

Likelihood ratios in psychological expert opinion, and their reception by professional judges

The International Journal of
Evidence & Proof
2022, Vol. 26(4) 325–341
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13657127221119545
journals.sagepub.com/home/epj



Eric Rassin 

ESSB, Erasmus Universiteit Rotterdam, Burgemeester Oudlaan 50, Rotterdam, Netherlands

Nurul Arbiyah

Clinical Psychological Science, Maastricht University, Universiteitssingel 40, Maastricht, Netherlands

Irena Boskovic 

Erasmus Universiteit Rotterdam, Rotterdam, Netherlands

Henry Otgaar

KU Leuven, Leuven, Belgium

Harald Merckelbach

Maastricht University, Maastricht, Netherlands

Abstract

In various countries, forensic scientists have begun to express their expert opinion in terms of the likelihood of observing the evidence under the primary and under an alternative hypothesis (i.e. the likelihood-ratio approach). This development is often confined to technical domains such as fingerprint analyses. In forensic psychological expertise, likelihood ratios are largely absent. In this contribution, we explain how forensic psychologists can employ likelihood ratios, and we describe two illustrating cases. We also present two studies in which we examined how (Dutch) professional judges appreciate psychological expertise framed in likelihood ratios. Findings suggest that judges ($N=39$) appreciate a fictitious expert witness report framed in likelihood-ratios similarly to an opinion framed one-dimensionally. Judges' ($N=79$) understanding of a psychological expert opinion framed in likelihood ratios was satisfactory as measured by self-report and an actual test. We conclude that, as is custom in forensic technical domains, psychological expert opinion can be expressed in likelihoods. Two of the hypothesised flip-sides, namely, lawyers' dislike of likelihoods, and their lack of proper understanding, may be surmountable.

Corresponding author:

Eric Rassin, ESSB, Erasmus Universiteit Rotterdam, Burgemeester Oudlaan 50, Rotterdam, Netherlands.

Email: rassin@essb.eur.nl

Keywords

Analysis of competing hypotheses (ACH), confession evidence, expert witness, eyewitness identification evidence, likelihood ratio

Courts may call upon psychologists to give an opinion on the validity, reliability, or trustworthiness of various pieces of evidence such as eyewitness identification evidence, witness statements, or disputed confessions. A crucial question that follows is how the psychological expert opinion should be framed. Consider a case in which two police officers drive a car at night, observe another car speeding, and try to catch up. Once the police have caught up, they signal the speeding car (now behind them) with their taillights to pull over. The car stops, and so does the police car. However, the police hit the brakes gently, and consequently there is now a distance of some hundred metres between them and the speeder. Therefore, the police drive in reverse, and it takes approximately a minute for the police to approach the people in the speeding car. It turns out to be an elderly married couple: The woman sits behind the wheel, and her husband is beside her in the passenger seat. He has had alcohol, and she has not. The police suspect that the couple swapped places in the short period needed to drive back to them. By doing so, the husband would evade penalties for drinking and driving. In fact, the police officer in the passenger seat testifies that he saw a man in the driver's seat when they passed the car. Eventually, the husband is charged with drunken driving, and the wife is charged with perjury. A psychologist is asked to give an opinion on the validity of the identification of the husband as the driver by the police officer.

Which arguments might the psychologist bring forward? (S)he might argue that the police officer expressed high confidence in his identification, and police officers generally may well be good observers (see Wells et al., 2020). Hence, these are reasons to assume that the identification is valid. Alternatively, one might argue that it was dark, both cars were driving at 100 m/h, the exposure time was brief, and the car had large headrests, so there is reason to argue that the identification is weak evidence. The psychologist might also discuss both pros and cons, and conclude that there is no scientific way of knowing how strong the identification is. Regardless of which direction the expert takes, this approach will always be one-dimensional, meaning that the expert weighs the arguments under one single hypothesis (how convincing is the identification of the male driver?). One problem with this approach is that the expert basically does the same as the judge (or, in adversarial systems, the jury). That is, they all wonder how strongly the identification supports the primary hypothesis that it was, in fact, the husband who drove the car. If the expert reports that the identification is (likely) valid, (s)he implicitly states that (s)he believes that the primary hypothesis is true. Should (s)he report that the identification is likely invalid, (s)he implicitly states that the alternative hypothesis (the wife drove the car) is true. Hence, this approach may lead to role confusion, in that the expert unintentionally takes the seat of the judge, and indirectly answers the question of whether the suspect is guilty.

Interestingly, more than thirty years ago, in 1988, Wagenaar already argued that this role confusion can and should be prevented if psychological experts take a likelihood-ratio approach. According to him, judges need to decide whether the primary hypothesis is true (probability of the primary hypothesis, given all the evidence; $P(H|E)$), whereas experts should limit themselves to an opinion on the likelihood of the evidence, given the hypotheses ($P(E|H)$). Applied to the identification case, this means that the expert has to consider two types of likelihoods. First, how likely is it that the police identifies a male driver when the driver was, indeed, male? Just for the sake of simplicity, one could settle that likelihood at 100%. Second, how likely is it that the police would identify a male driver when in fact the driver was female? Estimating this false positive error requires empirical study. When consulted as an expert witness in a case similar to the example presented above, Wagenaar carried out a study and subjected 210 participants to a simulated exposure to the female suspect (i.e. under suboptimal, darkened circumstances). Of these, 116 individuals (55%) reported having seen a man. Hence, Wagenaar reported that the maximum

likelihood ratio of the police identification of a male driver in this case is approximately $(100\% / 55\% =)$ 1.82, which by all standards is weak (see below). Obviously, this likelihood-ratio approach has limitations. For example, did the participants in the tailor-made study accurately represent the perceptual abilities of the police officer? Did the study model the actual incident? Does a study ever approach reality (see for an older version of the discussion, Walker and Monahan, 1987)? Nonetheless, Wagenaar presented one of the first examples of how the likelihood-ratio approach might be applied to psychological expertise. The approach is represented in general terms in Figure 1. In this figure, the evidence is approximately twice as likely to fit in with the primary than with the alternative hypothesis, and hence, the likelihood ratio is 2. When asked to give an opinion about the probative value of the evidence, the expert would report that the evidence is twice as likely to occur in the primary than in the alternative hypothesis.

Likelihood ratios in forensic sciences

At the beginning of the 21st century, accumulated accounts of errors in forensic technical expert witness evidence made people realise that forensic identification evidence might not be as strong and infallible as previously believed. Saks and Koehler (2005: 895) argued that it was time to ‘put the science into forensic identification science’. They advocated a radical paradigm shift ‘in the traditional forensic identification sciences in which untested assumptions and semi-informed guesswork are replaced by a sound scientific foundation and justifiable protocols. Although obstacles exist both inside and outside forensic science, the time is ripe for the traditional forensic sciences to replace antiquated assumptions of uniqueness and perfection with a more defensible empirical and probabilistic foundation.’

A milestone in this paradigm shift was the introduction of the likelihood-ratio approach in forensic sciences, as advocated by Wagenaar (1988) long before. By now, many forensic scientists in, for example, Europe and Australia have adopted this approach (Thompson, 2018; Thompson et al., 2018). Several professional associations and standard-setting bodies advise or even dictate the inclusion of likelihood ratios in expert reports (Association of Forensic Science Providers [AFSP], 2009; European Network of Forensic Science Institutes [ENFSI], 2016).

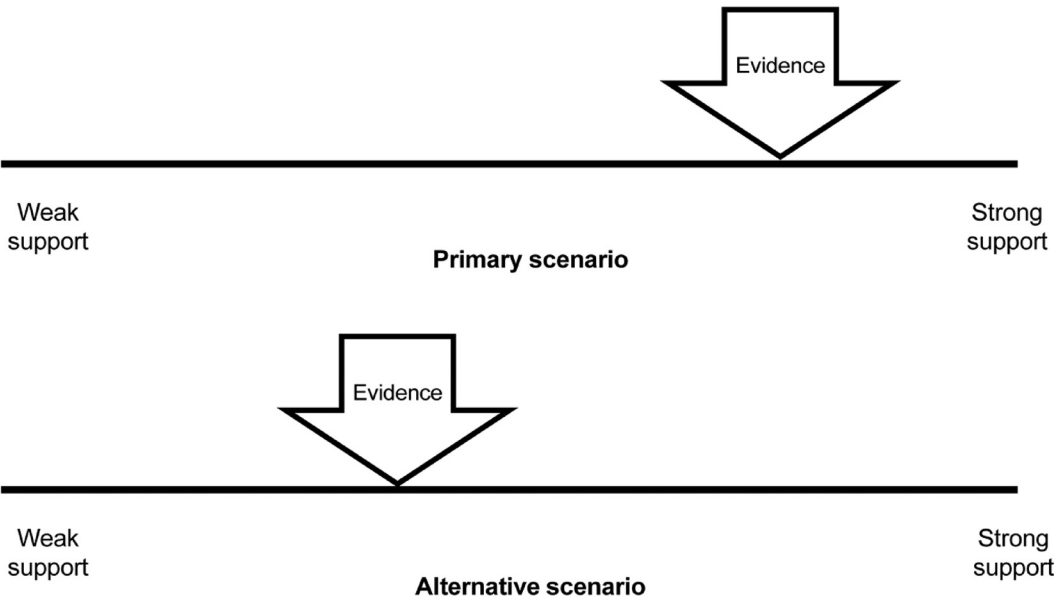


Figure 1. Schematic representation of the (two-dimensional) likelihood-ratio approach to evidence.

Advocates of likelihood ratios see various advantages of the approach. For example, the AFSP (2009) mentions as underlying guiding principles: balance (i.e. at least two competing hypotheses need to be included in the analysis), logic (likelihood ratios force the expert to reason prospectively), robustness (likelihood ratios ideally follow from scientific research, and they are fixed) and transparency (it is clear on what the expert's opinion is based). Implied in these guiding principles is a protection against various biases that threaten the validity of the expert opinion. By considering alternative hypotheses, the forensic expert is protected against tunnel vision, or confirmation bias, which can be defined quite broadly as 'the class of effects through which an individual's preexisting beliefs, expectations, motives, and situational context influence the collection, perception, and interpretation of evidence during the course of a criminal case' (Kassin et al., 2013: 45). Likelihood ratios help to reduce the risk of confirmation bias because the expert has to consider the possibility that the pertinent evidence occurs under an alternative hypothesis. Likewise, given that likelihoods are dictated by scientific research, the expert is protected against pressure by the commissioning party to phrase conclusions in a desirable fashion (cf. allegiance bias; Murrie et al., 2013).

A fundamental advantage of the likelihood ratio approach is that it sensitises experts and decision makers to the notion that in order to evaluate the strength of evidence one needs to realise that strong evidence not only fits well in the primary hypothesis, but simultaneously fits poorly in a competing alternative hypothesis. Further, it elucidates that (forensic) evidence is never error-proof. That is, there is always a likelihood that the evidence occurs under an alternative hypothesis, even if that likelihood is very remote. Therefore, the likelihood given in the denominator can never be zero (one cannot divide by zero). Hence, no matter how strongly the evidence supports the primary hypothesis (the likelihood in the numerator), it can never reach 100% certainty. In the past, the awareness of fallibility was often lacking in expert opinions, but is, by definition, implicated in the likelihood-ratio approach (Saks and Koehler, 2005; Thompson, 2018; Thompson et al., 2018).

How psychological testimony may benefit from likelihood ratios

Ironically, although Wagenaar (1988) already advocated the use of likelihood ratios long before their introduction in forensic technical sciences, they have become relatively common in technical disciplines (Thompson, 2018). By now, it would be considered inappropriate for a technical expert to frame a conclusion in a format of a probability of a fingerprint secured at the crime scene matching the suspect, *without* giving an estimate of a false positive likelihood. However, with few exceptions (Larrabee, 2008), likelihood ratios are still largely absent in psychological or psychiatric expert testimony. Psychological expert testimony is often still one-dimensional and ignores the importance of alternative hypotheses. For example, when it comes to verbal lie detection, the Criteria-Based Content Analysis (CBCA) is considered to be 'probably the most widely used veracity assessment technique in the world' (Oberlader et al., 2016: 441). CBCA is an entirely one-dimensional checklist, with anchor points being *testimony is not credible* and *testimony is credible*. With this instrument, the expert simply checks to what extent the provided statement meets specific criteria (such as detailedness and spontaneousness), and then advises on its credibility. Any alternative hypothesis (e.g. that the witness is mistaken, or lying) is merely addressed implicitly. Thus, if the expert concludes that the statement is credible, this implies that it is likely to be true, and thus unlikely to be false. In doing so, the CBCA seduces the expert to give an opinion about the probability that the primary hypothesis is true, which is actually the prerogative of the judge or jury. Wagenaar (1988: 499) warned that if social scientists do not employ a likelihood ratio format, there is a 'danger that social scientists will misrepresent the reliability of their knowledge and make biased statements'. Thus, expert opinion based on CBCA should be framed in terms of likelihoods, and this can, in principle, be done because the accuracy statistics are available from research. For example, the CBCA has a sensitivity of approximately 70% and a false positive rate of approximately 30% (Oberlader et al., 2016). Thus, it is conceivable that an expert concludes:

'My conclusion that the witness statement contains many cues of validity fits approximately (70% / 30%) = 2.33 times better in the primary than in the alternative hypothesis.' This way, the expert is open about the error rate of the CBCA.

However, some psychologists argue that likelihood ratios can, by definition, not be applied to psychological expert opinion. By their view, psychological insights are simply too frail to be captured in likelihood ratios. For example, psychologists Davies and Kovera (personal communication, May 18, 2018) argued as follows: 'Because it is likely impossible to generate a reliable population estimate that takes into account all the relevant variables, it is not appropriate to apply the likelihood ratio technique to this domain. Doing so imbues a level of certainty in the evaluation that is inappropriate and misleading.' Similarly, forensic scientist Peat argued that when applied to psychology, there is a 'lack of an adequate scientific basis for the methodology' (personal communication, August 2, 2021). Note that this argumentation goes beyond expressing concern that many topics in psychology still need to be investigated in order to reliably estimate likelihoods, or that the pertinent likelihood ratios tend to be modest in size. Here, a more principal argument against likelihood ratios is made based on the idea that psychology is inherently unreliable.

Admittedly, analyses of psychological topics (e.g. eyewitness identification, confession) involve more multicausality than, say, DNA or fingerprints at the identification level. This multicausality makes psychology a weak discipline, not in terms of methodology, but in terms of predictive power (see Manning et al., 2007). However, multicausality does not disqualify psychology as a scientific discipline suitable for likelihood ratio estimation. Multicausality merely becomes manifest in the relatively low likelihood ratios that are generally borne out by psychological research. These (low) likelihood ratios are informative because they shed light on the (limited) diagnostic power of the evidence at hand.

We argue that psychological expert testimony may benefit from likelihood ratios, and these benefits may exceed those already mentioned. For one thing, likelihood ratios make the exonerating power of evidence transparent. When taking a traditional one-dimensional approach, the exonerating power of negative evidence is oftentimes overlooked (Liebman et al., 2012). This is less likely to happen with the likelihood-ratio approach. In this approach, ratios smaller than 1 indicate that the evidence fits better in the alternative than in the primary hypothesis, and thus has exonerating power. Using Bayesian analyses, Wells and Olson (2002) argued that in the context of eyewitness identifications, 'the exonerating value of filler identification and "not there" responses can actually exceed the incriminating value of identifications of the suspect' (2002: 155). Or, in the words of Wells and Lindsay (1980: 777): 'There is no justifiable logic for approaching a lineup procedure with a set for considering an identification of the suspect to be informative while considering a nonidentification to be uninformative.'

The exonerating power of negative evidence provided by the likelihood-ratio approach may well be specifically relevant for psychology. Many laypeople have strong intuitive ideas about psychology, as opposed to, for example, fingerprint or DNA analyses (Manning et al., 2007; Lilienfeld et al., 2010). To illustrate, despite the importance of nonidentifications, Wells and Lindsay (1980: 776) argued that 'neither jurors and judges nor police investigators place much faith in the eyewitness who says "this is not the man"'. As to the diagnostic power of foil (i.e. filler) identifications, Clark and Wells (2008: 419) argued that it might be 'particularly easy for the casual observer to dismiss the diagnostic value of foil identifications because they are known immediately to be mistaken identifications; the eyewitness has made a mistake so the witness must have a bad memory. At one level this is true – the witness had made a mistake. But the nature of the mistake suggests that the eyewitness is saying "Your suspect in this lineup looks less like the perpetrator than does this foil," which should logically reduce the fact-finder's confidence that the suspect is the perpetrator.'

As is true for nonidentifications in a line-up, denials by suspects during police interrogations have exonerating value. Even though denials might easily be dismissed as strategic behavior, data by Russano et al. (2005) indicate that denying during interrogation when confronted with promises and minimisation is four times more likely (57%) in case of innocence than in case of guilt (13%), yielding a likelihood ratio of 0.23.

In sum, then, framing psychological expert opinion in terms of a likelihood ratio is a step forward in balance, robustness and transparency. It forces the expert to consult the scientific literature and to collect data rather than to base opinions on own personal insights or preferences. Hence, the likelihood-ratio format decreases the intrusion of subjectivity into the work of expert witnesses and in the interpretation thereof by commissioning parties.

The current studies

In addition to the concerns about the likelihood-ratio paradigm already mentioned, a problem is that readers of expert witness reports may not appreciate or even understand likelihood ratios well. An error is to mistake the likelihood for the probability that the suspect is guilty. This is known as the *transposed conditional* or the *prosecutor's fallacy*, even though prosecutors are obviously not the only ones who are susceptible to this fallacy (Thompson and Schumann, 1987). Another mistake documented in the literature is the *weak evidence effect*. That is, while likelihood ratios between 2 and 10 are verbally referred to as weak evidence for the primary hypothesis, it is important to realise that they do support the primary hypothesis. However, people have been found to believe that *weak evidence* actually implies that the evidence fits better in the alternative hypothesis (Martire et al., 2013).

A decade ago, de Keijser and Elffers (2012) explored understanding of forensic reports, framed in likelihood ratios, in a sample of 118 Dutch professional criminal trial judges. Participants were presented with a fictitious forensic expert witness report including a likelihood ratio. Participants rated their understanding of the report on a scale from one through seven. The mean score on this self-reported understanding was 5.28 ($SD = 1.47$). Furthermore, participants were given eight *true/false* items that tested their actual understanding. The mean score on this test was only 4.25 ($SD = 1.17$), indicating that participants had only half of the answers correct. The authors also tested whether a visual representation of the likelihood ratio, similar to the one in Figure 1, would increase understanding of the report, compared with a verbal presentation (e.g. 'The findings based on the selected visual materials of the facial comparison reported here are *much more likely* when the person depicted is one and the same person (hypothesis 1) than when they are different persons (hypothesis 2)' (de Keijser and Elffers, 2012: 196)), but they failed to find any incremental effect of such presentation.

In this contribution, we argue that likelihood ratios are a fruitful format to present psychological analyses in court. To support this claim, we apply a multimethod approach showing the advantage of likelihood ratios in psychological expert opinion. We first present two cases in which one of us included likelihood ratios in a forensic report, to illustrate how likelihood ratios can be used to give expert opinion. We then describe a first study that explored how likelihood ratios are received by professional judges. Based on previous research (de Keijser and Elffers, 2012; Eldridge, 2019), we expected that judges may dislike likelihood-ratio formats, and may prefer one-dimensional presentations. Finally, a second study is described in which the understanding of an expert witness report including likelihood ratios was explored. Again, based on previous research (de Keijser and Elffers, 2012; Eldridge, 2019), we anticipated that judges may not understand such reports sufficiently and prefer one-dimensional presentations.

Case I: Eyewitness identification

A few years ago, a defence attorney asked one of the authors to give an opinion on the strength of the eyewitness identification evidence against his client. The client was a prime suspect in an attempted murder case. At the time of the crime, the victim was walking in the street when he saw a car (Audi) approaching rapidly. He saw a gun pointed at him through an open car window. Although the victim was shot at repeatedly, he survived the attack. He claimed to have seen who pointed the gun and shot at him. He knew this person and spontaneously mentioned the nickname of the man who henceforth

became the suspect. The defence attorney asked for a (psychological) opinion on the strength of the identification evidence against his client.

Obviously, the strength, that is, diagnostic accuracy of an identification, is modulated by several factors. Many of these factors operate outside the realm of the police. This is also true for this case, where there were many so-called estimator variables beyond the control of the criminal justice system, such as the distance from which the victim saw the shooter, the illumination (the assault took place at night), emotional state of the victim, expectancy effects, and the exposure time that may have affected the victim's perception and recollection of the event. To estimate the strength of the eyewitness identification, the expert referred to a study by de Jong et al. in which the effect of distance and illumination on recognition of a familiar face was tested (de Jong et al., 2005). From the case file, it could be derived that the illumination at the time of the incident was around 5 lux (moderately illuminated city street at night), and the smallest distance between the witness and the perpetrator was at least 5 metres.

In their paper, de Jong et al. (2005) presented tables in which the likelihood of recognition under the primary hypothesis (the person the witness thinks to have seen, is in fact that person; i.e. a correct identification or true positive) and under the alternative hypothesis (the person seen by the witness is not who the witness thinks he is; i.e. a false identification or false positive) are listed. In short, the authors present true and false positives as a function of distance and illumination. Given a 5 metres distance and 5 lux illumination, there is a likelihood of 94% of recognising someone we know. Under these conditions, there is a likelihood of 14% to (mis)recognise someone who is, in fact, somebody else merely resembling our acquaintance. These two likelihoods can be used to calculate the likelihood ratio, which is $(94\% / 14\%) = 6.7$.

The expert witness report concluded the following:

The likelihood ratio of the identification, given the perceptual circumstances, is approximately 6.7 (...). However, there were additional circumstances that threaten the validity of the identification, that are not included in the estimation of this likelihood ratio, such as time pressure, stress, and movement of the perpetrator (...). Therefore, the precise likelihood ratio may well be smaller than 6.7 (...). It is impossible to give a more precise estimate.

The district's attorney responded as follows to the upper bound estimate of the strength of the eyewitness identification in terms of a likelihood ratio:

The expert's computation is based on the case information provided by the defense. Some information, supporting the identification, was lacking. For example, the expert was not informed about the important forensic finding that the suspect's DNA was secured on the headrest of the Audi. In our view, a correct computation of the likelihood ratio of the identification should have considered this important finding. Hence, by supplying the expert with limited information, the defense affected the computation. We cannot know to what likelihood ratio the expert would have concluded, had he been given information about the DNA evidence. Conclusion: the report has very limited value since it was based on incomplete case information.

Obviously, the likelihood ratio of eyewitness identification is *not* affected by the presence of other evidence. Only when deciding on the guilt of the defendant, the judge (or jury) will have to consider all the evidence conjointly. The district's attorney's argument would have probably more easily been recognisable as irrelevant, had he reversely argued that the DNA evidence should be considered to be stronger than the DNA expert had presented, merely because an eyewitness also identified the suspect. Of course, the strength of a DNA finding is not affected by whether or not the suspect was identified by an eyewitness, nor by any other evidence. Logically, this also works the other way around. Eventually, the suspect was convicted. The case illustrates that some legal professionals may have difficulty in understanding the idea behind likelihood ratios.

Case 2: Mister Big and a case of disputed confession evidence

In another case, one of us was asked to give an expert opinion on the validity of a confession. In this case, the suspect had confessed to a murder that had taken place almost twenty years ago. The victim was a man who, at night, accidentally encountered two people engaged in a drug deal. The drug dealers were frightened, shot the victim, laid him in the trunk of a stolen car, and set it on fire. Although the suspect was under investigation almost immediately, he was never interrogated custodially. Instead, years later, the police set up a long-lasting undercover operation in which the suspect was approached by two undercover agents pretending to be drug dealers. After several months of befriending, the suspect was given the impression that a lucrative drug deal was about to take place with which he could earn 75,000 euros. However, his new partners in crime required him to first come clean regarding his alleged involvement in the old murder. They created the illusion that if he confessed to them, they would bother to try to steer the police investigation away from him. This was allegedly possible due to their contacts inside the police. They also lulled the suspect into believing that he would only be included in the drug deal if he confessed. At that point, the suspect admitted his involvement in the murder. This type of undercover operation is known as ‘Mister Big’, referring to the fictitious boss of a fictitious criminal organisation who requires that the suspect confesses old sins and crimes, before the suspect is included in the organisation (Smith et al., 2009).

The expert witness referred to laboratory studies on false confessions. Particularly, he discussed Russano et al. (2005), who conducted a study in which participants were encouraged/pressed to confess to a transgression (i.e. having cheated on a pen-and-paper task) that either took place (guilty) or not (innocent). While questioning participants about the transgression, the researcher would, in some instances, include subtle tricks such as making promises (e.g. stating that ‘things could probably be settled pretty quickly’ if the suspect confessed), or minimising the severity of the to-be-confessed crime (‘I’m sure you didn’t realise what a big deal it was’). The authors found that in a short trickless interrogation, 46% of the participants who had actually cheated, confessed, but so did 6% of the participants who had not cheated. Taken together, these data suggest that in this study, the confession had a likelihood ratio of approximately $(46\% / 6\% =) 8$, meaning that a confession is eight times more likely to occur in case of guilt than of innocence. However, if a promise and minimisation were applied, no less than 87% of guilty participants confessed, but so did 43% of the innocent ones, yielding a likelihood ratio of about $(87\% / 43\% =) 2$, meaning that under these circumstances, a confession is only two times more likely to occur in case of guilt than of innocence. Given that in the current case, both promises and minimisation (i.e. the undercover agents repeatedly assured the suspect that they did not care whether he had been involved in the murder, nor would they think less of him if it was the case) were applied, the expert concluded that the likelihood ratio of the confession was quite limited.

Eventually, the court decided that the suspect’s confession could not be used as evidence, but the suspect was nonetheless convicted based on other evidence in the casefile.

Study 1: Professional judges’ appreciation of likelihood ratios

One of the problems with the application of likelihood ratios to the forensic domain is that the expert opinion, framed in likelihood-ratio format, does not answer the judge’s question. The judge seeks to answer the question whether or not the witness statement is true and hence the suspect is the perpetrator or not. The expert relying on likelihood ratios, however, will conclude that *if* the suspect is the perpetrator (or not), there is a certain likelihood that this witness statement will come about. Hopefully, the reader will by now appreciate that the mismatch between the judge’s and the expert’s quest might be perceived as a problem, but is in fact to be cherished.

The current study set out to explore to what extent professional judges appreciate a psychological expert opinion framed in likelihood-ratio format.

Method

Participants. Thirty-nine Dutch professional criminal trial judges (25 women, 66%) participated in this study. The mean age in the sample was 37.43 years ($SD = 7.04$). The judges took part in a psychology course that they were obliged to attend as a part of their continual professional education. At the onset of this course, they completed several assignments of which the current data were a part. Participants completed the assignment individually. At the time of data completion, participants were not informed about likelihood ratios by us, other than in the instruction (see below). Participants were asked permission to use the learning materials for scientific study. Those who agreed, handed in their booklet.

Materials and procedure. Participants were given the following written instruction.

Imagine that a witness testifies to have seen the perpetrator of a crime. The witness claims to have recognised the perpetrator as an individual with whom he is acquainted. This identification is a crucial piece of evidence in the case. However, the evidentiary value of the identification is put to question, and a psychologist writes an expert opinion about it. Let us assume that each of the following possible conclusions in the expert witness report is correct (or at least scientifically defensible). Please rate the extent to which you appreciate each conclusion.

1. The validity of the identification is subject to multicausality. Factors of interest are illumination, distance, and many others. Due to this multicausality, it is impossible to estimate the strength of the identification scientifically.
2. There is reason to argue that the validity of this identification is strong. Research suggests that the likelihood of a true identification under the circumstances as in the current case, evident from the case file (i.e. 5 meters distance and 5 Lux illumination), is approximately 94%.
3. There is reason to argue that the validity of this identification is weak. Research suggests that the likelihood of a false identification under the circumstances as in the current case, evident from the case file (i.e. 5 meters distance and 5 Lux illumination), is approximately 14%.
4. There is reason to argue that the validity of this identification is modest. Research suggests that the likelihood of a true identification under the circumstances as in the current case, evident from the case file (i.e. 5 meters distance and 5 Lux illumination), is approximately 94%, but also that under these circumstances, the likelihood of a false identification is approximately 14%. This implies that the positive identification fits approximately seven times better in the primary hypothesis (suspect is the perpetrator) than in an alternative hypothesis (suspect is innocent).

Participants rated their appreciation of each of the four conclusions by circling a number on a scale from 0 (*uninformative*) through 100 (*extremely informative*) with increments of 10. The four conclusions represent an uninformative opinion due to multicausality (1), a one-dimensional format that focuses on the fit of the evidence in the primary hypothesis (2), a one-dimensional format that focuses on the fit of the evidence in the alternative hypothesis (in technical domains, this is referred to as the *random match probability* (RMP)) (3), and a likelihood-ratio format (4).

Results

The mean appreciation ratings are shown in Figure 2. The data were analysed with both frequentist (SPSS) and Bayesian (JASP, free Bayesian software available at www.jasp-stats.org) approaches (see Dienes, 2008). Crucially, the latter analysis yields a Bayes factor which represents the likelihood ratio for the fit of the data in the null and in the alternative hypothesis. BF_{10} s smaller than 1 indicate that the data fit better in the null hypothesis than in the alternative hypothesis. BF_{10} s larger than 1 suggest

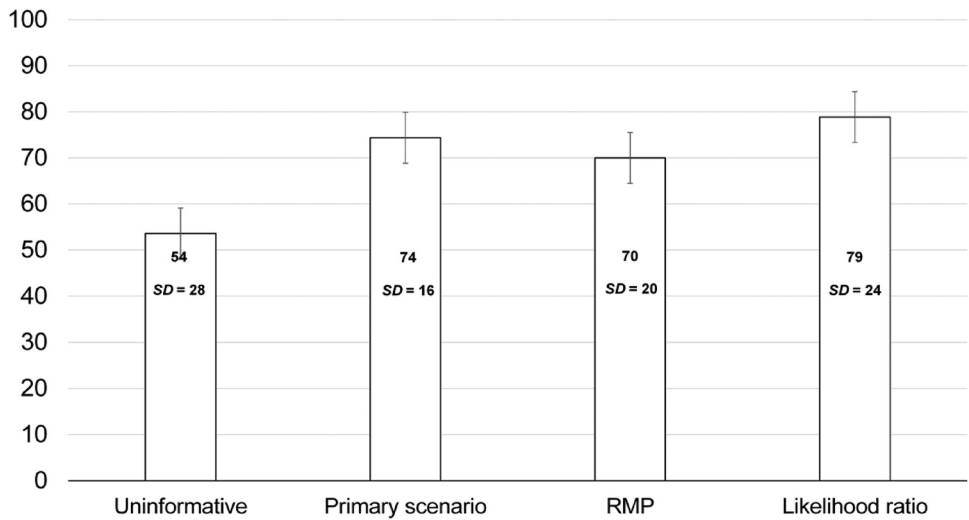


Figure 2. Mean appreciation scores (0–100) for four formats of psychological expert opinion.

that the alternative hypothesis predicts the data better. BF_{10} s larger than 3 can be interpreted as positive/substantial support for the alternative hypothesis. BF_{10} s larger than 10 represent positive/strong support, and BF_{10} s larger than 20 provide strong support for the alternative hypothesis (Jarosz and Wiley, 2014). In the current analyses, the prior odds were left undefined and thus set at 1.0.

A repeated measurements ANOVA showed that the conclusions were rated significantly differently $F(3) = 12.68, p < .001$, partial $\eta^2 = 0.25$. Post-hoc paired sample t -tests indicated that the appreciation for the uninformative format was significantly lower than that for all three other expressions: $t(38) > 3.17, ps < .003, BF_{10} > 11.65$. The one-dimensional primary hypothesis format was appreciated more than the one-dimensional RMP expression ($t[38] = 2.21, p = .033, BF_{10} = 1.51$), but similarly compared to the likelihood-ratio expression ($t[38] = 1.24, p = .224, BF_{10} = 1.64$). Finally, the likelihood-ratio format was appreciated more than the RMP format: $t[38] = 2.26, p = .030, BF_{10} = 1.64$).

Discussion

Apparently, our sample of professional judges was not displeased with the hypothetical psychological expert opinion framed in likelihood-ratio format. It seems that the problem that ‘jurists seem to dislike the use of rival hypotheses and long back to the simple old style formulations’ (de Keijser and Elffers, 2012: 205) can be overcome. It may well be the case that since de Keijser and Elffers published their paper, judges have become more familiar with statistical evidence in the form of likelihood ratios, precisely because technical forensic experts have been using this format from 2009 onwards (AFSP, 2009; ENFSI, 2016). Meanwhile, the current measurement differed vastly from that by de Keijser and Elffers.

An important limitation of the study is the small sample size. Post hoc power analysis (G*power; $d = .50, \alpha = .050$) indicated a power of .86. Also, experience of the participating judges was not registered. Further, the study had a within-subjects design in combination with a lack of counterbalancing. All participants rated their appraisal of all four formats (in the same order), and were thus implicitly given the opportunity to compare the formats. Hence, it might have become evident that the likelihood-ratio format is the most complete, which may have inflated appreciation scores. One might argue that the likelihood ratio expression is by definition most informative, because it contains the most information. Meanwhile, this cannot explain why the primary hypothesis expression was

appreciated equally, yet more than the RMP expression. Finally, the stimulus material was short, and hence not representative of what judges encounter in real cases. This obviously means that generalisations to real-life decision making should be made with caution.

Study 2: Professional judges' understanding of likelihood ratios

Besides the hypothesised lack of appreciation of likelihood ratios (de Keijser and Elffers, 2012; or even statistically framed expert opinion in general, see Eldridge, 2019), a lack of proper understanding might be an obstacle to using likelihood ratios. Not only do readers of forensic reports sometimes make mistakes such as the prosecutor's fallacy (Thompson and Schumann, 1987) and the weak evidence effect (Martire et al., 2013), but likelihood ratios seem to be generally hard to grasp (de Keijser and Elffers, 2012). In this study, we sought to explore to what extent professional judges understand the content (and limitations) of an expert opinion framed in a likelihood ratio. We followed the approach of de Keijser and Elffers (2012), which included a self-reported (supposed) and actual (test) understanding. Whereas de Keijser and Elffers employed a forensic technical expert opinion, and found that judges' understanding was limited, we used a psychological opinion akin to the one described in our first case study. Beforehand, it was expected that judges' understanding of an expert report that relies on likelihood ratios is limited (cf. de Keijser and Elffers, 2012; Eldridge, 2019).

Method

Participants. Seventy-nine Dutch professional criminal trial judges (48 women, 61%) participated in this study. The mean age in the sample was 41.42 years ($SD = 9.68$). The judges took part in a psychology course that they were obliged to attend as a part of their continual professional education. At the start of this course, they completed several assignments, of which the current data were a part. Participants were asked permission to use the learning materials for scientific study. Participants completed the assignment individually.

Measures and procedure. Participants were given a case file summary, loosely based on de Keijser and van Koppen (2007), about a young man, Anton de Koning, who, one evening, walked in the street with his girlfriend Corine de Jong. They encounter three young men: Joesef Abdullah, Sjon Tegelaar and Bas van Vliet, the man who will become the suspect. The latter made a remark about Corine, and Anton swiftly gave a witty reply. The parties follow their own path. Soon afterwards, the three men go separate ways. Next, Anton is attacked and physically mistreated. The police think that the suspect (aged 27 years), after having said goodbye to Joesef and Sjon, has followed Anton and Corine, and attacked Anton from behind, out of revenge for the witty remark that made Bas van Vliet look stupid. The stimulus materials included a one-page case summary, a report about a simultaneous photo-line-up in which Corine de Jong identifies van Vliet, and a report about a simultaneous photo-line-up given to an eyewitness, Alastair Offermans, who testifies that he does not recognise any of the shown photos as the perpetrator he saw.

The case file also included a psychological expert witness report about the positive and negative identification evidence. This report concluded:

Several factors affect the validity of a (non)identification, such as the circumstances during the witnessing, and retention time between the incident and the identification procedure. To date, no scientific analysis is available that includes all possible relevant factors when estimating the validity of a (non)identification. Wagenaar and van der Schrier (1996) studied two relevant factors that play a role in the current case, namely distance and illumination. From the case file, it is evident that for both witnesses, the illumination level was approximately 10 lux (i.e. the equivalent of a well-illuminated street at night). De Jong was in the immediate proximity of the

perpetrator, whereas Offermans stood at seven meters distance. It should be noted that the validity of a (non) identification can best be estimated by considering its fit in the guilt and in the innocence hypotheses. From the data by Wagenaar and van der Schrier, it can be concluded that if de Jong indeed saw the suspect (from close proximity and with 10 lux illumination), it is more likely that she would identify him later on, during the identification procedure, than if she saw someone other. *The true positive likelihood is estimated at 82%, the false positive likelihood is approximately 6%, and the likelihood ratio is therefore $(82\% / 6\% =) 14$.* The data by Wagenaar and van der Schrier further suggest that if Offermans witnessed the incident from seven meters and with 10 lux, and the suspect was not the perpetrator, a nonidentification would be more likely than if the suspect were the perpetrator. *The likelihood of a false negative is 29%, the likelihood of a true negative is 95%, and consequently the likelihood ratio (under the primary hypothesis) is $(29\% / 95\% =) 0.3$.* In this instance, the positive identification by de Jong is likely to be stronger than the nonidentification by Offermans. *The two likelihood ratios can be multiplied, yielding a combined evidentiary strength of $(14 * 0.3 =) 4$.* The combination of both pieces of evidence fits *four times* better in the primary hypothesis (i.e. the suspect is the perpetrator) than in the alternative hypothesis (i.e. the suspect is not the perpetrator).

Randomly, some participants ($n = 41$) received the report as quoted. This version is referred to as the 'numeric presentation' version. Others ($n = 38$) received a version from which the italic phrases were omitted. This version is referred to as the 'verbal presentation' version.

Participants completed a short questionnaire that tapped their understanding of the report. One item pertained to self-reported, supposed understanding, and was answered on a scale from 0 (*not at all*) through 10 (*completely*). Actual understanding was tested with eight statements adopted from de Keijser and Elffers (2012). Participants were asked to indicate whether or not each statement logically followed from the expert witness report (*true/false*). All items are displayed in Table 1.

Results

As shown in Table 1, participants who received the verbal version self-reported their understanding of the report to be better than did those who received the numeric version. However, actual understanding as borne out by the 8-items test revealed a more complex pattern. Some of the participants who were given the numeric version of the expert report (21%) fell prey to the defence attorney's fallacy (item 6; see also Thompson and Schumann, 1987), whereas none of those given the verbal version did. By contrast, whereas only 8% of participants in the verbal expression condition dared to conclude that the combination of (non)identifications construes incriminating evidence against the suspect (it does), 28% of those in the numeric condition did. Importantly, the numeric and verbal versions did not translate into differential performance on the 8-items test total score. Also, self-reported and actual understanding did not correlate ($r = .08$; $p = .508$; Cohen's $d = 0.16$; $BF_{10} = 0.19$).

Discussion

Wagenaar ended his contribution about likelihood ratios pessimistically: 'The strategy of providing likelihoods may not solve all problems of expert testimony ... the courts may interpret testimony given in the proper form as opinions thus abdicating their own responsibility of translating likelihoods into odds. A worse problem is that the experts themselves are not always fully aware of the distinction' (1988: 509). Contrary to our expectations based on Wagenaar's work and that of de Keijser and Elffers (2012), the professional judges had a fair understanding of the expert opinion framed in likelihood-ratio format. Whereas self-reported understanding seemed better in case of the verbal compared with the numeric presentation, performance on a test measuring actual understanding was not affected by the format, with a grand mean accuracy score of 68% (i.e. 5.45 out of 8). This pattern is in line with de Keijser and Elffers (2012), who also failed to find different levels of understanding as a function of presentation

Table 1. Self-reported and actual understanding of the expert witness report.

	Numeric (<i>n</i> = 41)	Verbal (<i>n</i> = 38)	<i>P</i>	<i>BF</i> ₁₀	Whole sample (<i>N</i> = 79)
How well do you understand the expert witness report? (0–10)	5.39 (2.18)	7.41 (1.19)	< .001	4226	6.35 (2.04)
1. It is (<i>four times</i>) ¹ more probable that the suspect is guilty than innocent. (False) ²	63%	74%	.295	0.51	68%
2. It is more probable that the suspect is the perpetrator than that someone else is. (False)	44%	58%	.215	0.60	50%
3. The report constitutes incriminating evidence against the suspect. (True)	28%	8%	.024	4.25	18%
4. The suspect is the perpetrator. (False)	97%	95%	.531	0.70	96%
5. It is concluded that the combination of the two identification outcomes is incriminating. (True)	53%	46%	.569	0.33	49%
6. The two identification outcomes (<i>combined likelihood ratio</i> = 4) ¹ are not the only relevant information. At any night, lots of overexcited youngsters are in the streets, many of whom fit the suspect's description. Hence, the probability that it was indeed the suspect who attacked the victim is small. (False)	79%	100%	.003	34.05	89%
7. The combination of the two identification outcomes fits better in the primary (guilty) hypothesis than in the alternative (innocent) hypothesis. (True)	88%	87%	.932	0.40	87%
8. Someone else than the suspect may be the perpetrator. (True)	90%	81%	.256	0.73	85%
Total of eight statements (0-8)	5.4	5.5	.647	0.27	5.45

Note. ¹ These quantifiers were only included in the numeric version; ² for the eight statements, percentages of correct answers are presented.

format (verbal vs. visual presentation). Likewise, Bali et al. (2021) found that presentation format (e.g. RMP, likelihood ratio or verbal label) did not affect understanding in lay samples (i.e. potential jury members). Meanwhile, it must be acknowledged that the two versions did not only differ in presentation mode, but also in quantity, in that the numeric presentation included more information.

Most importantly, compared with the average score of 4.25 on the 8-items test observed by de Keijser and Elffers (2012), the current mean score of 5.45 seems to be an improvement. A one-sample *t*-test comparing the mean of our judges to those of de Keijser and Elffers yielded a significant result: $t(70) = 9.26, p < .001; BF_{10} > 100,000$, indicating that judges' performance currently is indeed superior to that observed by de Keijser and Elffers in 2012. This increase may be attributed to the fact that de Keijser and Elffers collected their data right after the introduction of the likelihood ratio by the AFSP in 2009, whereas the current data were collected ten years later. Thus, judges may have become more familiar with likelihood ratios in the meantime. Another possibility is that the psychological evaluation of the identification evidence in the current fictitious case might have been perceived as clearer and easier to understand than was the technical evidence used in the study by de Keijser and Elffers. On the other hand, the likelihood-ratio principle remains the same regardless of which domain it is applied to. Moreover, the expert witness report in the present case targeted a combination of two identification outcomes, which is obviously more complex than the single

Table 2. Proposed likelihood ratio ranges (based on Jarosz and Wiley, 2014).

Verbal description	Likelihood ratio
Weak/anecdotal support	2–3
Positive/substantial support	3–10
Positive/strong support	10–30
Strong/very strong support	30–100
Very strong/decisive support	100 <

piece of evidence employed in the de Keijser and Elffers study. Either way, the current findings point at a fair understanding of the likelihood-ratio format.

Strikingly, the performance on the eight items differed vastly, with only 18% of participants answering number 3 correctly, and 96% number 4. Currently, we have no explanation for the wideness of this range. Finally, it is remarkable that supposed understanding did not correlate with actual understanding. Similarly, de Keijser and Elffers (2012) found weak to no correlations in their professional samples. Apparently, it is difficult to reflect on one’s insight in likelihood ratios.

General discussion

The considerations and data presented above underline two points. First, psychology as a forensic science should consider adhering to international agreements (e.g. AFSP, 2009; ENFSI, 2016) and start using likelihood ratios as a format of choice in expert reports whenever relevant scientific data are available. Second, the problem of readers of expert reports (particularly professional judges) not appreciating or even understanding likelihood ratios may not be insurmountable.

As to the first point, we presented two case studies illustrating the use of likelihood ratios in practice. In both cases, the likelihood ratios were copied from relevant scientific research, and their presentation was accompanied by the warning that precise estimates are virtually impossible in psychology. Recently, Bali et al. (2021: 370) argued: ‘Forensic science evidence...is inherently uncertain... As scientific evidence can be highly persuasive, practitioners have a responsibility to communicate this uncertainty clearly and accurately to lay and legal decision-makers... To this end, scholars and regulators encourage forensic practitioners to use statistical statements in their reports and testimony’. Although this quote pertains to technical forensic sciences, we argue that it is also relevant for psychological expertise.

As to the second issue, the extant literature on the understanding of statistical expert evidence has been limited to technical evidence and has largely ignored psychological and other non-technical domains of expertise. Furthermore, this literature has mainly focused on laypersons (potential jury members). Hence, while researchers have studied understanding of statistical expert evidence in laypeople ‘for decades’ (Bali et al., 2021: 371; Eldridge, 2019), little is known about professional judges’ understanding, which is crucial in inquisitorial systems. De Keijser and Elffers (2012) documented ten years ago that professional judges had limited understanding of likelihood ratios. Our data point towards a slightly more optimistic picture. Judges had a fair understanding of the likelihood-ratio report (Study 2), and appreciated it no less than they did other (one-dimensional) formats (Study 1). Our findings are in line with data obtained in lay samples. For example, Bali et al. (2021) reported that few (lay) participants displayed perfect comprehension, but the majority was able to make correct inferences based on expert reports, including likelihood ratios, RMPs or verbal labels.

A problem with applying the likelihood-ratio paradigm to psychological expertise that was not addressed in the present studies, but does deserve attention, is that likelihood ratios tend to be modest in size. For example, Wagenaar (1988) mentioned a value of 1.82 derived from his tailor-made study. The values in our case studies were 6.7 and 2. Wagenaar and van der Schrier (1996) argued that an

eyewitness identification with a likelihood ratio of 15 or more is to be considered strong evidence. In fact, psychological science rarely produces likelihood ratios that pass the threshold of 100 and would thus, most of the times at best be considered *moderately supportive* of the primary hypothesis, if one uses the verbal terms associated with likelihood ratio ranges (AFSP, 2009). This has much to do with multicausality. Hence, a behavior rarely fits perfectly in the primary hypothesis, and thus the numerator is rarely 100%. Likewise, the likelihood of a false positive (cf, RMP) is seldomly close to 0%, and therefore, the denominator is frequently relatively high, resulting in modest likelihood ratios. Notably, recent research suggests that technical forensic analyses at the activity level may also suffer from multicausality and may consequently not yield likelihood ratios of a similar magnitude as those of identity level analyses. Kokshoorn et al. (2017) presented an example suggesting that a full DNA-match may, at the activity level, yield a likelihood ratio of no more than 91. Mixed profiles may even yield likelihood ratios not exceeding 9. Perhaps, adjusted norms and verbal descriptions are needed for psychological expert opinion, other non-technical opinion and technical opinion at the activity level. One option may be to adopt the widely used ranges in science, as presented in Table 2 (Jarosz and Wiley, 2014).

As to the implication of our argument for legal decision making in general, we want to stress that we do not argue that criminal cases should be approached with a complete Bayesian analysis. In our view, and that of many others, a likelihood-ratio analysis of one or even several pieces of evidence does not imply a full Bayesian approach including the estimation of prior odds, even though likelihood ratios also occur in Bayesian analyses (de Keijser and Elffers, 2012; Dienes, 2008; Fenton et al., 2016; Royall, 1997). However, we do advocate that judges (and juries) adopt a scenario approach in that they do not simply accumulate evidence against the suspect, but analyse every piece of evidence as to its fit in competing scenario's, even without trying to actually quantify these fits with likelihoods. In the domain of intelligence analyses, Heuer (1999) termed this approach an Analysis of Competing Hypotheses (ACH). It has also been forwarded as a fruitful way of legal decision making (van Koppen and Mackor, 2019).

An important topic for future research is whether the use of likelihood ratios actually yields the hypothesised protection against various biases, such as confirmation bias (Kassin et al., 2013) and allegiance bias (Murrie et al., 2013).

Both Wagenaar (1988) and de Keijser and Elffers (2012) were rather pessimistic about legal decision-makers' appreciation and interpretation of likelihood ratios. Let us express the hope that we have moved forward in the right direction over the past decades. At a minimum, the current data indicate that professional judges are open to receiving psychological expert opinion in likelihood-ratio format. Hence, our colleagues should be further encouraged to include the likelihood-ratio rationale in their reports.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Eric Rassin  <https://orcid.org/0000-0002-0830-7758>

Irena Boskovic  <https://orcid.org/0000-0003-1675-1004>

References

Association of Forensic Science Providers (AFSP) (2009) Standards for the formulation of evaluative forensic science expert opinion. *Science and Justice* 49(3): 161–164.

- Bali AS, Martire KA and Edmond G (2021) Lay comprehension of statistical evidence: a novel measurement approach. *Law and Human Behavior* 45(4): 370–390.
- Clark SE and Wells GL (2008) On the diagnosticity of multiple-witness identifications. *Law and Human Behavior* 32(5): 406–422.
- De Jong M, Wagenaar WA, Wolters G, et al. (2005) Familiar face recognition as a function of distance and illumination: a practical tool for use in the courtroom. *Psychology, Crime & Law* 11(1): 87–97.
- De Keijser J and Elffers H (2012) Understanding of forensic expert reports by judges, defense lawyers and forensic professionals. *Psychology, Crime and Law* 18(2): 191–207.
- De Keijser JW and van Koppen PJ (2007) Paradoxes of proof and punishment: psychological pitfalls in judicial decision making. *Legal and Criminological Psychology* 12(2): 189–205.
- Dienes Z (2008) *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inferences*. New York: Palgrave Macmillan.
- Eldridge H (2019) Juror comprehension of forensic expert testimony: a literature review and gap analysis. *Forensic Science International: Synergy* 1: 24–34.
- European Network of Forensic Science Institutes (ENFSI) (2016) *Strengthening the evaluation of forensic results across Europe (STEOFRAE)*. <https://enfsi.eu>.
- Fenton N, Neil M and Berger D (2016) Bayes And the law. *Annual Review of Statistics and its Application* 3: 51–77.
- Heuer RJ (1999) *Psychology of Intelligence Analyses*. Washington, DC: CIA.
- Jarosz AF and Wiley J (2014) What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving* 7: 2–9.
- Kassin SM, Dror IE and Kukucka J (2013) The forensic confirmation bias: problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition* 2(1): 42–52.
- Kokshoorn B, Blankers BJ, de Zoete J, et al. (2017) Activity level DNA evidence evaluation: on propositions addressing the actor of the activity. *Forensic Science International* 278: 115–124.
- Larrabee GJ (2008) Aggregation across multiple indicators improves the detection of malingering: relationship to likelihood ratios. *The Clinical Neuropsychologist* 22(4): 666–679.
- Liebman JS, Blackburn S, Mattern D, et al. (2012–2013). The evidence of things not seen: non-matches as evidence of innocence. *Iowa Law Review* 98: 577–688.
- Lilienfeld SO, Lynn SJ, Ruscio J, et al. (2010) *50 Great Myths of Popular Psychology: Shattering Widespread Misconceptions About Human Behavior*. Chichester: Wiley-Blackwell.
- Manning R, Levine M and Collins A (2007) The Kitty Genovese murder and the social psychology of helping. *American Psychologist* 62(6): 555–562.
- Martire KA, Kemp RI, Watkins I, et al. (2013) The expression and interpretation of uncertain forensic science evidence: verbal equivalence, evidence strength, and the weak evidence effect. *Law and Human Behavior* 37(3): 197–207.
- Murrie DC, Boccaccini MT, Guarnera LA, et al. (2013) Are forensic experts biased by the side that retained them? *Psychological Science* 24(10): 1889–1897.
- Oberlader VA, Naefgen C, Koppehele-Gossel J, et al. (2016) Validity of content-based techniques to distinguish true and fabricated statements: a meta-analysis. *Law and Human Behavior* 40(4): 440–457.
- Royall R (1997) *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.
- Russano MB, Meissner CA, Narchet FM, et al. (2005) Investigating true and false confessions within a novel experimental paradigm. *Psychological Science* 16(6): 481–486.
- Saks MJ and Koehler JJ (2005) The coming paradigm shift in forensic identification science. *Science (New York, N.Y.)* 309(5736): 892–895.

- Smith SM, Stinson V and Patry MW (2009) Using the 'Mr Big' technique to elicit confessions: successful innovation or dangerous development in the Canadian legal system? *Psychology, Public Policy, and Law* 15(3): 168–193.
- Thompson WC (2018) How should forensic scientists present source conclusions? *Seton Hall Law Review* 48: 773–813.
- Thompson WC and Schumann EL (1987) Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behaviour* 11(3): 167–187.
- Thompson WC, Vuille J, Taroni F, et al. (2018) After uniqueness: the evolution of forensic science. *Judicature* 102(1): 18–27.
- Van Koppen PJ and Mackor AR (2020) A scenario-approach to the simonshaven case. *Topics in Cognitive Science* 12(4): 1132–1151.
- Wagenaar WA (1988) The proper seat: a Bayesian discussion of the position of expert witnesses. *Law and Human Behavior* 12(4): 499–510.
- Wagenaar WA and Van der Schrier JH (1996) Face recognition as a function of distance and illumination: a practical tool for use in the courtroom. *Psychology, Crime & Law* 2(4): 321–332.
- Walker L and Monahan J (1987) Social frameworks: a new use of social science in law. *Virginia Law Review* 73: 559–598.
- Wells GL, Bull Kovera M, Douglass AB, et al. (2020) Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior* 44(1): 3–36.
- Wells GL and Lindsay RCL (1980) On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin* 88: 776–784.
- Wells GL and Olson EA (2002) Eyewitness identification: information gain from incriminating and exonerating behaviours. *Journal of Experimental Psychology: Applied* 8(3): 155–167.