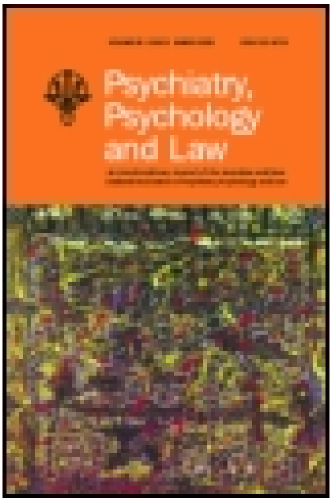
This article was downloaded by: [The UC Irvine Libraries], [Nicholas Scurich]
On: 13 November 2014, At: 10:28
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Psychiatry, Psychology and Law
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tppl20

The Differential Effect of Numeracy and Anecdotes on the Perceived Fallibility of Forensic Science

Nicholas Scurich[a]
[a] Psychology & Social Behavior; Criminology, Law & Society, School of Law, University of California Irvine, Irvine, USA

Published online: 11 Nov 2014.

To cite this article: Nicholas Scurich (2014): The Differential Effect of Numeracy and Anecdotes on the Perceived Fallibility of Forensic Science, Psychiatry, Psychology and Law, DOI: 10.1080/13218719.2014.965293

To link to this article: http://dx.doi.org/10.1080/13218719.2014.965293

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Psychiatry, Psychology and Law, 2014
http://dx.doi.org/10.1080/13218719.2014.965293

Routledge
Taylor & Francis Group

The Differential Effect of Numeracy and Anecdotes on the Perceived Fallibility of Forensic Science

Nicholas Scurich

Psychology & Social Behavior; Criminology, Law & Society, School of Law, University of California Irvine, Irvine, USA

Contrary to popular belief, forensic science – including forensic DNA testing – is not infallible. The rate at which errors occur exerts an inordinate impact on the probative value of a DNA match. Previous research indicates that jurors are insensitive to this effect. The current study tests two possible explanations for the observed insensitivity: (1) juror innumeracy or (2) quantified error rates are not sufficiently vivid. Jury-eligible adults ($n = 568$) read a synopsis of a rape trial in which the quantified error rate was manipulated (either 1-in-10 or 1-in-100), as was the vividness of an error (the laboratory technician was anecdotally portrayed as: sloppy, biased, both or none). Overall, both manipulations affected participants' verdicts. However, numerate participants were affected by the quantified error rate but not anecdotal information, whereas innumerate participants were affected by anecdotal information but not the error rate. The results indicate that the well-known effect on the use of abstract vs. concrete information is moderated by numeracy. On a practical level, the results suggest that, depending on their numerical proficiency, jurors rely on different types of information when evaluating the possibility that forensic scientific evidence is fallible.

Key words: forensic science; judgment and decision-making; numeracy.

Contemporary depictions of the criminal justice system are suffused with forensic science (Tyler, 2006). Gone are the days when conventional detective work solved criminal cases. Popular programing is now rife with scenarios in which forensic science unimaginably and dispositively solves an otherwise impossible-to-solve case. Such programming has undoubtedly contributed to the assumption that forensic science is infallible. Indeed, several appellate courts have propounded that forensic science – and forensic DNA testing in particular – is "failsafe" and that errors are "impossible" (see Koehler, Chia, & Lindsey, 1995).

The notion that forensic science is infallible is a myth (W. C. Thompson, 2012). Data collected by the Innocence Project indicate that forensic science was present in well over half of the exoneration cases (Saks & Koehler, 2005). In 2009, the National Research Committee – a committee empaneled by the National Academy of Sciences – released a report in which it criticized the scientific foundation of many forensic science techniques, and called for proficiency testing to evaluate the validity of the claims that are regularly proffered by forensic scientists in court. To date, these calls have gone largely unanswered (for an exception, see M. B. Thompson, Tangen, & McCarthy, in press).

Proficiency testing efforts have been met with stern resistance from forensic scientists

---

Correspondence: Nicholas Scurich, Psychology & Social Behavior; Criminology, Law & Society, School of Law, University of California Irvine, 4312 SBSG Irvine, CA 92617, USA. Email: nscurich@uci.edu

(Saks & Koehler, 2005). The result is a current lack of knowledge about the frequency with which forensic science techniques err (Koehler, 2013). However, several studies from the late 1980s to the early 1990s provide a ball park figure of the rate at which errors might occur in forensic DNA testing – the so-called "gold standard" of forensic science (W. C. Thompson, 2012). Koehler et al. (1995) synthesized these studies and suggest that a lower bound estimate places the error rate around 1-in-100 to 1-in-1000. In actual practice, they note, error rates are likely to be higher, given that the estimates from the studies were based on pristine samples under optimal conditions.

Although an error rate of 1-in-100 might seem small, it exerts an inordinate impact on the probative value of a forensic DNA match. Indeed, the error rate is largely determinative of the probative value of the DNA evidence. This potentially counterintuitive truism is the logical result of Bayes' theorem (Bolstad, 2007). An exposition of this principle follows.

Laboratory Error Rates and the Probative Value of Forensic DNA Evidence

The probative value of any piece of evidence can be described with a likelihood ratio (Lyon & Koehler, 1996). The likelihood ratio expresses the relative likelihood of observing the evidence given that the hypothesis is true to the likelihood of observing the evidence given that the hypothesis is not true. Formally, $p(E|H)/p(E|-H)$. A piece of evidence is probative to the extent that it is more likely to be observed when the hypothesis is true than when it is not true. In other words, it is probative to the extent the likelihood ratio is > 1.

To give an example, suppose the evidence (E) is a DNA match and the hypothesis (H) is that it came from a particular suspect. How probative is the DNA match? That depends both on the chances of declaring a match when in fact the samples match (the numerator) and the chances of declaring a match when the two samples do not match (the denominator). If the numerator is quite high and the denominator is quite low, then the match will be highly probative. By contrast, even if the numerator is high, the match will not be very probative if the denominator is also high.

There are several ways in which a false-positive match (i.e., the denominator) can occur. The first is purely by coincidence. Genetic profiles are not unique. Thus, two DNA profiles could match even though they are not from the same source. The prevalence of a profile within a given population is typically characterized by a random match probability (RMP). The random match probability is the reciprocal of the likelihood ratio. For example, if the RMP is 1 in 1,000,000, the likelihood ratio would be 1,000,000 to 1.

The other way in which a false-positive match might occur is by laboratory error, such as cross contamination, mislabeling or misinterpretation (see W. C. Thompson, 2012). The rate at which such errors occur is referred to as the laboratory error rate (LER). The denominator of the likelihood ratio must take into account both the RMP and the LER (Koehler, 1997), which changes the formal expression of the likelihood ratio to: $p(E|H)/p(RMP+(LER^{*}(1\text{-}RMP)))$ (W. C. Thompson, Taroni, & Aitken, 2003). Table 1 provides a few examples of the impact that the LER exerts on the likelihood ratio.

The values contained in Table 1 express the likelihood ratio associated with a particular

Table 1. The effect of random match probabilities and laboratory error rates on probative value.

Random match probability	Laboratory error rates (LER)		
	1 in 1000	1 in 100	1 in 10
1 in 100,000,000	999.99	100.00	10.00
1 in 1,000,000	999.00	99.99	10.00
1 in 100,000	990.11	99.90	10.00
1 in 1000	500.25	90.99	9.91

combination of a RMP and a LER. It should be apparent that the LER is mostly driving the probative value. This is because the LER is magnitudes larger than the RMP, and probative value is primarily determined by the larger of the two input probabilities. As described by Scurich and John (2013), the two probabilities are like links in a chain; increasing the strongest link will have little overall effect relative to the weakest link. Thus, even an impressively rare profile, one that exists in 1 person in 100,000,000, might not be very probative evidence if the LER is large by comparison.

The Current Experiment

Research indicates that jurors do not appreciate the relationship between RMPs, LERs and probative value. Koehler et al. (1995) seminally examined this issue and found that jurors were highly impressed by the RMP (which was 1 in 1,000,000,000) and almost completely insensitive to the LER (which was either 2 in 100 or 1 in 1000). This finding was replicated by Nance and Morris (2005). Schklar and Diamond (1999) provided jurors with instruction on how to appropriately combine the RMP and LER and again found that jurors were relatively impressed by the RMP and largely neglected the LER. Scurich and John (2013) recently found that jurors did attend the LER, although they used an unrealistically high LER of 1 in 10 and found that the adjustments made by jurors were still quite meager. All in all, it seems safe to conclude that jurors do not appreciate the disparate impact that LERs exert on probative value.

The purpose of the current study is to examine why jurors do not attend to LERs. Two hypotheses are proposed and tested. The first hypothesis is that jurors neglect LERs because of innumeracy. Numeracy refers to a person's comprehension of and proficiency with numerical information (Paulos, 1988). A burgeoning literature indicates that numeracy mediates medical decisions, estimates of risk, susceptibility to framing effects and so on (see generally, Peters et al., 2006; Reyna & Brainerd, 2007). Higher numeracy is generally associated with better-informed decisions (Peters, Hibbard, Slovic, & Dieckmann, 2007). In a similar vein, perhaps jurors misapprehend the effect of LERs on probative value because of innumeracy. After all, the relationship between RMPs, LERs and probative value is not intuitive, and it seems plausible that confusion is likely to be even more pronounced in individuals who are predisposed to disfavor numerical information. No research has examined whether numerate individuals are more likely to utilize LERs than innumerate individuals.

The second hypothesis is that jurors neglect LERs because a quantified error rate is pallid and abstract, whereas research suggests that concrete and anecdotal information is more likely to affect judgments and decisions. For example, Kahneman and Tversky (1973) provided participants with numeric information (i.e., base rates) about the relative frequency of lawyers and engineers, as well as a personality description that either seemed consistent with an engineer or a lawyer. The numeric information had no effect whatsoever on participants' judgments of the probability that a person was a lawyer or an engineer; the personality description did all the work. Similarly, Borgida and Nisbett (1977) found that university undergraduates did not rely on numeric course evaluations in determining what course to enroll in. However, anecdotal information from a peer student regarding the course had a substantial impact on students' decisions to enroll in the course. This finding led Borgida and Nisbett (1977) to conclude that "information is utilized in proportion to its vividness" (p. 285). The practical implications of this line of work have been explored extensively within the realm of healthcare intervention, in which anecdotal information – particularly patient testimonials – regularly accompany numeric risk information (Winterbottom, Bekker, Conner, & Mooney, 2008). Even in applied settings, anecdotal information appears to be far more persuasive than numeric information (Fagerlin, Wang, & Ubel, 2005; Ubel, Jepson, & Baron, 2001).

In order to test the hypothesis that jurors neglect LERs because they are abstract and pallid, this experiment manipulates the vividness of the possibility of an error. This is accomplished by providing anecdotal information about the possible ways in which an error might have occurred. It is predicted that jurors will respond to anecdotal information about the possibility of an error, and in the appropriate direction. Moreover, the degree of anecdotal information will moderate the effect such that more anecdotal information will increase the belief that an error occurred in this particular case. This subsidiary hypothesis is based on a study by Scurich, Monahan, and John (2012) which found that the degree to which a statistical risk estimate was unpacked (i.e., the risk factors on which the estimate was based were articulated) increased the perceived relevance of that risk estimate to a particular individual.

Methods

Participants

Five hundred and forty-two participants were recruited through Amazon's Mechanical Turk (see generally, Mason & Suri, 2012). Participants were eligible if they were aged over 18 years and a United States citizen. The median age of the sample was 28 years (interquartile range [IQR] = 12). The sample was comprised of 60% (n = 326) males; 38% (n = 205) of participants identified as Democrat, 13% (n = 70) identified as Republican, 35% (n = 191) identified as independent, and the rest indicated some "other" type of political affiliation. Most participants (70%, n = 377) indicated that they did not consider themself to be a religious person. Using a 9-point scale from 1 "extremely liberal" to 9 "extremely conservative", 15% (n = 82) considered themselves extremely liberal, whereas 3% (n = 16) considered themselves extremely conservative; the median score was 3 (IQR = 3).

Procedure

Participants completed an anonymous online questionnaire that included a summary of a criminal rape trial in Los Angeles County, California. The general facts of trial were adapted from an actual rape case in California (*People* v. *Johnson*, 2006). After reading all the materials, participants indicated whether they would convict the defendant. They also indicated the numerical likelihood that the defendant committed the alleged act on a 0–100% scale. The order in which participants made the dichotomous decision and the likelihood rating was randomized. Participants also answered several items probing various dimensions of the case. For example, they were asked, "How believable is the prosecution's case?" or "How trustworthy is the laboratory technician?" Participants then answered a reading comprehension question, which was designed to test whether participants had paid attention to the materials. Consistent with current practice, only those participants who correctly answered the question were included in the analysis (Oppenheimer, Meyvis, & Davidenko, 2009). Finally, participants provided demographic information and completed a numeracy scale (Lipkus, Samsa, & Rimer, 2001).

This experiment employed a 2 (Error rate: 1 in 10 or 1 in 1000) × 4 (Anecdotal information: none, sloppiness, bias, both sloppiness and bias) between-subjects factorial design. Participants were thus randomly assigned to one of eight experimental conditions. In all conditions, participants learned that a stranger rape had occurred and that a DNA profile was extracted from semen found in the victim. The prime suspect, who worked at a nearby restaurant, reluctantly submitted a blood sample for DNA testing. He was subsequently charged with sexual assault.

At trial, the prosecutor called a geneticist ("Dr Wong") who testified that the defendant's genetic profile matched the one recovered from the victim, and that such a match would occur at random in 1 in 200,000,000 Hispanics. On cross examination, the geneticist admitted that a proficiency test conducted at his laboratory did document an instance of laboratory error, namely declaring that two DNA samples match when in fact they did not. It was estimated that such an error occurred (1 in 10 or 1 in 1000) forensic DNA tests.

What followed next depended on the experimental condition. Participants in the sloppiness condition were then told:

> When questioned about the study, Dr Wong explained that the documented error occurred because the two samples were cross contaminated. "During the analysis, the samples were placed into plastic gels to make a comparison; at some point, there was fluid transfer between the two samples." Dr Wong admitted that he did not know – and in fact could not know – if cross contamination occurred in this case. But he stated that he strongly believed that no such error did occur.

Participants in the bias condition were told:

> Dr Wong ... stated that Bioluxe handles all the DNA testing for the Los Angeles Police Department, and just last year made over $2,000,000 from their cases alone ... Dr Wong testified that Detective Smith [the lead detective] brought the samples in question to Bioluxe, and told him that the samples are from a rape case that he's "having trouble closing." "I know this guy did it," Detective Smith said, "I just can't find any other evidence, and the creep is probably going to walk." ... Dr Wong admitted that there is a degree of subjectivity involved with DNA testing. "When comparing the samples, the technician must use his expert judgment to determine if the alleles are identical at certain marker points." Dr Wong denied that Detective Smith influenced his analysis.

Participants in the "both" condition were given both descriptions, whereas participants in the "none" condition received nothing. Participants were then told they have heard all of the evidence and admonished that they should only convict if the evidence presented to them proved the defendant's guilt beyond a reasonable doubt.

Results

Overall, 43% ($n = 234$) of participants voted to convict the defendant with a median likelihood estimate of 75 (IQR = 46). Table 2 contains the proportion of participants voting to convict and the likelihood ratings (with 95% confidence intervals [CI]), decomposed by their respective experimental condition.

A logistic regression with error rate and anecdotal information as the independent variables and verdict as the dependent variable ($\chi^2 = 25.39$, df = 7, $p < .001$) detected a significant main effect for error rate (Wald = 12.09, df = 1, $p < .001$) and anecdotal information (Wald = 10.05, df = 3, $p < .05$). The interaction was not significant ($p = .49$). Participants who received the 1-in-1000 error rate were three times more likely to convict

Table 2. *Proportion of guilty verdicts and likelihood of guilt ratings [with 95% CIs] decomposed by experimental condition.*

	Laboratory Error Rates (LER)	
Narrative	1 in 10	1 in 1000
None	.37, 66.7 [60.01, 73.38]	.64, 81.1 [74.79, 87.50]
Sloppiness	.43, 68.6 [62.12, 75.11]	.58, 79.7 [73.00, 86.30]
Bias	.31, 59.0 [52.50, 65.50]	.42, 66.8 [60.55, 73.09]
Both	.31, 60.0 [51.00, 63.26]	.42, 67.9 [61.35, 74.35]

(exp(b) = 3.0, 95% CI [1.48, 6.06]) than participants who received the 1 in 10 error rate. Participants who received the bias script or both the sloppiness and bias script were 2.5 times more likely to convict (exp(b) = 2.52, 95% CI [1.28, 4.96], $p < .01$, (exp(b) = 2.51, 95% CI [1.26, 4.99], $p < .01$, respectively) than participants who received no anecdotal information. There was no difference in the likelihood of convicting between participants who received no anecdotal information and participants who received only the sloppiness script ($p = .26$).

Two separate logistic regressions were conducted to examine whether the pattern of results is consistent across participants who are high and low in numeracy.[1] High and low numeracy was operationalized by conducting a median split (median = 0.91, IQR = .18) on numeracy scores. For participants who are low in numeracy ($n = 384$), a logistic regression with error rate and anecdotal information as the independent variables and verdict as the dependent variable ($\chi^2 = 16.02$, df = 7, $p < .05$) detected only a significant main effect for anecdotal information (Wald = 10.82, df = 3, $p < .01$). The main effect for error rate ($p = .12$) and the interaction ($p = .51$) were not significant. For participants who are high in numeracy ($n = 158$), a logistic regression with error rate and anecdotal information as the independent variables and verdict as the dependent variable ($\chi^2 = 21.66$, df = 7, $p < .01$) detected a main effect for error rate (Wald = 15.48, df = 1, $p < .001$). The main effect for anecdotal information ($p = .65$) and the interaction ($p = .36$) were not significant. Thus, error rates but not anecdotal information affected the verdicts of high numeracy participants whereas anecdotal information but not error rates affected the verdicts of low numeracy participants.

We next compared how high and low numeracy participants viewed various dimensions of the case. All ratings were made on a 9-point Likert scale. The only significant differences between high and low numeracy participants were on decisional confidence and the likelihood that an error occurred in this particular case, with low numeracy participants giving higher ratings to both ($t(540) = 2.6$, $p = .01$, $t(540) = 2.4$, $p < .05$, respectively). There were no differences between high and low numeracy participants with respect to their beliefs about the likelihood that the defendant committed the crime ($p = .275$), how strong the overall case is ($p = .616$), how believable the case is ($p = .781$), or how trustworthy the laboratory technician is ($p = .415$).

Discussion

Despite the fact that LERs exert an inordinate impact on the probative value of forensic DNA evidence, research finds that jurors are more persuaded by impressively small random match probabilities than quantified error rates. In fact, previous studies have found that jurors are almost entirely insensitive to error rates when evaluating DNA evidence (Koehler et al., 1995). Overall, the current results suggest that the likelihood jurors will attend to LERs increases with numeracy and with anecdotal information. These findings are consistent with previous research on the use of anecdotal information (Ubel et al., 2001), and have straightforward practical implications: jurors are more likely to appreciate the effect of LERs when they are accompanied with anecdotal information. They also support the contention that training jurors in elementary probability theory would increase the quality of decisions rendered by jurors (Koehler, 2006).

A more fine-grained analysis of the data, however, suggests that the overall analyses obscure fundamental differences in the way in which jurors utilize LERs. The verdicts of numerate participants were influenced by the quantified error rate but not anecdotal information, whereas the verdicts of innumerate participants were influenced by anecdotal information but not the error rate. This finding extends the voluminous literature on the effect of abstract vs. concrete information (Borgida & Nisbett, 1977). Specifically, there appear to be individual differences in the reliance on concrete information or numeric (i.e., base rate) information. To the best of our knowledge, this effect has gone heretofore undetected. The current findings suggest that it cannot be said that people universally prefer concrete, vivid information to numerical estimates, as some participants did in fact rely on the abstract numeric estimates but not anecdotal information.

It is important to note that no claims are being made regarding whether one form of reasoning is more appropriate than the other. In other words, it is not being suggested that reliance on anecdotal information is "irrational." Several features of the experimental design preclude such a claim. First, the experimental design does not permit a comparison to be made between participants' likelihood of guilt estimates and the appropriate value of the evidence according to Bayesian norms because participants' prior probabilities were not elicited. Second, it is not clear that the quantified error rate is the sufficient statistic and therefore that the anecdotal information is redundant information. Participants may have assumed that the quantified error rate reflected the technician's "honest effort", whereas the script about bias included an honest mistake plus chicanery. In short, it cannot be said that relying on the anecdotal information constitutes double counting evidence.

Several caveats must be noted. First, the stimuli were brief in terms of their level of detail. This could affect the extent to which the findings generalize to an actual criminal trial (but see Bornstein, 1999), and it could affect the internal validity of the findings. In particular, it is possible that austere materials accentuated the differential use of numeric and anecdotal information, and that when provided with materials containing richer information, participants would be apt to rely on both types of information. The results should be accepted cautiously unless or until replicated with different stimuli and within other contexts.

Final Thoughts

The public's fascination with forensic science and the way in which it has been portrayed in popular media can lead to unrealistic expectations. In reality, forensic DNA testing is neither automated nor infallible. This reality must certainly be conveyed to jurors. However, even if jurors accept that errors can and do occur during DNA testing, it does not follow that they will be able to appropriately evaluate the interaction of error rates and probative value. This article represents a nascent step in trying to understand why jurors are not particularly adept at assimilating LERs. Such an understanding is a necessary prerequisite to remediating the problem.

Note

1. This approach (splitting the sample based on numeracy scores) was utilized for the ease of interpretability of the results. The three-way interaction of error rate, anecdotal information and numeracy (as a log-transformed continuous variable) is statistically significant (Wald $= 7.432$, df $= 3$, $p < .05$). Note that the strong negative skew (-1.92, Kurtosis $=$ 4.91) of the numeracy scores necessitated the logarithmic transformation (Mosteller & Tukey, 1977).

References

Bolstad, W. M. (2007). *Introduction to Bayesian statistics*. Hoboken, NJ: Wiley.

Bornstein, B. H. (1999). The ecological validity of jury simulations: is the jury still out? *Law and Human Behavior, 23*, 75–91. doi: 10.1023/A:1022326807441

Borgida, E., & Nisbett, R. E. (1977). The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology, 7*, 258–271.

Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics? *Medical Decision Making*, *25*, 398–405. doi: 10.1177/0272989×05278931

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237.

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

Koehler, J. J. (2006). Train our jurors. In G. Gigerenzer & C. Engel (Eds.) *Heuristics in the law* (pp. 303−326). Cambridge MA: Massachusetts Institute of Technology.

Koehler, J. J. (2013). Proficiency tests to estimate error rates in forensic science. *Law, Probability, & Risk, 12*, 89−98. doi:10.1093/lpr/mgs013

Koehler, J. J., Chia, A., & Lindsey, S. (1995). The random match probability in DNA evidence: irrelevant and prejudicial? *Jurimetrics Journal*, *35*, 201−219.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37−44. doi: 10.1177/0272989×0102100105

Lyon, T. D., & Koehler, J. J. (1996). The relevance ratio: Evaluating the probative value of expert testimony in child sexual abuse cases. *Cornell Law Review, 82*, 43−83.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*, 1−23. doi: 10.3758/s13428-011-0124-6

Mosteller, F., & Tukey, J. (1977). *Data analyses and regression*. Reading: MA, Addison-Wesley.

Nance, D. A., & Morris, S. B. (2005). Juror understanding of DNA evidence: an empirical assessment of presentation formats for trace evidence with a relatively small random-match probability. *Journal of Legal Studies, 34*, 395−444. doi: 10.1086/428020

National Academy of Sciences, National Research Council, Committee on Identifying the Needs of the Forensic Science Community (2009). *Strengthening forensic science in the United States*. Washington, DC: National Academies Press.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867−872. doi: 10.1016/j.jesp.2009.03.009

Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences*. New York: Hill & Wang.

People v. Johnson (2006) 43 Cal.Rptr.3d. 587.

Peters, E., Hibbard, J., Slovic, P., & Dieckmann, N. (2007). Numeracy skill and the communication, comprehension, and use of risk−benefit information. *Health Affairs, 26*, 741−748. doi:10.1377/hlthaff.26.3.741

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*, 407−413. doi: 10.1111/j.1467-9280.2006.01720.x

Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: numeracy, risk communication, and medical decision making. *Learning and Individual Differences, 17*, 147−159.

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

Schklar, J., & Diamond, S. S. (1999). Juror reactions to DNA evidence: error and expectancies. *Law and Human Behavior, 23*, 159−184. doi: 10.1023/A:1022368801333

Scurich, N., & John, R. S. (2013). Mock jurors' use of error rates in DNA database trawls. *Law and Human Behavior, 37*, 424−431.

Scurich, N., Monahan, J., & John, R. S. (2012). Innumeracy and unpacking: bridging the nomothetic/idiographic divide in violence risk assessment. *Law & Human Behavior, 36*, 548−554. doi: 10.1037/h0093994

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (in press). Human matching performance of genuine crime scene latent fingerprints. *Law and Human Behavior*. doi: 10.1037/lhb0000051

Thompson, W. C. (2012). Forensic DNA evidence: the myth of infallibility. In S. Krimsky (Ed.), *Genetic explanations: Sense and non-sense* (pp. 227−255). Cambridge, MA: Harvard University Press.

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 Science, 48*, 47−54. doi: 10.1520/JFS2003315

Tyler, T. R. (2006). Viewing CSI and the threshold of guilt: managing truth and justice in reality and fiction. *Yale Law Journal, 115*, 1050−1085. doi: 10.2307/20455645

Ubel, P. A., Jepson, C., & Baron, J. (2001). The inclusion of patient testimonials in decision aids effects on treatment choices. *Medical Decision Making, 21*, 60−68. doi: 10.1177/0272989×0102100108

Winterbottom, A., Bekker, H. L., Conner, M., & Mooney, A. (2008). Does narrative information bias individual's decision making? A systematic review. *Social Science & Medicine, 67*, 2079−2088.