Evaluating testimony from multiple witnesses: single cue satisficing or integration?

Kirsty Phillips (k.phillips@bbk.ac.uk) & Ulrike Hahn (u.hahn@bbk.ac.uk)
Department of Psychological Sciences, Birkbeck, University of London,
Malet Street, London, WC1E 7HX, U.K.

Toby D. Pilditch (t.pilditch@ucl.ac.uk)
Department of Experimental Psychology, University College London,
26 Bedford Way, London, WC1H 0AP, U.K.

Abstract

Testimony is a fundamental feature of human life: typically, we receive testimonial evidence from others multiple times each day. Often, we have more than one source attesting to a particular claim. This paper examines the way people integrate testimonial evidence from multiple sources. We find evidence that participants deviate substantially from the normative expectation. Instead, results seem indicative of the operation of simple, non-compensatory heuristics, at least some of the time.

Keywords: Judgment; Reasoning; Decision Making; Evidence Evaluation

Introduction

The role of testimony in our everyday lives is pervasive: arguably, much of what an individual knows (or believes to know) relies on the word of others (Lipton, 1998). Crucially it provides a means of obtaining knowledge from events that we could not observe directly (Adler, 2006).

Nevertheless, in the philosophical literature, the importance of testimony was long neglected, due to the epistemological problem of "vulnerability": what justification do we have for taking a speaker's word for something, given that speakers may lie, deceive or make inadvertent errors (Adler, 2006)? As a result, testimony was long dismissed as a basis for genuine 'knowledge'. Such a view enjoyed longstanding support, all the way from ancient philosophers like Plato, through empiricists such as Locke to contemporary epistemology (Adler, 2006). Such dismissal of testimony echoes Socrates' claim that testimony is inherently uncertain: unlike knowledge which is absolute, testimony can only be considered to be a belief, (Walton, 2008). Testimony, on this view, is merely a means of transmission and does not speak to knowledge itself (Lipton, 1998).

This position seems incongruous with the prevalence of testimony in everyday life, and the fact that there are vast areas where an individual's *only* information is typically testimonial, yet we intuitively still consider ourselves to 'know' about them. In the last 25 years, philosophers have started to address this tension, seeking to identify ways in which testimony might justifiably give rise to 'knowledge' (Coady, 1994; Lipton, 1998; Hahn, Oaksford, & Harris, 2013). At the same time, the topic of testimony has become of increasing interest within psychology, spreading from applied domains such as forensic psychology (Winter & Greene, 2007), to 'core' areas such as cognitive (Harris & Hahn, 2009) and developmental psychology (Durfkin & Shafto, 2016) – reflecting an increasing recognition of the extent to which development and cognition itself are socially mediated.

Testimony as Evidence

Research on reasoning in forensic contexts has explored the impact of the perceived reliability of testimony on judgments of guilt, exploring sensitivity to internal inconsistencies within testimonies (Berman & Cutler, 1996; Spellman & Tenney, 2010) or the impact of evidence that contradicts testimony (Lagnado & Harvey, 2008; Lagnado, Fenton & Neil, 2013). The impact of consistency between witnesses has also been examined (Brewer, Potter, Fisher, Bond, & Luszcz, 1999), illustrating participant sensitivity to the coherence of multiple testimonies (Harris & Hahn, 2009).

Normatively, that is with respect to what a rational agent seeking to maximise the accuracy of her beliefs *should* do (Pettigrew, 2016), reasoners should be sensitive to both the number and reliability of witnesses (as well as any potential dependence between them, see Bovens & Hartmann, 2003). However, the capacity of lay reasoners to integrate accurately both the reliability and number of independent testimonies remains an open question. For instance, do lay reasoners weight individual reliability above the number of testimonies or vice versa, or do they show an appropriate integration of both cues? In particular, are they sensitive to the fact that multiple, lower reliability witnesses may together provide stronger evidence than one or more high-reliability source? The present paper seeks to shed light on these issues.

Corroborating Testimony & Evidential Reasoning

The focus of the present paper is on simple cases of corroborative witness testimony: cases where the essential details of the testimony of two or more witnesses are in agreement (Walton, 2008). Crucially, the corroborative testimony of two or more independent witnesses provides greater support for a given hypothesis than the testimony of one alone (Walton, 2008). This is borne out in coherence-based models of evidential reasoning, such as the Bayesian source credibility model (Bovens & Hartmann, 2003; Harris & Hahn, 2009). Here, the reliability and number of (independent) reports/sources are integrated using Bayes' theorem (Bayes, 1763). Accordingly, such Bayesian models provide a normative standard for accurate updating under uncertainty, and thus provide an expectation for how reasoners *should* infer the probative value of corroborative testimonies from multiple, independent witnesses. While it may be easy enough to see that multiple independent sources of equal reliability provide stronger evidence than one source alone, it is considerably less clear how varying degrees of reliability and varying numbers of witnesses are to be traded off. This too is specified by the Bayesian framework, but is intrinsically harder for people to track.

Moreover, work in judgment and decision-making, among other domains, has shown individuals often deviate from normative, computationally complex, calculations (see e.g., Tversky & Kahneman, 1974). Accurate integration of the conditional likelihoods across varying numbers of independent sources, as required by the normative response, is demanding (Ajzen & Fishbein, 1975), and appears ripe for 'satisficing' (Simon, 1956) heuristic (or 'approximate algorithm') strategies (Gigerenzer et al., 1999).

Similarly, behavioural models of persuasion, such as the Heuristic-Systematic Model (HSM; Chaiken, 1980) and Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986) have argued that individuals are often prone to focussing on individual salient cues (rather than the complete integration of available information): for example,, focusing on the credibility of a source when determining the strength of a message from a single source (Chaiken & Maheswaran, 1994). Hence, the question remains open as to when laypersons evaluate cases of corroborative eyewitness testimony in a normative, Bayesian manner, and when they deploy simpler, heuristic "operating rules".

Present Experiment

The aim of this study is to assess whether participants adhere to normative expectations when evaluating the degree of support provided by scenarios of varying corroborative value (in terms of witness reliability and number).

Participants were presented with a scenario in which a business was missing petty cash. The target hypothesis in question was whether the missing cash had or had not been stolen. To evaluate this hypothesis, participants were presented with testimonial evidence in the form of reports from five employees, all of whom stated that the cash was stolen. One employee, Chris, was much more reliable (at 95% hit rate), compared to the other four (Alan, Brad, David, and Edward; each at 15% hit rate). It was also stated that the false positive rate was low at 10% and consistent across all employee reports. Participants were presented with varying combinations of witness reports and asked to rank them from most to least convincing. Table 1 shows the witness combinations shown to participants, ranked according to likelihood ratio and posterior probability calculated via Bayes' Theorem[1], that is, in the normatively correct rank order.

Participants were also asked to provide a text explanation of how they determined their order and asked to rate their confidence in their ranking.

Our research questions can be summarized as follows:

1. Do participant rankings fit those implied by Bayesian inference (i.e. are responses normatively correct)?
2. If not, what alternative ranking preferences do participants have and why?

Table 1: Witness combinations ranked by likelihood ratio and resulting posterior probabilities for the target claim given the reports.

Witness Combination	Likelihood Ratio	P(Theft|E_{1-N})	Ranking
Chris & Alan	14.25	93.44%	1
Chris	9.5	90.48%	2
Alan, Brad, David & Edward	5.0625	83.51%	3
Brad & David	2.25	69.23%	4
Edward	1.5	60%	5

Methods

Participants 60 (38 female) US participants were recruited and participated online through the MTurk platform. Among the participants, 30 had been educated to the level of Bachelors Degree or above. The mean age of participants was 38.25 (*SD* = 10.18). Informed consent was obtained, and all participants were appropriately compensated for their time.

Procedure and Materials All participants completed the survey, conducted using the Qualtrics platform.
The survey consisted of 11 questions in total: Qs 1-3 obtained informed consent; Qs 4-6 obtained demographic information (age, gender and education level); Q7 obtained an MTurk ID for reimbursement; Qs 8 & 9 presented the scenario and obtained participants' rank order; Qs 10 & 11 obtained explanatory text and confidence ratings.

Analysis

The analysis is split into two parts. First, observed rankings are compared to normative prescriptions. Second,

[1] For the sake of Table 1 and Fig. 1, a prior, P(Theft), of 50% is assumed for illustrative purposes. Prior probability of theft was not given in problem statement, as the evidential value of evidence and relative ranks remain constant irrespective of the prior, and the prior is thus irrelevant for answering the ranking question.

Table 2: Descriptive statistics for participant rank selection. Also shown are mean rank, median rank and upper and lower quartiles of rank

Rank Position:	1		2		3		4		5		Mean Rank	Median Rank	Quartiles	
Witness Combination	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%			1st	3rd
Chris & Alan	*8*	13.33	*26*	43.33	*14*	23.33	*12*	20.00	*0*	0.00	2.50	2	2	3
Chris	*22*	36.67	*10*	16.67	*3*	5.00	*6*	10.00	*19*	31.67	2.83	2	1	5
Alan, Brad, David & Edward	*29*	48.33	*6*	10.00	*22*	36.67	*0*	0.00	*3*	5.00	2.03	2	1	3
Brad & David	*1*	1.67	*16*	26.67	*12*	20.00	*28*	46.67	*3*	5.00	3.27	4	2	4
Edward	*0*	0.00	*2*	3.33	*9*	15.00	*14*	23.33	*35*	58.33	4.37	5	4	5

variations in the observed ranking of options are analysed to determine possible ranking strategies among participants.

Normative Comparison In total, only 5 (8.33%) out of 60 participants gave the complete normatively correct rank order. Participant performance was measured against the normatively correct rank order, using Kendall's tau distance τ (as in Miller and Steyvers, 2017). This metric counts the number of pair wise disagreements between two orders, whereby smaller numbers are indicative of greater similarity and larger numbers are indicative of greater dissimilarity. Values of τ range from 0 to N(N-1)/2, where N is the number of items being ordered (in this experiment N=5). Therefore, when τ=0, the two orders are identical and when τ=10 the orders are maximally dissimilar (i.e., completely reversed). Random performance is indicated by an average score of 5. Normalizing the τ value (dividing τ by N(N-1)/2), represents the number of pair wise disagreements as a percentage. The median normalized Kendall's tau distance in our data was 0.3, that is, 30% (lower quartile = 0.1/10%, upper quartile = 0.5/50%). Therefore, at group level, differences between participants' given rank order and the optimal rank order were considerable, deviating by 30% on average.

Rank Order Analysis A Friedman test was conducted to determine if the observed ranking preferences significantly differed across options or whether selections were made at random. A statistically significant difference across options, $X^2(4)$ =75.627, p<0.001, was found. Post hoc analysis was then conducted using a Wilcoxon signed rank test for pair wise comparisons. To assist understanding of participant strategy, the top 3 options were compared in 3 pair wise comparisons: 'Chris & Alan (1) vs. Chris (2)' options (Z = -1.651, p =0.099); 'Chris & Alan (1) vs. Alan, Brad, David & Edward (3)' options (Z = -1.997, p =0.046); 'Chris (2) vs. Alan, Brad, David & Edward (3)' options (Z = -2.347, p =0.019). None of these comparisons were found to be significant once making the appropriate adjustment to the significance level (0.05/3=0.0167), to account for multiple comparisons (a result that is in keeping with the fact that the median rank for all 3 options was 2).

Discussion

This paper has sought to investigate how participants deal with considerations of witness reliability and number when integrating multiple testimonies. Participant rankings for the different witness combination generally failed to adhere to the correct rankings indicated by a normative (Bayesian) standard. Most surprisingly, only 8 participants (13.33% of the sample) correctly ranked 'Chris & Alan' above 'Chris', on his own. Moreover, of those 8 participants, only 5 (8.3%) ranked the entire order correctly. Finally, even though, at the group level, participants' rank orders differed substantially from the normative rank order (at 30% on average), participants were highly confident in their judgments, with the average confidence rating at 81.88%.

Given that the majority of participants did *not* adhere to the normative rank order, it is of interest to understand more precisely where and why deviations occur. The majority of participants had no difficulty ranking the last two options correctly ("Brad & David" as rank 4 and "Edward" as rank 5), as demonstrated by the dominant obtained rank preferences and median rank values (see Table 1). Given that witnesses "Brad" and "Edward" are of equal (modest) reliability, participants needed only to consider the number of witnesses in each scenario to correctly rank these options. Therefore, it is the top 3 options that are most instructive in explaining participants' selections. At a group level, there is

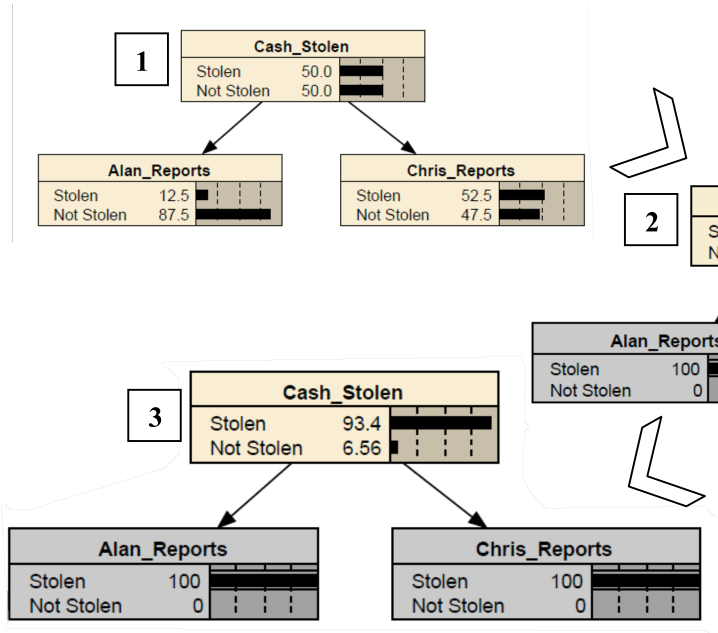

Figure 1: Bayes Net representation (using Netica) to demonstrate: (1) Probability of each witness reporting the cash "stolen" given only the prior of .5; (2) "Alan" has testified that the cash was stolen so both belief that the cash was stolen and that "Chris" will report it "stolen" increase; and (3) Final (normative) belief that the cash was in fact stolen, once testimony from both witnesses has been received.

no difference in how the top 3 options are ranked. This could reflect two possibilities: there are different strategies across distinct subgroups of participants, or, most participants genuinely cannot distinguish between these three options. Next, we tentatively consider potential explanatory strategies.

What can, however, be stated firmly is that participants have most difficulty with those parts of the rank order that require attending to *both* witness reliability and number of witnesses. It is here that orderings deviate most from the normative standard and most variable across participants. This suggests that at least some participants might be using simple heuristics based on just one of these factors.

Reasoning Shortcuts When considering the use of "single-cue" rules or heuristics, it seems important to distinguish between rules that have more or less ecological or adaptive fit. Although heuristics have long been argued to be adaptive given both cognitive constraints (Simon, 1956) and efficient use of time and attention (Gigerenzer et al., 1999), it is necessary to consider more precisely the conditions under which heuristics (e.g., Take-the-Best (TTB) – a single-dominant cue decision rule, see Gigerenzer & Goldstein, 1999) may outperform more computationally demanding processes (Czerlinski, Gigerenzer, & Goldstein, 1999).

How do the two cues relevant to our task -- witness number (or "mass") and high reliability ("reliability") fare? To what degree can they be considered ecologically adaptive? The preference for witness numbers (or "mass"), in the case of increasing numbers of independent witnesses, can be considered adaptive in the sense that (all else being equal) more independent pieces of data do, in fact, make a stronger case from a normative (Bayesian) perspective. Furthermore, picking this number cue over perceived reliability might be considered adaptive given the (arguably typical) absence of clear reliability cues for testimonial evidence in many real-world situations. In fact, recent modelling work has demonstrated how beliefs might be remarkably accurate even if witnesses are simply assumed to be fairly reliable, and how such a strategy may do only slightly less well, and on occasion better even, than one that seeks to estimate reliability (Hahn, von Sydow & Merdes, subm.). In particular, where there is more data, posterior degrees of belief will converge on the right value even where reliability is mis-estimated, as long as one is correct about whether the data provide evidence *for* or *against* the hypothesis in question (see Hahn, von Sydow & Merdes, subm. and references therein). In short, there is reason to pick number of witnesses as a simple, useful heuristic to evidential support, and to pick it over reliability.

However, participants do not universally endorse such a strategy: only a minority pick "Chris & Alan" over "Chris" in the way a "mass" strategy would suggest. In preferring "Chris" alone, participants seem to be displaying *avoidance* of an additional, less reliable witness. Not only is it hard to see an adaptive reason for this, it arguably renders participant responses inconsistent. "Alan" has the same reliability (accuracy) as all other witnesses bar the more reliable Chris. The responses on the lower ranks indicate clearly that participants correctly view those witnesses as providing positive support for the hypothesis: Only on that assumption does it make sense to view 4 of those witnesses ("Alan, Brad, David & Edward") as providing stronger support *for* the hypothesis than just two ("Brad & David") or one alone ("Edward"). However, the number of witnesses should dominate the choice between "Chris & Alan" and "Chris" as well. Yet the 22 who choose "Chris" over "Chris & Alan" 18 prefer "Brad & David" to "Edward" alone. In other words, they must be doing more than (consistently) picking the option with the greater number of witnesses.

This suggests that it is the differential reliabilities in the pairing "Chris & Alan" that are determining the response of those participants. Normatively, an additional, independent, corroborating witness will *always* add support in cases where that witness's hit rate is greater than their false positive rate[2] (as was the case in this study). In this case, a simple counting strategy will come to the same result.

The preference for "Chris" over "Chris & Alan" could be evidence of averaging (e.g., Lopes, 1985) as an alternative strategy: the addition of a less reliable witness would lower the average evidential strength relative to one high reliability witness alone. Alternatively, this preference could suggest either a more fundamental mis-understanding of reliability and/or the connections between the two reports.

Though the witnesses are described as independent, receiving a positive report from one will (normatively) influence the expectation of receiving a positive report from the other, as can be seen from the simple Bayesian network graphical representation summarising the relevant calculations in Fig. 1. In other words, corroborating evidence provides both evidence towards the hypothesis (i.e. Was the cash stolen?) and, indirectly, evidence about the other witness (i.e. Chris's testimony) (Walton, 2008), which is why coherence matters both for the truth or falsity of the target hypothesis (see Bovens & Hartmann, 2003) *and* for witness reliability (see Schubert, 2012). The present results suggest that some participants could be treating Alan's "corroboration" as tainting Chris's testimony, because Alan is less reliable than Chris.

Future work will need to distinguish between these two possibilities: averaging, or limited understanding of the effects of corroborating testimony on reliability. At the same time, however, this potential effect of reliability suggests that some participants do more than simple single cue aggregation via number.

[2] To clarify: the "hit rate" corresponds to the probability of receiving the evidence if the hypothesis is indeed true (P(e|H)), the "false positive rate" to that of receiving the evidence if H is in fact false (i.e., P(e|¬H). The quantity P(e|H)/P(e|¬H), the so-called likelihood ratio (LHR), determines the diagnostic impact of the evidence. As the posterior odds (P(H|e)/P(¬H|e)) are equal to the prior odds (P(H)/P(¬H)) times this likelihood ratio, an LHR > 1 (as obtains whenever the hit rate is higher than the false positive rate) will increase the posterior probability (that is, the probability of H in light of the evidence). By design, the hit rate exceeds the false positive rate for all witnesses in our study (see Methods above).

Conclusions

In conclusion, we find reasoners do not integrate testimonies in the manner expected by normative, Bayesian standards. Instead, simple cue preferences appear to dominate rankings of relative support provided by witness combinations for the majority of participants. However, for those who show an avoidance of supplementary, less reliable (but nevertheless diagnostic) testimony, a deeper seated reasoning error seems to be in play, warranting further research. Overall, what is striking is the extent to which the majority of participants struggle even in this very simple context. Given the fundamental importance of the accurate aggregation of multiple pieces of testimony to many real-world tasks (e.g. criminal investigation, intelligence services, military operations, medical professionals, auditors, etc) the issues raised merit further investigation. In particular, the simplicity of the paradigm used holds some hope that deviations from normative responding observed in more complex tasks (such as in the context of juror reasoning, see e.g., Winter & Greene, 2007; Greene, et al., 2002; Chen & Chaiken, 1999; Pennington & Hastie, 1993) might eventually be better understood.

Acknowledgments

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), under Contract [2017-16122000003]. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

This research was completed by the London Translation Team (LTT), of the Bayesian Argumentation via Delphi (BARD) project, funded by IARPA's CREATE program. Authors KP and UH formulated the study design, author KP conducted data collection and analysis, and KP, TP, and UH wrote the manuscript. The authors would like to thank the other members of LTT for helpful their comments. In particular, Nicole Cruz De Echeverria Loebell and David Lagnado are thanked for their contribution towards constructing the materials for this study.

Open Practices

The materials for this study and raw data have been made available at https://osf.io/kxf98/.

References

Adler, J. (2006). Epistemological problems of testimony. *The Stanford Encyclopedia of Philosophy*. Stanford University.

Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin, 82*, 261-277.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions, 53,*, 370-418.

Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *Journal of Applied Psychology, 81*, 170-177.

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology.* Oxford University Press on Demand.

Brewer, N., Potter, R., Fisher, R. P., Bond, N., & Luszcz, M. A. (1999). Beliefs and data on the relationship between consistency and accuracy of eyewitness testimony. *Applied Cognitive Psychology, 13,*, 297-313.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752-766.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*, 460-460.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken, & Y. Trope, *Dual process theories in social psychology* (pp. 73-96). New York: Guilford Press.

Christensen-Szalanski, J. J., Beck, D. E., Christensen-Szalanski, C. M., & Koepsell, T. D. (1983). Effects of expertise and experience on risk judgments. *Journal of Applied Psychology, 68*, 278-284.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & The ABC Research Group, *Simple heuristics that make us smart* (pp. 97-118). New York: Oxford University Press.

Durkin, K., & Shafto, P. (2016). Epistemic trust and education: effects of informant reliability on student learning of decimal concepts. *Child development*, *87*, 154-164.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In S*imple heuristics that make us smart* (pp. 75-95). Oxford University Press.

Gigerenzer, G., Todd, P. M., & ABC Research Group, T. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Greene, E., Chopra, S. R., Kovera, M. B., Penrod, S. D., Rose, V. G., Schuller, R., & al, e. (2002). Jurors and juries: A review of the field. In J. R. Ogloff, *Taking psychology and law into the twenty-first century* (pp. 225-284). New York: Kluwer Press.

Hahn, U., Oaksford, M., & Harris, A. J. (2013). Testimony and argument: A Bayesian perspective. In F. Zenker, *Bayesian argumentation: The practical side of probability* (pp. 15-38). Netherlands: Springer.

Hahn, U., Merdes, C. & von Sydow, M. (subm.) . How Good is Your Evidence and How Would You Know?

Harris, A. J., & Hahn, U. (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1366 –1373.

Joyce, E., & Biddle, G. C. (1981). Are auditors' judgments sufficiently regressive? *Journal of Accounting Research*, 120-145.

Lagnado, D. A., & Harvey, N. (2008). The impact of discredited evidence. *Psychonomic Bulletin & Review, 15*, 1166-1173.

Lagnado, D. A., Fenton, N., & Neil, M. (2013). Legal idioms: a framework for evidential reasoning. *Argument & Computation, 4*, 46-63.

Lipton, P. (1998). The epistemology of testimony. *Studies in History and Philosophy of Science Part A, 29*, 1-31.

Lopes, L. L. (1985). Averaging rules and adjustment processes in Bayesian inference. Bulletin of the Psychonomic Society, 23, 509–512.

Lynch, M., & Haney, C. (2000). Discrimination and instructional comprehension: guided discretion, racial bias, and the death penalty. *Law and Human Behavior, 24*, 337-358.

Miller, B., & Steyvers, M. (2017). Leveraging Response Consistency within Individuals to Improve Group Accuracy for Rank-Ordering Problems. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.* London: Cognitive Science Society.

Pennington, N., & Hastie, R. (1988). Explanation-based decision-making: effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory & Cognition, 14*, 521–533.

Pennington, N., & Hastie, R. (1993). The story model for juror decision-making. In R. Hastie, *Inside the Juror: The Psychology of Juror Decision-making* (pp. 192-221). New York: Cambridge University Press.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19*, 123-205.

Schubert, S. (2012). Is coherence conducive to reliability?. *Synthese*, *187*, 607-621.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138.

Spellman, B. A., & Tenney, E. R. (2010). Credible testimony in and out of court. *Psychonomic Bulletin & Review, 17*, 168-173.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Walton, D. (1998). *Ad hominem arguments.* University of Alabama Press.

Walton, D. (2008). *Witness Testimony Evidence. Argumentation, Artificial Intelligence and Law.* Cambridge: Cambridge University Press.

Winter, R. J., & Greene, E. (2007). Juror decision-making. In F. T. Durso, R. S. Nickerson, S. T. Dumais, S. Lewandowsky, & T. J. Perfect, *Handbook of applied cognition, 2* (pp. 739-762). Chichester, England: John Wiley & Sons, Ltd.