Forensic Science International 240 (2014) 61–68

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect

K.A. Martire *, R.I. Kemp, M. Sayle, B.R. Newell

School of Psychology, The University of New South Wales, Sydney 2052, NSW, Australia

ARTICLE INFO

Article history:
Received 8 October 2013
Received in revised form 28 March 2014
Accepted 1 April 2014
Available online 14 April 2014

Keywords:
Evidence interpretation
Expert evaluative opinions
Communicating uncertainty
Likelihood ratios

ABSTRACT

Likelihood ratios are increasingly being adopted to convey expert evaluative opinions to courts. In the absence of appropriate databases, many of these likelihood ratios will include verbal rather than numerical estimates of the support offered by the analysis. However evidence suggests that verbal formulations of uncertainty are a less effective form of communication than equivalent numerical formulations. Moreover, when evidence strength is low a misinterpretation of the valence of the evidence – a "weak evidence effect" – has been found. We report the results of an experiment involving $N = 404$ (student and online) participants who read a brief summary of a burglary trial containing expert testimony. The expert evidence was varied across conditions in terms of evidence strength (low or high) and presentation method (numerical, verbal, table or visual scale). Results suggest that of these presentation methods, numerical expressions produce belief-change and implicit likelihood ratios which were most commensurate with those intended by the expert and most resistant to the weak evidence effect. These findings raise questions about the extent to which low strength verbal evaluative opinions can be effectively communicated to decision makers at trial.

In many jurisdictions the practice of communicating forensic science expert opinions to courts is undergoing substantial change. Where it was once accepted practice that an expert would testify to categorical individualization [1,2], increasingly expressions reflecting the uncertain nature of forensic analyses are being demanded [3], recommended [4–7], and explored (e.g., [8–17], for detailed consideration).

The use of a likelihood ratio to express the subjectivity and uncertainty associated with forensic science evidence has been embraced by sectors of the forensic science community. In a position statement signed by 31 stakeholders and agencies these scientists declared likelihood ratios to be "the most appropriate foundation for assisting the court in establishing the weight that should be assigned . . ." [6]. The likelihood ratio (LR) is a statement which conveys the probability of the observations given each of the stated propositions or hypotheses (H). For example the likelihood ratio communicates the probability of obtaining the observed similarities between a fingerprint from a known origin and the fingerprint of questioned origin under the hypothesis that the two samples have the same origin (H_1) versus under the hypothesis that they have different origins (H_2) [18].

Critically, however, the signatories to the above mentioned statement appeared to suggest that the preferred form of expression for the likelihood ratio statement is verbal rather than numerical. This is a position supported by the Standards proposed by the Association of Forensic Science Providers [7] who proposed a scale for the translation of numerical likelihood ratios into verbal formats (see Table 1).

Accordingly, taking the approach recommended by Aitken et al. [6] it is preferred, for example, that the expert state: "In my opinion the correspondence between the fingerprint found at the crime scene and the fingerprint taken from the accused offers strong support if the two fingerprints originated from the same person than if the two fingerprints originated from different people"; Rather than: ". . . the correspondence between the fingerprint found at the crime scene and the fingerprint taken from the accused is 5500 times more likely if the two fingerprints originated from the same person . . ."; (see Table 1).

In addition to the perception that verbal expressions of the likelihood ratios are "the most appropriate basis for communication of an evaluative expert opinion to the court . . ." [6], it is also the case that the quantitative data necessary to compute a numerical likelihood ratio are unavailable in many domains of forensic science [12]. This means that, irrespective of their actual or perceived appropriateness, verbal expressions of uncertainty are likely to be observed with increasing frequency in the forensic sciences, at least until issues regarding data availability are resolved.

* Corresponding author. Tel.: +61 2 9385 8563; fax: +61 2 9385 3641.

http://dx.doi.org/10.1016/j.forsciint.2014.04.005
0379-0738/

It is also important to consider the intended audience for these expressions of uncertainty [19] and how their interpretations might be influenced by verbal and numerical expressions. Several avenues of research suggests that people often have difficulties understanding probabilities and statistics, and tend to produce suboptimal translations between verbal and numerical expressions of uncertainty. In particular, evidence suggests decision-makers tend to, but don't always [20] under-value probabilistic evidence compared with normative estimates (e.g., [8,10,11,13,21–23]). Furthermore, it is widely acknowledged that different people will understand the same verbal probability expression differently [24–29] leading to conclusions that verbal labels create an "illusion of communication" [30]. Consequently, it is not appropriate to simply assume a particular probability phrase will automatically and reliably result in a specific desired interpretation [9,24].

Despite these concerns it has been suggested that verbal expressions of uncertainty and evaluative labels can also be beneficial, remediating some of the misinterpretations associated with numerical probabilities [4,25,28,31–33]. In an attempt to resolve this ambiguity surrounding the relative "appropriateness" of verbal and numerical expressions, particularly in the context of likelihood ratios, Martire et al. [34] compared the amount of belief change resulting from expert forensic science opinions, expressed as verbal or numerical likelihood ratios of varying strength (low, moderate and high). To do this they used the labels and numerical equivalents recommended by AFSP for evidence offering "weak or limited", "moderately strong" and "very strong" support (see Table 1). Across two web-based studies involving 905 participants Martire et al. measured the extent to which participants' belief in the guilt or innocence of the accused changed after being presented with the testimony of an expert shoe impression examiner. The testimony varied in strength as described above, but always indicated that the likelihood of the observed similarity between the shoe print from the crime scene and the shoe print of the accused was more likely if the two prints shared a common origin (H_1) than if they had different origins (H_2) (i.e., was evidence in support of the prosecution case).

Three main effects emerged across the two studies: (1) A broad sensitivity to evidence strength was observed such that expert opinion evidence of greater strength resulted in significantly more belief change than did lower strength expert evidence; (2) a tendency to underweight the evidence compared to Bayesian norms; specifically, participants did not update their initial beliefs to the extent that would be predicted through the application of Bayes theorem; and (3) a weak evidence effect was observed for low strength verbal evidence when brought by the prosecution. That is, where participants were presented verbal evidence which "weakly" supported the prosecution's version of the case, rather than increase their belief in the guilt of the accused by a small amount as would be appropriate given the additional incriminating evidence, the majority of participants elected instead to decrease their belief in the guilt of the accused. This effect was not statistically significant where the evidence was presented numerically.

The weak evidence or "boomerang" [35] effect describes a situation where weak evidence supporting a proposition, in this case H_1, is wrongly interpreted as evidence supporting the alternate proposition H_2 [36]. In practice this meant that the expert's opinion which should have supported the prosecution's case was interpreted as supporting the defense case by a clear majority of participants in the low strength verbal conditions. Although not previously unknown [10,36–38], and to some extent context dependent [34], this inversion of the valence of the opinions of forensic scientists is somewhat concerning. Specifically, these weak evidence effects are of concern not only because they inaccurately reflect the valence of the expert's opinion, but also because of the stated belief that verbal expressions of evidence should be used by forensic science experts because they provide the most appropriate basis for communication [6]. Overall then, the observed undervaluing and weak evidence effects beg the question, if verbal expressions are not the most appropriate basis for communication, what possible alternative formulation might be?

Budescu and colleagues [25,28] suggest that presenting both verbal expressions and numerical values, in the context of the complete range of possible options (i.e., a table of values and expressions) can improve the interpretation of verbal expressions of uncertainty. In their 2009 study, Budescu and colleagues asked participants to read 13 sentences containing probabilistic terms from the 2007 Intergovernmental Panel on Climate Change (IPCC) report and provide a best estimate of the probability intended by the authors [25]. Participants were allocated to one of four conditions: (1) the control group were provided no instruction regarding the interpretation of the probabilistic term; (2) the translation group were provided a drop-down table including all verbal labels and their numerical equivalents (e.g., >99%); the verbal–numerical group (which was further split into two conditions) were provided with the table including either (3) a broad or (4) a narrow range of numerical values to accompany the verbal expressions which was presented alongside each sentence.

The researchers found that although consistency with the IPCC conversion table was generally low, consistency was significantly higher in the translation than control condition leading the authors to recommend the use of "both verbal terms and numerical values to communicate uncertainties" [25]. This conclusion was confirmed in a follow up study using a nationally representative US sample of 556 participants, which found that verbal–numerical scales (including numerical ranges) increased the differentiation between terms, the internal consistency of each term, and the correspondence with the IPCC report's intended message [28]. These results suggest that a dual form of expression not only provides more information, but also caters to a broad and heterogeneous group of decision-makers. What remains unclear however, is the extent to which the provision of similar verbal–numerical tables will improve interpretations of the more complex form of expression associated with likelihood ratios, specifically, where decision makers are explicitly asked to consider the likelihood of the observations under two competing hypotheses.

Research by de Keijser and Ellfers [12] begins to address this question by explicitly considering the interpretation of likelihood ratios in a forensic science context. Participants were 332 judges and justices, defense lawyers and employees of the Dutch Forensic Institute (NFI) who were presented with likelihood ratios reflecting expert evaluative opinions using either verbal probability statements or visual scales. Decision makers were presented with the expert's conclusion regarding the likelihood of the observations given two scenarios (S) corresponding with the two hypotheses underpinning likelihood ratio (e.g., S1: the tape used to restrain the victim originates from the roll of tape that was seized from the suspect's residence; S2: the tape used to restrain the victim originates from a random other roll of tape). In the visual conclusion condition the expert's opinion was indicated with an 'X' intersecting a horizontal line labeled from "Very strong in favour of Scenario 2" on the left, to "Very strong in favour of Scenario 1" on the right, with a "Neutral" point in the middle. Analyses showed that understanding of likelihood ratios was generally poor and that using visual scales as a "cosmetic attempt" to improve understanding neither improved or impaired participant performance.

It is, however, unclear how these visual scales and numerical–visual dual expressions of likelihood ratios might affect belief-change – rather than comprehension as in de Keijser and Elffers