Australian Journal of Forensic Sciences
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tajf20

The psychology of interpreting expert evaluative opinions

Kristy A. Martire [a], Richard I. Kemp [a] & Ben R. Newell [a]
[a] School of Psychology, University of New South Wales, Sydney, Australia
Published online: 18 Apr 2013.

Australian Journal of Forensic Sciences, 2013
http://dx.doi.org/10.1080/00450618.2013.784361

The psychology of interpreting expert evaluative opinions

Kristy A. Martire*, Richard I. Kemp and Ben R. Newell

School of Psychology, University of New South Wales, Sydney, Australia

(*Received 4 March 2013; final version received 6 March 2013*)

The standards for expressions of evaluative opinions in the forensic sciences are increasingly being challenged and refined. Where once categorical statements regarding the origin of a trace were standard practice, criminalists are now being encouraged to represent the uncertainty associated with their inferential process by using numerical or verbal likelihood ratios. Although there are valid reasons to support this shift, the approach is not without limitations. Decades of psychological research investigating the interpretation and integration of probabilistic expressions, and the equivalence of verbal and numerical formulations for uncertainty, reveals a disconnect between what is intended by experts and what is understood by decision-makers. In this paper we present an indicative review of the psychological evidence to foster communication and collaboration between forensic scientists and psychologists and reduce instances of miscommunication in our criminal justice system.

Keywords: expert evaluative opinions; psychology; likelihood ratios; probability; decision-making

Introduction

Forensic identification science involves inferential reasoning from the trace back to the source. This inductive logic does not permit categorical or certain conclusions that a trace (e.g. the latent fingerprint from a crime scene) originated from a specific known source (e.g. the finger of the police suspect). Rather, the observations made by the forensic scientist will either lend support to that hypothesis, or to the alternative hypothesis, that the trace originated from another source. Opinions offered by forensic scientists, which estimate the amount of support their analysis provides to one of these hypotheses as compared with the other, account for the inherent uncertainty associated with inductive inferences and have therefore been deemed logically appropriate[1]. However, evaluative opinions of this kind represent a challenge for those asked to interpret the meaning of such statements.

The push for probabilistic evaluative opinions in the forensic sciences

The forensic sciences are currently undergoing a period of significant upheaval. For many criminalists, until very recently, the pursuit of individualisation or 'absolute identification'[2] 'to the exclusion of all others' was at the core of their professional practice[3,4]. However, more recently, the notion of individualisation[5], its appropriateness for forensic

*Corresponding author. Email: k.martire@unsw.edu.au

science practice[6], and the use of inferential logic to reach such conclusions[7], have all received critical attention. Our intention is not to review these debates in detail here, but rather to highlight these issues as forces instrumental in motivating change in forensic science disciplines.

One of the most influential forces operating upon the practice of forensic science in recent times has been the National Academies of Science report 'Strengthening the forensic sciences in the United States'[8]. In this report, the general validity and reliability of expert analyses were questioned and concerns were raised regarding how the results of these analyses should be expressed. Specifically:

> There is a critical need in most fields of forensic science to raise the standards for reporting and testifying about the results of investigations. For example, many terms are used by forensic examiners in reports and in court testimony to describe findings, conclusions, and the degrees of association between evidentiary material (e.g., hairs, fingerprints, fibres) and particular people or objects. Such terms include but are not limited to 'match,' 'consistent with,' 'identical,' 'similar in all respects tested,' and 'cannot be excluded as the source of.' The use of such terms can have a profound effect on how the trier of fact in a criminal or civil matter perceives and evaluates evidence. Yet the forensic science disciplines have not reached agreement or consensus on the precise meaning of any of these terms. (Ref. 8, p.185)

To their credit, influential institutions and actors within the global forensic science community had foreseen these issues to some extent and responded by offering proposals for standardised and defensible approaches to the communication of evaluative forensic science opinions. In 2010, Berger spoke on behalf of the Dutch Forensic Institute (NFI) explaining the rationale behind its 'transition to a new logically correct conclusion scale for its forensic reports'. Their new direction specifically involved the formulation of evaluative opinions in line with Bayesian principles of belief updating. Namely, that forensic conclusions should convey the evidential value of an analysis with reference to two mutually exclusive hypotheses (e.g. Hypothesis 1: that X is true and Hypothesis 2: that X is not true) according to 'a new verbal conclusion scale' where:

> The findings of the examination are …[About equally likely / Somewhat more likely / more likely / much more likely / very much more likely] … when hypothesis 1 is correct, as / than when hypothesis 2 is correct. (p. 787)

Similarly, in 2009 the UK-based Association for Forensic Science Providers (AFSP) defined 'evaluative opinions' as 'opinions of evidential weight' and went on to propose guiding principles for the formulation of these opinions and standards for the expression of the opinions in court[9]. The AFSP advised that:

> The opinion or conclusion should express the degree of support for one proposition or the other depending upon the value of the likelihood ratio. (p. 163)

The AFSP also provided a scale for use by all AFSP organisations, where numerical values are presented along with their 'verbal equivalents' ranging from 'weak support for the proposition' to 'extremely strong' support. Thus, it is apparent that organisations within the forensic science community were willing to reform their procedures consistent with the criticisms contained in the NAS report, and to this end proposed various frameworks to guide the practice of criminalists offering evaluative expert opinions. These newfound practice directions did not, however, remain unchallenged for long.

In 2010 the England and Wales Court of Criminal Appeal judgment in the case of *R v T* threatened the admissibility of expert evaluative opinions reporting evidentiary value in the form of likelihood ratios, particularly where the formulation of that likelihood ratio could not be made transparent to the court [10]. Those instrumental in the formulation of standards, and others, were quick and critical in their response to this ruling[11–14], and in a position statement signed by 31 individuals and organisations, forcefully reasserted the view that probability theory provides the only 'coherent and logical foundation' for reasoning under uncertainty and that:

> 4… the likelihood ratio provides the most appropriate foundation for assisting the court in establishing the weight that should be assigned to … observations.
>
> 5. A verbal scale based on the notion of the likelihood ratio is the most appropriate basis for communication of an evaluative expert opinion to the court.

Although the position statement has been challenged with regard to its selection of likelihood ratios[15], there are signs that the position taken is acceptable to practitioners. For example research exploring the formulation of evidentiary likelihood ratios accounting for wear in interpreting shoemark evidence[16], and calculating the weight of fingerprint comparisons[17], has been recently published. Our primary concern, however, is not with the appropriateness of this approach (although that is the concern of some[18]), but rather with the question of whether the jurors and other triers of fact can interpret this form of evidence in the manner intended by the forensic science practitioners.

The psychology of juror interpretations of probabilistic evaluative opinions

Expert evaluative opinions, whatever their specific formulation, are presented to fact-finders in order to help them to reach an accurate resolution to a dispute in issue[19]. As such it is incumbent on the expert to have regard for the impact of their testimony on the decisions of judges and jurors. The recommended form – evidentiary likelihood ratios – is often presented in the context of Bayesian belief updating. In Bayesian terms, jurors hearing evidence in a trial can be thought of as updating 'prior beliefs' about guilt or innocence as new information is heard and integrated, thereby eventually reaching a 'posterior belief' (once all the evidence has been presented) which should guide a final judgment. One of the goals of the expert should be to assist in this belief updating process via the provision of testimony. This assistance should not necessarily go so far as providing jurors with the formula necessary for calculating a posterior belief (i.e. Bayes' Theorem), but rather simply make jurors aware that they should consider the impact of each piece of evidence on their current belief about guilt or innocence. Yet few, if any, in the forensic sciences have explicitly considered whether the proposed logically and scientifically appropriate form of testimony is also a valid and reliable means of conveying the intended value of the information to jurors.

Some psychologists are interested in the processes of belief updating and comprehension of probabilistic information, particularly within forensic contexts. A considerable body of empirical research has been amassed over decades, which is directly relevant to the likely impact of forensic science expert evaluative opinions on juror decision-making.

Probabilistic information and belief updating

Extensive research on the use of statistical evidence, within and outside the context of legal arguments, suggests that the logical consistency of an approach is unlikely to be sufficient to guarantee either comprehension or appropriate integration of probabilistic information by decision makers. Instead, the evidence suggests that people have a generally poor understanding of probabilities and statistics; tend to confuse likelihoods and posterior beliefs (as in the prosecutor and defence attorney fallacies)[20]; and tend toward an under-valuing of the evidence as compared to normative (i.e. Bayesian) models[21–24]. Moreover, meta-analyses of risk communication literature[25] in addition to numerous studies focusing specifically on probabilistic expressions in court[26–29] suggest that decision-makers are sensitive to the form in which statistics are presented and have a preference for natural frequencies or event rates.

To illustrate some of these issues; Gigerenzer and colleagues were interested in whether the general public correctly understood the meaning intended by weather forecasters when they state that there is 'a 30% chance of rain tomorrow' [30]. Researchers asked 750 citizens living in five cities (New York, Amsterdam, Berlin, Milan and Athens) to indicate which of the following three options gave the most accurate interpretation of the forecast: (a) it will rain tomorrow in 30% of the region; (b) it will rain tomorrow for 30% of the time; (c) it will rain on 30% of the days like tomorrow. Results indicated that while two-thirds of respondents in New York interpreted a probability of rain as intended by meteorologists (preferring option 'c'), only one-third to one-fifth of those in the remaining cities selected the correct interpretation. The authors attribute this miscommunication to the absence of a reference class in the statement as well as familiarity with the format (of the countries examined only the US regularly uses probabilistic forecasts). The authors conclude that the failure to consider and explicitly exclude all alternative interpretations when communicating single-event probabilities has far reaching consequences, including in the case of forensic science evidence in courts.

One of the arguments for the use of likelihood ratios in the presentation of forensic science evidence is that they explicitly state both of the hypotheses under consideration, and thus explicitly remove any ambiguity regarding the reference class as the context of the comparison and resultant conclusions is made explicit. However, this form of expression also has its limitations; de Keijser and Elffers[31] used realistic technical forensic reports to examine how well judges, lawyers and experts understood evaluative expert opinions expressed using the verbal scale recommended by the NFI[32]. Researchers found that although experts (members of the NFI) outperformed jurists, mistakes interpreting the likelihood ratios were common in both groups and insight into their own understanding was low. Specifically, in one case scenario, involving a comparison of chemical signatures of two samples of glue residue, 92.4% of judges, 80.9% of lawyers and 58% of experts endorsed erroneous interpretations of the evidence transposing the conditional (i.e. committing the prosecutor's fallacy). As a result, the authors called for more open-minded discussion between jurists and forensic experts to rid the criminal courts of such misunderstandings.

Other recent research focusing on the effects of subtle formulation components of statements about forensic statistics also highlights interpretational failures[33]. Koehler presented 315 jury-eligible participants with one of three types of expert shoeprint testimony describing a match between prints at the crime scene and the defendant's shoe. The match statistic reflected a combination of three threats to the validity of an inference

that the known and unknown samples share a common origin: coincidental match only; coincidental match plus lab error; or coincidental match, lab error and examiner error. The results indicated that participant-jurors were more persuaded by the evidence, had more confidence in the defendant's guilt, estimated higher probabilities of guilt and returned more guilty verdicts when they were presented evidence that did not account for threats to validity (i.e. the various errors) than the evidence which did.

Overall, then, a large body of psychological literature casts doubt upon the extent to which lay decision-makers, jury-eligible respondents, jurists and experts understand probabilistic information generally, and when presented specifically in the various recommended likelihood ratio formats. These challenges are further compounded by assumptions regarding the equivalence of verbal and numerical expressions for the evidentiary weight within expressions of uncertainty.

Interpretations of verbal expressions of uncertainty

The position statement[1] proposes an untested and unvalidated scale of verbal and numerical equivalence for the expression of evidentiary values. This would be an appropriate step towards standardising the expression of evaluative opinions only if it were reasonable to assume: (a) a consistent interpretation across decision-makers; and (b) an interpretation that is sufficiently similar to the meaning intended by the expert. Unfortunately, the literature suggests that neither of these conditions is likely to be met.

The meaning attributed to a single word can vary from person to person and from context to context. This is particularly true where there is ambiguity in the message. Research examining vague verbal expressions of probabilities or risk suggests that there are large differences between individuals in the attribution of numerical probabilities to the same verbal expressions of uncertainty[34,35]. Further, these interpretations can be moderated by various factors including the utility or disutility of the judged event[36,37] and the context within which the information is being presented[34,36,38]. Furthermore, although there may be some benefit associated with verbal rather than numerical communication[39,40,41], there is also a substantial potential for miscommunication[42,43]. Moreover, generally people tend to prefer, trust, understand and are more satisfied with numerical rather than verbal versions of the information (see review, Ref. 44).

Specifically, Budescu and colleagues[45,46] conducted two studies investigating the communication of uncertainty in the Intergovernmental Panel of Climate Change (IPCC) fourth report. In this report, a table for the likelihood of a particular environmental occurrence was matched with proposed verbal equivalents (e.g. >99% was 'virtually certain'). Participants in the first study were asked to read 13 sentences from the IPCC report, which included probabilistic pronouncements and to provide their best estimate of the probability intended by the author. Overall the responses across participants varied widely and generally represented an underestimate of the numerical value intended by the author. The second study attempted to extend and generalise these conclusions by using a random sample representative of the US population[46] and indeed demonstrated a consistent underutilisation of the evidence. The authors in this study also trialled alternative communication formats and found that a dual verbal-numerical scale, where the probabilistic statement was presented together with a copy of the scale indicating the full range of numerical and verbal expressions (i.e. from >99% or 'virtually certain' to <1% or 'exceptionally unlikely'), increased differentiation between the terms, increased the consistency of the interpretations of the terms, and increased correspondence with the experts' intentions.

Similar examinations of the verbal and numerical equivalence of uncertainty expression have also been conducted in the context of the forensic sciences. McQuiston-Surrett and Saks[47] examined undergraduate psychology students' ratings of the strength of a set of standardised terms proposed by the American Board of Forensic Odontology. The ratings provided by these mock-jurors were roughly the opposite of what the ABFO intended. For example participants attributed the greatest certainty to testimony of a 'match' (86 on a 100 point scale), but this was the phrase that the ABFO reserved for the lowest level of certainty among the four specified testimonial options. Conversely, the phrase the ABFO designated for the strongest expression of certainty, 'reasonable scientific certainty', was rated as second most uncertain by participants (70.7/100). Thus, in this instance there were large discrepancies between expert intentions and lay interpretations, leading the authors to conclude that 'forensic expert witnesses cannot simply adopt a term, define for themselves what they wish it to mean, and expect judges and juries to understand what they mean by it'.

We are currently examining the verbal and numerical correspondence of the scale proposed by the AFSP[9]. In one study[48], 75 undergraduate students in Australia and 545 community members from the United States read brief details of a case loosely based on *R v T*. Participant-jurors were asked to provide their belief in the likely guilt of the accused both before and after being presented with shoeprint evidence, which was considered against two alternative propositions: (1) that shoe has made the mark; (2) the shoe has not made the mark. The expert's opinion regarding the strength of the evidence in regard to these two propositions was then expressed in either a numerical or verbal format and at one of three strength levels based on those proposed by the Association of Forensic Science Providers[9] (low, moderate or high; see Table 1, adapted from Ref. 48). For example, a participant in the low evidence strength condition read the following expert opinion: 'In my opinion the correspondence between the footwear mark at the crime scene and the shoe of the accused [is 4.5 times more likely] (numerical) or [offers weak or limited support] (verbal) when proposition 1 is correct than when proposition 2 is correct'.

Results from these analyses indicate that although jurors have some sensitivity to the expert's opinion regarding the strength of the evidence (they believe guilt is more likely after hearing strong evidence than weak evidence), the amount of belief-change caused by the expert evidence was both consistently and markedly lower than the value intended by the expert. For example, where the expert indicated that a likelihood ratio of 4.5 should be adopted, the median participant response indicated a likelihood ratio of 1.2 was actually used. More remarkably, where the expert indicated a likelihood ratio of 495,000 should be applied, the median participant response was to update their beliefs using a likelihood ratio of just 1.5; a figure 353,571 times smaller than intended by the expert.

Table 1. Evidence strength (derived from Ref. 9) and presentation method.

	Presentation method	
Evidentiary strength	Numerical	Verbal (support)
Low	4.5	'Weak or limited'
Moderate	450	'Moderately strong'
High	495,000	'Very strong'

The results also indicated that the correspondence between the verbal and numerical expressions is low. Consistent with the findings of Budescu and colleagues[45,46] the changes in belief brought about by the numerical expressions of evidence were significantly larger than those brought about by the 'equivalent' verbal expressions. Specifically, numerical expressions of evidence resulted in greater adjusted mean belief-change (1.63 units) compared with verbal expressions (0.31 units).

Beyond this, there was also some suggestion that the precise wording for the lowest level of evidentiary strength ('weak or limited support') resulted in a majority of jurors (61.72%) treating evidence consistent with the prosecutions' case (i.e. that the trace from the crime scene originated from the shoe of the defendant) as evidence supporting the defence alternative (i.e. that the trace from the crime scene did not originate from the shoe of the defendant). Although this type of misinterpretation of evidence is particularly concerning in a forensic context, such inversions are not uncommon and have been observed by decision-makers in various contexts[49,50,51,52]. In sum, these studies suggest that the true meaning of the experts evaluative opinion is likely to be lost in translation.

Conclusions

Together, the research findings presented here call into question the appropriateness of both the specific verbal/numerical scale proposed by the AFSP and the use of likelihood ratios as the 'most appropriate basis for communication of an evaluative expert opinion to the court'. While a consensus appears to be developing regarding the logical and scientific appropriateness of the use of likelihood ratios in forensic work, many questions regarding the effect of this means of expressing evaluative opinions still require empirical investigation and thoughtful consideration – both from within forensic science and the criminal justice system more broadly. In particular, we would argue that, while logical and scientific appropriateness within expert evaluative opinions is certainly necessary, this should not be the only consideration when deciding how to express the results of an analysis to a juror or other decision-maker. It is also important to understand how factfinders think about and utilise evaluative opinions[32,53] and to strive towards a form of expression that is both accurate and likely to give rise to an appropriate interpretation.

Given the preceding review it would be appropriate to ask how the pursuit of appropriate and valid evaluative expert opinions in the forensic sciences should progress from here. A first and vital step is to increase communication between forensic scientists and a diverse range of interested researchers and scholars. We hope that this paper (together with the others in this special issue), will contribute to this dialogue. By engaging with each other's research, practice and points of view[31] it will then be possible to develop a broad, diverse and informed understanding of the complex array of factors contributing to the formulation of evaluative expert opinions and their interpretations by laypersons.

This increased communication may then facilitate the second step; drawing on multidisciplinary evidence to formulate and design research examining possible avenues for the refinement and improvement of validity and reliability of expressions of uncertainty. Within the literature cited above, various methods for improving the correspondence between expert intentions and lay interpretations, as well as verbal and numerical expressions of uncertainty, have been trialled. For example through the use of a dual verbal-numerical expressions[46], or visual representations of the evidence[31]. More broadly, the literature suggests methods for deriving evidence-based verbal equivalents[39,54] which could be applied to various numerical representations of evidence strength.

These two steps have the potential to significantly improve the correspondence between expert intentions and lay interpretations and in turn pave the way for a continuing and sophisticated dialogue to improve the functioning of our criminal justice system.

References

1. Aitken C, Berger CEH, Buckleton JS, Champod C, Dawid AP, Evett IW, Gill P, Gonzalez-Rodriguez J, Jackson G. Expressing evaluative opinions: a position statement. Sci Justice. 2011;51(1):1–2.
2. Stoney DA. What made us ever think we could individualize using statistics? J Forensic Sci Soc. 1991;31(2):197–199.
3. Inman K, Rudin N. Principles and practices of criminalistics: the profession of forensic science. 1st ed. Boca Raton, FL: CRC; 2001.
4. Osterburg JW. The evaluation of physical evidence in criminalistics: subjective or objective process? J Crim Law Crim. 1969;60(1):97–101.
5. Saks MJ, Koehler JJ. Individualization fallacy in forensic science evidence. The Vand L Rev. 2008;61:199–219.
6. Page M, Taylor J, Blenkin M. Uniqueness in the forensic identification sciences - fact or fiction? Forensic Sci Int. 2011;206(1–3):12–18.
7. Saks MJ, Faigman DL. Failed forensics: how forensic science lost its way and how it might yet find it. Annu Rev Law Soc Sci. 2008;4:149–171.
8. National Research Council Committee on Identifying the Needs of the Forensic Science Community. Strengthening forensic science in the United States: a path forward. Washington: The National Academies Press; 2009.
9. Association of Forensic Science Providers. Standards for the formulation of evaluative forensic science expert opinion. Sci Justice. 2009;49(3):161–164.
10. R v T, England and Wales Court of Appeal, Criminal. 2010.
11. Morrison GS. The likelihood-ratio framework and forensic evidence in court: a response to R v T. T Int J Ev Proof. 2012;16(1):1–29.
12. Berger CEH, Buckleton J, Champod C, Evett IW, Jackson G. Evidence evaluation: a response to the court of appeal judgment in R v T. Sci Justice. 2011;51(2):43–49.
13. Redmayne M, Roberts P, Aitken C, Jackson G. Forensic science evidence in question. Crim Law Rev. 2011;5:347–356.
14. Robertson B, Vignaux G, Berger CEH. Extending the confusion about Bayes. Mod Law Rev. 2011;74(3):444–455.
15. Facey OE, Davis RJ. Re: Expressing evaluative opinions; a position statement. Sci Justice. 2011;51(4):212–212.
16. Skerrett J, Neumann C, Mateos-Garcia I. A Bayesian approach for interpreting shoemark evidence in forensic casework: Accounting for wear features. Forensic Sci Int. 2011.
17. Neumann C, Evett IW, Skerrett J. Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. J Roy Stat Soc. 2011;175(2):371–345.
18. Ligertwood A, Edmond G. Expressing evaluative forensic science opinions in a court of law. Law Prob Risk. 2012;11(4):289–302.
19. Edmond G, Ligertwood A. Australian evidence: a principled approach to the Common Law and Uniform Acts, 5th ed. Sydney: Butterworths; 2010.
20. Thompson WC, Schumann EL. Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. Law Human Behav. 1987;11(3):167–187.
21. Smith LL, Bull R, Holliday R. Understanding juror perceptions of forensic evidence: investigating the impact of case context on perceptions of forensic evidence strength. J Forensic Sci. 2011;56(2):409–414.
22. Kaye DH, Koehler JJ. Can jurors understand probabilistic evidence? J Roy Stat Soc. 1991;154(1):75–81.

23. Goodman J. Jurors' comprehension and assessment of probabilistic evidence. Am J Trial Advoc. 1992;16:361–390.
24. Faigman DL, Baglioni A. Bayes' theorem in the trial process: instructing jurors on the value of statistical evidence. Law Human Behav. 1988;12(1):1–17.
25. Trevena LJ, Barratt A, Butow P, Caldwell P. A systematic review on communicating with patients about evidence. J Eval Clin Pract. 2006;12(1):13–23.
26. Koehler JJ. When are people persuaded by DNA match statistics? Law Human Behav. 2001;25(5):493–513.
27. Koehler JJ. On conveying the probative value of DNA evidence: frequencies, likelihood ratios, and error rates. Uni Colo Law Rev. 1996;67:859–886.
28. Koehler JJ, Macchi L. Thinking about low-probability events an exemplar-cuing theory. Psychol Sci. 2004;15(8):540–546.
29. Lindsey S, Hertwig R, Gigerenzer G. Communicating statistical DNA evidence. Jurimetrics. 2002;43:147–164.
30. Gigerenzer G, Herwig R, Van Den Broek E, Fasolo B, Katsickopoulos KV. 'A 30% chance of rain tomorrow': how does the public understand probabilistic weather forecasts? Risk Anal. 2005;25(3):623–629.
31. de Keijser J, Elffers H. Understanding of forensic expert reports by judges, defense lawyers and forensic professionals. Psychol Crime Law. 2012;18(2):191–207.
32. Berger CEHCriminalistics is reasoning backwards. Nederlands Juistenblad. 2010;85:784–789.
33. Koehler JJ. If the shoe fits they might acquit: the value of forensic science testimony. J Empir Legal Stud. 2011;8(s1):21–48.
34. Brun W, Teigen KH. Verbal probabilities: ambiguous, context-dependent, or both? Organ Behav Hum Dec. 1988;41(3):390–404.
35. Shaw N, Dear P. How do parents of babies interpret qualitative expressions of probability? Arch Dis Child. 1990;65(5):520–523.
36. Weber EU, Hilton DJ. Contextual effects in the interpretations of probability words: perceived base rate and severity of events. J Exp Psychol Human. 1990;16(4):781–789.
37. Patt AG, Schrag DP. Using specific language to describe risk and probability. Climatic Change. 2003;61(1):17–30.
38. Wogalter MA, Young SL, Brelsford JW, Barlow T. The relative contributions of injury severity and likelihood information on hazard-risk judgments and warning compliance. J Safety Res. 1999;30(3):151–162.
39. Wallsten TS, Budescu DV. A review of human linguistic probability processing: General principles and empirical evidence. Knowl Eng Rev. 1995;10(1):43–62.
40. Fox CR, Irwin JR. The role of context in the communication of uncertain beliefs. Basic Appl Soc Psych. 1998;20(1):57–70.
41. Politi MC, Han PJK, Col NF. Communicating the uncertainty of harms and benefits of medical interventions. Medi Decis Making. 2007;27(5):681–695.
42. Budescu DV, Wallsten TS. Dyadic decisions with numerical and verbal probabilities. Organ Behav Hum Dec. 1990;46(2):240–263.
43. Karelitz TM, Budescu DV. You say 'probable' and I say 'likely': improving interpersonal communication with verbal probability phrases. J Exp Psychol-Appl. 2004;10(1):25–41.
44. Visschers VHM, Meertens RM, Passchier WWF, De Vries NNK. Probability information in risk communication: a review of the research literature. Risk Anal. 2009;29(2):267–287.
45. Budescu DV, Broomell S, Por HH. Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. Psychol Sci. 2009;20(3):299–308.
46. Budescu DV. Por HH. Broomell S. Effective communicating of uncertainty in the IPCC reports. Climatic Change. 2011;DOI: 10.1007/s10584-011-0330-3.
47. McQuiston-Surrett D, Saks MJ. Communicating opinion evidence in the forensic identification sciences: Accuracy and impact. Hastings LJ. 2007;59:1159–1190.

48. Martire KA, Kemp RI, Watkins I, Sayle MA, Newell BR. The expression and interpretation of uncertain forensic science evidence: verbal equivalence, evidence strength and the weak evidence effect. Law Human Behav. In press.
49. Fernbach PM, Darlow A, Sloman SA. When good evidence goes bad: the weak evidence effect in judgment and decision-making. Cognition. 2011;119(3):459–467.
50. McKenzie CRM, Lee SM, Chen KK. When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. J Behav Decis Making. 2002;15(1):1–18.
51. Friedrich J., Smith P. Suppressive influence of weak arguments in mixed-quality messages: an exploration of mechanisms via argument rating, pretesting, and order effects. Basic Appl Soc Psych. 1998;20(4):293–304.
52. Harris A, Corner A, Hahn U. 'Damned by faint praise': a Bayesian account. 2009. Proceedings of the Cognitive Science Society.
53. Fenton N, Neil M. On limiting the use of Bayes in presenting forensic evidence. 2012; https://www.eecs.qmul.ac.uk/~norman/papers/likelihood_ratio.pdf
54. Mosteller F, Youtz C. Quantifying probabilistic expressions. Stat Sci. 1990;5(1):2–12.