BRIEF REPORT

Does DNA evidence in the form of a likelihood ratio affect perceivers' sensitivity to the strength of a suspect's alibi?

Gianni Ribeiro[1] · Jason Tangen[1] · Blake McKimmie[1]

Abstract

Forensic expert testimony is beginning to reflect the uncertain nature of forensic science. Many academics and forensic practitioners suggest that forensic disciplines ought to adopt a likelihood ratio approach, but this approach fails to communicate the possibility of false positive errors, such as contamination or mislabeling of samples. In two preregistered experiments ($N_1 = 591$, $N_2 = 584$), we investigated whether participants would be convinced by a strong DNA likelihood ratio (5,500 in Experiment 1 and 5,500,000 in Experiment 2) in the presence of varying alibi strengths. Those who received a likelihood ratio often concluded that the suspect was the source of the DNA evidence and guilty of the crime compared with those who did not receive a likelihood ratio—but they also tended to conclude that an error may have occurred during DNA analysis. Furthermore, as the strength of the suspect's alibi increased, people were less likely to regard the suspect as the source of the evidence or guilty of the crime, and were more likely to conclude that an error may have occurred during DNA analysis. However, people who received a likelihood ratio were actually *more* sensitive to the strength of the suspect's alibi than those who did not, driven largely by the low ratings in the strongest alibi. Interestingly, the same pattern of results held across both experiments despite the likelihood ratio increasing by two orders of magnitude, revealing that people are not sensitive to the value of the likelihood ratio.

Keywords Forensic evidence · Juror decision-making · Likelihood ratio · Alibi

For more than a century, forensic examiners were permitted to testify that two prints, samples, or marks *match* to the exclusion of all other possible sources. However, numerous misidentifications and exonerations on the basis of faulty or misleading forensic science have demonstrated that forensic comparison evidence is fallible (National Registry of Exonerations, 2019; Saks & Koehler, 2005). One of the most highly publicized cases of forensic misidentification is that of Brandon Mayfield, an American lawyer who was wrongly accused of orchestrating the 2004 Madrid train bombings on the basis of a fingerprint that the FBI determined to have originated from Mayfield (Thompson & Cole, 2005).

Interestingly, Mayfield had a strong alibi; he had no record of ever traveling to Spain, had not left his state of Oregon in 2 years, and had not been abroad since 1993 (Gumbel, 2004).

In light of these numerous high-profile misidentifications and considerable scrutiny, forensic testimony is beginning to reflect the uncertain nature of forensic science. Several authoritative reports (National Academy of Sciences, 2009; President's Council of Advisors on Science and Technology, 2016) have called for research to establish evidence-based standards for reporting forensic analyses to ensure that examiners' testimony is scientifically sound, including that examiners "should always state that errors can and do occur" (President's Council of Advisors on Science and Technology, 2016, p. 19). Despite these calls for evidence-based testimony, there is uncertainty about what this new model should look like.

Many researchers and forensic governing bodies alike have welcomed the use of likelihood ratios, suggesting that they are "the logically correct framework for the evaluation of evidence" (Morrison, 2016, p. 371). Likelihood ratios communicate the probability of an observation given two competing hypotheses: that the two samples share the same source (H1),

✉ Gianni Ribeiro
g.ribeiro@uq.edu.au

[1] School of Psychology, The University of Queensland, St Lucia, Australia

or that the two samples originate from different sources (H2). However, a considerable body of work has demonstrated that people are simply not good at understanding and evaluating probabilities. When asked to evaluate what the statement "a 30% chance of rain tomorrow" means, the majority of pedestrians surveyed came to an incorrect interpretation, concluding that it would rain tomorrow in 30% of the *area* or for 30% of the *time* (Gigerenzer, Hertwig, van den Broek, Fasolo, & Katsikopoulos, 2005). When told "if you take the medication, you have a 30%–50% chance of developing a sexual problem" before being prescribed fluoxetine for depression, many patients incorrectly understood that a sexual problem would occur in 30%–50% of their own sexual encounters (Gigerenzer, 2002). Finally, in the legal context, when forensic psychologists and psychiatrists were asked to judge the likelihood that a patient with a mental disorder would harm someone within 6 months of hospital discharge, their responses were heavily dependent on the format of the response scale, judging patients as less likely to be harmful when the probabilistic response scale was small (1% to 40%) compared with large (1% to 100%; Slovic, Monahan, & McGregor, 2000). But these incorrect interpretations of probabilistic statements cannot be attributed to participants' (in)numeracy (Gigerenzer & Galesic, 2012); the problem lies in the way that the information is communicated.

Here, we propose that because likelihood ratios are difficult to understand and interpret (Martire, Kemp, Sayle, & Newell, 2014; Martire, Kemp, Watkins, Sayle, & Newell, 2013), large likelihood ratios, which are typically encountered in DNA evidence, may be incorrectly interpreted and serve as a simple judgment heuristic of guilt, even in the face of strong exculpatory evidence such as a compelling alibi. This is often referred to as the "prosecutor's fallacy" or the fallacy of the transposed conditional, whereby fact-finders may incorrectly interpret a likelihood ratio as being the probability that the defendant is guilty of the crime. For example, (incorrectly) concluding that a likelihood ratio of 5,500 means that the odds of the defendant being guilty are 5,500:1. This fallacy has been demonstrated experimentally (Thompson & Newman, 2015) and in the context of a real trial. For example, in the Australian case, *R v. Keir* (2002), a trial Judge did exactly that. The prosecution argued that bone fragments found in the defendant's backyard belong to his missing wife, calling forth a forensic expert who stated that it was 660,000 times more likely that the DNA profile found in the bone could have been obtained if it came from a child of the missing woman's parents than from a child of a random mating in the Australian population. However, in summing up, the trial Judge incorrectly stated that the evidence showed that there was a 660,000:1 chance that the bones belonged to the missing woman, and thus also a 660,000:1 chance that alleged sightings of the woman after her disappearance were false (Australian Law Reform Commission, 2010). An appellate

court found that the prosecutor's fallacy had taken place, and a new trial was ordered.

As others have suggested (see, for example, Koehler, 1997; Thompson, Taroni, & Aitken, 2003), likelihood ratios could also be problematic as they do not—but *should*—take into account the possibility of false positive errors. An error that results in a false positive, such as mislabeling of samples or contamination, affect the *validity* of the likelihood ratio. The President's Council of Advisors on Science and Technology (2016) report notes that the possibility of a false positive error is far more likely than the possibility of two DNA samples from two different sources sharing the same profile (p. 7). As the likelihood ratio quantifies the strength of the hypothesis that two samples share the same source against the strength of the hypothesis that two samples come from different sources, false positives are not accounted for. Indeed, others have shared this view, arguing that communicating precise numbers to indicate the strength of evidence is "unnecessary due to the always-greater probability of an error somewhere in the obtaining, handling, or evaluation of the sample" (Spellman, 2018, p. 830).

Taken together, if fact-finders mistakenly infer that a likelihood ratio communicates the odds that the defendant is guilty *and* receive no information regarding the human nature of forensic decision-making and the chance of a false positive error having occurred, fact-finders may be overwhelmingly convinced by the DNA evidence and disregard other exculpatory evidence, such as a strong alibi.

Prior research by Scurich and John (2011) seems to suggest that this may be likely. In the first phase of the experiment, participants were told only about the presence of nongenetic evidence. In the strong nongenetic evidence condition, participants were told that the defendant did not have an alibi, had a moustache, which was consistent with the perpetrator, and that witnesses had seen the defendant driving a truck made by the same manufacturer as the one in which the crime had taken place. In the weak nongenetic evidence condition, participants were told that the defendant had a partial alibi, which was corroborated with his employer, no witnesses could recall him ever having a moustache, and police could not find any witnesses who had seen him driving a truck. Participants were asked to decide whether or not they would convict the defendant, and provide a judgment regarding the probability that the defendant committed the crime (guilt likelihood), but they were not yet told about any DNA evidence. For participants who heard the strong evidence, 29.8% were willing to convict the defendant and the average guilt likelihood judgment was 0.47, whereas for participants who heard the weak evidence, only 3% were willing to convict the defendant, and the average guilt likelihood was 0.16.

All participants were then provided with information about a DNA test, which revealed that there was a random match probability of 1 in 200 million, meaning 1 in 200 million

people randomly sampled from the population would match the evidence sample. Participants were then asked again whether or not they would convict the defendant and to provide an estimate of guilt likelihood. Unsurprisingly, the number of participants presented with strong nongenetic evidence who chose to convict increased to 72.7% and the estimated guilt likelihood increased to 0.8. However, these judgments also increased for those presented with weak nongenetic evidence, with 26.8% of participants choosing to convict and an increased estimated guilt likelihood of 0.61. These results clearly show that, even when a partial but verified alibi and other evidence pointing towards evidence is provided, fact-finders may still be overly convinced by the presence of DNA evidence. Although this study looked at random match probabilities, we see no reason why these results should differ for likelihood ratios, which are also difficult to interpret and prone to the prosecutor's fallacy.

In this paper, we investigate how mock jurors evaluate a strong DNA likelihood ratio as a function of the strength of a suspect's alibi. We are particularly interested in seeing whether or not participants are convinced by a strong likelihood ratio when the defendant's alibi is strong—where one could be very convinced that the suspect could not have committed the crime and thus the incriminating DNA evidence is likely the result of a false positive error, such as through sample mislabeling or contamination. We predict that our participants would be more likely to conclude that the DNA evidence comes from the suspect and that the suspect is guilty of the crime, but less likely to conclude that an error could have occurred during DNA analysis, when a likelihood ratio is present (rather than absent) and as the strength of the suspect's alibi decreases. Overall, we expect that, in all conditions where a likelihood ratio is presented, participants will be overly convinced by the DNA likelihood ratio and provide high ratings of source and guilt likelihood regardless of the strength of the suspect's alibi. Thus, participants not presented with a likelihood ratio will be more sensitive to the strength of the suspect's alibi.

Method

Participants and design

Our preregistered predictions for both experiments are available on the Open Science Framework, along with all materials, deidentified data, and analysis scripts (Experiment 1: https://osf.io/yngxk/, Experiment 2: https://osf.io/2hmvz/). We recruited 600 participants for each experiment consisting of a 2 (likelihood ratio: absent vs. present) × 6 (alibi strength: 6 levels) between-groups design. Sample size was determined a priori as having 95% power to detect a small-to-medium effect ($f = 0.2$) at an alpha level of .05. We recruited participants

residing in England or Australia, due to similarities in legal systems and spelling, from the Prolific online testing platform, who agreed to receive £1.00 for their participation.

Following our preregistered exclusion plan, we excluded nine participants in Experiment 1 and excluded 16 participants in Experiment 2. The final sample for Experiment 1 ($N = 591$) consisted of 391 women, 199 men, and one other with an age range of 18–72 years ($M = 38.65$, $SD = 11.84$), and the final sample for Experiment 2 ($N = 584$) consisted of 383 women and 201 men with an age range of 18–77 years ($M = 37.57$, $SD = 11.89$). The majority of participants in both Experiment 1 and 2 reported residing in England (99.7% and 99.5%, respectively). Other participant information, such as highest level of education and prior jury service, are presented in the supplemental materials.

Procedure and materials

Participants completed the experiment on their own devices. They started by reading a short case vignette detailing the facts of a murder. They were told that a DNA sample was retrieved from underneath the victim's fingernails and that police have interviewed various suspects, including a man with the initials B.M., loosely based on the Mayfield case. Participants randomly assigned to the likelihood-ratio-present conditions received information about the comparison between the DNA sample retrieved from the victim's fingernails and B.M.'s DNA sample, including a likelihood ratio of 5,500 (Experiment 1) or 5,500,000 (Experiment 2). We selected the value of 5,500 as it falls in the middle of the 'strong' category and 5,500,000 as it exceeds the strongest category of 'extremely strong' according to the Association of Forensic Science Providers (2009) standards for numerical and verbal expression of likelihood ratios. Participants in the likelihood-ratio-absent conditions were simply told that the DNA sample retrieved from the victim's fingernails is currently being analyzed by a forensic DNA examiner.

Finally, participants were randomly assigned to read one of six alibis ranging in strength. The four weakest alibis were adapted from Olson and Wells (2004), and the two strongest alibis were created by the first author. All six alibis were pilot tested with 26 participants to ensure they represented a range of strengths. Participants in the pilot test were asked "How likely is it that the suspect could have committed the crime?" on an 11-point scale from 0 = *definitely could not have committed the crime* to 10 = *definitely could have committed the crime*. Responses ranged from $M = 0.63$ for the weakest alibi to $M = 8.39$ for the strongest alibi, with responses for all other alibis increasing in a linear fashion.

In the current study, participants who were randomly assigned to the weakest alibi condition were told that initially the suspect could not remember where he was on the evening in question, but later in the interview claimed that he had been

out for a walk in his neighbourhood. Participants in the strongest alibi condition were told that the suspect claimed he was overseas for 1 week, including the evening in question, for his friend's wedding, in which he was a groomsman. They were also told that border security records confirmed he was out of the country, and that the newlyweds provided time-stamped images of the suspect in the bridal party.

After reading the scenario, participants were asked to respond to three key dependent variables: source likelihood ("On a scale of 0–100, how likely is it that the DNA evidence retrieved from underneath the victim's fingernails belongs to B.M.?"), guilt likelihood ("On a scale of 0–100, how likely is it that B.M. committed this murder?"), and error likelihood ("On a scale of 0–100, how likely is it that there could be an error during the analysis and comparison of B.M.'s DNA to the DNA sample retrieved from underneath the victim's fingernails?"). Finally, participants responded to one attention check question, three manipulation check questions, and five demographic questions (see full experimental materials on the Open Science Framework).

Hypotheses

We predict a main effect of likelihood ratio, such that participants will be more likely to conclude that the suspect is the source of the DNA evidence and committed the crime, but less likely to conclude that an error occurred during DNA analysis, when a likelihood ratio is present compared with absent. We also predict a main effect of alibi, such that participants will be less likely to conclude the suspect is the source of the DNA evidence and committed the crime, but more likely to conclude that an error occurred during DNA analysis, as the strength of the suspect's alibi increases. Finally, we predict a Likelihood Ratio × Alibi interaction, where we expect that participants will not be sensitive to the strength of the alibi when a likelihood ratio is present (compared with absent). In other words, regardless of whether the suspect's alibi is weak or strong, participants will be convinced by the likelihood ratio and be equally as likely to conclude that the suspect is the source of the DNA and that the suspect committed the crime in each alibi condition.

Results

Analyses

To assess our hypotheses, we conducted 2 (likelihood ratio: present vs. absent) × 6 (alibi strength: 6 levels) between-groups factorial analyses of variance (ANOVAs) on each of our dependent measures: source, guilt, and error likelihood. As the data violated the assumptions of normality and homogeneity of variance, we also conducted robust ANOVAs using 20% trimmed means

(Wilcox, 2012). Unless otherwise stated, the results of both approaches were the same, so we will report the original ANOVA results for ease of interpretation. We conducted polynomial planned contrasts to follow up any main effects of alibi or any significant Likelihood Ratio × Alibi interactions and report the associated contrast estimate, C (Haans, 2018). Tables containing means and standard deviations for each condition can be found in the supplemental materials.
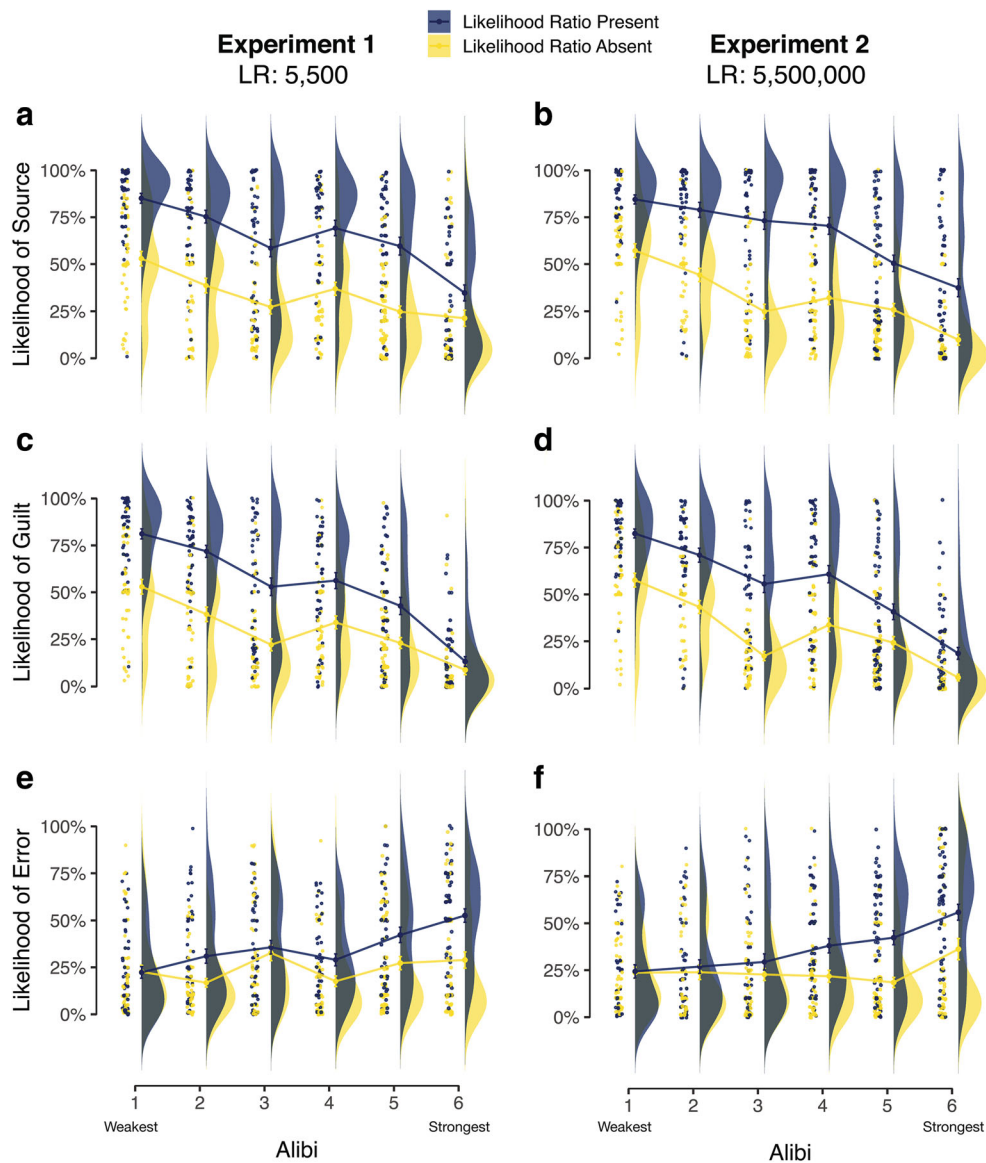
Source

As predicted, participants were more likely to conclude that the suspect was the source of the DNA evidence when a likelihood ratio was present compared to absent in both Experiment 1, $F(1, 579) = 184.50, p < .001, \eta_p^2 = .242$, and Experiment 2, $F(1, 572) = 219.03, p < .001, \eta_p^2 = .277$. Furthermore, there was a significant main effect of alibi on likelihood that the suspect was the source of the DNA evidence in both Experiment 1, $F(5, 579) = 27.19, p = < .001, \eta_p^2 = .190$, and Experiment 2, $F(5, 572) = 36.33, p = < .001, \eta_p^2 = .241$. Polynomial planned contrasts revealed a significant linear trend in both Experiment 1, $C = -28.57, p < .001$, and Experiment 2, $C = -36.38, p < .001$, showing that participants were less likely to conclude that the suspect was the source of the DNA evidence as the strength of the suspect's alibi increased.

As shown in Fig. 1, we also found a significant Likelihood Ratio × Alibi interaction in Experiment 2, $F(5, 572) = 2.60, p = .024, \eta_p^2 = .022$, but not Experiment 1, $Q = 7.58, p = .199$. In Experiment 2, simple effects revealed that alibi influenced the likelihood that the suspect was the source of the DNA evidence both when the likelihood ratio was present, $F(5, 572) = 22.63, p < .001$, and absent, $F(5, 572) = 17.19, p < .001$. Planned linear contrasts revealed that a negative linear trend was present both when the likelihood ratio was present, $C = 323.55, p < .001$, and absent, $C = 285.15, p < .001$, but the difference between them was nonsignificant, $C = -38.41, p = .409$.

Guilt

Consistent with our predictions, participants were more likely to conclude that the suspect committed the murder when a likelihood ratio was present compared to absent in both Experiment 1, $F(1, 579) = 129.52, p < .001, \eta_p^2 = .183$, and Experiment 2, $F(1, 572) = 143.05, p < .001, \eta_p^2 = .200$. There was also a significant main effect of alibi on likelihood that the suspect committed the crime in both Experiment 1, $F(5, 579) = 60.29, p = < .001, \eta_p^2 = .342$, and Experiment 2, $F(5, 572) = 64.34, p = < .001, \eta_p^2 = .360$. Polynomial planned contrasts revealed a significant linear trend in both Experiment 1, $C = -40.64, p < .001$, and Experiment 2, $C = -42.13, p < .001$, showing that participants were less likely to conclude that the suspect committed the murder as the strength of the suspect's alibi increased.

Fig. 1 Panels **a** and **b** depict participants' source likelihood ratings, panels **c** and **d** depict participants' guilt likelihood ratings, and panels E and F depict participants' error likelihood ratings for Experiments 1 and 2, respectively. The raincloud plots depict a half violin plot of participants' mean guilt likelihood ratings overlaid with jittered data points from each of the 591 participants in Experiment 1 and 584 participants in Experiment 2 who were randomly assigned to the six alibis (1 = weakest alibi, 6 = strongest alibi) along with the standard error of the mean per condition

As shown in Fig. 1, we also found a significant Likelihood Ratio × Alibi interaction in both Experiment 1, $F(5, 579) = 4.46$, $p = .001$, $\eta_p^2 = .03$, and Experiment 2, $F(5, 572) = 3.29$, $p = .006$, $\eta_p^2 = .028$. In both experiments, alibi influenced the likelihood that the suspect committed the crime both when the likelihood ratio was present, $Fs > 41.54$, $ps < .001$, and absent, $Fs > 18.55$, $ps < .001$. However, contrary to our predictions, planned linear contrasts revealed that, in both experiments, the negative linear trend was greater when likelihood ratio was present in both Experiments 1 and 2, $C_{Experiment\ 1} = 424.27$, $p < .001$, and $C_{Experiment\ 2} = 404.86$, $p < .001$, compared with absent, $C_{Experiment\ 1} = 255.76$, $p < .001$, and $C_{Experiment\ 2} = 300.13$, $p < .001$. This effect appears to be driven particularly by the lower ratings of guilt likelihood for the strongest alibi. To investigate this possibility, we conducted an exploratory analysis comparing the slope of the linear trends when a likelihood ratio was present or absent excluding the strongest alibi. When doing so, we found the difference between the linear trends was nonsignificant, $C_{Experiment\ 1} = -28.16$, $p < .089$, and $C_{Experiment\ 2} = -17.28$, $p = .292$, demonstrating that the apparent sensitivity to alibi strength when a likelihood ratio is presented is driven primarily by the strongest alibi.

Error

Contrary to our predictions, participants were more likely to conclude that an error could have occurred during the DNA

analysis when a likelihood ratio was present compared with absent in both Experiment 1, $F(1, 579) = 30.07$, $p < .001$, $\eta_p^2 = .049$, and Experiment 2, $F(1, 572) = 28.43$, $p < .001$, $\eta_p^2 = .047$. However, as predicted, there was a significant main effect of alibi on likelihood that an error could have occurred during the DNA analysis in both Experiment 1, $F(5, 579) = 9.56$, $p = < .001$, $\eta_p^2 = .076$, and Experiment 2, $F(5, 572) = 9.18$, $p = < .001$, $\eta_p^2 = .074$. Polynomial planned contrasts revealed a significant positive linear trend in both Experiment 1, $C = 13.54$, $p < .001$, and Experiment 2, $C = 15.44$, $p < .001$. Thus, participants were more likely to conclude that an error could have occurred during the DNA analysis as the strength of the suspect's alibi increased.

As shown in Fig. 1, we also found a significant Likelihood Ratio × Alibi interaction in both Experiment 1, $F(5, 579) = 3.06$, $p = .01$, $\eta_p^2 = .026$, and Experiment 2, $F(5, 572) = 3.20$, $p = .007$, $\eta_p^2 = .027$. Simple main effects revealed that alibi influenced likelihood that an error occurred during DNA analysis both when the likelihood ratio was present, $Fs > 9.42$, $ps < .001$, and absent, $Fs > 2.47$, $ps < .001$. However, contrary to our predictions, planned linear contrast revealed that the positive linear trend was greater when a likelihood ratio was present in both Experiments 1 and 2, $C_{\text{Experiment 1}} = 178.86$, $ps < .001$, and $C_{\text{Experiment 2}} = 212.56$, $ps < .001$, compared with absent, $C_{\text{Experiment 1}} = 47.68$, $p = .107$, and $C_{\text{Experiment 2}} = 45.83$, $ps < .158$. Again, this effect appears to be driven by the higher ratings of error likelihood for the strongest alibi. To investigate this possibility, we conducted an exploratory analysis comparing the slope of the linear trends when a likelihood ratio was present or absent excluding the strongest alibi. When doing so, we found the difference between the linear trends was nonsignificant in Experiment 1, $C = -27.96$, $p = .068$, but still significant in Experiment 2, $C = -59.49$, $p < .001$, demonstrating that there is sensitivity to alibi strength when a likelihood ratio of 5,500,000 is presented, but for a likelihood ratio of 5,500 the apparent sensitivity to alibi strength is driven primarily by the strongest alibi.

Exploratory analyses: comparing likelihood ratio conditions across experiments

We were interested to see whether participants' judgments differed depending on the value of the likelihood ratio across the two experiments. Since the two likelihood ratios are two orders of magnitude apart and intended to reflect different degrees of evidential strength, we should see an increase in participants' judgments about source and guilt likelihood in Experiment 2 (with a likelihood ratio of 5,500,000) compared with Experiment 1 (with a likelihood ratio of 5,500). Thus, we conducted an exploratory 2 (likelihood ratio value: 5,500 vs. 5,500,000) × 2 (likelihood ratio: present vs. absent) × 6 (alibi strength: 6 levels) between-groups ANOVA on source, guilt likelihood, and analysis error to see whether the value of the

likelihood ratio across experiments moderated the effects. There was no main effect of likelihood ratio value on the likelihood that the suspect was the source of the DNA evidence, $F(1, 1151) = 0.06$, $p = .813$, $\eta_p^2 < .001$, guilty of the crime, $F(1, 1151) = 0.66$, $p = .417$, $\eta_p^2 = .001$, or that an error could have occurred during DNA analysis, $F(1, 1151) = 0.07$, $p = .788$, $\eta_p^2 < .001$. These results demonstrate that participants' ratings did not differ depending on whether they received a likelihood ratio of 5,500 or 5,500,000.

Discussion

The forensic science community is considering moving towards using likelihood ratios to communicate uncertainty, which communicates the strength of two competing source hypotheses. However, as likelihood ratios are difficult for fact-finders to interpret (Martire et al., 2014; Martire et al., 2013), they may be overwhelmed by large likelihood ratios commonly encountered in DNA evidence or (incorrectly) interpret the likelihood ratio as being the probability of guilt and ignore exculpatory evidence, such as a compelling alibi.

In this paper, we investigated how perceivers would evaluate a strong likelihood ratio in the face of varying strengths of a suspect's alibi. Participants were more likely to conclude that the suspect was the source of the evidence and guilty of the crime when presented with a likelihood ratio compared with no likelihood ratio—but were also more likely to conclude that an error could have occurred when a likelihood ratio was present compared with absent. This finding demonstrates that likelihood ratios may be a positive step towards communicating uncertainty, as, contrary to our predictions, participants did not simply believe that a strong likelihood ratio was error free. Further, we found that participants were less likely to conclude that the suspect was the source of the evidence and guilty of the crime, but more likely to conclude that an error could have occurred during DNA analysis as the strength of the suspect's alibi increased. Finally, we predicted that participants presented with a likelihood ratio would be *less* sensitive to the strength of the suspect's alibi compared with participants who were not presented with a likelihood ratio, such that participants in the likelihood ratio conditions would provide high estimates of source and guilt likelihood regardless of the strength of the suspect's alibi. However, in both experiments we found that participants who received a likelihood ratio were actually *more* sensitive to the strength of the suspect's alibi than participants who did not receive a likelihood ratio, shown by a greater positive linear trend in the likelihood-ratio-present conditions compared with absent.

While overall it may appear that likelihood ratios increase sensitivity to the alibi evidence, this is largely driven by the low ratings in the strongest alibi. Indeed, exploratory analyses comparing the slopes of the linear

trends when a likelihood ratio was present or absent without the strongest alibi seem to support this. The strongest alibi was the only alibi in which the suspect was nowhere near the vicinity of the crime on the date of the murder (as he was confirmed to be overseas for 1 week, including the date of the murder) and thus was likely the only alibi in which participants believe there was clearly reasonable doubt that he could have committed the crime. It is certainly promising that participants took the strongest alibi into account and significantly reduced their estimates of the likelihood that the suspect was the source of the evidence and guilty of the crime; however, participants were no more sensitive to the strength of the suspect's alibi for the remaining five alibi conditions when a likelihood ratio was present. The only instance in which the linear trend remained significant after removing the strongest alibi was participants' ratings of the likelihood of error in the second experiment only; however, given the exploratory nature of this analysis, this effect may not be robust. It was, however, alarming to find that despite increasing the value of the likelihood ratio by two orders of magnitude (5,500 in Experiment 1 to 5,500,000 in Experiment 2), we found near identical patterns of results. Consistent with prior research (Martire et al., 2013), participants' judgments about source, guilt, or error likelihood were not significantly different depending on which likelihood ratio they received. This is troubling, as the Association of Forensic Science Providers (2009) standard on likelihood ratios proposes six categories relating to the strength of the evidence. These categories are only useful insofar as jurors can actually distinguish between them, but our results suggest that mock jurors cannot. Thus, adopting likelihood ratios as a means to quantify the strength of evidence may not be the best path forward. Our results seem to suggest that the presence of a likelihood ratio may make participants more aware of the potential for error, even if they are not sensitive to the strength of the likelihood ratio. Thus, we should determine whether alternative methods of communicating testimony may result in fact-finders being better able to differentiate the strength of evidence. One such approach may be to directly communicate the error rates of forensic disciplines which, in some circumstances, are lower than what laypeople believe (Ribeiro, Tangen, & McKimmie, 2019).

Open practices statement Experiment 1 was preregistered prior to data collection, and all materials, data, and analysis scripts are available on the Open Science Framework (https://osf.io/yngxk/). Experiment 2 was preregistered prior to data collection, and all materials, data, and analysis scripts are available on the Open Science Framework (https://osf.io/2hmvz/).

References

Association of Forensic Science Providers. (2009). Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice*, *49*(3), 161–164. https://doi.org/10.1016/j.scijus.2009.07.004

Australian Law Reform Commission. (2010). *Presentation of DNA evidence.* Retrieved from https://www.alrc.gov.au/publication/essentially-yours-the-protection-of-human-genetic-information-in-australia-alrc-report-96/44-criminal-proceedings/presentation-of-dna-evidence/#_ftn28

Gigerenzer, G. (2002). *Reckoning with risk: Learning to live with uncertainty.* London, England: Penguin.

Gigerenzer, G., & Galesic, M. (2012). Why do single event probabilities confuse patients? *British Medical Journal*, *344*(1), e245. https://doi.org/10.1136/bmj.e245

Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis*, *25*(3), 623–629. https://doi.org/10.1111/j.1539-6924.2005.00608.x

Gumbel, A. (2004). Madrid suspect 'never went to Spain'. Retrieved from: https://www.independent.co.uk/news/world/americas/madridsuspect-never-went-to-spain-563188.html

Haans, A. (2018). Contrast analysis: A tutorial. *Practical Assessment Research & Evaluation*, *23*(9), 1–21.

Koehler, J. J. (1997). Why DNA likelihood ratios should account for error (even when a National Research Council report says they should not). *Jurimetrics*, *37*(4), 425–437.

Martire, K. A., Kemp, R. I., Sayle, M., & Newell, B. R. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International*, *240*(61). https://doi.org/10.1016/j.forsciint.2014.04.005

Martire, K. A., Kemp, R. I., Watkins, I., Sayle, M. A., & Newell, B. R. (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. https://doi.org/10.1037/lhb0000027

Morrison, G. S. (2016). Special issue on measuring and reporting the precision of forensic likelihood ratios: Introduction to the debate. *Science & Justice, 56*(5), 371–373. https://doi.org/10.1016/j.scijus.2016.05.002

National Academy of Sciences. (2009). *Strengthening forensic science in the United States: A path forward.* Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf

National Registry of Exonerations. (2019, January 16). *% exonerations by contributing factor.* Retrieved from http://www.law.umich.edu/special/exoneration/Pages/ExonerationsContribFactorsByCrime.aspx

Olson, E. A., & Wells, G. L. (2004). What makes a good alibi? A proposed taxonomy. *Law and Human Behavior*, *28*(2), 157–176. https://doi.org/10.1023/B:LAHU.0000022320.47112.d3

President's Council of Advisors on Science and Technology. (2016). *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods.* Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf

R v. Keir. (2002, February 28). NSWCCA 30 (Unreported, Giles J. A., Greg James and McClellan, J.).

Ribeiro, G., Tangen, J. M., & McKimmie, B. M. (2019). Beliefs about error rates and human judgment in forensic science. *Forensic Science International*, *297*(1), 138–147. https://doi.org/10.1016/j.forsciint.2019.01.034

Saks, M., & Koehler, J. (2005). The coming paradigm shift in forensic identification science. *Science*, *309*(5736), 892–895. https://doi.org/10.1126/science.1111565

Scurich, N., & John, R. S. (2011). Trawling genetic databases: When a DNA match is just a naked statistic. *Journal of Empirical Legal Studies*, *8*(S1), 49–71. https://doi.org/10.1111/j.1740-1461.2011.01231.x

Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, *24*(3), 271–296. https://doi.org/10.1023/A:1005595519944

Spellman, B. A. (2018). Communicating forensic evidence: lessons from psychological science. *Seton Hall Law Review, 48*(3), 827–840.

Thompson, W. C., & Cole, S. A. (2005). Lessons from the Brandon Mayfield case. *The Champion, 29*(3), 42–44.

Thompson, W. C., & Newman, E. J. (2015). Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents. *Law and Human Behavior*, *39*(4), 332–349. https://doi.org/10.1037/lhb0000134

Thompson, W. C., Taroni, F., & Aitken, C. G. G. (2003). How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences*, *48*(1), 47–54.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). Boston, MA: Academic Press.