UNIVERSITY OF CENTRAL OKLAHOMA

Edmond, Oklahoma

Jackson College of Graduate Studies

A Test of Bayesian and Story Models on Juror Decision

A THESIS

In partial fulfillment of the requirements

For the degree of

MASTER OF ARTS IN FORENSIC PSYCHOLOGY

by

Blake L. Nesmith

Edmond, Oklahoma

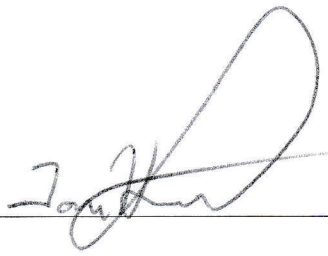
2019

A TEST OF BAYESIAN AND STORY MODELS ON JUROR DECISION

A THESIS

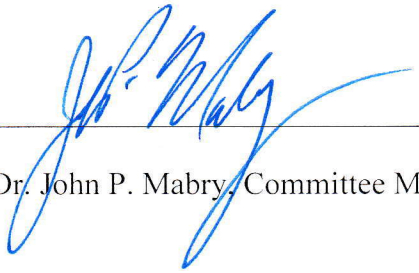
APPROVED FOR THE DEPARTMENT OF PSYCHOLOGY

July 12, 2019



Dr. Thomas Hancock, Committee Chairperson

Dr. Robert Mather, Committee Member



Dr. John P. Mabry, Committee Member

Abstract

Juror decision-making involves the individual's evaluation of evidence in court. There are two major types of models: mathematical and explanation. Both models make specific, contrasting assumptions of juror decision-making, but few studies have compared both models with the same experimental paradigm. The current study compared information integration/Bayesian models mathematical) to story models (explanation) using the same paradigm. A bias was introduced before evidence evaluation began. Participants read one of three narratives designed to bias their perception of the defendant towards a guilty, innocent, or neutral stance. Participants freely selected among six items of evidence (three eyewitnesses and three physical items), and each item had four additional sub-evidence items (two suggest guilt and two suggest innocence) they rated for guilt using a 10-point Likert scale. Results from the study suggest jurors use a combination of both models when evaluating evidence, but the methodology forced participants to use Bayesian methods, making it difficult to interpret. The bias introduced at the beginning of the study did not affect final verdict choice, and selection order did not affect the final verdict. However, there was a difference in selection order by condition, suggesting bias was present and influencing what items were selected. These findings suggest participants may have been inoculated against bias due to study methodology of making guilt ratings after each sub-evidence item.

Keywords: Decision-making, Bayesian, information integration, story models, juror, predecisional distortion

A Test of Bayesian and Story Models on Juror Decision

Most decision-making models originated from economics where the goal was to predict/explain how people choose among a set of options. One of the most popular models is expected utility theory, and it focuses on cost-benefit models where the decision maker should choose the option that results in the highest payoff (expected utility/value) for themselves (von Neumann & Morgenstern, 1944; Schoemaker, 1982). Models like expected utility assume the decision maker uses logic when assessing available options, especially when the outcome of the decision impacts them. However, these models are a poor fit for decisions where the decision maker must choose an option for another individual (e.g., person A must decide whether person B does X or Y). They were not designed for situations like this, and they raise questions about how a decision maker may choose for another person, such as in criminal trials. The Court Statistics Project (CSP) reported that 17.8 million criminal cases were reported in the United States during 2016 (CSP, 2018). Not all reported cases will go to trial. For each reported criminal case, there was an individual deciding if enough probable cause existed to bring the case to trial, if the case should be dismissed, whether the defendant was guilty or not guilty, or what the sentence should be if found guilty. Decision makers in these situations (i.e. jurors, judges) must decide on limited information and the decision will affect another person. Because the final decision has no obvious impact on the decision maker, it is important to understand how jurors evaluate evidence and what factors influence their evaluation and decision process.

Juror Decision Making Models

Traditional juror decision making models propose that jurors begin with an initial assessment and adjust the assessment as each new item is introduced. Anchoring and adjustment are common when the decision involves ambiguity, and each new item adjusts the initial

assessment (the anchor) toward certainty or uncertainty between two options (Einhorn & Hogarth, 1985). This is done in an attempt to remove ambiguity and determine the best decision given the information. However, not all juror decision-making models are based on this anchor and adjustment technique. There are two main juror decision making models that attempt to explain how jurors evaluate evidence: mathematical models (Bayesian and information integration models) and explanation models (story model) (McCullough, 2007). Bayesian, information integration, and story models are the three of the most known decision-making models, and they differ in their explanation of juror decision-making and how item weight (or perception of guilt) is combined to choose a verdict.

Bayesian models assume each item of evidence is independent of the next with no influence on subsequent items (Rust, Inman, & Zahorik, 1998). The weight of evidence is used to adjust the overall average of information and is evaluated independently, but Bayesian models consider the accuracy of items of evidence to represent the actual event (Einhorn, Hogarth, & Klempner, 1977; Schum & Du Charme, 1971). Jurors are faced with the question of whether the evidence presented in court is representative of events that transpired in commission of the crime. Jurors have no way to assess the credibility of evidence other than from testimony provided by both sides of the court. They may pull from prior knowledge or opinion to determine the weight of items to remove ambiguity about the information, but each item is assessed independently of the others (Marshall & Wise, 1975; Schum & Du Charme, 1971; Einhorn & Hogarth, 1985). Prior knowledge and context for some items may influence the weight of the item differently between jurors, thus affecting how they weigh the evidence (Kaye & Koehler, 2003). For example, juror A may weigh the testimony of an eyewitness less than juror B because they know eyewitness testimony is not very reliable. Although Bayesian models

consider the accuracy of evidence presented in court and juror attempts to make accurate determinations, they do not consider the influence of previous items on subsequent items presented during trial (Carlson & Russo, 2001); items are independent and assume no causality between them.

Information integration models assume that an averaging effect happens as items are introduced with the average value for guilty/not guilty being adjusted as additional items are introduced (Einhorn & Hogarth, 1985; Kaplan & Kemmerick, 1974; Ostrom, Werner, & Saks, 1978). Criminal trials usually involve some ambiguity about the crime, and jurors are often tasked with integrating individual pieces of information to determine the defendant's role, which makes use of the anchoring and adjusting process. Information integration models assume individual items of evidence are weighted differently in their indication of guilt/not guilty, and it is the varying weight of individual items that contribute to the overall average computed by jurors (Kaplan & Kemmerick, 1974; Schum & Du Charme, 1971). Jurors adjust their average as additional items are introduced to integrate multiple items, but item A does not influence the weight of item B or D, or vice versa. Each item's weight contributes to the overall average weight of guilty/not guilty without influencing the next item. Mathematical equations are a hallmark of Bayesian and information integration models and are predictive of juror decisions, but they do not incorporate any psychological concepts that may explain how jurors choose their final outcome.

Story model appears to address many of the factors ignored by information integration and Bayesian models and is one of the current leading models for juror decision making (Carlson & Russo, 2001; Pennington & Hastie, 1981; Willmott, Boduszeck, Debowska, & Woodfield, 2018). Other juror decision making models address different components of decision making

and tend to ignore courtroom procedures and juror tasks (Pennington & Hastie, 1981). Pennington and Hastie (1986) proposed the story model of juror decision making, suggesting jurors construct mental narratives of events that took place leading to the crime as items of evidence are introduced in trial to create a story of likely events that occurred. Multiple narratives may be constructed by jurors (i.e. guilty or innocent narrative) from evidence introduced in court, and each narrative is continuously adjusted as additional information is provided. As the trial progresses, jurors compare each narrative with possible verdicts and match the best explanation to a final verdict. Evidence presented in a story format has been shown to lead to better understanding of events, suggesting jurors attempt to combine individual items of evidence into a cohesive story to assist with decision making in criminal trials (Pennington & Hastie, 1992; Wiener, Richmond, Seib, Rauch, & Hackney, 2002). Although there is support for story model, it may be too simple to account for the complex nature of juror decision making because jurors have been found to use multiple story structures to construct narratives about mock murder cases (Wiener et al., 2002). The story model also cannot account for early verdict determinations in trials (Carlson & Russo, 2001). Additionally, items in court generally are not presented in chronological order and typically follow a question/answer format, which does not allow for easy or logical story construction.

Predecisional Distortion and Order Effects

There are other factors that may provide an explanation for how jurors arrive at their final verdict outside those proposed by decision making models. Graphic images, emotions, detail of eyewitness testimony, and the presence of facial tattoos affect jurors' decision making and influence final verdicts and sentences (Bright & Goodman-Delahunty, 2006; Douglas, Lyon, & Ogloff, 1997; Bell & Loftus, 1985). Many of these factors influence jurors without their

conscious awareness, functioning as implicit biases and affecting their motives and decisions throughout the trial. One such factor influencing jurors is the concept of predecisional distortion: a dominant stance towards guilty or not guilty is assumed early in trial by jurors and evaluation of subsequent evidence introduced following the development of the stance are done in such a manner where it corresponds with and supports the dominant stance (Carlson & Russo, 2001; Dahlstrand & Montgomery, 1984; Russo, Medvec, & Meloy, 1996). Predecisional distortion is a concern in trials because jurors are typically instructed to avoid making determinations of guilt until all evidence has been presented, and varying decision selection strategies may exist when the goals of each juror vary (Payne, Bettman, & Johnson,1988). If jurors create a dominant stance or preference towards one verdict at the beginning of a trial, decision strategies used later when evaluating evidence may differ between jurors with an opposite or neutral stance and may affect the final verdict.

It is important to note most legal systems in the United States follow set procedures that dictate when each side of a criminal trial is permitted to present their case and subsequent evidence, which poses challenges to the story model. It is uncommon for either side of a criminal trial to present evidence in chronological order, so courtroom procedures are not inherently conducive to the story model's goal of a cohesive narrative. Courtroom procedures appear to favor information integration and Bayesian models. However, courtroom procedures do allow jurors to combine individual pieces of evidence and testimony with their own thoughts and feelings to construct a narrative that makes sense to them (Pennington & Hastie, 1986), even if they must do so on their own during deliberation. The procedures are designed to eliminate any bias or advantage for either side of a criminal trial to ensure fairness, but presentation order has been found to influence verdicts in mock trials. Recency effects have been found when guilt

provoking evidence is shown last (Charman, Carbone, Kekessie, & Villalba, 2016; Enescu & Kuhn, 2012), suggesting it may be advantageous to present the most convincing evidence last, and some procedures may benefit one side over the other and may affect the final verdict or sentence (Englich, Mussweiler, & Strack, 2005).

Order effects become increasingly concerning when predecisional distortion is considered. Information integration and Bayesian models of juror decision making generally assume independence between items and do not place much weight on prior opinions in determining the average weight of guilty/not guilty information (Kaplan & Kemmerick, 1974; Marshall & Wise, 1975; Rust et al., 1998; Schum & Du Charme, 1971). They cannot account for predecisional distortion in their basic design because they do not assume causality between items and treat each item as an individual contributor. Decision selection strategies are also known to differ depending on the goal of the decision maker (Payne et al., 1988), and it is possible that predecisional distortion affects decision strategies used by the decision maker when the presentation order of evidence is varied. Story model may be able to explain order effects and predecisional distortion, although it has been argued the story model's reliance on prior verdict category knowledge cannot account for distortion effects until the very end (Carlson & Russo, 2001). Bayesian models do explain and predict individual item evaluation, but it cannot explain how the weight of item C influences item D's weight. Additionally, it is unclear how freedom of selection affects evidence evaluation and predecisional distortion when considering decision making models.

Current Study

The current study aims to investigate decision strategies that may arise during the evidence evaluation process while also investigating predecisional distortion within a mock

criminal trial. Bayesian, information integration, and story models will serve as the basic theories for comparison to determine whether one better explains mock juror decision making and can account for predecisional distortion. Because information integration and Bayesian models make similar predictions of independence, they will be combined for the purpose of this study, whereas the story model predicts an interdependence between items. Additionally, the study addresses previous literature regarding order effects on juror decision making and how freedom of choice may affect juror decisions regarding evidence evaluation. It addresses whether initial information about a criminal case prior to evidence evaluation affects subsequent evaluations and decision strategies while assessing freedom of choice on selection order. Participants served as mock jurors and evaluate main items of evidence and sub-evidence items from a mock court case used previously by Skolnick and Shaw (2001) and Seckinger (1992). Based on the aforementioned theories, if selection patterns are found within each narrative type the story model would be supported; however, if no selection patterns are found information integration and Bayesian models would be supported. In a similar vein, if a relationship is found between evidence guilt and selection order (i.e. guilt rating for the first selected item predicts subsequent items' guilt rating) the story model would be supported; no relationship between evidence ratings and selection order would support information integration and Bayesian models. It is predicted that the story model will be used over the Bayesian model, and that the story model will reflect predecisional distortion and order effects.

Method

Participants

One hundred and three participants were recruited from introductory psychology courses at the University of Central Oklahoma through Sona Systems for partial fulfillment of course

credit. Software crashes during data collection resulted in 93 participants for inclusion in data analysis (29 males, 64 females). Participants ranged in age from 18 to 28 years of age (M = 19.38, SD = 1.73). Forty-eight participants reported their ethnicity as Caucasian/White (51.6%), 16 reported being African American/Black (17.2%), 10 reported being Hispanic/Latino (10.8%), 7 reported being Native American/Alaskan Native (7.5%), 5 reported being Asian/Pacific Islander (5.4%), and 7 reported as Other (7.5%). Only one participant indicated they previously served on a jury.

Materials

A modified narrative from a hypothetical criminal case based on the work of Skolnick and Shaw (2001) and Seckinger (1992) is used as the mock crime for this study (see Appendix A). The criminal case involves a man (the defendant) who is charged with murdering his ex-wife, and six items are presented (three eyewitnesses and three physical items of evidence) that are used to assist with making a guilty/not guilty verdict. The narrative is modified to create three different stances towards the defendant by altering the length of his divorce from his ex-wife (recently divorced, divorced several years ago, or no mention of divorce). The narratives do not contain many details for the purpose of allowing participants to base their decision on the provided evidence. Actual photos are not included to eliminate any potential bias known to originate from graphic images (Bright & Goodman-Delahunty, 2006; Douglas et al., 1997). No images accompany the witnesses. Each item of evidence has four sub-evidence items that provide additional information about the specific item. Two sub-evidence items suggest guilt, while the remaining two sub-evidence items suggest the defendant may be innocent. A 10-point Likert scale is used to rate each sub-evidence item for guilt by sliding a bar either to the left (Not Guilty) or the right (Very Guilty). The sub-evidence is no longer visible after it is rated, and the

process is repeated until all sub-evidence items are selected and rated. The item of evidence relating to the sub-evidence is also no longer visible once its sub-evidence items are reviewed. A 10-point Likert scale is also used to rate decision confidence (left Not Confident, right Very Confident) once all items and sub-evidence have been reviewed. The ratings represent how their stance towards the defendant and overall guilt is changing throughout the study when viewing different items of evidence (see Appendix B for an overview).

Procedure

Participants are randomly assigned to read one of three narratives about the defendant and his preliminary guilt (Guilty, Innocent, or Neutral condition) prior to beginning the study. Informed consent is obtained from participants, and the researcher provides participants with verbal instructions to follow all instructions presented on the computer and to ask questions at any time. Participants are directed towards the computer to begin, and the researcher remains in the room while participants complete the study.

Participants advance to the next screen and read a brief narrative about the crime surrounded by six icons representing the six items of main evidence (stepmother's testimony, neighbor's testimony, bartender's testimony, knife, car, and footprint). These icons are visible at once to allow participants to choose freely among them. Participants decide the order icons are selected based on their perceived importance, beginning with the most important and ending with the least important. Each icon has four sub-evidence items that provide additional information about the selected item. Participants select sub-evidence in any order they wish, beginning with the most important and ending with the least important. Participants make a guilt rating for each sub-evidence on a Likert scale (left Not Guilty, right Very Guilty) before they may select the next sub-evidence. Once all four sub-evidence is viewed and the fourth guilt rating is made,

participants return to the original screen with the five remaining evidence icons. Participants repeat the process until all six icons are viewed and all sub-evidence items are rated.

Participants complete a distractor paper folding task after viewing all items and have two minutes to complete the task. Participants are directed back to the computer after the task and continue to the next screen to indicate how confident they are in their decision on a Likert scale (left Not Confident, right Very Confident). Participants advance to the next screen where they make a final guilty/not guilty verdict, followed by a debriefing statement. Participants complete a demographic questionnaire after reading the debriefing and asks the researcher any questions before they are dismissed. Selection order, guilt rating, final verdict, and response time are recorded throughout the study.

Results

Condition and Final Verdict

When collapsed across all three conditions, 59.14% of participants found the defendant guilty and 40.86% found the defendant innocent. A two-tailed chi-square test of independence was performed to analyze the relationship between condition (Guilty, Innocent, and Neutral) and final verdict outcome (Guilty and Innocent). There was no significant association between narratives read by participants and the final verdict outcome, $\chi^2(2) = .44$, $p = .803$ (see Appendix C for verdict outcomes). The bias introduced in the narrative at the beginning of the study did not sway participants toward one verdict over the other.

Main Evidence, Selection Order, and Condition

A two-tailed chi-square test of independence was performed to analyze the relationship between selection order (1^{st} – 6^{th}) and condition (Guilty, Innocent, and Neutral) among items of main evidence (see Appendix A). No significant associations were found between selection

order and condition (Selection Order 1: $\chi 2$ (10) = 13.96, p = .175; Selection Order 2: $\chi 2$ (10) = 16.34, p = .090; Selection Order 3: $\chi 2$ (10) = 18.31, p = .050; Selection Order 4: $\chi 2$ (10) = 4.76, p = .906; Selection Order 5: (10) = 10.83, p = .371; Selection Order 6: $\chi 2$ (10) = 12.45, p = .256). To correct for missing cell frequencies in the chi-square analysis, items of main evidence were collapsed by eyewitness testimony (stepmother, neighbor, and bartender) and physical evidence (knife, footprint, and car). A chi-square test of independence was performed to analyze the relationship between eyewitness testimony and physical evidence selection order (1^{st} – 6^{th}) by condition (Guilty, Innocent, and Neutral). A significant association was found between the second selected eyewitness testimony and physical evidence ($\chi 2(2) = 7.388$, $p < .05$; Cramer's V = .282). Appendix D shows the percentage that eyewitness testimony and physical evidence were selected by condition. In all three conditions, physical evidence was selected first; however, a different pattern was found for participants' second selection. For this second selection, a one-way MANOVA with a Gabriel post hoc test was performed to analyze the difference between condition (Guilty, Innocent, and Neutral) and type of main evidence (eyewitness testimony and physical evidence). A significant difference between condition and the second selected item was found ($F(2, 90) = 3.883$, $p < .05$; Wilks' $\Lambda = .921$). Physical evidence was selected second 34% of the time in the Guilty condition compared to 68% of the time in the Innocent condition, but there was no significant difference between the Neutral condition (52% physical evidence, 48% eyewitness testimony) and either group. The difference in selection patterns between the Guilty and Innocent conditions suggest that the bias introduced at the beginning of the study influenced how items were selected.

Selection Order, Sub-Evidence, and Guilt

Sub-evidence guilt ratings were combined to create an average rating for each selection order (1^{st} – 6^{th}) by sub-evidence type (Guilty and Innocent), resulting in the creation of 12 variables for comparison. The average guilt ratings were used for the following analyses. A 2 (sub-evidence type: Guilty and Innocent) x 6 (selection order: 1^{st} – 6^{th}) x 3 (condition: Guilty, Innocent, and Neutral) mixed MANOVA with a Gabriel post hoc test was performed to analyze the difference in guilt ratings. Mauchly's Test of Sphericity was violated for selection order (Mauchly's W = .702, $p < .05$) and sub-evidence type by order (Mauchly's W = .756, $p < .05$), and Greenhouse-Geisser was used. There was a significant interaction between sub-evidence type and selection order ($F(2, 90) = 3.57$, $p < .05$, $\eta^2 = .03$, Wilks' $\Lambda = .863$) on sub-evidence guilt, but no other significant interactions were found between the remaining variables. There was no significant difference by condition ($F(1, 90) = 1.467$, $p = .236$) or selection order ($F(4.43, 90) = 2.28$, $p = .053$; Wilks' $\Lambda = .9$) on sub-evidence guilt. There was a significant difference by sub-evidence type ($F(1, 90) = 298.66$, $p < .001$, $\eta^2 = .77$; Wilks' $\Lambda = .232$) on sub-evidence guilt. Guilty and innocent sub-evidence were rated differently within the conditions, confirming that they were successful in suggesting guilt or innocence of the defendant. Condition (Guilty, Innocent, and Neutral) and selection order (1^{st} – 6^{th}) did not affect guilt ratings, which shows no support for bias or order effects. However, selection order did affect guilt ratings when sub-evidence type was considered; innocent sub-evidence items increased in guilt across selection order (see Appendix E).

To see if guilt ratings differed for sub-evidence, a 2 (sub-evidence type: Guilty and Innocent) x 6 (selection order: 1^{st} – 6^{th}) x 2 (final verdict: Guilty and Innocent) mixed MANOVA with a Gabriel post hoc test was performed. Mauchly's Test of Sphericity was significant for

selection order (Mauchly's W = .71, p <.05) but not for selection order by sub-evidence type (Mauchly's W = .775, p = .065), and Greenhouse-Geisser was used. There was a significant difference by verdict (F(1, 91) = 33.49, p < .001, η^2 = .27) and sub-evidence type (F(1, 91) = 293.17, p < .001, η^2 = .76; Wilks' Λ = .237) on sub-evidence guilt. Guilty verdicts had higher guilt ratings (M = 5.38) than Innocent verdicts (M = 4.37), which is expected. Guilty and innocent sub-evidence were rated differently within the conditions, revealing the guilt and innocent sub-evidence influenced perception toward the defendant. Selection order, verdict, and sub-evidence type did not affect guilt ratings, revealing there were no order effects when considering verdict and sub-evidence type.

To determine whether a relationship existed between sub-evidence guilt ratings and selection order, a two-tailed Spearman's rho correlation was performed. Significant relationships were found between guilt ratings and selection order (see Appendix F). In general, guilty items tended to correlate with other guilty items, and innocent items tended to correlate with other innocent items. Bayesian and information integration models assume items have no relationship with each other, and the presence of relationships between multiple items of evidence is opposite of what is expected with these models. The presence of a significant relationship supports the story model and provides support for the likelihood of predecisional distortion being present.

Confidence and Final Verdict

A 2 (final verdict: Guilty and Innocent) x 3 (condition: Guilty, Innocent, and Neutral) ANOVA with a Gabriel post hoc test was performed to analyze the difference between confidence ratings. A significant difference was found between verdict (F(1, 87) = 5.926, p < .05), but not condition (F(2, 87) = .251, p = .778). Guilty verdicts had higher confidence ratings

(M = 6.6) than Innocent verdicts (M = 5.58). Condition (Guilty, Innocent, and Neutral) did not affect confidence ratings, suggesting the bias introduced at the beginning of the study did not influence participant confidence.

Discussion

The first goal of this study was to compare Bayesian/information integration and story models of juror decision-making with the same experimental paradigm to determine which is used by mock jurors. It was predicted that the story model is used over Bayesian and information integration models because it is better equipped to handle bias and assumes items of evidence are related to one another (Carlson & Russo, 2001; Pennington & Hastie, 1992; Wiener et al., 2002). Different selection patterns among items of main evidence were expected between the conditions, and relationships were expected to exist between item guilt ratings and selection order. There was no significant difference between conditions when looking at the first selected item, but there was a significant difference in selection pattern between the Guilty and Innocent conditions. Physical evidence was selected second 34% of the time (eyewitness testimony was selected 66%) in the Guilty condition, while physical evidence was selected 68% of the time (eyewitness testimony was selected 32%) in the Innocent condition. This suggest that different decision strategies were used during the decision-making process (Payne & Bettman, 1988), likely due to the bias originating from the narrative. Different selection patterns suggest participants had different goals, or possibly different narratives, between the Guilty and Innocent conditions. Story models assume multiple narratives are constructed during the decision-making process in an attempt to consolidate the information presented to them (Pennington & Hastie, 1992; Wiener et al., 2002), and this finding supports the story model. However, no significant differences were found between the first and third selection, and this suggests an independence

of items and condition, supporting the Bayesian/information integration assumption of independence (Einhorn & Hogarth, 1985; Marshall & Wise, 1975; Rust et al., 1998; Schum & Du Charme, 1971). Significant relationships were found between item guilt ratings and selection order across all conditions, suggesting an item's weight influenced the evaluation of another items. This also supports story over Bayesian/information integration models. Guilty and innocent sub-evidence items were rated differently within each condition and differed by verdict, supporting the Bayesian/information integration assumption that items have their own weight that contributes to the final average (Kaplan & Kemmerick, 1974; Rust et al., 1998; Schum & Du Charme, 1971). The results from this study suggest a combination of mathematical and explanation models are used, but the interpretation of results is more complex and may be attributable to the design of the study.

The methodology of the study required participants to make decisions after viewing each sub-evidence item (24 guilt ratings per participant). Story models make no assumption about item evaluation and how item weights are combined to reach a final verdict, but Bayesian and information integration models assume an averaging effect takes place as items are introduced. The results support the story model, but participants had to provide individual ratings for each sub-evidence item, thus requiring them to use the Bayesian model. The methodology allowed participants to select items of main evidence and sub-evidence in any order they chose, and this allowed participants to construct their own narrative and use the story method. However, the story model was overshadowed by the continued task of rating items. To test Bayesian models, individuals have to give ratings after each item, so the methodology used in this study was predisposed to support the Bayesian model. Interpretation of results is not clear because it cannot be determined with certainty what model was used by participants.

The second goal of the study was to investigate predecisional distortion and how it changed over time, while also accounting for order effects. The juror decision-making models assessed in this study make no overt claims about bias or how it might be reflected in their design. Bayesian and information integration models do not appear capable of accounting for bias (e.g., items are independent of one another), while story models, on the surface, appear better suited for it through the construction of narratives. No significant associations were found between condition and verdict, and guilt ratings also did not significantly differ by condition. The bias introduced prior to evidence evaluation did not distort participant perception of the defendant, suggesting that predecisional distortion was not present (Dahlstrand & Montgomery, 1984; Russo et al., 1996).

These findings suggest bias systematically varies throughout the decision-making process. One possible explanation is that the bias introduced in the narratives prior to evidence evaluation was not strong enough to sway verdict outcomes or guilt ratings. However, the difference in selection order between physical evidence and eyewitness testimony suggest the bias was present following the initial narrative. Another explanation stems from the methodology of the study. Because participants provided guilt ratings after each sub-evidence item (24 guilt ratings per participant), it is possible that the bias presented at the beginning of the study was weakened or eliminated. Bias occurs when individuals use quick judgements and do not critically evaluate the information presented to them. Participants had to evaluate each sub-evidence item, and it is possible that required ratings forced participants to engage in effortful thought and critically evaluate items, inoculating them from bias. The inoculation effect has been compared to antibodies and disease resistance (Grande, 2016), and it is the idea that exposing oneself to a bias builds a tolerance and removes future influence from it (McGuire &

Papageorgis, 1962). Frequent guilt ratings could have acted as an inoculation against the bias introduced at the beginning of the study, and it is possible that by the third selection order participants became inoculated (eight ratings are given by the third item of main evidence).

An interesting observation was the change in average guilt for the innocent sub-evidence from the first selection to the last selection (Appendix E). Guilty sub-evidence was rated consistently across selection order, and innocent sub-evidence had an increase in guilt across selection order (1^{st} – 6^{th}). Condition did not have any significant effects on the guilt ratings, but guilty and innocent sub-evidence were rated differently within the conditions, revealing they were perceived differently by participants. The increase in innocent sub-evidence guilt is likely not attributable to the bias introduced prior to evidence evaluation. Bilz (2010) argued for an exclusionary aspect of the story model, stating that information that serves no evidential value or is an attempt to create empathy towards the defendant should be/is excluded by the juror. The sub-evidence used in this study (see Appendix A) included contradictory information and information that could be perceived as an attempt to gather empathy. Participants may have evaluated the innocent sub-evidence in this manner and chose to ignore information that appeared to be unrelated to the crime itself (such as the defendant claiming he still loved his wife and wanted her back), and conflicting information between the eyewitness testimonies may have contributed to the increase in guilt. Conflicting information has been found to impact decisions (Simons & Green, 2013), which could be related to multiple narratives being constructed during evidence evaluation (Wiener et al., 2002). This could also be related to the difference in confidence ratings between Guilty and Innocent verdicts, but additional research is needed before any determinations can be made. However, this doe not explain why Guilty sub-evidence

remained stable in ratings, and the difference between them may stem from criteria needed to make a final verdict. Further research should explore this difference.

There are some limitations of the present study that should be addressed. First, the final verdict decision had Guilty/Innocent options, not Guilty/Not Guilty options. There is a legal difference between being found Guilty, Not Guilty, Innocent, and other verdict options. It is possible that this difference in verdict options negated the relationship between the narrative read prior to evidence evaluation and verdict. Second, participants did not receive any instructions about verdict criteria or how the verdicts differ, which is always provided to a jury before the start of a court trial. The effect of jury instructions should be investigated in future studies, especially due to the difference in verdict options (Connelly, 1999; Hope, Greene, Memon, Gavisk, & Houston, 2008) and between reasonable doubt and confidence (Cicchini & White, 2017; Horowitz & Kirkpatrick, 1996). Finally, limitation in interpreting the results of this study stem from the methodology. There is no other way to assess Bayesian decision-making models other than to have participants make ratings after each item. Doing this in the current study forced participants to use Bayesian models, so it is difficult to separate Bayesian and story models. To investigate whether the bias introduced at the beginning of the study is mitigated by conscious and effortful evaluation, guilt ratings should be removed.

Juror decision-making is complex with multiple interacting components. It is important to understand decision-making processes at the individual level so that group decision-making can be understood. There are many factors known to influence juries, and it is not known how juror decision-making models account for these influences or if they accurately can. Results from this study suggest a combination of mathematical and explanation models are used by mock jurors when evaluating evidence, but the methodology used does not allow for easy separation

between them. Results also suggest bias introduced prior to evidence evaluation may be diminished or removed through effortful thought via guilt ratings, and a comparison study with guilt ratings removed is needed to know for sure if an inoculation effect occurs.

References

Bell, B. E., & Loftus, E. F. (1985). Vivid persuasion in the courtroom. *Journal of Personality Assessment, 49*(6), 659–664.

Bilz, K. (2010). We don't want to hear it: Psychology, literature and the narrative model of judging. *University of Illinois Law Review, 2*, 429–487.

Bright, D. A., & Goodman-Delahunty, J. (2006). Gruesome evidence and emotion: Anger, blame, and jury decision-making. *Law and Human Behavior, 30*, 183–202.

Carlson, K. A., & Russo, J. E. (2001). Biased interpretation of evidence by mock jurors. *Journal of Experimental Psychology: Applied, 7*(2), 91–103.

Charman, S. D., Carbone, J., Kekessie, S., & Villalba, D. K. (2016). Evidence evaluation and evidence integration in legal decision-making: Order of evidence presentation as a moderator of context effects. *Applied Cognitive Psychology, 30*, 214–225.

Connelly, C. (1999). Courts. In P. Duff & N. Hutton (Eds.), *Criminal justice in Scotland* (pp. 146–165). Aldershot, England: Dartmouth Publishing.

Court Statistics Project (CSP). (2019). Total incoming criminal caseloads reported by state courts, all states, 2007 – 2016. Retrieved fromhttp://www.courtstatistics.org/~/media/ Microsites/Files/CSP/ Criminal/PDFs/EWSC-2016-CRIM-Page-1-Trend.ashx

Dahlstrand, U., & Montgomery, H. (1984). Information search and evaluative processes in decision making: A computer based process tracing study. *Acta Psychologica, 56*, 113–123.

Douglas, K. S., Lyon, D. R., & Ogloff, J. R. P. (1997). The impact of graphic photographic evidence on mock jurors' decisions in a murder trial: Probative or prejudicial? *Law and Human Behavior, 21*(5), 485–501.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review, 92*(4), 433–461.

Einhorn, H. J., Hogarth, R. M., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin, 84*(1), 158–172.

Enescu, R., & Kuhn, A. (2012). Serial effects of evidence on legal decision-making. *The European Journal of Psychology Applied to Legal Context, 4*(2), 99–118.

Englich, B., Mussweiler, T., & Strack, F. (2005). The last word in court – a hidden disadvantage for the defense. *Law and Human Behavior, 29*(6), 705–722.

Grande, E. (2016). Legal transplants and the inoculation effect: How American criminal procedure has affected continental Europe. *American Journal of Comparative Law, 64*(3), 583–618.

Hope, L., Greene, E., Memon, A., Gavisk, M., & Houston, K. (2008). A third verdict option: Exploring the impact of the not proven verdict on mock juror decision making. *Law and Human Behavior, 32(*3), 241–252.

Horowitz, I. A., & Kirkpatrick, L. C. (1996). A concept in search of a definition: The effects of reasonable doubt instructions on certainty of guilt standards and jury verdict. *Law and Human Behavior, 20*(60), 655–670.

Kaplan, M. F., & Kemmerick, G. D. (1974). Juror judgment as information integration: Combining evidential and nonevidential information. *Journal of Personality and Social Psychology, 30*(4), 493–499.

Kaye, D. H., & Koehler, J. J. (2003). The misquantification of probative value. *Law and Human Behavior, 27*(6), 645–659.

Marshall, C. R., & Wise, J. A. (1975). Juror decisions and the determination of guilt in capital

punishment cases: A Bayesian perspective. In Wendt, D., & Vlek, C. (Eds.), *Utility, probability, and human decision making*. Dordrect, Holland: Reidel.

McCullough, G. W. (2007). Function of text structure in jurors' comprehension and decision making. *Psychological Reports, 101*, 723–730.

McGuire, W. J., & Papageorgis, D. (1962). Effectiveness of forewarning in developing resistance to persuasion. *Public Opinion Quarterly, 26*(1), 24–34.

Ostrom, T. M., Werner, C., & Saks, M. J. (1978). An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology, 36*(4), 436–450.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534–552.

Pennington, N., & Hastie, R. (1981). Juror decision making models: The generalization gap. *Psychological Bulletin, 89*(2), 246–287.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *52*(2), 242–258.

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), 189–206.

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavioral and Human Decision Processes, 66*(1), 102–110.

Rust, R. T., Inman, J., & Zohorik, A. (1998). A Bayesian model of quality and customer retention. *Marketing Science, 18*, 77–92.

Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence and

limitations. *Journal of Economic Literature, 20*(2), 529–563.

Schum, D. A., & Du Charme, W. M. (1971). Comments on the relationship between the impact and the reliability of evidence. *Organizational Behavior and Human Performance, 6*, 111–131.

Seckinger, J. H. (1992). *State v. O'Neill, problems and case file* (5th ed.). Notre Dame, IN: National Institute for Trial Advocacy.

Simons, J. P., & Green, M. C. (2013). Distracted by details: Narrative influence following conflicting stories. *Media Psychology, 16,* 221–243.

Skolnick, P., & Shaw, J. I. (2001). A comparison of eyewitness and physical evidence on mock-juror decision making. *Criminal Justice and Behavior, 28*(5), 614–630.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wiener, R. L., Richmond, T. L., Seib, H. M., Rauch, S. M., & Hackney, A. A. (2002). The psychology of telling murder stories: Do we think in scripts, exemplars, or prototypes? *Behavioral Sciences and the Law, 20*, 119–139.

Willmott, D., Boduszek, D., Debowska, A., & Woodfield, R. (2018). Introduction and validation of the Juror Decision Scale (JDS): An empirical investigation of the Story Model. *Journal of Criminal Justice, 57*, 26–34.

Appendix A

Narratives and Items of Evidence

N1 (Guilty): A 911 call was made at approximately 10:00pm from a woman stating her daughter had been attacked. The victim died due to multiple stab wounds to the body. No suspect is immediately found. During a preliminary interview with the stepmother, detectives learn the victim recently went through a divorce. The ex-husband is later charged with the crime. Three eyewitnesses agreed to give testimonies, and there are three physical pieces of evidence related to the case.

N2 (Innocent): A 911 call was made at approximately 10:00pm from a woman stating her daughter had been attacked. The victim died due to multiple stab wounds to the body. No suspect is immediately found. During a preliminary interview with the stepmother, detectives learn the victim had been separated from her husband for a long time and they were divorced three years ago. However, the ex-husband is later charged with the crime. Three eyewitnesses agreed to give testimonies, and there are three physical pieces of evidence related to the case.

N3 (Neutral): A 911 call was made at approximately 10:00pm from a woman stating her daughter had been attacked. The victim died due to multiple stab wounds to the body. No suspect is immediately found during a preliminary interview. However, the ex-husband is later charged with the crime. Three eyewitnesses agreed to give testimonies, and there are three physical pieces of evidence related to the case.

Items of Evidence

Eyewitness #1: Victim's stepmother

Victim returned home around 10:00 p.m. from the movies the night of the murder. Her stepmother heard commotion outside and saw a man attack her daughter. He immediately fled in a dark blue automobile identified as one similar to what the defendant owns.

- The porch light was on and there was a full moon lending to an easier identification of the suspect
- Due to the extended period of time that the stepmother has spent with her step-daughter's ex-husband, she confidently identified his voice and stature
- When the stepmother took a physical examination, it was found that she needed glasses. She reported that she was not wearing them during the time of the murder; therefore, her vision was not perfect
- She was emotionally distressed during questioning which potentially influences the ability to engage in efficient recall.

Eyewitness #2: Neighbor

The neighbor from across the street corroborated the stepmother's story. She heard screaming from the stepmom and got out of bed and ran to the window to look. She got to the window in time to see a man carrying an object run to a dark car and drive away.

- The neighbor also took physical examinations and was found to have excellent vision.
- The neighbor reported clearly recognizing the defendant's build and hair style. He had lived across the street for a number of years, and they have had numerous encounters.
- The neighbor was groggy during questioning due to sleeping pills taken earlier that night. The grogginess may affect perception.
- Due to the time of night, she could not positively identify the exact color or make of the car from her vantage point across the street.

Eyewitness #3: Bartender

The bartender said that he had seen the defendant in the bar multiple evenings a week around the same time. There was nothing unusual about tonight. He was there for a few hours and left around 9:15 p.m. He didn't mention where he was going once he left.

- The defendant complained about his failed marriage throughout the night at the bar.
- The bartender was very confident in his identification due to the fact that the defendant had become a regular.
- The bartender said the defendant was wearing a different outfit than what the stepmom and neighbor identified the suspect as wearing.
- The bartender said that the defendant had previously mentioned how he wanted his wife back and still loved her.

Physical evidence #1: Knife

- Based on entrance wounds in the body, the knife was determined to be a 6-inch serrated hunting knife. The same sort of knife was found at defendant's house.
- The knife found at his house was just recently cleaned.
- The entry wounds in the body had been probed during an earlier autopsy so the precise comparison with the serrated knife was impossible.
- The defendant was an avid hunter, and a knife of that type would not be uncommon to find in the house of any hunter.

Physical evidence #2: Footprint

- A photograph of a bloody shoeprint was taken from the crime scene, and the prints matched shoes that the defendant owns.
- Partial fingerprints were taken from the railing around where the shoeprint was found, and some matched the defendant's. However, he used to live there, so it can't be determined to be from that night.
- This type of shoe could also belong to thousands of other men.
- No traceable blood was found on the defendant's shoes.

Physical evidence #3: Car

- Tire marks found on the road outside of the victim's house match the tread from factory installed tires, which are the same type of tires found on the defendant's car.
- Soil taken from crime scene had similar characteristics with dirt that was taken from his car a few days after the murder.
- Someone else had stated seeing his car across town around the same time as the murder which would place him in different location and unable to commit the crime.
- The defendant claims that the dirt could have been from any time before or after the murder. He's always around that area hunting.

Appendix B

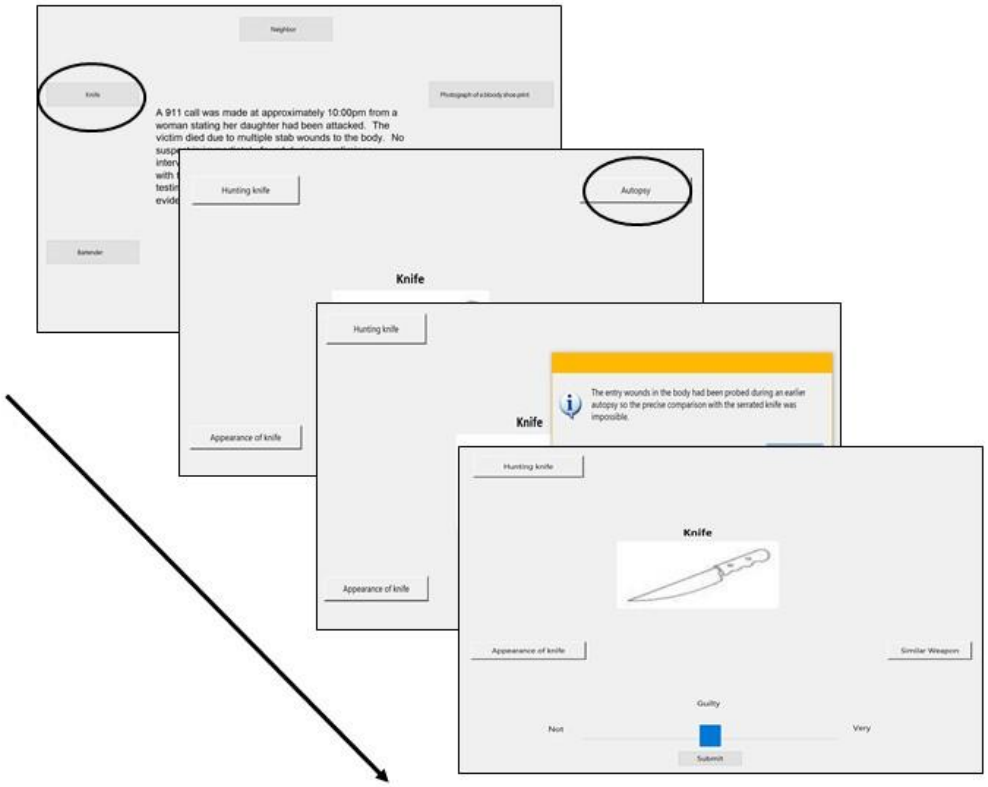


Figure 1. Example experimental overview of the neutral narrative with the knife selected first, followed by the autopsy sub-evidence. The process is repeated for all sub-evidence and main items.

Appendix C

Table 1
Verdict Outcomes by Condition

	Condition			
Verdict	Guilty (n = 35)	Innocent (n = 31)	Neutral (n = 27)	Total (N = 93)
Guilty	22	17	16	55
Innocent	13	14	11	38

Note. χ^2 (2) = .438, p = .803 (two-tailed).

Appendix D

Table 2
Percent Physical and Eyewitness Evidence Selected by Condition

	Selection Order					
Type	1	2	3	4	5	6
Guilty						
Physical	86%	34%	57%	34%	49%	40%
Eyewitness	14%	66%	43%	66%	51%	60%
Innocent						
Physical	68%	68%	45%	48%	35%	35%
Eyewitness	32%	32%	55%	52%	65%	65%
Neutral						
Physical	70%	52%	48%	41%	48%	41%
Eyewitness	30%	48%	52%	59%	52%	59%

Note. Guilty (n = 35); Innocent (n = 31); Neutral (n = 27).

Appendix E

Table 3

Sub-Evidence Guilt Rating by Selection Order

	Selection Order											
	1		2		3		4		5		6	
Type	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Innocent	3.73	1.62	3.56	1.77	3.97	1.7	3.85	1.62	3.92	1.69	4.35	1.64
Guilty	5.82	1.94	5.89	1.66	6.09	1.85	6.15	1.78	6.28	1.74	6.02	1.53

Notes. $N = 93$.

Appendix F

Table 4

Correlation Coefficients between Selection Order Guilt Ratings Across Conditions

	1st Guilty	1st Innocent	2nd Guilty	2nd Innocent	3rd Guilty	3rd Innocent	4th Guilty	4th Innocent	5th Guilty	5th Innocent	6th Guilty
1st Innocent	.262*										
2nd Guilty											
2nd Innocent		.312**	.284**								
3rd Guilty			.301**								
3rd Innocent				.445**	.380**						
4th Guilty			.362**		.282**	.242*					
4th Innocent		.295**		.567**		.423**	.262*				
5th Guilty			.243*	.243*	.391**		.458**	.304*			
5th Innocent		.295**		.386**		.430**		.522**	.328**		
6th Guilty	.209*						.213*	.247*	.372**	.225*	
6th Innocent		.249*		.395**		.346**		.438**	.31**	.490**	.517**

Note. $N = 93$, $^{*}p < .05$ (2-tailed), $^{**}p < .001$ (2-tailed).