesothelioma even without the defendant’s negligent exposure, and in many personal injury cases issues will arise regarding the plaintiff’s prognosis, into the *future*. In *Gregg v Scott* Lord Nicholls noted that ‘[t]he theory underpinning the . . . proof of past facts appears to be that a past fact either happened or it did not and the law should proceed on the same footing’,[53] adding that ‘the underlying certainty, that a past fact happened or it did not, is absent from hypothetical facts’.[54] In the Australian High Court decision, *Malec v JC Hutton Pty Ltd*, Deane, Gaudron, and McHugh JJ suggested that where future and hypothetical events are concerned, ‘proof is necessarily unattainable’.[55] Of course, even where past events are concerned, certainty may appear elusive. In *Gregg* Lord Nicholls acknowledged that ‘[w]hether an event occurred in the past can be every bit as uncertain as whether an event is likely to occur in the future’.[56] But he added that ‘by and large this established distinction works well enough. It has a comfortable simplicity which accords with everyday experience of the difference between knowing what happened in the past and forecasting what may happen in the future.’[57]

A belief that future or hypothetical events are inherently uncertain may lead to a sceptical and defeatist attitude to proof. If there can be no expectation that fresh evidence would bring greater certainty, what motivation is there for considering it? One response is to reject this damaging belief. Lord Hoffmann in *Gregg* said that ‘[t]here is no inherent uncertainty about what caused something to happen in the past or about whether something which happened in the past will cause something to happen in the future. Everything is determined by causality. What we lack is knowledge’.[58] On this view, fresh evidence would be welcomed even in respect of future and hypothetical events in the hope of reducing uncertainty. Indeed, even if determinism appears unjustified by our experience of the world—and in many contexts, as Lord Nicholls observes, the future does appear far less knowable than the past—the deterministic attitude may have the methodological benefit of displacing the sceptical

[49] Ibid 1021.

[50] Ibid 123; see also Jaynes (n 45) 1021.

[51] Michael S Pardo, ‘Reference Classes and Legal Evidence’ (2007) 11 International Journal of Evidence & Proof 255, 255.

[52] David Hamer, ‘“Chance Would Be a Fine Thing”; Proof of Causation and Quantum in an Unpredictable World’ (1999) 23 Melbourne University Law Review 557.

[53] [2005] 2 AC 176 [14].

[54] Ibid.

[55] (1990) 169 CLR 638, 642–3 (Deane, Gaudron and McHugh JJ).

[56] Above (n 53) [13]

[57] Ibid [13].

[58] Ibid [79]. See also Richard Wright, ‘Causation, Responsibility, Probability, Naked Statistics and Proof: Pruning the Bramble Bush by Clarifying the Concepts’, (1988) 73 Iowa Law Review 1001, 1041; Jaynes (n 45), 116.

and defeatist attitude.[59] However, as Lord Hoffmann notes, the law is unable to adopt universal determinism. 'One striking exception to the assumption that everything is determined by impersonal laws of causality is the actions of human beings. The law treats human beings as having free will and the ability to choose between different courses of action'.[60] And this is a fairly big exception given that human beings play a considerable role in the events that give rise to juridical investigation.[61] And so the question remains, what motivates the consideration of further evidence if pursuing certainty is recognized as futile?

6. Resilience

A number of commentators have suggested another benefit of increasing the weight of evidence, one which is consistent with the probabilistic measure, but does not demand an assumption of determinism. As the weight of evidence increases, it is suggested that the probability assessment increases in 'resilience'[62] or 'stability'[63] and is 'less likely to be shaken by potential additions to its information base'.[64] This could be an empirical observation of what happens to probability assessments as the weight of evidence increases. But the claim is sometimes supported by reference to the probabilistic calculus, in particular the statistical concept of confidence intervals.[65] If, for example, there is a mixture of red and white balls in an urn and we are interested in the probability of drawing a red ball, we can estimate this by repeatedly drawing balls[66] from the urn and noting the frequency of red balls. As the number of draws increases, we can place greater confidence that the true answer lies somewhere close to the estimate, and this will be reflected in the narrowing of the confidence intervals. With 5 red balls in 10 draws, the 95% confidence interval is 0.5 ± 0.32. For 300 red balls in 600 draws it is 0.5 ± 0.04. Both sets of data give rise to a probability assessment of 50% that the next ball will be red, but the assessment based on the larger data set appears to provide a firmer assessment. This expectation of increased stability flows directly from the mathematics of statistics and probability. Confidence interval width is calculated from the standard deviation which measures the variation of sample data from a mean point.[67] As sample size increases, the standard deviation decreases, and the estimate based on the sample will tend to become more resilient.

[59] J Dupré, *The Disorder of Things: Metaphysical Foundations of the Disunity of Science* (Harvard University Press, Cambridge, Mass 1995) 184–5.

[60] Above (n 53) [82].

[61] The exception grows still larger considering the potential for humans to interact with, and introduce uncertainty into, what would otherwise be deterministic segments of the world: Dupré (n 59) 190.

[62] Stein (n 6) 48; James Logue, 'Resiliency, Robustness and Rationality of Probability Judgments' (1997) 11 International Studies in the Philosophy of Science 21.

[63] Ho (n 8) 278.

[64] Stein (n 6) 88. See also Neil Cohen, 'Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge' (1985) 60 New York University L Rev 385. See also Barbara Davidson and Robert Pargetter, 'Weight' (1986) 49 Philosophical Studies 219, who define 'weight' in terms of the probability assessment's resistance to further change rather than by reference to the quantity of evidence that has been considered.

[65] Cohen (n 64); Logue (n 62) 91; Stein (n 6) 48 fn 51.

[66] With replacement, unless the number is very large.

[67] The 95% confidence interval is most commonly used. Briefly, this can be roughly calculated as $p \pm 2\sigma$. The symbol σ represents the standard deviation, which can be estimated by $\sqrt{(pq/n)}$, where n is the number of samples, p is the frequency of 'successes' (eg red balls) and q the frequency of 'failures' (eg non-red balls): W Mendenhall, *Introduction to Probability and Statistics* (Duxbury Press, Mass 1979) 239–40. This formula is based upon the mathematics of the normal function which can be taken as an approximation of the binomial probability distribution. The normal approximation is better for larger sample sizes, however, Mendenhall suggests that where the frequency is centred, as it is in the present case with $p = 0.5$, the normal approximation of a binomial distribution of only 10 samples is still 'reasonably good': at 203.

But the ball and urn experiment and the construction of confidence intervals provides a poor analogy for juridical proof.[68] First, there will often be a major difference in what we know about the correct answer. The object in the urn and ball example is to obtain an accurate estimate of the proportion of balls in the urn that are red. This is a figure somewhere between 0 and 100%, apparently in the region of 50%. However, most often a trial is concerned with a past event which, as Lord Nicholls points out, either happened or did not happen. Objectively, the probability that the defendant sexually assaulted the victim, or that the defendant's bus hit the plaintiff, is either zero or one. As discussed in the previous section, objective certainty can motivate the consideration of fresh evidence in the pursuit of certainty. It follows that, in respect of past events, any probability figure that is removed from either end point of the unit interval cannot be considered resilient because there is clearly scope for it to change. For a determinist, this observation would apply equally to future events that are bound by the laws of causality.

It might be thought, though, that the confidence interval notion has application to non-determined future or hypothetical events for which, on one view, certainty is unattainable. But the analogy remains weak. The kinds of predictions that courts engage with are markedly different from urn and ball experiments. Confidence intervals are constructed through the statistical calculus of repeated independent trials. These calculations have no application to the diverse body of evidence that courts rely on at trial, and there is no reason to think that a fact-finder's probability assessment will become increasingly stable. Suppose in a wrongful death case the court was seeking to determine whether, without the defendant's negligence, the deceased would have continued working until 65 years.[69] Unlike the repeated draws of a ball from the urn, there appears no basis to expect the evidence or the probability assessment it generates to conform to a regular pattern. The deceased smoked a packet of cigarettes a day, suggesting low life expectancy. But he exercised regularly, suggesting a high life expectancy. But he worked as an industrial chemist with the risk of exposure to highly toxic substances. But his parents are still living and in good health and his grandparents all lived to 90 years. But he raced motorbikes. But his diet was extremely good. And so on. Unlike the ball and urn experiment, where the urn *does* contain a particular proportion of red balls, in this longevity hypothetical there is no apparent reason to expect the probability assessment to converge towards a particular point on the unit interval as the weight of evidence increases.

7. The probabilistic motivation for increasing weight

If the strength of evidence is measured probabilistically, why bother gathering fresh evidence? The answer provided by some philosophers of probability is that fresh evidence pays in expectation.[70] A fact-finder should examine fresh evidence and bring it to account because this can be *expected* to result in an increase in certainty and a decrease in the risk of factual error.[71] The expectation result is

[68] See also David Kaye, 'Apples and Oranges: Confidence Coefficients and the Burden of Persuasion' (1987) 73 Cornell Law Review 54; Hamer (n 10) 518–20.

[69] See also LJ Cohen, *An Introduction to the Philosophy of Induction and Probability* (Clarendon Press, Oxford 1989) 99–108.

[70] Good (n 41) 178; see also B Skyrms, *Choice and Chance* (Belmont: Wadsworth, 4th ed, 2000) 155.

[71] The relationship between probabilistic certainty and the risk of error is mediated by the standard of proof. If the standard of proof is 50%, the relationship is direct—the apparent risk of error is the complement of the degree of certainty. With a higher standard of proof, the relationship is less straightforward. Eg, in a criminal case, a fact-finder's certainty may increase from 50 to 90% while still not eliminating a reasonable doubt. Certainty has increased, but this has only brought an increase in the apparent risk of a mistaken acquittal. But while the asymmetric standard of proof is a complicating factor, it remains true that, at the limits, the risk of error is reduced by increasing certainty. Note also that fresh evidence offers the expectation of making a decision with greater utility: see (n 73).

inherent in the probability calculus. This expectation of certainty is quite different from the external one, discussed above, based on one's experience of proof and one's belief in the objective certainty of past events.

As discussed above, prior to its examination, the content of fresh evidence cannot be known. In a criminal case, for example, a forensic examination of a sample from a crime scene may favour the prosecution or the defendant, increasing or lowering the probability of guilt. The *expected* impact of the evidence is an average of these possible values, weighted by the probability that the evidence will take either form. As suggested above,[72] the expected probability assessment flowing from the consideration of a piece of fresh evidence is equal to the current probability assessment. Not yet knowing what form the fresh evidence will take, the various possibilities appear to cancel out. But, despite that, probabilistic certainty can be expected to increase. In this context, 'certainty' is defined as the probability of a proposition or its negation, whichever is greater. If the fact-finder assesses the probability of the defendant's guilt as 0.1, then the fact-finder is 90% certain with regard to the issue of the defendant's guilt. While the *expected probability* of guilt from fresh evidence can be expected to equal present probability, *expected certainty* is greater than or equal to present certainty.[73] This does not mean that certainty cannot decrease, but in the long run it increases more than it decreases, and so on average and in expectation it does not decrease. This result is stated with greater precision in Box 1 and is proved in Appendix A.

8. Modelling the relationship between weight and probability

The Blue Bus Case raises concerns about the probability measure's supposed failure to take account of the low weight of so-called 'naked statistical evidence'. Courts have expressed similar concerns in actual low weight cases arising from toxic torts and delayed complaints. From a probabilistic point of view, to decide a case contrary to the probabilities is to increase the expected cost of error. However, a failure to respect the weight of evidence would express scepticism about the value of evidence and the entire fact-finding enterprise. The preceding parts have considered various views as to the probabilistic value of increasing the weight of evidence.

In this section I present a formal model of the relationship between weight and probability. This model does not incorporate an assumption of objective certainty or that 'the truth will out'. As will be seen, the probability does not show any clear trend towards either end of the unit interval as the weight of evidence increases. And yet, in line with the expected certainty result, on average and in expectation, certainty does increase as the weight of evidence increases.

The model also provides support for the argument above regarding the inapplicability of the resilience notion to heterogeneous bodies of evidence. The model, in fact, displays a marked anti-resilience effect—as weight increases, the probability becomes increasingly unstable. As will be discussed, to some extent this results from the particular aspects of the model—aspects that may appear unrealistic. Nevertheless, while the anti-resilience may be amplified in the model, the model still gives reason to believe that this is a feature of real-world proof.

[72] At the end of Section 4, and see proof in Appendix A.

[73] Good's analysis (n 41) focused on the expected utility of a decision based on a probability assessment. Expected certainty behaves in essentially the same way as expected utility—both are V-shaped linear functions of probability. Certainty has the advantage of being a purely epistemic measure. '[W]eight . . . is undoubtedly an epistemic concept, and attempts to account for it or measure it in terms of non-epistemic criteria must be suspect': James Logue, *Projective Probability* (OUP, Oxford 1995) 84.

Box 1 *The expected certainty result*

Definitions:

(1) Certainty is defined as the maximum of the probability of a proposition and its negation.

$$\mathrm{EC(G)} = \max(P(\mathrm{G}), P(\mathrm{NG}))$$

(2) The expected value of a measure from considering fresh evidence is the average of the different possible values that the measure may take depending upon the content of the fresh evidence, weighted by the probability that the fresh evidence will have that content. So if fresh evidence, E, may take only two forms, E′ and E″, certainty expected from considering E is:

$$\mathrm{EC(G|E)} = P(\mathrm{E}') \cdot \mathrm{C(G|E')} + P(\mathrm{E}'') \cdot \mathrm{C(G|E'')}$$

The result:
The certainty expected from considering fresh evidence is greater than or equal to current certainty:

$$\mathrm{EC(G|E)} \geq \mathrm{C(G)}$$

A proof appears in Appendix A.

Note:
Here and elsewhere I make the simplifying assumption that fresh evidence is limited to only two possible forms. This is unrealistic, but it simplifies the proof, and the operation of the model below. Furthermore, the two-form case generalizes to the multi-form case through combination/iteration. For example, suppose a piece of fresh evidence may take four possible forms, $E_1, \ldots E_4$. A correspondence can be established between these and combinations of evidence A and B, each of which may take only two possible forms:

$$E_1 \leftrightarrow A'B' \qquad E_2 \leftrightarrow A'B'' \qquad E_3 \leftrightarrow A''B' \qquad E_4 \leftrightarrow A''B''$$

And then, since $\mathrm{EC(G|A)} \geq \mathrm{C(G)}$, $\mathrm{EC(G|A'B)} \geq \mathrm{C(G|A')}$ and $\mathrm{EC(G|A''B)} \geq \mathrm{C(G|A'')}$:

$$\mathrm{EC(G|AB)} \geq \mathrm{C(G)}$$

Which is to say:

$$\mathrm{EC(G|E)} \geq \mathrm{C(G)}$$

where E has four possible forms.

The model[74] operates by reference to a notional frequency table with each cell containing a randomly generated number between 0 and 100. The table may, for example, represent the numbers of individuals that are guilty and not guilty, G and NG, given a number of pieces of evidence, A, B, etc, with particular content, A’ or A”, B’ or B”, etc.[75] From this table the prior probability of G, P(G), can

[74] This model was implemented on an Excel spreadsheet with programming in Visual Basic. My thanks to Joshua Klose for the programming.

[75] Because the numbers in the table were generated randomly, the ’ and ” marks do not signify whether the evidence is favourable or unfavourable to the conclusion, G.

Box 2 *Actual and Expected Certainty for Three Pieces of Evidence*

			G	NG
A’	B’	C’	78	19
		C”	92	38
	B”	C’	91	1
		C”	67	2
A”	B’	C’	77	40
		C”	5	4
	B”	C’	44	41
		C”	45	65

Prior certainty, C(G) = 0.70

Actual certainty on first run through evidence (shaded) C(G|A’B”C”) = 0.97

Actual certainty on second run, C(G|A”B”C’) = 0.52

Expected certainty given three pieces of evidence, EC(G|ABC)=0.73

be calculated, and the prior certainty regarding G, C(G). The table can also be used to calculate probability and certainty as each piece of evidence, in turn, is examined, and brought to account, for example *P*(G|A’), *P*(G|A” B’). Movements in probability and certainty for a series of different runs through the evidence can be generated and compared. And the expected certainty—which is also the average certainty on all possible runs through the evidence—can also be calculated from the table.

Box 2 shows a notional frequency table for three items of evidence. The prior probability is the ratio of the sum of the figures in the first column to the sum of all figures in the table. The prior probability and prior certainty are both 0.7. A couple of runs through the evidence lead to different results. On the first run, which is shaded, the three items of evidence took the forms A’, B” and C” respectively, and the probability and certainty rose to 0.97. On the second run, A”B”C’, certainty decreased to 0.52. But on average, and in expectation, certainty rises to 0.73.

The notional frequency table for three pieces of evidence contains sixteen cells. In general, for n pieces of evidence, there are 2^{n+1} cells. Where a larger body of evidence is being modelled the table becomes unwieldy and graphs give a clearer picture. Figures 1–3 relate to a model involving 15 items of evidence. (The frequency table, which contains more than 65 000 cells, is not reproduced here.) Ten runs were taken through the evidence with the form taken by each piece of evidence randomly determined. Figure 1 illustrates the way that probability varied as the weight of evidence increased, and Figure 2 illustrates the variation in certainty, with the smoother darker line indicating how certainty was expected to vary as increasing evidence is considered.

These graphs conform to the expectation results discussed in the previous section. Probability is expected to remain at an identical level. As Fig. 1 illustrates, the consideration of evidence may increase the probability or decrease it, depending on the form the evidence happens to take, whether it supports the proposition or the negation. Of course, prior to considering the evidence it is not known which way the evidence will go. These possibilities cancel each other out. On average and in expectation probability remains steady.

However, Fig. 2 illustrates that certainty can be expected to increase. A comparison between the two graphs illustrates why this occurs. Whereas probability covers the complete unit interval, certainty

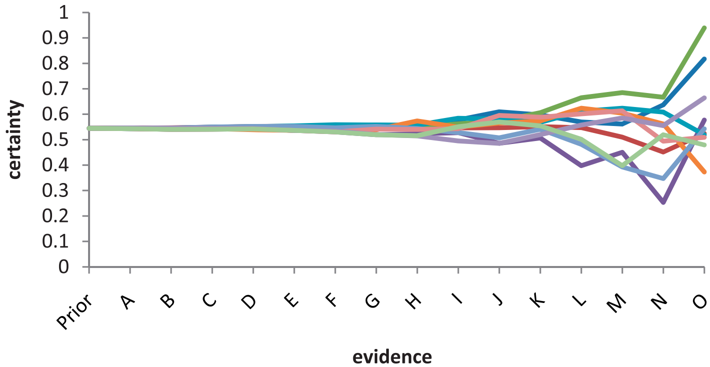

FIG. 1. Probability, 15 items of evidence, 10 runs.

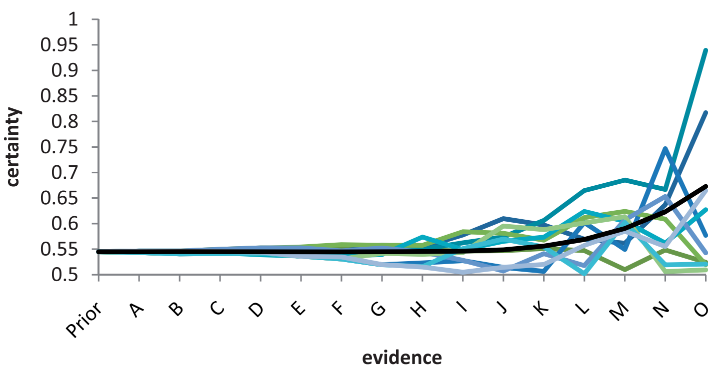

FIG. 2. Certainty, 15 items of evidence, 10 runs.

only covers the region from 0.5 to 1. An *increase in probability* above 0.5 brings an *increase in certainty*. But *certainty also increases* as *probability drops* below 0.5. It is this difference that explains why certainty increases in expectation while probability remains constant. Where a piece of evidence offers the possibility of a probability assessment crossing the 0.5 level, expected certainty will be greater than expected probability, and thus greater than prior certainty. If a piece of evidence does not present this possibility, expected certainty, like expected probability, will remain at the prior value. (This point emerges more clearly in the proof of the expected certainty result in Appendix A.)

Certainty can be expected to increase, but this is not to say that it necessarily will increase. The unexpected can happen. In fact, of these 10 runs through this body of evidence, certainty unexpectedly decreased on five of the runs, and on a further three of the runs the increase was less than expected. Overall, this body of evidence on these 10 runs provided less of an increase in certainty than expected. Certainty was expected to increase from 0.54 to 0.67, but on average over the ten runs, it only increased to 0.62. (There are a total of $2^{15} = 32{,}768$ different possible runs through this evidence.)

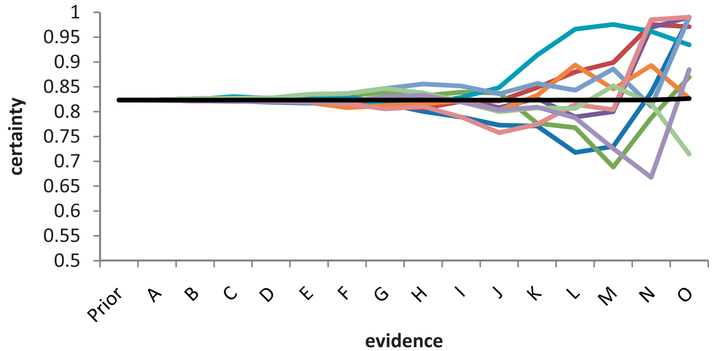

FIG. 3. Certainty, 15 items of evidence, 10 runs, high prior.

9. The prior's role

It should not be surprising that the body of evidence discussed in the previous section offered an increase in expected certainty. The starting point was at almost maximum uncertainty—a prior of 0.546. This leaves considerable scope for an increase in certainty; from 0.5 any change in probability will produce increased certainty. With a higher prior level of certainty, there is less scope for fresh evidence to increase certainty, and greater scope for it to decrease certainty.

The model demonstrates this. While the expected certainty result still holds—expected certainty is greater than or equal to prior certainty—the expected increase is deferred and diminished. Figure 3 is generated from a body of evidence which produces prior certainty of 0.823.[76] As shown, expected certainty (the smoothest darkest line) remains equal to prior certainty until the final two pieces of evidence, and scarcely increases then. Recall that, for expected certainty to increase, evidence must offer scope for the probability assessment to cross the 0.5 level. Otherwise expected certainty behaves the same as expected probability, and does not increase above the prior level. On none of the 10 runs through the evidence did the probability level cross 0.5. Surprisingly, perhaps, out of 10 runs, by the 15th piece of evidence, certainty increased in all but one. But until the final few pieces of evidence, certainty decreased in at least half the runs through the evidence.

If one starts with an extremely high level of certainty, then, obviously, there is less scope for certainty to increase.[77] The motivation provided by the expectation of an increase will be reduced. From one point of view, this makes good sense. If you are already virtually certain, then why bother seeking out fresh evidence? As Jaynes suggests, a low level of certainty prompts the search for further evidence; a high level of certainty may indicate a decision can be made without further evidence.[78] And yet, as hypotheticals like the Blue Bus Case show, if there are obvious categories of missing

[76] Cell entries in the original table were randomly selected numbers between 0 and 100. Because each number has an equal chance of being selected, as the size of the table increases, the prior probability (the ratio of the sum of half the table divided by the sum of the whole table) tends towards 1/2 as table size increases. (This is related to the anti-resilience effect discussed in Section 10.) The cell entries in this notional frequency table were again randomly generated but, to generate a higher prior probability, the G entries were skewed up and the NG entries skewed down.

[77] At least this is true on the probabilistic unit interval. The odds and log-odds measures are a different matter. There is insufficient space here to explore the implications of the different measures.

[78] Above (nn 45–46).

evidence, one's existing level of high certainty may appear fragile. In the Blue Bus Case, objections to making a decision on the basis of existing evidence are couched in terms of possible further evidence—damage to a particular bus, an admission from a particular bus driver—that could point directly at the involvement of a red bus. This possibility—which would involve the probability crossing the 50% threshold—means that such evidence offers an expectation of increased certainty.

The presumption of innocence provides a further example of a high prior level of certainty. While the presumption has a normative dimension,[79] it also has a factual operation and basis. The fact-finder should commence with a very low prior probability of guilt and a high certainty in innocence. This appears factually appropriate. At the outset of the trial there is no inculpatory evidence before the court.[80] It is as though the defendant has been selected at random. And the prior probability that a randomly selected individual committed a particular offence must be at a very low level.[81] The court begins with a high level of certainty that the defendant is innocent. And yet, in reality, given that this probability is based upon no specific evidence at all, the court would be highly motivated to consider fresh evidence. It would not be inconsistent with the presumption of innocence to anticipate that the prosecution may be able to present evidence providing significant support for a finding of guilt. Again, where evidence has the potential to carry the probability across the 50% line, it will offer the expectation of greater certainty.

At this point, a distinction should be drawn between the expected certainty result and the frequency table model. The expectation result is a mathematical truth of probability theory. The model, however, is a particular version of how proof may operate. The model provides a way of exploring certain features of proof, but the extent to which the model's behaviour reflects that of proof in the real world is up for debate. It may be that more can be expected of evidence in the real world than in the model. Each piece of evidence in the model only takes two possible forms, and with a high prior it may be that neither form will offer the possibility of the probability crossing the 50% level. In the real world, however, a piece of evidence may offer far broader range possibilities, one or more of which will almost invariably take the probability across the 50% level even with a high prior level of certainty. Moreover, even if a piece evidence does not offer this possibility by itself, a party's body of evidence may have this cumulative potential. In the real world, the expected certainty result may apply more strongly, and be less inhibited by the existence of a high prior.

10. The opposite of resilience

A noticeable feature of the model, highlighted by the graphs, is that the probability level is initially highly resistant to change as new evidence is brought to account. The early pieces of evidence bring very little change in probability or certainty. But scope for change grows as the final pieces of evidence are considered. At first glance, this behaviour may seem surprising. The increasing *instability* is quite the opposite of that described by advocates of resilience.[82] They suggest that the first few pieces of

[79] Eg, Hock Lai Ho, 'The Presumption of Innocence as a Human Right', in Roberts and Hunter, Ch 11, p. 259 (2012) (n 21).

[80] Significance should not be attached to the fact that the police have laid charges and the prosecution service has brought the matter to trial. Whatever incriminating evidence influenced them, if admissible, can be presented to the court in due course: *Bell v Wolfish* 441 US 520 (1979), 533.

[81] Just as crime rates are relatively low in probabilistic terms. Eg among serious crimes, sexual assault has the one of the highest crime rates. The victimization rate for 2010 for Australia was 79.5 per 100 000 (as against 1 in 100 000 for murder): Australian Bureau of Statistics, *4510.0 Recorded Crime - Victims* (Commonwealth of Australia, Canberra 2001) 9. This relatively high crime rate is less than 1/10 of 1%.

[82] Above (nn 62–64).

evidence may bring quite dramatic changes in probability, but as the weight of evidence increases, the stability of the probability assessment also increases. Is anti-resilience something that we might expect to find in the real world or is it just an artefact of the model?

If we take a closer look at the model the source of anti-resilience can be identified. Interestingly, this can be explained in terms of the statistics of sample sizes. But in this context, unlike the ball and urn experiment, the statistics suggest that increasing the weight of evidence increases rather than decreases the standard deviation and confidence-interval width. Consider how the probability assessments for the model are calculated as each piece of evidence is considered. The prior probability of guilt, P(G), is the sum of the entries in the first column of the frequency table—the G column—divided by the sum of all the entries in the entire table—both the G column and the NG column. As each piece of evidence is considered, half of the frequency table is eliminated. If A is examined and has the form A', those cells relating to A", the bottom half of the frequency table, can be disregarded (see Box 2 or Box 3). If B is then examined, and takes the form B', the B" entries can be disregarded, and the frequency table is halved again.

Now we can understand why the probability assessment may fluctuate more dramatically as the weight of evidence increases. Each probability calculation can be considered an estimate of the probability calculation for an infinitely large frequency table. As more evidence is brought to account, the probability calculation will be based on a smaller part of the table. The estimate is based on a smaller sample and the standard deviation for the probability calculation will increase.[83] This means there is greater scope for variation in the probability assessment as the weight of the evidence increases.

Actually, the relationship between the number of items of evidence and the scope for variation in the probability assessment is not quite this straightforward. While the potential for variation does increase with the weight of evidence, it has a more direct connection with the number of evidential items left to be considered. It is as one approaches the end of the evidence series that the probability calculation involves fewer entries, creating the greatest potential variation. With the final piece of evidence, the probability calculation involves a single row of the frequency table. It is a single entry divided by the sum of two entries. With the second last piece of evidence, the probability calculation is the sum of two entries divided by the sum of four entries. In general, with m pieces of evidence remaining, the probability calculation is the sum of 2^m entries divided by the sum of 2^{m+1} entries.

This identification of the source of variation suggests that the anti-resilience phenomenon is, to some extent, an artefact of the model which may not be reflected real world proof. According to the model, with many items of evidence to consider, early probability calculations will be based on the sum of many entries. Individual differences in entries will tend to average out, and there will be little variation. This early stability is contrary to the experience of real-world proof. It results from the randomly generated entries in the frequency table that do not capture the impact of crucial lines of evidential and inferential argument. In reality, early pieces of evidence clearly can have a significant impact on a probability assessment.

[83] The standard deviation provides a measure of the variation of sample data around a mean. The probability calculation for the frequency table is not a simple calculation of a mean, but instead involves the ratio of means. The probability—the sum of the first column divided by the sum of both columns—is equal to the mean of the first column divided by two times the mean of both columns. The standard deviation of functions of means such as this, like the standard deviation of a simple mean, is inversely proportional to the square root of the sample size. See eg, US Commerce Department, *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/ [2.5.5.2].

Box 3 *Increasing weight, increasing detail, and homing in*

						G	NG
A’	B’	C’	D’	E’	F’	53	60
					F”	2	45
				E”	F’	92	37
					F”	80	86
			D”	E’	F’	58	24
					F”	100	58
				E”	F’	45	11
					F”	60	28
		C”	D’	E’	F’	82	10
					F”	0	67
				E”	F’	74	54
					F”	17	69
			D”	E’	F’	90	79
					F”	61	26
				E”	F’	86	27
					F”	89	53
	B”	C’	D’	E’	F’	45	78
					F”	7	8
				E”	F’	19	96
					F”	77	80
			D”	E’	F’	59	1
					F”	64	81
				E”	F’	59	16
					F”	74	64
		C”	D’	E’	F’	60	73
					F”	70	83
				E”	F’	62	68
					F”	97	26
			D”	E’	F’	42	42
					F”	97	67
				E”	F’	42	23
					F”	64	23
A”	B’	C’	D’	E’	F’	20	6
					F”	82	55
				E”	F’	100	50
					F”	88	3
			D”	E’	F’	83	43
					F”	37	85
				E”	F’	96	29
					F”	89	30
		C”	D’	E’	F’	14	9
					F”	53	35
				E”	F’	59	28
					F”	100	2
			D”	E’	F’	94	41
					F”	3	31
				E”	F’	64	36
					F”	63	5
	B”	C’	D’	E’	F’	94	41
					F”	56	43
				E”	F’	85	19
					F”	97	50
			D”	E’	F’	72	9
					F”	28	60
				E”	F’	40	74
					F”	59	83
		C”	D’	E’	F’	85	96
					F”	38	28
				E”	F’	34	61
					F”	89	15
			D”	E’	F’	75	36
					F”	85	20
				E”	F’	13	89
					F”	33	26

P(G) = 0.59

P(G|A”) = 0.62

P(G|A”B’)= 0.68

P(G|A”B’C”) = 0.70

P(G|A”B’C”D’) = 0.75

P(G|A”B’C”D’E”) = 0.84

P(G|A”B’C”D’E”F”) = 0.98

Box 3 *Continued*

						G	NG
A′	B′′	C′′	D″	E′	F′	42	42
					F″	97	67
				E″	F′	42	23
					F″	64	23
A″	B′	C′	D′	E′	F′	20	6
					F″	82	55
				E″	F′	100	50
					F″	88	3
			D″	E′	F′	83	43
					F″	37	85
				E″	F′	96	29
					F″	89	30
		C″	D′	E′	F′	14	9
					F″	53	35
				E″	F′	59	28
					F″	100	2
			D″	E′	F′	94	41
					F″	3	31
				E″	F′	64	36
					F″	63	5
	B″	C′	D′	E′	F′	94	41
					F″	56	43
				E″	F′	85	19
					F″	97	50
			D″	E′	F′	72	9
					F″	28	60
				E″	F′	40	74
					F″	59	83
		C″	D′	E′	F′	85	96
					F″	38	28
				E″	F′	34	61
					F″	89	15
			D″	E′	F′	75	36
					F″	85	20
				E″	F′	13	89
					F″	33	26

P(G) = 0.59

P(G|A″) = 0.62

P(G|A″B′)= 0.68

P(G|A″B′C″) = 0.70

P(G|A″B′C″D′) = 0.75

P(G|A″B′C″D′E″) = 0.84

P(G|A″B′C″D′E″F″) = 0.98

Note: Many of the A′ entries have been omitted from the table for reasons of space.

Moreover, according to the model, anti-resilience increases, not with weight, but with proximity to the end of the evidence series. If one compares the certainty fluctuation in a series of 15 pieces of evidence with the certainty fluctuation in a series of five pieces of evidence, the latter would exhibit the same fluctuation as the last five pieces in the former (see Fig. 4). But in the real world of proof, the evidence series may be potentially endless. There may always be the theoretical potential for further items of evidence to be brought to account. The conditions bringing a dramatic increase in probabilistic variation in the model may never be present in the real world.

But despite the artificiality of the model, its portrayal of the formal process of bringing further items of evidence to account can be related to real world proof. The model offers a realistic representation of a fact-finder building up detail and homing in on an event of interest (see Box 3). As each piece of evidence is considered, it further partitions the universe of possibilities, placing the individual case into a narrower class with fewer members. This is the opposite of increasing the number of repeated independent trials. The direction of movement in the urn and ball experiment is from the specific to the general as the event of interest is compared with a broader range of other events. But in juridical proof, as the diverse body of evidence grows, an increasingly detailed picture of the case is built. The

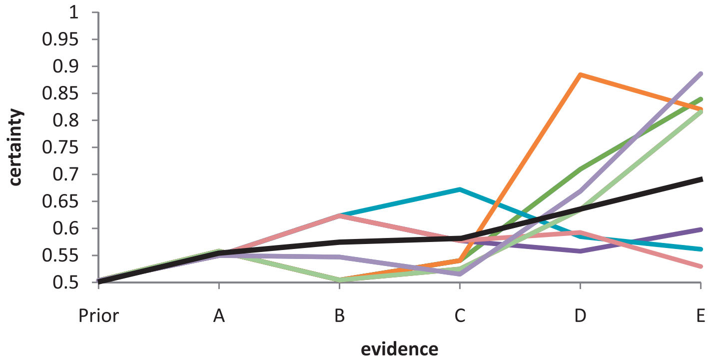

FIG. 4. Certainty, 5 items of evidence, 10 runs.

present case is distinguished further from other cases. The movement is from the general to the specific.

11. Conclusion

Hypothetical cases like the Blue Bus Case, and real cases—toxic tort compensation claims and delayed sexual assault prosecutions—reveal that fact-finders are reluctant to decide cases on the basis of a probability assessment stemming from a slight body of evidence. Some commentators argue that the standard of proof tests not only the probabilistic strength of evidence but also its completeness or weight. However, from the probabilistic perspective, introducing a weight requirement makes no sense. It would give the burden of proof too great a force. It would needlessly increase the cost of expected error, with a bias against plaintiffs and the prosecution.

This article lends support to the probabilistic position. The standard of proof should not incorporate a weight requirement. Missing evidence may influence the fact-finder's probability assessment through a spoliation inference. But if evidence has been lost without spoliation, as may be the case, for example, in toxic tort and delayed complaint cases, the fact-finder should decide the case on the basis of the probabilistic strength of the remaining evidence. The case of the prosecution or plaintiff should not be discounted on account of the slight weight of evidence.

But in rejecting a weight requirement in juridical proof, I have acknowledged that weight is epistemically crucial. To deny this would be to express scepticism about the entire fact-finding enterprise. Probability theory must provide an explanation for why fresh evidence should be considered. And it does. Fresh evidence should be considered because it offers the expectation of increased certainty. This expectation is inherent in the calculus of probability.

This article has explained and modelled the expected certainty result. An increase in certainty can be expected from the consideration of fresh evidence. But the relationship between weight and certainty is not simple and direct. On average, certainty will increase as the volume of evidence increases, but this is not guaranteed. Fresh evidence may bring increased doubt. This seems to be in line with our experience of proof in the real world. It is hard to totally dismiss a sceptical view towards the value of evidence.

This article has rejected an alternative motivation for increasing the weight of evidence—the notion that it will bring probability assessments with greater resilience. In a ball and urn experiment, a statistical probability assessment based upon an increasing sample size may become more resilient as individual variation gets averaged out. But this situation provides no analogy for a growing body of heterogeneous evidence in juridical proof. As the model demonstrates, each piece of evidence adds detail, homing in on the instant case, and increasing individualization. The reference class containing the instant case is partitioned and narrowed, increasing rather than decreasing the scope for variation.

Appendix A: The expected impact of fresh evidence

Expected probability

The impact of fresh evidence on a probability assessment will depend upon the form the evidence takes. In expectation, these different possibilities are balanced. The potential increase in probability cancels out the potential decrease. This can be shown as follows.

Assume that the missing evidence can only take two forms, pro-prosecution or pro-defendant. (This generalizes to multi-form evidence through iteration.[84]) If the missing evidence takes a pro-prosecution form, E^+, it will lift the probability of guilt from a prior probability of $P(G)$ to a posterior probability of $P(G|E^+)$. Similarly, if the missing evidence takes a pro-defendant form, E^-, the probability of guilt would be lowered to $P(G|E^-)$. It is not known which form the evidence takes, so the expected posterior probability of guilt from the missing evidence, EP(G|E), is the average of these possible posterior probabilities, weighted by the probability of the missing evidence taking either form, $P(E^+)$ and $P(E^-)$.

$$\begin{aligned} \mathrm{EP(G|E)} &= P\left(\mathrm{E}^+\right)\cdot P\left(\mathrm{G|E}^+\right) + P(\mathrm{E}^-)\cdot P(\mathrm{G|E}^-) \\ &= P\left(\mathrm{G}\,\&\,\mathrm{E}^+\right) + P(\mathrm{G}\,\&\,\mathrm{E}^-) && \text{(by the product rule)} \\ &= P(\mathrm{G})\cdot\left(P\left(\mathrm{E}^+|\mathrm{G}\right) + P(\mathrm{E}^-|\mathrm{G})\right) && \text{(by the product rule)} \\ &= P(\mathrm{G}) && \text{(by the complementary negation principle)} \end{aligned}$$

So the expected probability of guilt from the missing evidence, the content of which is unknown, is equal to the prior probability of guilt.

$$\mathrm{EP(G|E)} = P\left(\mathrm{E}^+\right)\cdot P\left(\mathrm{G|E}^+\right) + P(\mathrm{E}^-)\cdot P(\mathrm{G|E}^-) = P(\mathrm{G}) \tag{1}$$

Expected certainty

Whereas the probability assessment expected from fresh evidence is equal to the prior probability, expected certainty is greater than or equal to prior certainty.

The inequality holds where there is a possibility that the fresh evidence will result in the probability assessment moving to the other side of the 50% level. That is, expected certainty from evidence E will be greater than prior certainty where $P(G|E^+)$ and $P(G|E^-)$ lie on different sides of the probability level 0.5.

The expected certainty result follows from the expected probability result and the relationship between probability and certainty.

[84] Box 1.

* * *

Certainty is defined as follows:

$$C(G) = \max(P(G), P(NG)) \tag{2}$$

First, consider the situation where the prior probability is greater than 0.5, and whatever form the fresh evidence takes, the probability will remain greater than 0.5. That is,

$$P(G|E^+) > P(G) > P(G|E^-) > 0.5$$

Now, given the definition of certainty, (2), these probabilities are equal to the corresponding certainties.

$$C(G|E^+) = P(G|E^+) \tag{3a}$$

$$C(G) = P(G) \tag{3b}$$

$$C(G|E^-) = P(G|E^-) \tag{3c}$$

Expected certainty is defined as:

$$EC(G|E) = P(E^+) \cdot C(G|E^+) + P(E^-) \cdot C(G|E^-)$$

Substituting in the certainty figures from (3a) and (3c).

$$EC(G|E) = P(E^+) \cdot P(G|E^+) + P(E^-) \cdot P(G|E^-)$$

By the expected probability result, (1), the right-hand side of the equation equals $P(G)$, and from (3b), this equals prior certainty C(G). And so the equality also holds for expected certainty.

$$EC(G|E) = P(E^+) \cdot C(G|E^+) + P(E^-) \cdot C(G|E^-) = C(G) \tag{4}$$

Where the prior probability is greater than 0.5, and there is no possibility that the fresh evidence will shift the probability to a level below 0.5, expected certainty equals prior certainty.

* * *

The equality also holds where the prior probability is lower than 0.5, and there is no possibility that the fresh evidence will shift the probability to a level above 0.5. That is, where:

$$P(G|E^-) < P(G) < P(G|E^+) < 0.5$$

Now, given the definition of certainty, (2), the certainties are the complements of the corresponding probabilities:

$$C(G|E^+) = 1 - P(G|E^+) \tag{5a}$$

$$C(G) = 1 - P(G) \tag{5b}$$

$$C(G|E^-) = 1 - P(G|E^-) \tag{5c}$$

Expected certainty is defined as:

$$EC(G|E) = P(E^+) \cdot C(G|E^+) + P(E^-) \cdot C(G|E^-)$$

Substituting in the corresponding probabilities from (5):

$$\begin{aligned}\mathrm{EC}(\mathrm{G}|\mathrm{E}) &= P\big(\mathrm{E}^+\big)\cdot\big(1-P\big(\mathrm{G}|\mathrm{E}^+\big)\big)+P(\mathrm{E}^-)\cdot(1-P(\mathrm{G}|\mathrm{E}^-)) \\ &= \big(P\big(\mathrm{E}^+\big)+P(\mathrm{E}^-)\big)-\big(P\big(\mathrm{E}^+\big)\cdot P\big(\mathrm{G}|\mathrm{E}^+\big)+P(\mathrm{E}^-)\cdot P(\mathrm{G}|\mathrm{E}^-)\big) && \text{(expanding, regrouping)} \\ &= 1-P(\mathrm{G}) && \text{(by the negation rule and the expected probability result)} \\ &= \mathrm{C}(\mathrm{G})\end{aligned}$$

And so, in this situation too, the equality holds. Where the prior probability is less than 0.5, and there is no possibility that the fresh evidence will shift the probability to a level above 0.5, expected certainty equals prior certainty.

* * *

Consider now the situation where the prior probability is greater than 0.5, and the fresh evidence, if it takes a negative form, will produce a probability less than 0.5. In this situation prior certainty equals prior probability, and certainty, given positive evidence, will equal the probability given positive evidence. But certainty given the negative evidence will be equal to the complement of, and greater than, probability given negative evidence.

$$\mathrm{C}(\mathrm{G}) = P(\mathrm{G}) \tag{6a}$$

$$\mathrm{C}\big(\mathrm{G}|\mathrm{E}^+\big) = P\big(\mathrm{G}|\mathrm{E}^+\big) \tag{6b}$$

$$\mathrm{C}(\mathrm{G}|\mathrm{E}^-) = 1-P(\mathrm{G}|\mathrm{E}^-) > 0.5 > P(\mathrm{G}|\mathrm{E}^-) \tag{6c}$$

Now, expected certainty is defined as:

$$\mathrm{EC}(\mathrm{G}|\mathrm{E}) = P\big(\mathrm{E}^+\big)\cdot\mathrm{C}\big(\mathrm{G}|\mathrm{E}^+\big)+P(\mathrm{E}^-)\cdot\mathrm{C}(\mathrm{G}|\mathrm{E}^-)$$

Substituting from (6b):

$$\mathrm{EC}(\mathrm{G}|\mathrm{E}) = P\big(\mathrm{E}^+\big)\cdot P\big(\mathrm{G}|\mathrm{E}^+\big)+P(\mathrm{E}^-)\cdot\mathrm{C}(\mathrm{G}|\mathrm{E}^-)$$

Given inequality (6c):

$$\mathrm{EC}(\mathrm{G}|\mathrm{E}) > P\big(\mathrm{E}^+\big)\cdot P\big(\mathrm{G}|\mathrm{E}^+\big)+P(\mathrm{E}^-)\cdot P(\mathrm{G}|\mathrm{E}^-)$$

Now by the expected probability result, (1), the right-hand side of the equation equals $P(\mathrm{G})$, and from (6a), this equals C(G). So:

$$\mathrm{EC}(\mathrm{G}|\mathrm{E}) > \mathrm{C}(\mathrm{G})$$

In this situation—where the prior probability is greater than 0.5, and the fresh evidence, if it takes a negative form, will produce a probability less than 0.5—expected certainty is greater than prior evidence.

* * *

The final situation to consider is where the prior probability is less than 0.5, and the fresh evidence, if it takes a positive form, will produce a probability greater than 0.5. In this situation:

$$C(G) = 1 - P(G) > P(G) \tag{7a}$$

$$C(G|E^+) = P(G|E^+) > 1 - P(G|E^+) \tag{7b}$$

$$C(G|E^-) = 1 - P(G|E^-) > 0.5 > P(G|E^-) \tag{7c}$$

Now, expected certainty is defined as:

$$EC(G|E) = P(E^+) \cdot C(G|E^+) + P(E^-) \cdot C(G|E^-)$$

Substituting from (7c):

$$EC(G|E) = P(E^+) \cdot C(G|E^+) + P(E^-) \cdot (1 - P(G|E^-))$$

Given inequality (7b):

$$EC(G|E) > P(E^+) \cdot (1 - P(G|E^+)) + P(E^-) \cdot (1 - P(G|E^-))$$

$$\begin{aligned} EC(G|E) &> (P(E^+) + P(E^-)) - (P(E^+) \cdot P(G|E^+) + P(E^-) \cdot P(G|E^-)) && \text{(expanding, regrouping)} \\ EC(G|E) &> 1 - P(G) && \text{(by the negation rule and the expected probability result)} \\ EC(G|E) &> C(G) && \text{(from(7a))} \end{aligned}$$

In this situation—where the prior probability is less than 0.5, and the fresh evidence, if it takes a positive form, will produce a probability greater than 0.5—expected certainty is greater than prior evidence.