

Consistency and credibility in legal reasoning: A Bayesian network approach

Saoirse Connor Desai¹ (saoirse.connor-desai@city.ac.uk), Stian Reimers¹ (stian.reimers.1@city.ac.uk), David Lagnado² (d.lagnado@ucl.ac.uk), ¹Psychology, City University, London, EC1R 0JD UK, ²Experimental Psychology, UCL, London, WC1H 0AP UK

Abstract

Witness credibility is important for establishing testimonial value. The story model posits that people construct narratives from evidence but does not explain how credibility is assessed. Formal approaches use Bayesian networks (BN) to represent legal evidence. Recent empirical work suggests people might also reason using qualitative causal networks. In two studies, participants read a realistic trial transcript and judge guilt and witness credibility. Study 1 varied testimonial consistency and defendant character. Guilt and credibility assessments were affected by consistency but not prior convictions. Study 2 constructed a BN to represent consistency issues. Individual parameter estimates were elicited for the corresponding BN to compute posterior predictions for guilt and credibility. The BN provided a good model for overall and individual guilt and credibility ratings. These results suggest people construct causal models of the evidence and consider witness credibility. The BN approach is a promising direction for future research in legal reasoning.

Keywords: Legal reasoning, Evidential reasoning, Bayesian networks, Evidence, Reliability, Credibility

Introduction

Evaluating witness credibility is crucial to establishing the inferential value of testimony in criminal trials. The juror's task is to assess the truthfulness, reliability, and accuracy of the witnesses whose evidence is at issue, and evaluate how well the evidence supports the claims of the prosecution and defence in order to reach a verdict (Crown Court Bench Book, 2010). Lawyers can use many strategies to undermine the credibility of witnesses thereby challenging the reliability of their evidence. Drawing out testimonial inconsistencies under cross-examination, introducing contradictory testimony by other witnesses, disclosing information of previous convictions considered relevant to issues of credibility, and evidence relevant to a witness' reputation or truthfulness, are all important methods of assessing credibility (Spellman & Tenney, 2010). Legal reasoning studies confirm that people perceive inconsistent prosecution eyewitnesses as less accurate and credible, reducing the likelihood of conviction and increasing defendant credibility (Berman & Cutler, 1996; Berman, Cutler & Narby, 1995). Further, guilt judgments are sensitive to whether evidence contradicting an alibi shows a witness has been intentionally deceptive or made a genuine error (Lagnado & Harvey, 2008; Lagnado, 2011; Lagnado, Fenton & Neil, 2013). Mock juror studies also show that disclosing similar previous convictions affect judgments about the testifying defendant's credibility making them

appear more likely to lie under oath and/or more likely to have committed the alleged crime (Lloyd-Bostock, 2000; Wissler & Saks, 1985).

Empirical studies of how people reason about legal evidence show that people reason about different types of evidence in complex ways (Pennington & Hastie, 1992). How people represent the credibility of witnesses and reliability of their evidence is important for understanding legal reasoning and determining what inferences are permissible given these representations. This paper presents a framework for analyzing the integration of testimonies whose sources vary in credibility and reliability. This approach builds on extant descriptive accounts of juror decision-making and employs the Bayesian network framework to model inferences about witness credibility and evidential reliability. Though intended mainly as a normative and prescriptive model of evidential reasoning this paper will show that the framework also captures peoples' ability to draw probabilistic conclusions from interrelated bodies of evidence.

Legal reasoning The story model of juror decision making is the leading cognitive model of how people reason about legal evidence (Pennington and Hastie, 1986, 1992). According to the story model, jurors organize and interpret the mass of evidence presented during the trial by constructing narrative explanations from the evidence. They use causal schemas – such as scripts of typical human thought and behavior – to fill gaps in the evidence and develop a causal 'situation' model of what transpired. The story ultimately adopted is the one that provides the best 'fit' for the evidence and is most plausible, complete and coherent.

The story model has achieved broad empirical support and has considerably advanced understanding of juror decision making. It qualitatively describes the constructive nature of people's explanations and emphasizes the interdependencies between trial evidence. One weakness of the story model is that it does not model how people reason about the credibility and reliability of different types of evidence and how this affects their story evaluation. In addition to reasoning about the crime itself, jurors (or fact-finders in general) must also reason about how well the evidence presented supports the arguments put forward by the prosecution and defence (i.e., what is the evidential support for a given story).

A further shortcoming of the story model is that it rejects the idea that people reason probabilistically. Even if people cannot estimate the precise probabilities of events, they can

often draw probabilistic conclusions from the evidence they hear in the courtroom (e.g. the presence of suspect at crime scene increases the likelihood that he is guilty).

Causal models of witness evidence are constructed in a similar way to causal models of the crime itself. Causal schemas about the nature of witnesses are used to draw inferences about the motivations and beliefs of people giving testimony; these inferences can also modify beliefs about the crime. Without a way to represent the relations between different types of evidence and how they interact it is difficult to elicit and test individual causal models constructed from the evidence. Further, without a formal analysis of how evidence items relate it is impossible to ascertain which inferences are permissible given the evidence.

Modeling evidence reliability The Bayesian network framework provides a potential solution to this problem. This approach makes it possible to test people's causal models of the reliability and credibility of witnesses giving testimony and compare inferences to a normative standard.

Bayesian networks (BN) use graph structures to represent the probabilistic relations between hypotheses and uncertain evidence, showing what inferences are rationally permitted from a given model of the evidence (Pearl, 1988, 2000). BNs have proved valuable for modeling relations in bodies of uncertain evidence in forensic contexts (e.g., Garbolino & Taroni, 2002) and have also been applied to legal contexts (Fenton et al., 2014; Lagnado et al., 2013). Fenton et al., (2014) claim that fact-finders (e.g., jurors) could use small-scale causal building blocks (*legal idioms* corresponding to common inference schemas) that make it possible to reason about complex and interrelated bodies of evidence. These idioms are customized to the legal context, capture generic patterns of legal inference, and can be re-used to make large-scale inferential problems tractable.

The basic idiom consists of the relation between a hypothesis and an item of evidence, corresponding to the relation between the legal proposition that needs to be proved (e.g., the defendant is guilty) and the submitted evidence. The evidence idiom can be supplemented with a reliability idiom enabling the modeling of potential causes of an evidence report that are vital for establishing the reliability of evidence from human sources. Reliability can be separated into issues of: i) observational sensitivity, ii) objectivity, and iii) veracity (Schum, 1994). The graph can be used to represent the fact that these different causes serve to explain the evidence. For example, the victim's testimony in an assault case depends both on whether or not the defendant assaulted the victim, and whether or not the victim is trustworthy and/or inaccurate. These factors are directly related to assessments of witness credibility.

Graphical models have been successfully applied to research in a number of areas of causal cognition (for review see Danks, 2014). Recent empirical work also suggests that people may reason about testimonies using

qualitative causal networks (Lagnado, 2011; Lagnado, Fenton & Neil, 2013).

Study 1

Study 1 investigates some of the conditions under which witness credibility can be challenged, and how this impacts assessment of guilt. In particular, we explored how the consistency of the victim's testimony with other key pieces of evidence, and evidence disclosing the defendant's previous convictions, affected judgments of guilt and credibility, and how these factors combine and interact.

Real trial dialogue (*R v. Capel*) including cross-examination of witnesses was used to accentuate issues of credibility. These factors were specifically chosen because: 1) they address issues of witness credibility and reliability, and were raised in closing arguments and judge's directions, and 2) could be subtly manipulated in order to maintain ecological validity. The stimulus case has been used in previous mock jury research and typically results in a hung jury.

There were two main aims of Study 1: to establish whether people's judgments were affected by changes in witness consistency and whether it is feasible to model these changes using causal schemas of witness credibility and reliability. Given the findings of previous legal decision making research it was hypothesized: 1) that consistent testimony would result in greater belief in guilt, weaken the credibility of the defendant, and bolster the credibility of the victim, relative to inconsistent testimony, and 2) that disclosing a similar prior conviction to the current crime would result in greater belief in guilt, weaken the defendant's credibility and bolster the victim's credibility.

Methods

Participants 126 U.S. and U.K. based participants (64 female, mean±SD age 29.12±10.80, randomly split between six conditions) were recruited from <https://www.prolific.ac/> a site for recruiting participants for web-based studies, and were paid £2.40(\$3.49). Average completion time was 31 minutes.

Design, materials and procedure We investigated the impact of prior character (three levels: no prior, different prior, similar prior) and inconsistency (two levels: consistent, inconsistent) on evidential reliability. Participants read one of six versions of a realistic courtroom transcript. Consistency was manipulated by varying the consistency of key pieces of testimony with the victim's testimony (i.e., consistent or inconsistent). Prior conviction evidence was manipulated by substituting evidence of good character (i.e., revealing the fact that the defendant has no prior convictions) with evidence of bad character (i.e., disclosure of a prior conviction).

The transcript was divided into 20 evidential statements and judgments about the probability of guilt and credibility of the victim and defendant were elicited after each

statement. All participants saw the evidential statements in the following order: the charge and plea, the prosecution’s opening statement, direct and cross-examinations of three prosecution and three defence witnesses, prosecution and defence closing arguments, the judge’s summary, and instructions on the law. They read through the evidential statements at their own pace and updated their ratings the probability of guilt, and credibility of the victim and defendant, in light of each statement. After reading all the evidence, participants indicated their final judgments about guilt, and credibility of the victim and defendant.

Table 1: Manipulation of Key Pieces of Evidence

	Victim’s Friend	Police Officer	Bartender
Inconsistent	8 pints	JD drunk JD not drunk	Saw push
Consistent	4 pints	drunk	No push

Case summary and consistency manipulations The defendant (SC) is charged with assaulting the victim (JD), and pleads not guilty. The prosecution argues that SC punched JD in an unprovoked attack and calls three witnesses: JD, JD’s friend, and a police officer who was at the scene. The defence argues that the punch was an act of self-defence; JD was drunk, aggressive and had pushed SC first. The defence also calls three witnesses: SC, SC’s friend, and a local bartender who was also at the scene. JD makes three important claims in his testimony: 1) that he had only had 4 pints to drink, 2) he was not drunk, and 3) he did not provoke SC first. In the inconsistent version, these claims were contradicted by three witnesses: JD’s friend testifies that JD drank 8 pints, the police officer testifies that JD appeared drunk, and the bartender states he saw JD push SC first. In the consistent version, these claims were not contradicted: JD’s friend corroborates JD’s claim that he only drank 4 pints; the police officer states that JD did not appear drunk, and the bartender does not mention that he saw the push (see Table 1). Inconsistencies between the victim and subsequent testimonies regarding these key issues could make the victim appear likely to have lied about the incident because he was drunk and aggressive or show that his recollection of events was inaccurate.

Study 1: Results

Probability of guilt Mean probability of guilt judgments given at the end of the trial, across all six conditions, are shown in Fig. 1. A prior conviction for a similar crime ($M = 67.60, SD = 28.27$) increased belief in guilt relative to prior conviction for a different crime ($M = 65.16, SD = 29.95$), or no prior convictions ($M = 64.32, SD = 27.58$), but these differences were not significant $F(2, 98) = 1.90, p = .16$. Introducing key pieces of evidence that were inconsistent with the victim’s testimony ($M = 58.98, SD = 29.03$)

resulted in significantly diminished belief in guilt relative to a consistent testimony ($M = 72.20, SD = 26.49$), $F(1, 98) = 12.35, p < .001$.

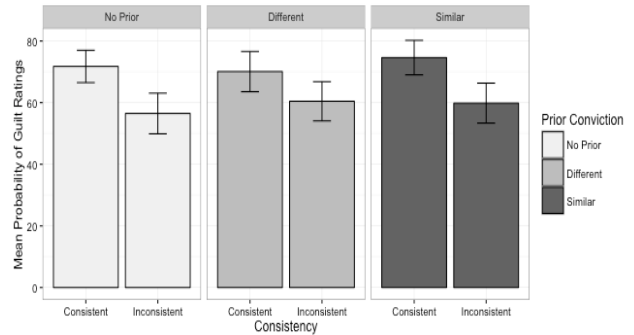


Figure 1: Mean final probability of guilt judgments (\pm standard error) for each condition of the study. Range from 0 -100.

Defendant credibility The defendant was considered to be more credible after an inconsistent testimony ($M = 44.69, SD = 22.77$), than a consistent one ($M = 35.55, SD = 21.40$), but this difference was not significant, $F(1, 98) = 1.94, p = .17$. Defendant credibility ratings were in the expected direction, defendant appeared more credible when he had no prior convictions ($M = 43.59, SD = 21.20$), than a prior conviction for a different ($M = 39.40, SD = 26.37$), or similar ($M = 37.26, SD = 19.17$) crime, but the differences were not significant, $F(2, 98) = .08, p = .92$.

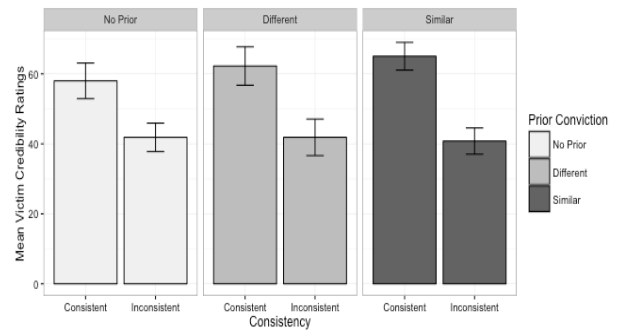


Figure 2: Mean victim credibility judgments at the end of the trial (\pm standard error) for each condition of the study. Range from 0 -100.

Victim credibility Mean victim credibility judgments across the six conditions are shown in Fig. 2. The victim appeared more credible when the defendant had a prior conviction for a similar ($M = 53.50, SD = 21.40$), than different ($M = 51.81, SD = 26.60$) offence, and rated least credible when the defendant had no prior convictions ($M = 50.12, SD = 22.26$), but these differences were not significant, $F(2, 98) = .14, p = .87$. The victim appeared more credible when evidence was consistent with his testimony ($M = 61.81, SD = 22.31$) than when it was inconsistent ($M = 41.52, SD = 19.93$), $F(1, 98) = 16.97, p < .001$. Evidence relating to the defendant’s credibility did not influence the perceived credibility of the victim.

Inconsistencies between the victim’s testimony evidence mentioned in other witness testimonies substantially undermined the victim’s credibility.

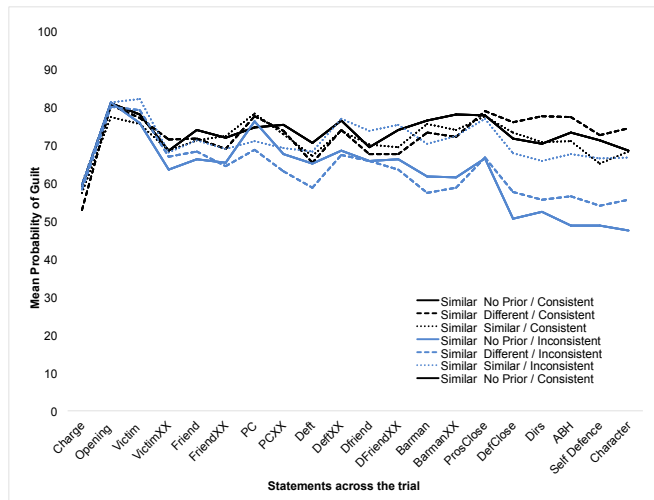


Figure 3: Mean probability of judgments across statements for each condition of in Study 1. Range from 0 -100. The graph shows that the bartender testimony is important for establishing probability of guilt.

Discussion

The results suggest that people draw inferences about witness credibility from subtle inconsistencies between their testimony and the testimonies of other witnesses, which in turn influence their beliefs in the defendant’s guilt. In this study, there was no clear impact of evidence aimed at undermining the defendant’s character, and therefore his credibility, or inferences about guilt. One reason that disclosing prior convictions had no impact on guilt in the current study could be due to methodological differences between current and previous studies. In previous studies participants read descriptions of hypothetical cases in which prior conviction evidence outweighed other evidence in the case. In this study prior conviction information was balanced with other issues in the case, which could account for the lack of difference between conditions. These results suggest that people do more than constructing a plausible story from the evidence; also factoring in the consistency and credibility of witnesses. The key effect of consistency on people’s judgments of credibility and guilt was replicated in a laboratory study too.

Study 2

Study 2 was designed to replicate the consistency effect and to examine whether the idiom-based model provided a good fit for participants’ guilt and credibility ratings. We constructed a BN to represent the key pieces of testimony that were varied in Study 1 (see Fig. 4). The BN captures the impact of the three witness testimonies that were varied

on the probability that JD was drunk, is a credible witness, and provoked SC. Individual parameter estimates were elicited for the corresponding CPT in the BN in order to compute posterior predictions for each of the aforementioned probabilities.

Methods

Participants 137 participants (65 female, mean \pm SD age 32.52 ± 10.67 , randomly split between two conditions) were recruited from <https://www.prolific.ac/> and completed the study for monetary compensation (£2.40/\$3.49). The average completion time was 25 minutes.

Design, materials and procedure We manipulated the consistency of key evidence with the victim’s testimony in the same way as Study 1 (Table 1), thus the materials were identical except character evidence (i.e., defendant prior convictions not manipulated). In Study 2 we also took participants’ estimates of the conditional probabilities to complete individual parameterizations of the BN model, after they provided posterior guilt and credibility ratings. These questions were asked in the following format:

- a) Suppose that JD had drunk 8 pints.
What’s the probability that the policeman would testify that JD appeared drunk/did not appear drunk?
- b) Suppose that JD had drunk 4 pints.
What’s the probability that the policeman would testify that JD appeared drunk/did not appear drunk?

Conditional probability ratings were indicated using a slider ranging from 0 = *exceptionally unlikely* to 100 = *virtually certain*, and pre-set to the midpoint. Participants in both Experimental groups completed these questions, however, questions that related to specific details in the testimonies of the three witnesses were adapted according to condition (e.g., JD appeared OR JD did not appear drunk).

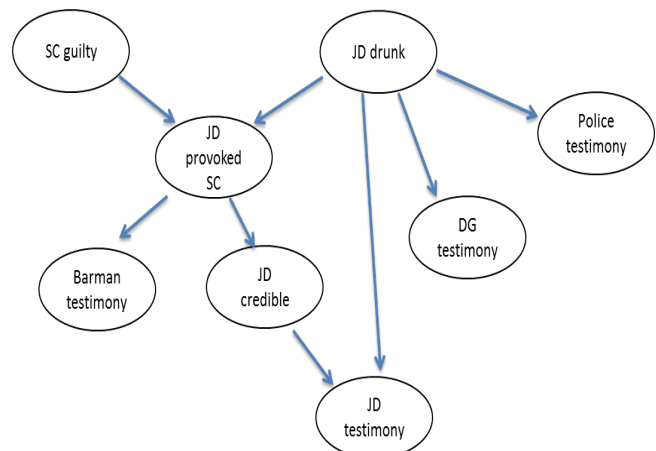


Figure 4: A Bayesian network of *R v. Capel* using evidence- reliability idiom to capture JD credibility.

Study 2: Results

Observed posterior probability judgments After considering all the evidence, a consistent testimony resulted in greater belief in guilt ($M = 73.87$, $SD = 27.42$) than an inconsistent testimony ($M = 68.34$, $SD = 29.13$), but this difference was not significant, $t(135) = -1.41$, $p > .05$. When key pieces of evidence were inconsistent with the victim's testimony ($M = 44.67$, $SD = 24.19$), the defendant appeared more credible than when evidence was consistent ($M = 35.58$, $SD = 22.98$), $t(135) = 2.52$, $p < .05$. The victim was also considered less credible after an inconsistent testimony ($M = 50.14$, $SD = 22.78$) than a consistent one ($M = 64.58$, $SD = 24.83$), $t(135) = -3.55$, $p < .001$. The results almost replicate the effect of consistency in Study 1 and were in the same direction. However, observed posterior guilt judgments showed an overall stronger tendency toward guilt than in Study 1, which could be explained by variation in people's prior expectations and assumptions about criminal proceedings or defendants in general.

Modeling participants' inferences using the BN To test the BN model more formally, we analyzed model predictions using the conditional probabilities provided by participants. The conditional probabilities concerned key issues relating to the reliability of JD's testimony (shown in Table 1)¹.

First, participants' mean probability judgments were used to parameterize separate graphs for each condition. The posteriors generated by the model showed that the probability JD was drunk was higher for the inconsistent (.79) than the consistent condition (.24), likewise the probability that JD provoked SC was higher for the inconsistent (.77) than the consistent (.17) condition, and the probability that JD is a credible witness was lower for the inconsistent (.42) than consistent condition (.70). This shows that the BN provided a good model for participants' judgments of guilt and credibility. The discrepancy between the predicted probabilities estimates and observed posterior probability judgments also suggest that participants are reasoning about factors other than consistency.

To examine whether the BN captured participants' judgments at an individual level, we used each participants' unique conditional probabilities to parameterize the BN. We then used this model to compute posteriors for JD credibility, the probability JD provoked SC, and the probability JD was drunk, and compared these model predictions with participants' actual posterior judgments of guilt and credibility. Observed posterior judgments of guilt and credibility were correlated with the model posterior prediction for the probability that JD provoked SC. In this instance, the model prediction for 'JD Provoke' served as a proxy for the probability of guilt. More precisely, a higher posterior probability that JD provoked SC, would be associated with lower observed guilt ratings, lower ratings

of JD's credibility, and higher ratings of SC's credibility, and vice versa.

The predictions derived from the BN were supported by the data. The model predictions for provoke were negatively correlated with observed posterior probability of guilt ratings, $r = -.37$, $p < .001$, $R^2 = .14$. Provoke was also negatively correlated with victim credibility ratings $r = -.40$, $p < .001$, $R^2 = .16$, and positively correlated with defendant credibility ratings $r = .56$, $p < .001$, $R^2 = .45$. The model predictions for JD credible were also positively correlated with ratings for guilt and JD credibility and negatively correlated with SC credibility. The model prediction for JD drunk was positively correlated with SC credibility but was not correlated with observed posteriors for guilt or JD credibility.

These results show that the BN model provides a good fit for individual participants (despite the simplicity of the BN model). The model predicts that the more credible JD is considered to be, the more unlikely it was that he was drunk and provoked SC, and this was upheld by the data. The R^2 value shows that the posterior predictions for provoke share some of the variability in the guilt and credibility judgments. The relatively small R^2 is most likely due to noisy participant' judgments, because we used a rich set of materials, and because provoke was used as a proxy for guilt. In addition, we only predicted the impact of consistency of selected pieces of evidence with the victim's testimony and did not model other elements of the case. Other factors were identified in closing arguments (e.g., the force of the punch) that might have been equally important for reasoning about the defendant's guilt. Nonetheless, the results of Study 2 demonstrate that it is possible to construct causal BNs to represent evidence that includes critical aspects of witness credibility, and then compare the model predictions with actual inferences.

General Discussion

The studies reported in this paper examined how the consistency of testimony affects inferences about witness credibility and judgments of guilt in a realistic legal reasoning task. Our results suggest two main conclusions. Firstly, that the consistency between key pieces of evidence and the victim's testimony, as identified in closing arguments, affect judgments about witness credibility and the degree of belief in the defendant's guilt. More specifically, belief in guilt was lower when key pieces of evidence mentioned in the testimonies of subsequent witnesses was inconsistent with details given in the victim's testimony. These inconsistencies damaged the credibility of victim and strengthened the credibility of the defendant. Secondly, the pattern predicted by the BN model captured qualitative patterns of inference displayed by participants at a general and individual level. In the consistent version of the case, participants thought the victim was credible, unlikely to be drunk, and unlikely to have provoked the defendant. Inconsistent statements lead to the opposite

¹ A prior of drunk = .50 was used for all the modeling.

pattern, the victim appeared less credible, more likely to be drunk and more likely to have provoked the defendant. Study 2 provides direct support for the reliability model described earlier.

The results of this study strengthen the claim that people's reasoning is sensitive to interrelations between testimony, credibility, and reliability that are predicted by the qualitative aspects of Bayesian network models. The BN captures people's intuitions that credibility is related to whether or not we think the victim's testimony is caused by the fact that he was indeed assaulted as he claimed or that he was drunk and provoked the defendant into a fight. It is therefore possible to model the strength of the links between the different items of evidence using the Bayesian network supplemented with legal idioms.

These results complement the story model and show that in addition to constructing a causal model of the crime people also construct causal models of the witnesses giving evidence and this affects their story evaluation. The results show that people do more than construct a plausible story to explain the evidence, but also take into account issues of credibility and reliability, such as the consistency of testimony. These results suggest that it is possible to extend the story model to include issues relating to credibility and reliability of evidence (Lagnado et al., 2013). In fact, these results strengthen Pennington and Hastie's claim that evidential reasoning is not a straightforward updating process as has been proposed by belief-adjustment models (Hogarth & Einhorn, 1992). This methodology fits with previous Bayesian approaches to modeling belief updating (e.g., Hahn & Harris, 2009) and by using Bayesian networks adds a richer structural account of people's reasoning.

The findings reported in this paper demonstrate that it is possible to model testimony integration using BNs. One potential shortcoming of the approach used here is that we used the model posterior probability for whether the victim provoked the defendant as a proxy for inferences about guilt. This could explain why the correlations between the model predictions and the observed posterior judgments were not stronger. Another reason that the model was not better able to capture people's inferences is that we only modeled some of the evidence in the case. More specifically, the BN focused on issues relating to the victim's testimony and did not include critical evidence relating to the defendant's testimony (e.g., the defendant was seen running away after the assault). Furthermore, the observed posterior judgments showed that consistency also influenced perceptions of the defendant's credibility, which was also not included in the BN. The BN model could readily be extended to include these factors.

Conclusions

In sum, this research shows that it is possible to model legal arguments using the BN framework and that this approach describes the qualitative patterns of inference exhibited in legal reasoning. Although it appears that people use

network-like structures to reason about evidence it is still necessary to develop a fuller psychological model of how people represent hypotheses and reason about uncertain evidence. Building upon the story model, and the insight that people use causal representations and inference, is a promising direction for further research.

Acknowledgements

We thank Prof. Cheryl Thomas who provided the transcript used in these studies, and Jacob de Zoete who assisted with modeling the data.

References

- Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *Journal of Applied Psychology, 81* (2), 170.
- Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior, 19* (1), 79.
- Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. MIT Press.
- Fenton, N., Neil, M., & Lagnado, D. A. (2013). A general structure for legal arguments about evidence using Bayesian networks. *Cognitive science, 37*(1), 61-102.
- Garbolino, P., & Taroni, F. (2002). Evaluation of scientific evidence using Bayesian networks. *Forensic Science International, 125*(2), 149-155.
- 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*(5), 1366.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology, 24*(1), 1-55.
- Judicial Studies Board (2010). *Crown Court Bench Book: Directing the Jury, London, UK*
- Lagnado, D. (2011). Thinking about Evidence. In *Proceedings of the British Academy*, 171, 183-223.
- Lagnado, D. A., Fenton, N., & Neil, M. (2013). Legal idioms: a framework for evidential reasoning. *Argument & Computation, 4*(1), 46-63.
- Lagnado, D. A., & Harvey, N. (2008). The impact of discredited evidence. *Psychonomic bulletin & review, 15*(6), 1166-1173.
- Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems*, Palo Alto, CA: Morgan Kaufmann.
- Pearl, J. (2000). *Causality*. New York: Cambridge University Press (2nd edition).
- Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51*(2), 242.
- Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of personality and social psychology, 62*(2), 189.
- Schum, D.A. (1994), *The Evidential Foundations of Probabilistic Reasoning*, Northwestern University Press.
- Spellman, B. A., & Tenney, E. R. (2010). Credible testimony in and out of court. *Psychonomic bulletin & review, 17*(2), 168-173.