MODELING JUROR DECISIONS: A COMPARISON OF PERCEPTIONS OF INNOCENCE AND GUILT

by

Sami El-Sibaey, MA, Ryerson University, Toronto, June 2010

B.Sc. (Hons.) University of Toronto Scarborough Campus, Toronto, June 2007

A dissertation

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Program of

Psychology

Toronto, Ontario, Canada, 2014

© Sami El-Sibaey, 2014

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A DISSERTATION

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

Modeling Juror Decisions: A Comparison of Perceptions of Guilt and Innocence

Doctor of Philosophy, 2014

Sami El-Sibaey

Psychology, Ryerson University

The research in this dissertation investigates the consistency of juror decision models when evaluating incriminating and exonerating evidence. Many stochastic and psychological models indicate that an interaction between a person's prior beliefs and their evaluation of the evidence contribute to their verdict decision. However, less is known regarding how this interaction occurs for different forms of evidence. In particular, a pervasive assumption is that jurors use the same or similar models to evaluate exonerating and incriminating evidence. The data from this dissertation indicates that this may not be the case. Participants adjust estimates of probability of guilt in a Bayesian inference problem more when case specific evidence is incriminating versus exonerating. Further, their response patterns and reported and observed measures of the type and variety of information they are using to produce probability of guilt estimates indicate that they engage in a process of mental estimation more often than they report that they do. The findings indicate that jurors may potentially use different decision models to evaluate different forms of evidence. Further, the framing of the search for culpability provides a plausible explanation for differences in the decision models that are used. Specifically, a juror's selection criteria and perceived importance of a given piece of evidence will vary depending on its relevance to their decision task. Thus, asking jurors to estimate likelihoods of guilt may lead to their underutilization of evidence implying innocence.

Acknowledgments

I think I may be the luckiest person that I know, and it's due to the amazing people I've had the privilege of sharing my experience with. No words of thanks will suffice, but I will do my best.

Thank you to my blood family. My mother (Jomana), father (Samir), and brother (Rami) have been an infinite source of love, learning and growth. My family in Syria and Egypt and other parts of the world are a constant reminder of the value of life and the passion it takes to smile every morning. You are forever in my heart.

Thank you to my academic family. Thank you to my advisor Dr. Tara Burke whose support put me in a position to write this letter (the probabilistic likelihood would have been infinitely smaller otherwise). Thank you to all the wonderful educators that have influenced my world view. Dr. Ben Dyson, Dr. Todd Girard, Dr. Arne Kislenko, Dr. Gary Wells, Dr. Stephen Want, Dr. Frank Russo, Dr. John Turtle, Dr. Maria Gurevich, Dr. David Day, Dr. John Kennedy, Dr. Steve Joordens, and many more who nurtured my thirst for learning. To my extended "I got your back" academic family, Tara Stallberg, Lisa Vuong, Francesca Grillo, and Carson Pun, thanks for, well, having my back. You are forever in my heart.

Thank you to my world family and my sports family. Cici, Adam, Maryam(x2), Abdo, George, Meera, Kathleen, Steph, Sara, Stuart, Holly, Andrea, Sarah, Raj, Pete, Bob, Umar, Abdul, Mike, Angela, Mirza, Rami, Pierre, and everyone I couldn't name here, the laughs that we share are the fuel to my existence. You are forever in my heart.

I could not have imagined a better village and I can't wait to meet the rest of the family. You are all forever in my heart.

Dedication

For my uncle Ali Issa

"Better to light a single candle, than to curse the darkness"

- Common proverb

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1: A Case of Wrongful Conviction

The research in this dissertation was motivated by stories of wrongful conviction and the research pertaining to analyzing, understanding, and minimizing this phenomenon of legal decision making. In turn, the literature review will begin with a summary of a case study that was integral to bringing many issues of wrongful conviction to light in Canada and follow with a more specific focus on certain aspects of the case pertaining to witness testimony. Specifically, the literature review will focus on investigating various models used to explain juror decisions in the context of understanding their impact on how jurors select and integrate information, and how this process may contribute to wrongful conviction.

In 1985 Guy Paul Morin was charged with the rape and murder of Christine Jessop, the 9-year old daughter of his next door neighbour. Morin, 25 years old at the time, was a furniture maker with no previous criminal record who lived with his parents. The charges against Morin led to him spending over two years in prison and more than ten years dealing with the legal, psychological and social stressors of his case, after which he was finally exonerated and acquitted in 1995 through the use of new DNA evidence. Morin was convicted primarily based on four factors: 1) the testimony of two jailhouse informants who claimed they overheard him confess to the crime while in prison, 2) fiber analysis linking Morin to the crime, 3) evidence related to his "consciousness of guilt" (essentially implying that his demeanor suggested that he was guilty of the crime), and 4) testimony from Christine Jessop's mother and brother that contradicted Morin's alibi and allowed for the possibility that he committed the crime. He was exonerated almost solely based on the DNA analysis of a semen sample obtained from Jessop's dress which did not match with his DNA. In 1996 a public inquiry into the wrongful conviction of

Guy Paul Morin was launched with the help of the “Justice for Guy Paul Morin Committee”, a group that eventually became the Association in Defense of the Wrongly Convicted(AIDWYC). The inquiry revealed several troubling factors.

The results of the fiber analysis that implicated Morin were found to be from a contaminated lab sample which the analyst knew was contaminated but failed to report to the court at the time the evidence was presented. The jailhouse informants also recanted their testimony several times at Morin’s appeal trial, leading the presiding judge to determine that their testimony was unreliable. Indeed, one of the informants later admitted to fabricating his testimony. The admissibility of “consciousness of guilt” as evidence is still a highly debated issue and entirely dependent on the subjective perceptions of the triers of fact. In the specific case of Guy Paul Morin, the “Report of the Kaufman Commission on Proceedings Involving Guy Paul Morin” clearly indicates that much of the evidence used to imply that Mr. Morin’s personality, testimony, and interactions with witnesses made him look suspicious should not have even been admitted (Ministry of the Attorney General [MAG], 2014). In addition, knowing that certain actions are illegal and may lead to the deduction of a guilty verdict does not necessarily imply a desire to commit those illegal actions on the part of the suspect. In turn, any evaluation of a suspect’s conduct or demeanor being consistent with guilt may be the result of tunnel vision due to the charges the suspect is facing rather than an evaluation of their actual guilt (MAG, 2014). In Morin’s case, this is further highlighted by the fact that his acquittal at his first trial was appealed by the Crown on grounds that “reasonable doubt” was not well defined to the jury. The institutional pressure to search for and find guilt in Morin’s case was so strong

that this appeal was upheld despite the presence of other suspects who were known to the police and the fact that Morin's alibi evidence should theoretically have exonerated him.

Finally, the original testimony obtained from Christine's mother and brother presented a different timeframe than the testimony presented by the police as evidence at his first trial. While Janet and Ken Jessop's testimony indicated that they arrived home at 4:10 p.m., Morin's work time card indicated that he left work at 3:32 p.m. A timing run from his work to his house took the police 42 minutes putting Morin at his house at 4:14 p.m. Morin also claimed he had gone grocery shopping with his parents the afternoon that Christine was reported missing. However, if he left work when he said, it would have been impossible for him to have returned home in time for him to murder Christine Jessop. Morin testified that he arrived home that afternoon between 4:30 p.m. and 5 p.m. Although this account matched the family's original testimony, Christine's mother and brother revised their testimony's time-line after being questioned a second time by the police and indicated that they arrived home between 4:30 p.m. to 4:35 p.m. or even later. This made it possible for Morin to have committed the crime. Further, the police had no recorded or written account of the second interview and Christine's family reiterated at Morin's appeal trial that they were quite confident in their original statement but coerced by the second police interrogation to provide a window of opportunity.

A particular oversight in Morin's case that is of interest for this dissertation is the failure of the police to appropriately evaluate and admit alibi evidence that implied his innocence 10 years before his DNA exoneration. As a result of the conflicting testimonies of Janet and Ken Jessop, Morin's alibi was overlooked as evidence for his innocence. The significance of this oversight was not apparent until DNA evidence provided further support for his case and

Christine's family indicated that they were coerced by investigators to provide testimony that challenged Guy Paul Morin's alibi. Guy Paul Morin's story highlights many of the challenges that legal scholars are attempting to identify and address in order to minimize the chances of wrongful conviction and general miscarriages of justice. Many authorities now require police interviews to be taped or recorded and interrogation strategies have received a lot of interest from researchers hoping to limit the likelihood of coerced confessions, false statements, or fabricated memories (e.g., Kassin et al., 2007). General interest in the utility and proper evaluation of diagnostic tests like DNA analysis has also increased and there is more awareness in courts of their strengths and weaknesses (Hampikian, West, & Akselrod, 2011). Yet Morin's case also presents a deeper and somewhat more abstract challenge that relates to the manner in which evidence for his case was evaluated, and the decision to convict him was reached. If compelling evidence indicating Morin's innocence (in the form of his alibi statement) existed at the start of his trial, then the question is why it took over ten years and additional DNA evidence to prove his innocence. I believe this relates to the difficulty in integrating the many forms of evidence presented to a trier of fact to form a coherent argument for guilt or innocence. Some aspects of this challenge will be further discussed later in the context of juror decision models.

While there has been great effort in advancing policy change through a focus on procedural modifications to the way we provide and evaluate various types of evidence (e.g., Wells et al.'s 1998, recommendations for lineup identification and photospread procedures), only recently have researchers also started to gather momentum on understanding how the nature and type of evidence can have an effect on its use and impact by and on jurors (e.g.,

Bornstein & Greene, 2011; Devine, 2012). Specifically, the research for this dissertation will examine the kind of decision models that jurors use when evaluating evidence for innocence and guilt and whether they could be using different models depending on the nature of the evidence. Thus, the focus of the empirical research will be first to understand what kind of decision models (e.g., logical, Bayesian, story based) jurors may use in their quest to evaluate and integrate different forms of evidence, and second to understand why they choose the models that they do. The document will begin with an overview of current research on decision models and how they can be applied to a jury context. This will include a discussion of the trial process at three distinct levels: 1) the juror, 2) the evidence, and 3) the trial.

Chapter 2: Literature Review

Notes on wrongful conviction

The current research project was motivated by the desire to explore our understanding of the leading causes of wrongful conviction. To that end, two primary sources for case file analysis include AIDWYC, a Toronto based organization founded in 1993 (AIDWYC, 2013), and the Innocence Project, a New York based organization founded in 1992 (Innocence Project, 2013). Between them, the two organizations have contributed to the exoneration of over 300 victims of wrongful conviction over the past 20 years (311 in the USA and 20 in Canada with over 100 cases currently being reviewed). With an average of 15 exonerations per year, the focus of these organizations is addressing the egregious errors in the trial, police, and evidence processes that contribute to wrongful conviction. Thorough analyses of over 200 Innocence Project cases reveals that the primary contributors to wrongful conviction are misidentification (by a witness or the victim), false confessions, faulty forensic evidence and expert testimony, and jailhouse or government informant testimony (Garret, 2008; Hampikian et al., 2011).

One less reported but important contributing factor is the presence of a "weak alibi". Alibi evidence refers to testimony by the suspect or a witness on their behalf that the suspect was present somewhere other than the scene of the crime at the time it occurred. Generally, alibi testimony is considered weak if it changes over time, is perceived as inconsistent, lacks any physical evidence (e.g., security camera footage), or is provided by a witness that is related to the suspect (Culhane & Hosch, 2004; Olson & Wells, 2004; Olson 2013). In their analysis of 40 Innocence Project cases Wells et al. (1998) found that a suspect's statement that he/she was somewhere other than the scene of the crime at the time it occurred (and so could not possibly

be responsible) was not strong enough evidence to overturn eyewitness testimony against them. In fact, the presence of a "weak alibi" was often cited as a reason for conviction. A review of cases involving individuals convicted of crimes and later exonerated through DNA evidence found that of 157 cases, over 25 percent included an alibi that was not believed (Burke, Turtle & Olson, 2006). These findings highlight an interesting point that I will return to in that the observation that a suspect has no argument for innocence does not necessitate that that suspect must be guilty.

Defining the scope of wrongful conviction in North American legal systems is challenging because most efforts to do so focus on cases involving violent crime that contain some form of DNA evidence (Garrett, 2008). More recently, cases including other forms of forensic evidence such as fabric analysis have also been investigated (e.g., Hampikian et al, 2011) but our understanding of the reasons and scope of wrongful conviction is still limited to a narrow area of legal cases. These limits are primarily due to the difficulty in identifying cases of substantive (as opposed to procedural) wrongful conviction and to inconsistencies in record keeping and analysis methodologies related to the study of appeal cases, exonerations and inmate surveys (Poveda, 2001). That being said however, Morin spent over 2 years in prison and over 10 years in various criminal and appeal courts and the Innocence Project (2013) estimates the average sentence served by their exonerees is almost 14 years. While the psychological and social impact on an individual who spends that amount of time behind bars is enormous, there is also an economic impact of maintaining such a margin of error.

The fact that organizations exist to help these individuals makes it clear that an effort is required to monitor, identify, control, and address error in legal decision making. This relates to

limiting the incidence of wrongful conviction by controlling the number of innocent individuals found guilty and the number of guilty individuals set free. Incidentally, convicting a substantively innocent person increases the rate of both these errors. An innocent person is sent to jail while a guilty person is still free to roam the streets. This may in part explain why case files from AIDWYC and The Innocence Project have provided the most relevant platform for discussing issues relating to wrongful conviction. Beyond the extensive analysis of normative data, there is also a drive for focus on applied problem solving that can lead to the implementation of policy change and on empirical research guided by theoretical frameworks that can help us understand the trial process and mechanisms that decisions of guilt or innocence operate through. Such frameworks will help clearly define the context of wrongful conviction within our understating of models of legal decision making. This will comprise the main component of the discussion to follow. Specifically, the current dissertation aims to contribute to a growing body of literature investigating the interaction between the juror as a trier of fact and the evidence as the main body of information contributing to evaluations of culpability. The research will be discussed in the context of how this interaction may affect a juror's decision models and decisions.

Equation Based Models for Understanding Juror and Jury Decisions

Understanding the interaction between juror and jury decisions provides some context for interpreting optimal models of juror decisions. For example, Penrod and Hastie (1979) provide an interesting review and analysis of several proposed models for jury decision making in which initial juror ballots in combination with factors such as jury size, and implicit or explicit

decision rules relating to the unanimity of a jury panel can be used to predict the accuracy of jury decisions. In their review, Penrod and Hastie (1979) outline a series of models that "explore the relationship among the initial distribution of individual juror opinions, individual difference in persuadability, rates of juror vote changes, coalition sizes within a jury, deliberation times, and the distribution of jury verdicts" (p.462). In relation to jury decisions, they emphasize that a key factor for researchers interested in better understanding this process is to apply some form of decision rule that can relate pre-deliberation juror ballot distributions with final verdicts. Particularly, they stress the utility of a decision rule that can consistently predict final verdicts under varying initial distributions such as when a panel may be unanimous, split 50/50, or under majority influence.

In an analysis of mock jury verdicts for a simulated rape trial, for example, Davis and colleagues (Davis, Kerr, Atkin, Holt & Meek, 1975) found that a model which uses a two-thirds decision rule seems to provide the best fit for predicting verdicts from initial ballots. Such a model predicts that if at least two-thirds of jurors agree on a verdict (conviction or acquittal), this will likely determine the verdict in a case (otherwise a jury is hung). Further, Davis et al. found that the mock jurors reached a verdict decision that matched the initial ballots of a two-thirds majority even when they were instructed to only give a verdict if everyone was in agreement, and this was true of juries composed of six or twelve members. Interestingly, the research found that even though verdict distributions did not vary depending on jury size or the decision scheme assigned to the jury, they were affected by the time taken for deliberation and the number of polls taken during this process. Specifically, twelve person juries took longer to deliberate and took more polls than six person juries when both groups were asked to use

unanimity rule. However, when the groups were asked to use a two-thirds majority rule, twelve person juries actually took less time to deliberate than six person juries and took fewer polls. This may be a reflection of the effects of conformity or group influences that are dependent on the actual size of the group (rather than just the verdict distributions) and could also imply that a primary manner in which juries reflect the opinions of individual jurors is by allowing a context through jury deliberation for those individuals (or at least a two-thirds majority of them if its exists) to persuade their peers of a certain perspective. In their analysis of the implications of jury decisions models for and from psychology, Bornstein and Greene (2011) also indicate that the distribution of pre-deliberation verdicts is the strongest predictor of jury verdicts and that in approximately 90% of trials, the position favoured by the majority when a trial begins eventually becomes the jury verdict.

Penrod and Hastie (1979) further outline that probabilistic models for jury verdicts, such as that proposed by Davis et al. (1975), tend to consistently rely on the key assumptions that individual juror decisions are independent of each other and random (in the sense that a juror randomly selected from a large jury pool will have an equal chance as any other juror to vote for guilt). Further, the applicability of these models is reliant on the binary nature of the verdict decision in that a juror only has the choice to vote either guilty or not guilty. Walbert's (1971 as cited in Penrod & Hastie, 1979) analysis of such probabilistic models indicates that they need to primarily address only two key factors: 1) the probability that a jury will produce a majority on the first ballot, and 2) the disposition of cases in which the jury evenly splits on the first ballot. This, according to Walbert's analysis, is because majority persuasion operates in the vast margin (93%) of cases and juries that split tend to evenly divide between conviction and

acquittal. In other words, the above mentioned parameters seem to provide the most ecologically valid assumptions for a binomial probabilistic model for juror/jury verdicts. Further, Walbert found these assumptions while moderated by the decision rules implied by Davis et al. lead to a model that is highly influenced by jury size. Particularly, this model predicts that a bigger jury is quite advisable as smaller juries will not only convict more low-guilt suspects, but also acquit more high-guilt suspects, if the probability of any randomly selected juror to vote for guilt is held constant.

This is challenged to some extent by a recent report funded through the National Center for State Courts and the National Institute of Justice in the United States of America. The report indicates that while hung jury rates in state courts from thirty different jurisdictions and all federal courts averaged at around 6.2% (in other words around 94% of trials reach a verdict), these rates can range between 5-35% across different jurisdictions (Hannaford-Agor, Hans, Mott & Munsterman, 2002). It is clear that while Walbert's analysis may apply in certain contexts, further research is needed to accurately understand the criteria for the influence of juror ballots on jury verdicts and the application of these influences in specific contexts.

Penrod and Hastie (1979) are also quick to point out that the assumption that the probability of any randomly selected juror to vote for guilt is constant is a concern when attempting to model juror or jury decisions. Specifically, they indicate that while many models provide insight into how juror decisions can be used to predict jury verdicts, little is known about the individual juror's ability to accurately determine a defendant's guilt or innocence and there is no reliable method to determine the objective guilt or innocence of a defendant either.

Some attempts at understanding how jurors evaluate evidence have relied on Bayesian models of reasoning under the assumption that the best predictor of the guilt of a defendant is agreement among jurors in initial ballots. However, this still limits a discussion regarding an evaluation of the acuity of a juror in identifying guilt or innocence from evidence. Bayesian models for juror decisions do necessitate the consideration of two parameters. First, one must consider the probability before trial that an accused individual brought to trial is convictable. Second, one must also consider the probability that a juror will accurately assess, and vote with, the weight of the evidence (Devine, 2012). With these parameters in mind, it is possible to predict the probability a juror will make a correct verdict based on the proportion of defendants who are guilty and the proportion of defendants acquitted, from an analysis of normative data on pre-deliberation ballots at trials. This type of reasoning assumes that initial juror ballots are a good indicator of guilt and also that jurors are equally skilled at determining guilt or innocence. Both of these assumptions will be discussed in further detail later but it is important to note that research has quite mixed findings relating to the relevance of using initial ballots as a reference point for a Bayesian model and research is also mixed with respect to the consistency of juror's evaluations of guilt and innocence. Some analyses, however, indicate that jurors are more tolerant of letting a suspect who has a low but present chance of guilt go free on a first ballot than they are of letting a suspect who has a low but present chance of innocence go to jail.

Ostrom, Werner and Saks (1978) demonstrated this tendency to presume innocence by having participants listen to conflicting evidence regarding murder, rape, or theft crimes and then asking them to integrate that information into a single subjective estimate regarding

whether the defendant in each scenario was guilty beyond a reasonable doubt. As a test of their participants' initial dispositions, Ostrom et al. predicted four possible states that allowed a juror to try to be objective in their evaluation of evidence and still incorporate their prior opinion. First, the juror can assume the chances of a suspect being guilty or innocent are equally likely. Second, they could try to maintain objectivity by attempting to base their belief on the outcome of similar trials. This would require knowledge of a fair sample of trial outcomes and also takes the risk of placing the burden of proof heavily on the Defense or the Crown if initial beliefs significantly deviate from a 50% chance of guilt. Third, the juror could adopt a presumption of innocence and maintain that the defendant is "innocent until proven guilty", a stance jurors in many legal systems are asked to take. In this case initial probabilities will be close to 0%. Fourth, a juror may be aware that their prior beliefs could influence their verdict decision but feel that they should only rely on the evidence to make a decision and so ignore their initial disposition and give it no weight in their decision. In this case, their decision is assumed to be based entirely on the effect of the evidence.

Ostrom et al.'s (1978) data indicated a mock juror's tendency to use an arithmetic model for integrating prior beliefs and evidence. Namely, their data supports the notion that the participants presumed innocence and then averaged their initial beliefs with their perceived weight of the evidence to come to a final decision of guilt. This was consistent with Anderson's (1971) seminal framework of information integration theory which favors averaging over summation as an arithmetic explanation of juridical judgement tasks. Surprisingly, Ostrom et al.'s data also indicated that anti-defendant participants (those who gave stronger judgements of guilt) actually showed a greater willingness to presume innocence before encountering the

evidence. On the other hand, pro-defendant participants gave lower probability of guilt judgements to evidence sets than anti-defendant participants. This implies that judgements of guilt are not necessarily related to initial dispositions of guilt or innocence but rather to the malleability of those opinions and the weight a juror might place on them and the evidence. This finding also presents an intriguing dilemma relating to the presumption of innocence that is often promoted in criminal trials. If participants with a greater willingness to presume innocence before the presentation of evidence are ultimately more likely to be influenced by evidence for guilt, perhaps that is a result of the framing of the judgement task. In other words, and as suggested by Ostrom et al.'s model, a strong presumption of innocence may place the burden of proof heavily on the Crown or Prosecution in which case a juror may quite naturally place more weight on evidence presented for an argument of guilt. It is also possible that the discrepancy between an initial disposition of innocence and evidence for guilt simply allows more room for change to operate, as opposed to a disposition of guilt that may be closer to the weighted average of the evidence. Either way, the findings indicate that the interaction between initial dispositions and the evidence is integral to understanding the judgement of jurors, and is important to consider for all mathematical models of juror decision making.

Viewing the juror-jury interaction through the framework of Bayesian probability theory (Tversky and Kahneman, 1980) necessitates that a juror's probability estimate of the guilt of a suspect is derived from calculating the posterior odds that that individual is likely guilty. Posterior odds are a function of the interaction between prior odds and likelihood ratios. Prior odds refer to the probability that a suspect is guilty before considering any case specific information. They generally represent the juror's initial dispositions towards guilt or innocence

before encountering any evidence. Likelihood ratios represent evaluations of the weight of each piece of evidence. They are calculated by considering the probability of evidence existing if an event under consideration did occur over the probability of its existing if the event did not occur. In the case of an eyewitness' testimony for example, the likelihood ratio of the testimony can be calculated by dividing the probability of a true testimony by the probability of a false testimony. These probabilities can be calculated based on the properties of case specific evidence. For example, normative data on an eyewitness's accuracy can be used to estimate whether he/she will testify to an event given that that event actually occurred.

Wells and Lindsay (1980) used a Bayesian model of information gain to demonstrate the informativeness of eyewitness identifications and non-identifications. Based on their observation that identifications and non-identifications are exhaustive of witness behaviour in an identification task (they can only do one or the other), and the assumption that the probability of identification is always greater when the suspect is the culprit than when he/she is not, Wells and Lindsay showed that if an eyewitness identification increases the probability of a suspect being the culprit then a non-identification should decrease that probability. From an evidential perspective, a witness testifying that the suspect is not the person they saw at the scene of a crime should decrease estimates of guilt in a similar manner that identification at the scene of the crime would increase them. In fact, Wells and Lindsay demonstrated that the informativeness of identifications versus non-identifications is determined by the probability of obtaining one versus the other and argued that eyewitness non-identifications are underutilized in criminal investigations as they are actually more informative than

identifications (since they are less frequent) but are not as readily admitted as evidence relating to the culpability of the suspect.

Similar findings were obtained by McAllister and Bregman (1989). Participants in their study were asked to make verdict decisions for mock armed-robbery cases that varied in strength of evidence and whether an eyewitness and alibi witness provided identification, non-identification, or no testimony. Participants in this case also did not use witness non-identifications to their maximum potential. For alibi witnesses, identification lead to significantly lower estimates of guilt than non-identification and no testimony. Alibi non-identification showed no statistical difference from no testimony. Even though limited data on the frequency of alibi identifications and non-identifications makes it difficult to ascertain their diagnosticity, the Bayesian model of information gain should still predict that both identification and non-identifications differ significantly from no testimony. In the eyewitness condition however, identification led to judgements of guilt that were no different than the no testimony condition and non-identification led to significantly lower judgements than the control condition. Thus, in this study, alibi non-identifications and eyewitness identification were underutilized. McAllister and Bregman hypothesized that since mean guilt ratings indicated a bias for innocence in their sample (the overall proportion of guilty votes for the experiments was .38), the data may actually represent a tendency to underutilize evidence that counters (or at least does not confirm) a held hypothesis about guilt or innocence. Thus, while Well's and Lindsay's analysis was based on real case samples where the initial disposition of guilt was high, McAllister and Bregman's sample showed the opposite trend. Bayesian analysis of both sets of data however indicate a non-optimal strategy for the integration of evidence

and also highlights a major challenge for Bayesian models in delineating the effects of extraneous factors such as previously held juror beliefs that are not directly related to a case.

In an attempt to investigate the applicability of Bayesian models to juror decisions, Hinsz, Tindale, Nagao, Davis and Robertson (1988) conducted a study to examine how well jurors might use this type of model in order to assess guilt in a series of inductive inference problems. Specifically, Hinsz et al. were interested in how well their participants would be able to appropriately integrate base rate information and case specific information in order to produce estimates of guilt and whether these estimates would match the predictions made by Bayesian models for the same inference problems.

The inference problems used by Hinsz et al. were modeled after Kahneman and Tversky's 'cab problem' (Kahneman & Tversky, 1974). In these problems, one of two entities is implicated in a crime or work negligence scenario. Base rate information regarding the involvement of each entity is presented (as an estimate of prior odds) in addition to case specific information in the form of an eyewitness who identifies the implicated entity and varies in his/her accuracy (for calculating likelihood ratios). The classical implementation of this problem involves a cab involved in a hit and run accident in a city. There are two cab companies that operate in the city, a blue cab company and a green cab company. 85% of the cabs in the city are blue and the remaining 15% are green. A witness identifies the cab involved in the accident as green and the court tests the reliability of the witness with a diagnostic test that determines their hit rate and false alarm rate (i.e., how often they say a cab is green when it is actually green, and how often they say it is green when it is actually blue). Participants are then

asked to estimate the probability that the cab involved in the accident was green rather than blue given the witness' testimony.

Hinsz et al. (1988) asked their participants to use the information provided in their scenarios to estimate the likelihood of guilt of the implicated entity. A major theoretical concern of Hinsz et al. (1998) was whether their participants would exhibit base-rate neglect in that they would ignore or underemphasize the distributional or base rate information and overemphasize diagnostic or individuating eyewitness information when evaluating the problem scenarios. Hinsz et al. indicate that in order to solve the problems being presented, their participants would likely rely on a combination of statistical and human inference strategies. While statistical inference strategies rely on quantitative data, human inference strategies will be mediated by cognitive heuristics. Thus, their participants' probability judgements will likely result from a combination of their reliance on the evidence, their own personal perceptions and biases and their understanding and application of instructions.

As noted by Tversky and Kahneman (1974) individuals presented with both distributional (base rates) and individuating (eyewitness) information will often rely more on the individuating information when making probability judgements. In the case of an eyewitness, this may be due to a stronger affinity to information obtained directly from human sources as opposed to information presented in the form of statistics or other normative data. For example, it is widely observed in psycho-legal literature that a confident eyewitness is the strongest predictor of a jury's final verdict (Cutler, Penrod, & Stuve, 1988; Garret, 2008). Further, evidence from other fields indicates that the tendency to orient to information from human sources may be a pervasive human characteristic. For example, home owners are likely

to use real estate agents even when normative data indicate that the real estate agents have no significant impact on the price of a house (Zumpano, Elder & Baryla, 1996). While using real estate agents may simply be a matter of convenience, it still seems in all cases that an affinity to individuating information is quite attractive.

Hinsz et al. (1988) formulated three theoretical frameworks for the interpretation of their data. They concluded the interaction between base rate and case specific information in their scenarios would likely be mediated by a) a relevance-via-specificity approach that predicts individuating information will always dominate base-rate information because it is always more specific, b) a diagnosticity approach that predicts that in cases where the source is highly diagnostic (20%, 80%) individuating information will dominate but that this effect will disappear as soon as the source becomes non-diagnostic (50%), and c) a source credibility approach where highly credible sources will lead to individuating information dominating to the point that the information is consistent (50%). Beyond this point (below 50%), the source will be discarded and more emphasis will be placed on the base-rate information. Hinsz et al. did indeed find that as the accuracy of a source increased, participants found base rate information less relevant and individuating information more relevant. Thus, their results supported the source credibility approach. Further, participants were more confident in their responses when they relied on only once piece of information (either the base rates or the witness) in order to make a probability judgement. Hinsz et al. (1988) note that the large proportion of responses from their participants did not approach the normative standards suggested by Bayes theorem. In this case, participants' intuitive processes did not seem to match the mathematically

proposed model though concerns regarding the applicability of Bayesian solutions to the "cab problem" prototype have been given (e.g., Birnbaum, 1983).

Research has indicated that the occurrence of base rate neglect is not a universal phenomenon and studies have demonstrated that participants can be alerted to base rate information and in some cases even prefer it to individuating information (e.g., Barbey & Sloman, 2007; Koehler, 1996, 2002). This is apparent in the Hinsz et al. study when participants orient to base rate information as an anchor for their estimates when they encounter eyewitness testimony that is deemed highly inaccurate. Many studies support the notion that a highly diagnostic witness will only be utilized if they are also perceived as accurate (Barbey & Sloman, 2007; Lyon & Slovic, 1976; Thagard, 2004; Hinsz, Tindale & Nagao, 2008). In addition, manipulating the perceived relevance and specificity of base rate information in inference problems by either stressing the direct link between the base rates and the specific problem at hand or reducing these properties in individuating information also leads to greater utilization of base rate information in probability estimates (Bar-Hillel, 1980). Wells (1992) summarizes the likelihood of finding utility in statistical evidence as being dependent on the bidirectional effect of the statistical facts and the evidence on each other. The statistical facts must be relevant to and affect perceptions of the evidence they relate to, and interpretations of the evidence and its utility must also affect perceptions of the value of the statistical facts.

A more fundamental argument against the utility of Bayesian models for understanding juror processes is that jurors simply do not have the ability to frame base rate information in the correct mathematical context in order to use it appropriately. In addition to the findings above, Barbey and Sloman (2007) demonstrated that rephrasing base rates in frequentist terms

(e.g., 10 in 100 vs 10%) drastically increases the ratio of correct responses to Bayesian inference problems by a factor of 30-60% across several studies. Similarly, Koehler and Macchi (2004) showed that phrasing DNA-match statistics in specific frequentist terms that were easier for participants in a mock murder trial to imagine led those participants to be more willing to accept the notion that the defendant's DNA may have been matched due to chance because they could more easily imagine other individuals whose DNA could also match the sample. For example, reading that "1 in 100,000 people in town who are not the source of a DNA sample will nonetheless match it" reduced the impact of this DNA evidence as opposed to reading that "the chance the suspect would match the sample if he was not the source was .001%".

Another factor that needs to be carefully considered in Bayesian models is the juror's judgement criteria when integrating initial dispositions with the weight of the evidence to produce an evaluation of guilt. This is often referred to as the juror's threshold of "reasonable doubt". Combining the issue of defining subjective thresholds of reasonable doubt with the juror's challenge of selecting the relevant values for prior odds and likelihood ratios leads to a serious reference class problem when evaluating the validity of Bayesian statistics as a model for juror decisions. Specifically, the task of identifying all the relevant assumptions for a normative estimate of guilt given a specific case and evidence set becomes quite challenging. In fact, there is no real world normative standard with which to compare the accuracy of juror decisions (Devine, 2012). Comparing juror decisions to those of a judge or other legal professional may give some insight into the level of agreement among these parties regarding the strength of the evidence but does not speak to the objective quality of the decisions and

makes the strong assumption that the source of comparison (i.e., the judge) is in fact evaluating the case accurately.

It is impossible to establish a standard of comparison before the facts of a trial have been presented and evaluated because these facts are what create the appropriate set of assumptions for a normative solution (Allen & Pardo, 2007). This is further compounded by the fundamental assumption in most Bayesian juror decision models that estimates of the likelihood ratios of evidence are independent of base-rates, which is not true for real eyewitnesses. For example, it is not realistic to assume that an eyewitness' ability to identify a green cab at a hit and run accident is independent of her impression of the prevalence of green cabs in her area. In more general terms, it makes little sense to assume that the ratio of hit rates (correct identifications) to false alarms (incorrect identification) is independent of signal probability (base rates) (Birnbaum, 1983). Birnbaum brings to light in this argument that it is necessary to be sure that the normative analysis of our Bayesian models is correct before comparing human responses to it. Further, Birnbaum shows that when adjusting for the assumption that juror judgement criteria will shift depending on the relationship between signal strength and likelihood ratios, normative solutions to the cab problem very closely match participant estimates of the solution. This leads him to the conclusion that Bayes theorem does not imply an effect of the base rates in an inference problem unless a specific theory of the participant's response criteria is assumed. Some probabilistic models address this by anticipating that a number of different jury decisions are possible from a given starting point and that we can estimate the chance that each one will occur (Devine, 2012). Ultimately, these outcomes are still based on the assumed relationships between initial beliefs and the weights

given to different sets of evidence and so will face the same reference class problem as other stochastic models.

In summary, mathematical models do seem to provide some robust prescriptive potential in terms of understanding the probative value of evidence and its interaction with initial beliefs of guilt or innocence. However, this is challenged by the difficulty of identifying a framework for a normative mathematical solution and the difficulty in identifying the correct parameters and assumptions for the appropriate mathematical solution. The reference class problem presents a barrier in that it causes the information gain of probabilistically interpreted evidence to change with the selected reference class. For example, interpreting that blue cabs are responsible for 15% of deliveries in a specific street versus in the entire city for the "cab problem" may change the perceived impact and usefulness of that information in a mathematical solution (i.e., the independence of the observation and the base-rate is not clear). Further, every reference class leads to different assumptions about the information gain presented by a single piece of evidence and the correct class (the one that makes the correct assumptions for the case being evaluated) is exactly the one we are trying to discover.

The reference class problem also speaks to how certain forms of evidence like witness testimony can be highly influenced by initial dispositions. Since most decisions can be made from an infinite number of perspectives, inevitably some will always point out that a decision maker is reliable and others that he/she is not. Also, interests in the conclusions that can be drawn out from given reference classes (i.e., due to confirmation bias or previous hypotheses of guilt or innocence) help identify classes as more or less relevant for a decision at hand (Allen and Pardo, 2007). This highlights two important aspects of decisions made in this context. First,

decision makers can be right without actually knowing that they are right. In the case of a trial, they may reach an appropriate and true verdict by correctly approximating an appropriate reference class, or they may reach an inappropriate but true verdict from an inappropriate reference class. For example, a witness asked to identify a suspect from an array of eight pictures in which only one is coloured may select the suspect in the coloured picture because he/she is actually the person they saw, or simply because the coloured picture stood out in the array. Thus, decisions may be influenced by systematic variables outside of our control that will sometimes lead to correct responses and sometimes not, regardless of the robustness of the decision strategies used (even a broken clock is right twice a day). Second, beliefs about certain hypotheses will affect how the relevance of different reference classes is evaluated, which will in turn affect the evaluation of the validity and importance of different kinds of evidence. To add to this picture, careful attention is required to account for how these effects will operate across different jurors with the note that the interaction between jurors will further complicate the decision process, especially if we cannot assume that juror decisions are independent of each other.

Thus, while mathematical models provide an intriguing and potentially useful framework for explaining what jurors should do, it is unlikely that many of the strategies described mimic what a juror actually does. The challenges of making assumptions regarding the substantive guilt or innocence of a suspect at trial present a barrier when trying to appropriately apply mathematical models. Further, these models also need to address the issue of how individual juror decisions translate to jury verdicts in a consistent enough fashion so that it is possible to make accurate and informative predictions when evidence quality, jury

size and composition, implicit and explicit decision rules, and the timing of deliberation and decisions can vary. The weakness of the links that relate initial juror ballots to final jury ballots creates a strong barrier to the utilization of mathematical models though it is a barrier that could be overcome with a keener understanding of juror specific decision models and a better delineation of the cognitive factors that may be contributing to the manner in which a juror assesses and uses evidence in order to reach a decision. In turn, better understanding the nature and variety of models used by jurors to estimate the objective value and impact of different forms of evidence will contribute significantly to our understanding of judgements of culpability.

The Story Model and Agent-based Explanations of Juror and Jury Decisions

Most discussions relating to jury decision processes have been driven by the need to address specific policy questions such as the minimum number of jurors needed on a panel, or the level of complexity of instructions that a juror is capable of processing (Pennington and Hastie, 1992). In other words, a primary focus is *how* juries come to a decision rather than *why*. Psycho-legal research suggests that juries should a) be composed of twelve members, b) be asked to deliberate to a unanimous decision, and c) get simplified instructions from judges and other legal professionals. Some data also implies that jurors may benefit from note taking throughout a trial (Devine, 2012). This allows for an array of social and cognitive influences within a jury deliberation that may ultimately affect verdicts. In their Social Interaction Scheme Model, Stasser and Davis (1981) note that along with the fact that initial ballot distributions will have a strong effect on final verdicts, the probability of any juror shifting between a positive or

negative ballot and their confidence in the guilt or innocence of the defendant largely explains when changes in a verdict are likely to occur due to informational gain or peer pressure. Specifically, their model indicates that changes to confidence in guilt are often due to information gain during deliberation while changes in preferred verdict are a result of the combination of information gain and peer pressure.

Some of the effects of peer pressure can be mitigated by using large rather than small juries. In a meta-analysis of seventeen studies examining the differences between six and twelve person juries, Saks and Marti (1997) found that large juries are significantly more representative of a community sample as they were more likely to contain an individual from a minority group. Though having a diverse jury may not directly reduce the likelihood of pressure due to a polarized jury panel, general sampling theory would suggest that a larger jury panel is more likely to contain a wider variety of perspectives on the evidence and defendant that are representative of distributions in the general population. This is supported by Saks and Marti's (1997) findings that larger juries take approximately twenty minutes longer than small juries to deliberate the facts of a case and are more likely to hang due to an inability to reach a unanimous decision. Importantly, Saks and Marti also found that members of larger juries remember more of the facts of evidence in a post-deliberation test and discuss trial testimony more accurately than members of smaller juries. This is supported by Hastie, Penrod, and Pennington's (1983) extensive content analysis of mock jury deliberations. Participants were asked to reach majority or unanimity decisions after viewing a three hour re-enactment of a real homicide case. As in Saks and Marti's meta-analysis, Penrod et al. found that jurors in the unanimity condition did deliberate for longer periods of time and also discussed more relevant

key facts from the case, corrected mistaken assertions more often, elicited the opinions of all jury members during deliberation and expressed greater satisfaction with the process.

There is also good evidence that simplifying jury instructions would be a good idea. Generally, mean comprehension of jury instructions ranges between 50-70% (Lieberman, 2009) even though jurors report completely understanding their instructions. If legal and psychological models are based on the assumption that jurors use and evaluate evidence based on the instructions they are given, then this informational gap presents a problem in accurately modeling this process. For example, jurors are instructed to ignore inadmissible evidence yet research often demonstrates that they will use it in their final decisions. In a meta-analysis of 48 studies on the matter Steblay, Hosch, Culhane and McWethy (2006) showed that participants exposed to inadmissible evidence continuously gave more judgements of guilt than participants in a no-inadmissible evidence control group. Evidence that was contested but deemed admissible produced the strongest shifts in perceptions of guilt and further, the effects of inadmissible evidence were apparent regardless of whether participants were given instructions to ignore it or limit its use before or after evidence presentation. Similarly, Allison and Brimacombe (2010) found that mock jurors in their study correctly recalled instructions that prior convictions hold no bearing on the current trial, and yet the presentation of prior conviction evidence showed strong effects on both the believability of alibi evidence and guilt ratings of the suspects. In essence, participants remembered the content of judicial instruction but did not use it when rating the defendant's guilt. Given this, careful consideration needs to be given to the parameters of any decision model attempting to explain juror and jury models.

Perhaps inadmissible evidence and other forms of information that are usually assumed to not influence jurors should be factored into a juror decision model.

Thus, there appears to be a gap in our understanding of what jurors should do and what they actually do. To explain this, Pennington and Hastie (1991) proposed the Story Model as a unified and coherent discussion of the general behaviour of jurors. This model is premised on the key assumptions that jurors use explanation-based coherence models to organize and understand large and complex bodies of evidence, and that the stories they create as a result of these explanations of evidence are the key determinant in their decisions of guilt or innocence. The aptly named Story Model proposes that jurors organize evidence in a manner which imposes a narrative story, or several plausible stories, and then select the most plausible story to reach a decision of guilt or innocence. According to Pennington and Hastie, the construction of these story alternatives will depend on the extent to which each story accounts for the varying pieces of information and evidence available. Further, a story must be devoid of internal contradictions and provide a unique representation of the events being evaluated. If these conditions are met in a particular story alternative, then it will likely be identified as the most plausible representation of true events (Pennington and Hastie, 1991). In order for a story to then influence juror decisions, the juror must first be familiar with the different verdict alternatives that are applicable to a specific case (e.g., issues of guilt or innocence, liability, specific instructions from a judge) and try to find the best match between the constructed stories and the verdict categories. Thus, the story model provides two key insights. First, narrative story structures are spontaneously created as explanatory evidence summaries, and

second, story structures co-vary with and mediate juror decisions (Pennington and Hastie, 1991).

In a test of the Story Model, Pennington and Hastie (1992) found that the presentation mode of evidence had an effect on the way in which their participants recalled evidence from a trial summary. While the total recall of evidence items did not differ between participant groups, participants presented with testimony from multiple witnesses ordered to present a story tended to recall the evidence through a narrative and participants presented with testimony ordered based on issue (e.g., motive, opportunity, character, relationship to suspect) recalled the evidence based on different issue categories. In addition to finding that the recall mode of the participants corresponded to the mode of presentation of the evidence, Pennington and Hastie (1992) also found that the response mode of their participants had an effect on the manner in which they coded the evidence. When participants were asked to provide step by step evaluations of their verdicts after the presentation of each piece of evidence they tended to follow an anchor-and-adjust mechanism whereby they sequentially updated their beliefs on guilt or innocence after each presentation of evidence. On the other hand, when participants were asked to make a single global judgement after all the evidence was presented, they tended to organize the evidence into a narrative story. Further, these story based assessments also led to more extreme judgements than the item-by-tem based assessments even though they were based on exactly the same information. In turn, Pennington and Hastie concluded that unless jurors were explicitly instructed to make item by item judgements, their process would likely follow the prescriptions of the Story Model.

A major shortcoming of the Story Model is that it does not account for the objective value of any single piece of evidence. Thus, while the model provides a quite useful means for understanding the way that jurors assess and use evidence, it does not address the issue of optimizing juror decisions as many mathematical models attempt to do. Further, the model posits that the primary factors that will contribute to the formation of a narrative are how much of the evidence presented it covers and whether it provides a cohesive story that is devoid of internal contradictions. However, each of these principles requires careful consideration. First, it is unclear to what extent the construction of a narrative is reliant on the incorporation of all the evidence presented at trial. For example, we have seen that inadmissible evidence can affect juror decisions even when they are instructed to ignore it. Even in this limited context it is apparent that certain factors deemed outside of the trial evidence will affect the narrative of guilt. Other research has shown that factors such as the perception of a previous criminal record (Allison and Brimacombe, 2010), negative pretrial publicity (Ruva, Guenther &Yarbrough, 2011), or general confidence in the legal system and social justice (Lecci & Myers, 2008) all also have an effect on increasing a juror's willingness to judge a defendant guilty.

Further a juror's narrative is not only affected by positive information (i.e., information that contributes to the story) but by negative information as well (i.e., information that is left out). As discussed earlier, jurors often underutilize non-identification evidence despite its probative value (McAllister & Bregman 1989; Wells & Lindsay, 1980;). Jurors can also be quite selective with regards to the evidence that they choose to include in a story scheme as a result of their confirmation bias. The term "confirmation bias" was first coined by Peter Wason in the

early 1960s to describe how participants continuously used hypothesis affirming questions to figure out a rule that linked a series of 3 numbers. For example, if the number set (6, 12, 18) was presented and a participant believed that the binding rule was “even numbers”, they were more likely to state that “all the numbers are even” than that “all the numbers are multiples of 3” (Gilovich & Griffin, 2002). A lasting interpretation of Wason’s findings is our tendency to seek out confirming information and avoid disconfirming information once we construct a hypothesis about a given situation. For example, Hogarth and Einhorn (1981) asked a group of 23 trained statisticians what the minimum amount of information needed to test a market consultants’ claim that “the market rises every time he predicts it will”, and found that the large majority of them were content with affirming information relating to an actual rise in the market or a favorable report from the consultant. A small portion of their participants did ask for information pertaining to a fall in the market and only 5 of the 23 statisticians indicated the need of a disconfirming scenario with a favorable report from the consultant and a subsequent fall in the market. Thus, it appears the drive to confirm the hypothesis is quite strong even in individuals whose training necessitates the disconfirmation of a null hypothesis (i.e., statisticians).

The application of this for jurors is that the hypotheses they form regarding guilt or innocence can have a strong effect on the evidence they choose to include or ignore when they are constructing their narratives. Though Ostrom et al. (1978) note that jurors are open to accepting the innocence of a defendant, their hypothesis of guilt or innocence may be revised at the early stages of evidence presentation. In other words, while jurors may not be operating through a guilty bias, a hypothesis of guilt is more likely to form before a hypothesis of

innocence since evidence of guilt presented by the Crown is available before evidence of innocence presented by the Defense. In essence jurors may, as Ostrom et al., suggested, have the option of presuming innocence, but not actually use it in real courtroom situations.

Dahl, Brimacombe and Lindsay (2009) suggest that the effects of the order of presentation of evidence can be quite complex. In an elaborate study of the underlying effects of witness identification on mock police-investigator's estimates of the guilt of a suspect, Dahl et al. found that in situations when an eyewitness identified the same suspect as the investigator, judgements of likelihood of guilt were near maximum levels whether witness IDs were given before or after independent evaluation of other evidence. Alternatively, when the eyewitness implied that the suspect was not present in the lineup, this information was more likely to be ignored when it appeared before the evaluation of other evidence that implied guilt. Specifically, participants were significantly more likely to arrest if a "not present ID" occurred before their evaluation of other evidence for guilt than after. This finding poses an interesting challenge to the confirmation bias hypothesis in that it implies that the operation of this principle may be highly contingent on the specific contexts of an investigation or trial. Dahl et al. also found that a strong alibi testimony which implied the innocence of a suspect only reduced judgements of probability of guilt if it was presented after a witness identification that implied guilt but not if the order of the evidence was reversed. Thus, for highly contradictory forms of evidence, the order of presentation may have a strong effect on the decision to utilize specific sets of information in final verdict decisions. To add to this, the nature of the judicial task in an adversarial criminal trial is to affirm guilt beyond a reasonable doubt, which may further the effect of early hypothesis formation and any subsequent confirmation bias. While it

is not clear from this research at what point a hypothesis may form, it is evident that jurors can be selective about the evidence they utilize once it is. In fact, the effects of confirmation bias can be so powerful that a piece of acknowledged counterevidence can even weigh for the hypothesis it rejects, such as when perceived inconsistencies in a 'weak' alibi story are perceived as evidence that a suspect is lying about their involvement in a crime (Olson & Wells, 2004). The inconsistencies may discredit the alibi story and its argument for innocence, but they should not also credit an argument for guilt.

Thus, despite the appeal of the story model as a descriptive tool for explaining how jurors may evaluate evidence, it still lacks the necessary prescriptive boundaries to help aid jurors in actually making better decisions. A strong contributor to this is the adversarial nature of the trial process. Triers of fact are faced with arguments from competing parties whose sole responsibility is to present evidence that supports their arguments. In criminal trials in Canada, the Crown represents the case for guilt, and the Defence represents the case for innocence. This is however an idealistic perspective on the matter. In actuality, while the Crown presents an argument for guilt, the Defence is usually engaged in the task of negating guilt rather than proving innocence. This notion is supported by findings regarding the order effects of exonerating evidence (e.g. Dahl et al., 2009; Price & Dahl, 2014). This is partly promoted by the burden of proof placed on the Crown to "prove guilt beyond a reasonable doubt" but ultimately both parties (the Crown and Defence) are responsible for presenting an argument of legality rather than argument of truth (Givelber, 2005). This means that either party's responsibility falls on providing the greatest amount of evidence possible to argue whether a suspect is *legally* guilty or not rather than whether they are *actually* guilty or not.

Many cases from AIDWYC and The Innocence Project highlight situations where individuals were found legally guilty despite the presence of evidence suggesting their innocence. As mentioned earlier, the fact that it can be very challenging to identify the true guilt or innocence of a suspect presents a strong challenge here and necessitates the need to abduce a verdict from the evidence. In the case of the story model, when jurors are faced with the choice of deciding whether a suspect is guilty or not guilty and operate through an initial disposition (whether it is for innocence or guilt), then the evidence identified as the most likely to explain a perceived verdict will also be perceived as the most diagnostic. In more general terms, this strategy for information evaluation is referred to as affirming the consequent or inference to the best explanation (Govier, 2005). For example, a juror operating under the belief that only suspects who the legal system has probable cause to believe are guilty are brought to trial may deem evidence for guilt as more diagnostic when they actually encounter a suspect at trial. Similarly, a juror with a strong belief that a large amount of innocent individuals are brought to trial may orient to evidence implying innocence. The major challenge in these situations is that any evidence set needs only to be *sufficient* to explain a given story alternative. There is no requirement that a verdict must *necessarily* follow from all the evidence.

Naturally in an adversarial system when a juror accepts one argument they must refute the alternative. It cannot logically follow for example that a juror believes an eyewitness who places a suspect at the scene of the crime and an alibi witness who places the suspect away from the scene at the same time. Therefore, it is important to achieve some evaluation of the criteria used by jurors to select the evidence that forms their narrative and to better

understand not just how jurors form these narratives but why as well. Influencers like the confirmation bias or the framing of the juror decision task have strong effects on the way a juror may decide on a verdict and also have implications for the future influence of those decisions as they relate to policy reform or addressing the issue of wrongful conviction. At a more basic cognitive level, there is strong empirical support for the notion that recalling specific list items can lead to the suppression of other list items. For example, cuing recall for specific list items by providing a category clue (e.g., fruits) leads to better recall of those items but also to poorer recall of the remaining list items (Aslan, Bauml & Grundgeiger, 2007; Bauml & Aslan, 2004). In cognitive psychology, this effect is commonly referred to as retrieval-induced forgetting and is underscored by the assumption that in order to retrieve target items from memory, competing items must be selected against or actively inhibited (e.g., Storm & White, 2010). Thus, retrieval of target items leads to stronger recall of those items (i.e., they become more salient memory sets) and to the active inhibition of the remainder of the original set. This is likely not far off from the experience a juror has when encountering a list of evidence items and composing a narrative to link those evidence items together into a coherent story. With the motivation of specific hypotheses of guilt or innocence, jurors will naturally find that some evidence items are more salient than others. This will lead to the active retrieval of this evidence and also to the suppression of competing or non-retrieved evidence (as mentioned earlier, one of the advantages of having larger juries is the fact that a wider set of evidence items will be discussed). For example, Shaw III, Bjork and Handal (1995) found that cuing recall for the details of a crime scene by questioning their participants about specific aspects of the scene led to better recall of the items questioned about but to poorer recall of other items. In

their study, participants were asked to retrace their steps through a slideshow of pictures of a party they were asked to imagine they had attended where their wallet had gone missing. The primary memory test involved recall for a set of university sweaters and a set of university textbooks that were in one of the slides. Participants were asked about the presence or absence of some of the items in an "interrogation" and later asked to recall as many items as possible in a free-recall test. Participants cued for the recall of specific items in the interrogation phase showed significantly stronger memory for those items and significantly poorer memory for non-cued items than participants who were not cued during the interrogation phase. Similarly, a juror, through their own personal discourse or through the influence of jury deliberation, will likely find that some evidence items will become more salient and others will be actively inhibited.

In summary, even though the story model provides an appealing and intuitive framework for understanding how jurors may mentally organize and use evidence, it lacks the prescriptive utility to comprehensively explain why or when specific narratives should be prioritized. Additionally, research on the effects of pretrial publicity, juror attitudes, and other social cognitive factors contributes significantly to understanding how trial related experiences will interact with a juror's prior knowledge and beliefs but these models still provide a limited insight on how to help a juror make an optimal decision. The story model has an intuitive simplicity but it is fairly challenging to objectively measure its effectiveness as a framework for juror decisions. In spite of this, many attorneys (particularly in the USA) have taken to utilizing social scientists as trial consultants to aid in jury selection. This practice is aptly referred to as "scientific jury selection" (Lieberman, 2011).

In US criminal courts, the jury selection process (known as voir dire) allows attorneys to eliminate potentially biased jurors. In most commonwealth countries the term voir dire is used to describe the process of determining the competency of a witness or the admissibility of evidence in addition to referring to an evaluation of a juror's competency (Lieberman, 2011). In either case, it is debatable whether the utilization of scientific jury selection aids in the selection or rejection of specific jurors. While most court appeals are based on the grounds of removing potentially biased individuals (e.g., due to racial bias) the methods used for scientific jury selection imply the pursuit of a jury panel that will produce a favorable outcome (i.e., jury with an increased likelihood of accepting the argument of the Defence or Crown). For example, the process often involves thorough evaluations of community perspectives on the effectiveness of the arguments or evidence to be presented which is supplemented by carefully constructed demographic profiles that can be used to identify individuals that may be more responsive to specific arguments (Lieberman, 2011; Seltzer, 2006). Research on the effectiveness of scientific jury selection indicates that much more evidence is needed to accurately evaluate its utility. Due to the variety of methods used in the process much more research is needed to understand when and how scientific jury selection may be effective. In addition the high cost of hiring a trial consultant clouds any safe judgement regarding her/his effectiveness as it is confounded by the strong likelihood that a client who can afford a trial consultant can also afford strong legal representation (Lieberman, 2011). In other words, concluding that juries who are "scientifically selected" fit the predictive outcome model of the social scientists who aid in their selection may be misleading if the outcome is due to the efforts of a strong lawyer rather than the juror model itself. These shortcomings highlight some very

important factors though. For one thing, it is clear that there is a need and an application for clearly understanding juror decision models. Gaining a clearer perspective on how and why jurors make their decision will help reduce the likelihood of a biased jury and ultimately the likelihood of wrongful conviction. In addition, enhancing our understanding of juror decision models will aid in identifying situations where attorneys aim to "stack" a jury in order to gain a favorable outcome. Granted the function of the court is to come to a legal decision, but if there is potential to arrive at a just decision then such avenues are and should be utilized to their maximum potential. In order to do this however, we need to present a comprehensive and adaptable understanding of juror decisions and how they impact jury deliberation and final verdicts.

On one hand, mathematical models provide a keen insight into the probative value of evidence although the assumptions necessitated by these models often ignore psychological influences that can affect how a juror utilizes evidence. On the other hand, psychological models neatly address the manner a juror may assess and process evidence but often don't account for the objective value of any single piece of evidence. Thus, it seems that these two approaches, though quite distinct, have the potential to be rather complimentary. In turn a hybrid model that addresses both the probative value of evidence and the influence of the juror may provide a coherent and applicable model for not only understanding, but also optimizing the juror decision process. Specifically, understanding when jurors may act through the influence of personal bias and when they may act through the influence of evidence in a comprehensive model will add significantly to our understanding of the legal decision process. Further, conceptualizing juror decisions in the context of the relevance, applicability, and

uniqueness of specific models provides further opportunity to study the utility of framing a single, comprehensive, optimal or rational model for juror decision. It seems quite likely that a combination of several cooperative models may provide a more comprehensive insight into the process of juror decisions.

Comprehensive and Hybrid Models for Juror and Jury Decisions

To date, most published research implies that juror level decision models have had little influence on jury level decision models (Devine, 2012). In addition, more research is needed to understand the context of specific moderators of individual jurors' influence on group deliberation. For example, while Hinsz et al. (2008) imply that certain individual biases, such as the tendency to ignore or emphasize base-rate rate evidence depending on the accuracy of an eyewitness, are accentuated at the group level, Hastie et al. (1983) indicate that deliberation tends to reduce bias and enhance the opportunity to cover a wider range of case specific evidence (especially in larger juries). Thus conclusions on any specific effects of juror models on jury decisions need further revision and must evaluate the ecological validity of research from mock jurors and juries (Devine, 2012). Further, in attempting to understand the interaction between juror and jury models, it is instrumental to consider the framework and context of aggregating juror models and not just juror decisions. In a similar vein to how the reference class problem presents challenges in understanding the value of individual bits of information, unique reference classes with respect to the aggregation of juror decision models will also lead to different interpretations of their utility.

This is partly due to the diversity of the models proposed and to the legal and scientific motivations for understanding the diagnostic potential of different models. For example, individual juror models are applicable at voir dire while comprehensive jury models are more applicable at the trial and deliberation stages. Further, motivations for understanding what a juror *should* do in terms of integrating and assessing evidence and what a juror *actually* does have also driven research in various directions. While mathematical models tend to focus on the first part of this equation, psychological models tend to explain the latter. Despite these convergent approaches there is evidence to suggest there is room for integrating both approaches to understand at least some aspects of a juror's decision in a more comprehensive manner. Harris and Hahn (2009) provide some intriguing evidence that Bayesian models can be used as a quantitative measure of the evaluation of multiple witness testimonies in a story like context.

In their research, Harris and Hahn (2009) indicate that accounting for the coherence of multiple testimonies in a Bayesian model may provide a quantitative model for understanding Pennington and Hastie's (1992) coherence criteria for story construction in the Story model. Particularly, Harris and Hahn emphasize that coherence is integral to belief revisions and information integration when receiving information from multiple witnesses. They also suggest that even when witnesses are deemed equally reliable (i.e., they are equally likely to be accurate and truthful), highly coherent witness testimonies are more likely to be integrated into a juror's information set or story than incoherent witness testimonies. For example, they indicate that information from three witnesses where the first indicates an individual was speaking French, another indicates that he was wearing a French football jersey, and the last

indicating that he was waving a French flag, is more likely to be integrated than testimony indicating that that individual was speaking German, wearing a French jersey, and waving a Union Jack flag.

Harris and Hahn asked their participants to estimate the adjustment in the likelihood that a police officer would use testimony from multiple witnesses that moved from high to low coherence and compared their responses to estimates produced by a Bayesian model that accounts for the coherence of the testimonies. According to the strong correlation found from this comparison, they concluded that their participants' judgements seemed to reflect a noisy estimate of the Bayesian model's outcomes. Thus, in contrast to research implying that jurors do not tend to follow Bayesian probability models in updating beliefs, Harris and Hahn (2009) imply that Bayesian probabilities might provide a quantitative measure for the coherence of witness testimonies and can be used as a predictor of juror behaviour as well. So, while Bayesian models may not be applicable in delineating the assessment of all forms of evidence, they may provide vital insights in specific contexts.

Further, a common ground for almost all juror models is the assumption that verdict decisions are the result of the interaction between the prior beliefs or hypotheses of the juror, and the evidence being presented. While differing models vary in their attributions of the sources of prior beliefs and the nature and extent of the effects of evidence, there is a common agreement that the juror decision is a result of the influence of the presented evidence on prior beliefs. An influential understanding of these concepts was proposed by Hogarth and Einhorn (1992) in their Belief Adjustment Model which builds on Anderson's (1971) work on information integration.

According to Hogarth and Einhorn (1992), evidence that needs be evaluated and integrated into an information set can either be coded as a deviation relative to the size of a previous anchor that the decision maker has established, or as positive or negative feedback regarding a hypothesis under consideration. Further, Hogarth and Einhorn indicate that the process of this information management can occur in one of two modes. Either jurors will update their beliefs in a step-by-step (SbS) fashion and adjust their opinions after the presentation of each piece of evidence, or they will attempt to aggregate all the evidence presented in one attempt after it has been presented in its entirety at the end of a sequence (EoS). These modes of processing (SbS and EoS) also act as response modes for jurors. Thus, a juror can express their belief in the guilt or innocence of a suspect after integrating each piece of evidence or they may report their opinion once after all the evidence has been presented. "The Belief-Adjustment-Model assumes that people handle belief-updating tasks by a general, sequential anchoring-and-adjustment process in which current opinion, or the anchor, is adjusted by the impact of succeeding pieces of evidence" (Hogarth and Einhorn, 1992, p.8).

In addition, the model makes significant assumptions regarding the interaction between response modes and order effects relating to the presentation order of evidence. Specifically, Hogarth and Einhorn's models posits that for short and simple evidence series, an EoS response mode will contribute to a primacy effect, where early evidence items will have a larger effect on beliefs, and a SbS response mode will contribute to recency effects where recently encountered items of evidence will have greater effects on beliefs. Further, long simple tasks will lead to primacy effects and complex tasks (short or long) will lead to recency effects independent of response modes. This is due to the decrement in attention that is experienced as evidence sets

become more challenging to process as a result of their complexity. Since a SbS process where integrating evidence one piece at a time is assumed to be less cognitively taxing than an EoS process where all the evidence must be aggregated at once, the model predicts that the chances of a SbS process being used will increase as decision tasks become more complex and in turn recency effects will be more prominent. Thus, when a response mode is SbS, a SbS process model will always be used. However, when a response mode is EoS, an EoS process will be used for short or simple decision tasks and a SbS process model will be used for more complex evidence items.

This provides a partial explanation for Pennington and Hastie's (1992) findings (discussed earlier) regarding their test of the Story Model. Like Hogarth and Einhorn (1992), Pennington and Hastie (1992) found that participants asked to provide SbS evaluations of their verdicts used an anchor-and-adjust mechanism whereby they sequentially updated their beliefs on guilt or innocence after each presentation of evidence and participants asked to make an EoS judgement tended to organize the evidence into a narrative story. Pennington and Hastie went on to conclude that jurors would follow the prescriptions of the Story Model unless explicitly instructed to make item by item judgements but did not account for the effects of the complexity of evidence that are predicted by the Belief Adjustment Model. Further, they seem to discount the possibility that story narratives may be evaluated in a step by step manner.

Hogarth and Einhorn (1992) also indicate that their order effect predictions will likely depend on whether the sequence of evidence being processed is consistent or mixed. This is particularly relevant when considering the adversarial nature of the verdict decision and the likelihood that a juror will encounter evidence both for and against a hypothesis. Thus, when

there is no reference point to predict the impact of a piece of evidence against, the SbS process will predict no order effects for consistent evidence (because an anchor is unidentifiable) but will always predict a recency effect for mixed evidence (except if evidence for or against a hypothesis is completely ignored). On the other hand, since the EoS process is characterized by a single adjustment that symbolizes the aggregate impact of all the evidence on an original anchor a primacy effect will likely occur since in the absence of an external reference point, an anchor will be derived from the first piece(s) of evidence encountered. Current evidence implies that jurors may orient to SbS processes as suggested by Dahl et al.'s (2009) findings that investigators' judgements of guilt were only reduced when contradicting evidence (e.g., No ID or alibi testimony) was presented after independent evidence evaluations. In other words, in cases with mixed evidence, strong evidence denying a hypothesis for guilt was only effective when it acted through a recency effect.

Further, Hogarth and Einhorn (1992) indicate that in the context of mixed evidence, jurors tend to persevere in a new state of knowledge even when the source of that knowledge is subsequently discredited. This suggests that a juror's sensitivity to positive information may be greater than his/her sensitivity to negative information in general. In fact, the evaluation of non-preferred information may even trigger more effortful cognitive processing on the part of the decision maker (Ditto, Scepansky, Munro, Apanovitch, & Lockhart, 1998). In accordance with this notion, a study by Gilbert, Tafarodi and Malone (1993) found that participants who read false statements either extenuating or exacerbating a suspect's involvement in a crime were significantly more likely to be influenced by these statements (they gave longer or shorter sentences) when they were distracted by a number search task than when they were not.

Participants in their study were instructed to read a mixture of brief evidence statements (e.g., Kevin had a gun) that were color coded as true or false (red for true, black for false) while concurrently working on a numbers search task. Gilbert et al. found that participants distracted by the number search task were more likely to believe that false statements were true, but not that true statements were false. This led to higher judgements of guilt when the false evidence was exacerbating and lower judgements of guilt when it was extenuating. In other words, when subjects were cognitively distracted they instinctively believed the information that was being presented to them and could not engage in a process to evaluate and "un-trust" the false statements if need be and so accepted them as true. In turn, this misplaced belief in false statements led to differences in their sentencing judgements. Similar to Hogarth and Einhorn's perspective, the notion here is that presented evidence is automatically accepted as true and integrated into information sets, and it can be quite challenging to "un-trust" this evidence at a later point.

This was further supported but qualified by Ask, Rebelius and Granhag (2008) when they observed that participants generally rate the reliability of evidence in a manner that is consistent with the their overall decisions (i.e., verdict and beliefs of culpability) relating to a case. The same evidence could be evaluated as consistent with guilt or innocence depending on the participants' perspective. The mere presentation of the evidence and its interaction with prior beliefs seems to play a stronger role in its integration for decisions of culpability than the possibility of its objective evaluation as a description of true events. Ask et al. also found that mock investigators in their study showed great malleability in their evaluations of identical evidence that either confirmed or disconfirmed a hypothesis of guilt. They presented their

participating investigators with fictional murder scenarios designed to plant the hypothesis that a specific suspect was guilty and then asked them to read confirming or disconfirming information in the form of DNA, photo, or witness evidence. Participants rated disconfirming evidence as significantly less reliable than confirming evidence and this effect translated to evaluations of the type of evidence in general. So, when participants rated disconfirming witness testimony in the case with low reliability, they also tended to view witness testimony in general as an unreliable form of evidence. This adds to the argument that a decision maker's sensitivity to information with confirming implications for an established belief may be greater than his/her sensitivity to disconfirming information in general.

Further, discrediting a witness can sometimes lead to rebound effects wherein the overall weight of an initially positive testimony may become negative. For example, Hatvany and Strack (1980) demonstrated that even when participants did change their beliefs about a case upon hearing witness testimony, their beliefs were readjusted to such an extent when the witness was discredited that the overall weight of the witness' testimony was actually negative. In their study, mock jurors evaluating liability in several civil cases logically ignored discredited testimony in most cases but also overcorrected (in their judgements of liability) in some cases, particularly if the testimony of was in the form of an accusation that was later discredited. This is in a similar vein to Olson and Wells' (2004) findings discussed earlier that inconsistencies in a 'weak' alibi story can be perceived as evidence that a suspect is lying about their involvement in a crime. In other words, discredited evidence in this case acted as counter-evidence to the initial belief it implied. Despite these trends, Hogarth and Einhorn do indicate that it is possible to nullify order effects by instructing participants that they need to justify their decisions. This

implies that the appropriate instructions from a judge may lead to more balanced assessments of evidence by jurors, particularly in light of appreciating how these assessments will affect prior beliefs of guilt or innocence. Some of the previously discussed research also implies that jury instructions can have a positive effect on juror decisions although findings are still quite mixed on their efficacy (e.g., Lieberman, 2009; Steblay et al., 2006).

Though a variety of approaches have been proposed, most researchers agree that juror decisions are a result of the shared effects that prior beliefs and the presented evidence have on each other. Some models provide a descriptive overview of how a juror might actually manage these effects (i.e., the story model) while others give a more prescriptive insight on what a juror should do (i.e., Bayesian models). Also, some models make extended assumptions relating to how individual juror decisions can affect and relate to jury verdicts. A comprehensive perspective of the trial experience of a juror from selection to jury verdict and beyond still requires more insight but recent research on the topic seems to be motivated by the philosophy that a hybrid theory may be most appropriate. This model needs to account for the psychological underpinnings of juror decisions that are motivated by natural and intuitive processes and also factor some objective evaluation of the probative value of different forms of evidence. Two contemporary approaches to modeling legal decision making that have taken this approach are Bex, van Koppen, Prakken and Verheij's (2010) hybrid formal theory of arguments, stories and criminal evidence and Dennis Devine's (2012) integrative multi-level theory of jury decision making.

An integral part of Bex et al.'s model is an understanding that evaluations of criminal evidence are always defeasible in nature. This means that models that account for the

interactions between evidence, personal knowledge, and verdicts (whether they are logical, mathematical, story based, etc.) can never provide a complete or final account of the truth of a case. Rather, these models will provide plausible explanations for the observed outcome of a case through evidential support and these plausible explanations will all have an inherent fallibility. One of the main reasons for this perspective is that models of evidence evaluation always require an evaluation of the causes after an effect has been observed. For example, in Guy Paul Morin' case, an evaluation of the evidence was only required after Christine's body was found. Thus, the relation between cause and effect must be inferred after the fact. There are no grounds to believe that the presence of any kind of evidence is completely necessary and sufficient to explain whether Morin killed Jessop though some forms of evidence will provide stronger justification than others.

Bex et al. (2010) argue that we generally approach evidence evaluation from either an argument or story-based perspective and that these approaches are separate but related. The argument based approach relies on different pieces of evidence to build a justified argument for a given case or observation. For example, if an individual is suspected of robbing a store, an evaluation of evidence placing the suspect at the scene of the crime, such as eyewitness testimony, security footage, or DNA, will be used to evaluate the veracity of that claim. In this way, the evidence is used to justify the observation through a form of defeasible argumentation. On the other hand, stories have the main function of establishing causal connections between the facts of a case that need to be explained and achieve this through a process of abductive reasoning by inference to the best explanation. As discussed earlier, in this process if an effect is observed and we are aware of a specific cause and effect relationship,

then we can infer that our known cause is a plausible explanation for the observed effect. For example, if we know that a thief must be physically present at the store for items to be stolen, then we can infer this cause from the observation that the store was robbed (the effect). Further, if we believe that only individuals against whom the Crown has significant evidence to imply guilt are brought to trial, we may infer that a suspect is likely guilty due to their observed presence at court.

Each approach has its benefits and drawbacks and both help a potential juror structure and analyze their reasoning in order to make sense of a specific case. According to Bex et al., arguments help establish a connection between the evidence and the facts of a case that need explaining. Thus, if each fact is considered a conclusion of an argument, then the premises of that argument are the relevant pieces of evidence. In turn, depending on the nature of the evidence, arguments can be made to support or refute certain facts. For example eyewitness evidence could support the argument that someone was present at the scene of a crime while alibi testimony would refute this argument. This leads arguments to fall in one of three states. A "justified" argument survives competition with any counterargument; an "overruled" argument loses competition to a counter argument; and a "defensible" argument is plausible but challenged by an equally plausible counterargument.

Stories on the other hand establish a coherent sequence of events that relates different aspects of the case together. They help a juror build plausible scenarios for the facts of a case by inferring the causal connections between one piece of evidence and another. As with arguments, Bex et al. maintain that stories need to uphold certain standards of coherence as well. First, stories need to be internally consistent. For example, a story that claims that a

suspect was both present at the scene of a crime and at a different location at the same time is not consistent. Second, stories need to be plausible in the manner with which they conform to our general knowledge. For example the generalization that jailhouse informants may testify for their own benefit may seem more plausible than the generalization that a suspect who is maintaining innocence at court would confess to a crime when behind bars. Third, stories need to be complete in that they cover the motivations for an event, the actions taken, and the resulting consequences.

So, Bex et al. (2010) explain that arguments are very good at carefully analyzing each piece of evidence while stories help organize the evidence in a coherent and sequential manner that may also help jurors fill in the gaps in a case. In turn, these two strategies can be combined in a hybrid model where the advantages of each compensate for the disadvantages of its counterpart. The hybrid approach relies on stories that construct intuitive hypotheses about what happened in a case and then uses arguments to support or attack these stories. A juror relies on their intuitive story making strategy, but then meticulously checks the quality of each story alternative through a careful evaluation of the veracity and scope of the evidence it covers. Specifically, a juror needs to evaluate the quantity and quality of evidential support for a story, account for any evidential contradictions, and rely on their common sense knowledge to address any evidential gaps, or at least address their presence. Though Bex et al. acknowledge that a purely logical framework for the juror decision is likely unrealistic and unattainable, they do stress that their model can at least be used as a guideline for well thought out and rational decisions about a case. In a sense, this approach combines the descriptive advantages of traditional psychological models like the story model with the prescriptive

advantages of more regimented models like Bayesian reasoning or formal arguments. While the constraints of the reference class problem still apply to building a model within the constraints of a purely logical theory, the evaluation criteria for plausible arguments and coherent stories proposed by this hybrid model can still be used as objective measures to allow for rational evaluations of evidence. In a sense, this may be a more realistic way of modeling jury decisions if we concede the impossibility of maintaining a purely logical model that has some ecological validity as well.

Naturally, the hybrid model does have some limitations. For one thing, as with many mathematical models, since the chance of a story forming is highly contingent on the type of evidence present, it is difficult to assume that evaluations of the likelihood or veracity of evidence will be independent from evaluations of the story alternatives formed. In other words, there is still a risk for confirmation bias resulting from preferred stories leading to the marginalization of certain aspects of the evidence. Bex et al. (2010) also note the risk that this issue may be further exacerbated by "good" stories that could marginalize "true" stories. Coherent stories that seem intuitive but exclude some of the evidence may overshadow less coherent alternatives that aim to account for more of the information presented. This can be mediated by a careful evaluation of the evidence and careful study of the arguments for and against story alternatives but it is a present limitation nonetheless.

Further, the model does not provide an evaluation for when a standard of proof has been met. Thus, it holds great value in the comparative investigation of alternative story options but gives little input to a potential juror on how their evaluations of the stories could influence their verdict. The model as it stands also provides little predictive ability on the likely

outcome of a juror's evidence evaluation although Bex et al. suggest that this issue could be addressed through the extension of the model to account for the absolute (as opposed to comparative) quality of a story. However, as Bex et al. point out, setting such a standard of proof is a matter of legal concern rather than a quality of any formal model of reasoning. Further, research on the criteria for standards of proof remains mixed and inconclusive and the variability across potential jurors may provide a significant barrier (Saunders, 2005). Additionally, Bex et al.'s model does not account for the adversarial nature of many court systems. While the model accounts for the effects of a given story choice once it is selected, it does not account for how a decision can be made between arguments that are based on contradictory evidence. For example, it may account for the consequence of upholding or rejecting witness testimony, but provides no criteria for the evaluation of these alternatives. This again is a matter of legal issue that is outside of the framework of a formal model of reasoning and will be characterized by variability across different jurors. As discussed earlier, many factors may play a role in a juror's tendency to orient towards evidence for guilt or evidence for innocence and their choice will have an inevitable influence on the arguments and stories they subsequently form to evaluate a case. Nonetheless, Bex et al.'s model provides a promising step in the direction of a comprehensive model of juror decision making that integrates both natural and optimal decision tools in a realistic manner. In a similar vein, Devine's (2012) integrative multi-level theory of jury decision making also aims to capture this philosophy with the addition of explaining how individual juror decisions might play into the broader picture of jury verdicts.

Devine's (2012) multi-level theory is divided into two sections. The first, "The Director's Cut Model", covers decisions at the juror level while the second, "The Story Sampling Model", covers decision at the jury level. In his presentation of the model, Devine is adamant that any integrative model of jury decisions should cover four basic criteria. First, the model needs to incorporate major empirical findings. Second, the new model needs to be consistent with current theoretical models unless there is an empirically supported call for a change. Specifically, Devine is referring to the pervasive and accepted theoretical assumptions that jurors likely follow a story creation model when evaluating evidence, and rely on mental models in general to understand and evaluate information at trial. Third, an integrative model needs to address decisions at both the juror and jury levels. Fourth, the model needs to extend the current literature in some way.

According to Devine (2012), the empirical findings on juror and jury research can be summarized into several thematic elements. 1) Individual differences matter at almost all levels of the trial. Thus, many demographic characteristics of the juror, defendant, or legal representation may have an effect on how, why and when decisions are made. Further, appreciating the interactions between these individual differences is integral to understanding the decision process. 2) Jurors use relevant information regardless of its source. As discussed earlier for example, jurors are generally quite poor at following instructions to ignore inadmissible evidence or previous criminal records. 3) Jurors use story structures to organize and understand the evidence in a trial. 4) The deliberation process is not egalitarian and representative of all points of view. Members of the jury who are more vocal or assertive will

have a stronger influence on the outcomes of jury deliberations. Further, majority factions can have a large impact on the proceedings of a deliberation.

The Director's Cut Model. Devine uses the metaphor of a director creating a final cut of her movie in order to emphasize his assumption that the final stage of the decision process at the juror level represents the collection of post-trial story structures that every individual juror will carry with them into deliberation. Thus, before a trial a juror may have access to some information in the form of existing knowledge about the case (through pre-trial publicity for example), their own beliefs and assumptions and the opening statements provided by the Crown and Defense. Devine emphasizes that the impact of opening statements could be quite large on the formation of initial story structures, particularly when we consider that the opening statement of the Crown will likely take the form of a narrative connecting the evidence to imply a Defendant's guilt. On the other hand, the opening statement of the Defense may take the form of a counter narrative that implies innocence, but may also take the approach of presenting several narratives to imply that there is no legal or logical reason to believe the Crown's story is true, attempt to poke holes in the Crown's narrative by highlighting specific oversights, or imply the Defendants' presence at trial is the result of prejudice or legal error (Devine, 2012). Devine very interestingly notes that the case for the Defense may possibly be the only incident in common argumentation where a party may present no counterargument at all and still expect to win. If the Crown's story is not believed, then the Defense wins the argument by default.

In any case, Devine indicates that jurors likely form a premise for their story alternatives at this early stage and update their story model as evidence is presented. Like a director might

do when editing the final cut of their film, jurors may edit some evidence out of their final stories that doesn't fit the story structure they find most appealing or causes too much conflict. Ultimately, this will leave a juror in one of four states. A "believer" favors the Crown's case and is content with a story structure that assumes the guilt of the defendant. A "doubter" favors the Defense's case and either sees no merit in any story structures that imply guilt or sees great merit in a story structure that implies innocence. A "muller" may side with the Defense or the Crown but has a hard time considering which of several possible story structures best explains their decision. For example, they may believe that the defendant is guilty, but remain unsure whether they should integrate inconclusive DNA evidence into their story structure. A "puzzler" is a juror who remains undecided after all the evidence has been presented. This is not very far from the more traditional representation of the story model but Devine does integrate many empirical findings in his model with the aim of better explaining and understanding why a juror may end up at one of these four states. Devine posits that the combination of the narrative opening statements, specific characteristics of the case (e.g., number of charges, pre-trial publicity), and the individual characteristics of the juror and the defendant (e.g., race, gender, political views) will all have an effect on a juror's sense of a defendant's pretrial culpability. Presentations of evidence by the Crown and Defense will then affect the likelihood that a juror will believe the Crown's story or attempt to create or believe an alternative story and this will ultimately determine whether each juror's individual post-trial story structure connects or extricates a defendant from the charges presented.

The Story Sampling Model. Jury deliberation mainly consists of jurors sharing their post-trial story structures with each other during the "story sampling" phase. According to Devine, the most critical aspect of this process is the number and quality of the stories offered by jurors to explain the evidence. He argues that arguments presented by jurors have a hierarchical standing that inversely relates the frequency of a type of argument with its influence. At the bottom of the pyramid, isolated statements about specific events occur most frequently but are the least influential type of juror interaction. Following this are factual assertions. These are statements that may not necessarily be true but are founded in the facts of the evidence and the trial such as the assertion that a witness identified the defendant at the scene of the crime. The testimony may be false but it is part of the trial evidence nonetheless. Following factual assertions are arguments relating to a juror's verdict preference. These arguments stem from a juror's belief in the innocence or guilt of a suspect and aim to persuade other jurors to accept the presented verdict as the appropriate choice for the case. Finally, fully fledged story presentations are the least frequent but most influential of the argument classes. In this case, a juror will provide a coherent narrative to relate all the evidence in a manner that implies a specific verdict may be appropriate. These types of arguments, together with the influence of jury deliberation type (to be explained shortly), polling frequency and timing, and the presence or absence of a majority faction influence, will all have an effect on a juror's preferred verdict choice during deliberation. The amalgamation of juror verdicts will in turn determine the jury verdict.

In this model, jury deliberation is usually verdict-driven or evidence-driven. Verdict-driven juries focus their discussion on choosing a verdict while evidence-driven juries focus on

carefully reviewing each piece of evidence or testimony to establish what the facts of a case are. In turn, this may lead to early and frequent verdict polls if a verdict-driven deliberation style is operating and to postponed and less frequent polls if deliberation is evidence-driven (Hastie et al., 1983). Based on Petty and Cacioppo's (1986) "Elaboration Likelihood Model", Devine indicates that the interactions between the elements described above will usually operate by affecting either a juror's cognition or his/her emotions. Petty and Cacioppo's model outlines two main routes for persuasion. The central route to persuasion operates through information and factual influence and relies on a careful and logical evaluation of the coherence and plausibility of given arguments. On the other hand, the peripheral route to persuasion is less effortful and relies on mental heuristics and shortcuts or other normative influences that help evaluate the appeal of an argument. In this vein, Devine posits that jurors will influence each other through informational or normative routes. Thus, in the "Story Sampling Model" jurors' post-trial stories will influence their inclination to provide arguments in the form of isolated statements, factual assertions, verdict preferences, or fully fledged stories. These arguments are mediated by the deliberation style of the jury and can have either an informational influence (through the central route of persuasion) or a normative influence (through the peripheral route) on a juror's final story. The likelihood of informational influence increases the higher the argument presented is in the pyramid and the likelihood of normative influence mainly operates through the faction power of majority factions.

One major contribution that Devine's multi-level model makes is that it allows for the individual evaluation of separate jurors. While many of the traditional models touch on some of the same principals, jurors are generally portrayed as interchangeable decision makers.

Devine's model on the other hand allows us to account for individual differences that may have a large impact on the decision models and strategies that each juror may employ. Further, though Devine's model, like Bex et al.'s model, does not provide an outright evaluation of decision criterion, an observation of a juror's state after the director's cut portion of the model will give an indication whether they have reached that criterion or not. For example, "believers" and "doubters" have likely established their beliefs beyond a subjective decision threshold while "mullers" and "puzzlers" may still fall below those thresholds. Again, the issue here is somewhat beyond the constraints of the model although Devine's approach does provide an opportunity to integrate decision thresholds at the level of each individual juror rather than treat the concept of reasonable doubt as a universal variable. This approach seems logical considering that notions of the threshold of reasonable doubt are varied and ill-defined (Dhami, 2008). Devine's model also has broad applicability over different case types (criminal and civil) and charges and provides a map for testing and understanding some of the underlying cognitive structures involved in jury decision making.

In summary, current research models show the potential for integrating mathematical and psychological approaches to understanding juror decisions. Specifically, this integration speaks not only to how jurors and juries reach verdict decisions, but also to how policy, instructions, and evidence presentation (among other factors) may help to optimize this decision process. As suggested by Bex et al. (2010) psychological models can be used to illuminate the storyboarding process utilized by jurors while logical and mathematical models can be used to evaluate the validity of different stories. Devine elaborates on this approach to understanding juror decisions by delineating a structural map of the potential influences on

jurors beyond just the evidence presented at a case. Some of these factors include exposure to pre-trial publicity, and the demographic characteristics of the defendant and jury. Devine expands his model to also include a structural demonstration of how individual jurors can affect each other through normative or informational routes. This influence in turn affects jury deliberation and verdict decisions.

The present theoretical framework on juror decisions suggests that they likely update their beliefs about the culpability of a defendant in a step-by-step manner, especially when evidence sets are of a complicated or contradictory nature. This means that evaluations of culpability are updated after the presentation of each argument or piece of evidence rather than after all the information has been presented. Research has demonstrated that the SbS belief update system does lead to some challenges in the process of evaluating evidence. These primarily include the effects of the order of evidence presentation, the fact that jurors will use information that they deem to be useful regardless of its source (e.g., inadmissible evidence), and the anchoring effects of exposure to certain arguments of culpability. Luckily, some of these effects can be observed and evaluated by using the appropriate models. Research also demonstrates that some of the biases in decision making that result from these issues can be mitigated by the appropriate task instruction or by educating decision makers on their effects.

Thus, current models of juror decision making demonstrate that understanding when and how errors of judgement may occur is instrumental to the goal of optimizing the juror decision process. Devine's model in particular provides a broad perspective on many of the theoretical, practical, and empirical influences on this process. Even though the model does not propose a quantitative estimate of the effects of its different factors, this could be achieved

through the diagnostic analysis of juror decisions over time. The qualitative nature of the model does however highlight that the operation of many factors which affect juror decisions can and will vary dramatically across different individuals. In other words, this slightly reframes the task from “identifying an optimal juror decision strategy” to “identifying the optimal decision strategy for *this specific* juror”.

The limitations of even the most elaborate juror decision models highlight that the task of relating juror beliefs to juror decisions is still in need of careful observation. Many models for example indicate that jurors may shift their beliefs of culpability on a continuum between innocence and guilt yet it is unclear how their standing could determine a verdict of guilt. This relates to the pursuit of identifying an objective standard of proof for guilt or innocence and the question of whether this is even an attainable goal for any theoretical decision model. Nonetheless, our current understanding of some of the aspects that are influencing juror and jury decisions provides great promise towards the optimization of the decision process and the study of its utility as an indication of true culpability.

Proposed Research

The research for this dissertation aims to expand our current knowledge of jury decision models and strategies. While many different approaches have been taken to address this challenge, there is general agreement among researchers that juror decisions likely result from the interaction between the juror's beliefs before and after exposure to evidence and the trial. In other words, jurors all have intrinsic biases and beliefs, which influence and are influenced by the evidence they encounter during the trial and their interaction with other jury members. In the Bayesian model of juror decision making, these are referred to as the juror's estimate of "prior odds" and the "likelihood ratio" of the evidence. Stochastic models refer to these factors as the juror's initial "anchor" and the "adjustment" which results from exposure to evidence. Psychological models including the story model refer to a juror's "prior beliefs, biases and heuristics" and the "defeasible story alternatives" that form as a result of a juror's inference to the best explanation based on the evidence available.

Traditionally, juror decision models have treated jurors as interchangeable factors in the model that interact with the evidence in a uniform manner. Recent efforts however, such as Devine's (2012) multi-level theory, begin to outline the importance of considering the individual differences between jurors. Further, many models continue to operate under the assumption that jurors (as well as judges, lawyers, and any other legal decision makers) are able to evaluate the objective utility of different forms of evidence and that these effects will be uniform across different conditions. Some recent research critically highlights the relevance of considering individual differences between jurors and the importance of integrating some form of absolute (rather than comparative) mechanism for valuating different forms of evidence. For example,

much of the research discussed earlier implies that the efficacy of perceivably strong evidence like a DNA match can be highly influenced by whether it is presented as a statistic or a frequency. Similarly, other research indicates that while eyewitness testimony can be very effective, it only seems to maintain its integrity as a form of evidence for guilt (e.g., Dahl & Brimacombe, 2009). For example, while a confident eyewitness who identifies a suspect at the scene of a crime is the strongest predictor of verdict, an alibi witness who identifies a suspect away from the scene of a crime is often ignored or underutilized. Even alibi testimony that is effective at shifting perceptions of guilt is rated with relatively low credibility and often underutilized. Similar trends are also seen with non-identification evidence when a witness indicates that a suspect in a line-up is not the person they saw. Interestingly, in their proposed taxonomy of what makes a good alibi, Olson and Wells (2004) found that even the strongest forms of alibi corroboration only produced believability ratings from their participants of seven on a ten point scale. In fact, alibi corroboration in their study only affected believability ratings of the alibi when there was no form of physical evidence present. In other words, if an alibi provider presented physical evidence for their alibi, like a receipt or security camera footage, the alibi witness' effect on believability was nullified.

These findings raise some intriguing questions regarding how the effect of evidence can be integrated into a juror decision model. Specifically, they give concern over the notion of evaluating evidence that implies guilt versus evidence that implies innocence. Some of the theoretical underpinnings for suggesting that evidence for guilt may be evaluated differently than evidence for innocence have been discussed earlier. For one thing, the argument for *legal* culpability as opposed to *substantive* culpability means that the Crown and the Defense will

adopt different strategies to defend their positions and in turn activate different decision strategies in jurors. Further, the framing of the question of legal culpability can have a large effect on the manner in which evidence is perceived. Specifically, the notion of assuming that a suspect is "innocent until proven guilty" places the burden of proof on the Crown and frames the task of the juror as investigating whether there is enough evidence to prove guilt beyond a reasonable doubt. This may bias jurors with a drive to search for evidence for guilt and undermines the implicit directive for a balanced evaluation of incriminating and exoneration information. Traditional verdict categories also highlight that the framing of the juror decision task may require further study if we wish to balance the evaluation of evidence. Even when a suspect is acquitted they are found "not guilty" rather than "innocent".

The notion that evidence for innocence is underutilized has been discussed in the literature but it remains unclear how this may manifest in a juror decision model and what the primary factors contributing to such an effect may be. A primary goal of the research in this dissertation is to investigate the function of decisions of culpability when a participant encounters evidence for guilt and evidence for innocence. To investigate the function of estimates of guilt, participants will complete a series of Bayesian inference problems where the prior odds that each of two suspects is involved in a crime or work negligence scenario will be presented. Following this, information from an eyewitness who insinuates or extenuates one of the parties from the problem scenario will be presented with the results of a diagnostic test that determines their accuracy as an eyewitness for the scene in question. Specifically, study 1 investigates the similarities and differences in participants' estimates of guilt to these Bayesian inference problems where a witness either insinuates or extenuates a suspect from a criminal

or work negligence activity. Studies 2 and 3 expand on the findings from study 1 by further investigating how manipulating the prior odds of guilt in the same Bayesian inference problems affects the functions and models used by participants to estimate the probability of culpability.

Hypotheses

The main hypotheses are:

1) The function or model used by participants to evaluate the probability of culpability will be the same for scenarios involving crimes and work negligence. This is consistent with Devine's (2012) summary and estimate that absolute evaluations of culpability will be the same or similar in criminal (represented by crime scenarios in the study) and civil cases (represented by work negligence scenarios in the study) though output decisions in the form of sentencing or liability may differ. In other words, perceptions of culpability will be the same but the decisions to act on those perceptions may vary.
2) The accuracy of eyewitness testimony will have a significant effect on estimates of the probability of guilt. This is consistent with Hinsz et al.'s (2008) research indicating that highly accurate witnesses will polarize estimates in their direction while low accuracy witnesses will cause bias towards presented base-rates in the inference problem.
    a. The effect of the accuracy of the witness will be smaller in scenarios where the witness exculpates a suspect than when the witness implicates them. In other words, an incriminating eyewitness will lead to bigger shifts in judgements of probability of guilt as the witness gets more accurate than an exonerating eyewitness.

b. The function of the decision model will be the same for exculpating and implicating testimony (i.e., they will both be linear or quadratic, etc.) but decision thresholds may vary. For example the observed results may be linear in both cases but differ in their intercept and slope.

3) Evaluations of probability of guilt in the scenarios where the witness provides exculpating testimony may demonstrate an added variable that is representative of the extra step in accounting for evidence for innocence when the decision task is framed in the context of estimating guilt.
4) The functions for the evaluation of evidence will remain consistent regardless of the initial presentation of prior odds. Specifically, implicating and exculpating witness testimony will have consistent effects on judgements of probability of guilt regardless of the initial base-rate of participation of the implicated suspect.
5) Participants will show higher confidence ratings in their judgements when witnesses have high accuracy as opposed to when they have low accuracy
    a. Participants will show higher confidence ratings when the witness implies guilt as opposed to when he/she implies innocence.
    b. Confidence ratings will be highest when witnesses are either highly accurate or highly inaccurate. This implies that participants many feel safer in completely ignoring low accuracy witness or completely embracing high accuracy witnesses and so have more confidence in the information they are using for a decision.
    c. Confidence ratings will be consistent across different base-rates.

d. Confidence ratings will be higher when participants indicate they relied on only one piece of evidence.

Chapter 3: Study 1 Methods, Results and Discussion

Study 1 is an expansion on Hinsz, et al.'s (2008) research on judgements integrating base rate and case-specific information discussed earlier. To summarize, Hinsz et al. investigated how probability of guilt judgements varied in cases where participants integrated base-rate and implicating witness information of varying degrees of accuracy in a probabilistic inference problem. They were interested in how group decision processes accentuate base-rate neglect or any other effects that are found in individual decision tasks that involve the integration of base rate and case-specific (i.e., eyewitness) information.. Hinsz et al. used a template of the Kahneman and Tversky's cab problem to create ten scenarios that were presented to participants, five involving crimes and five involving work negligence. They then asked their participants to give judgements of the probability of guilt for a given party in each scenario and found that a highly accurate eyewitness led to a strong effect of base rate neglect where probability judgements were polarized towards the given accuracy of the witness and did not match the calculated Bayesian solutions for the given scenarios. Further, this effect was accentuated when decisions were made in groups. Their participants seemed to attempt to integrate the base rate and case specific information most often when accuracy levels were at or around the 50% mark. At low accuracy levels participants actually polarized their responses towards the base rate information.

Hinsz, et al.'s research provides valuable insight into the potential strategies that their participants used in attempting to integrate the evidence they were presented with and neatly highlights some of the benefits and limitations of using Bayesian models of reasoning when dealing with evidence evaluation. However, their research only addresses the issue from the

perspective of case-specific evidence that incriminates a certain party. Thus, study one of the dissertation replicates certain aspects of Hinsz et al.'s work with the added manipulation of introducing case specific information that exonerates a given party. As with Hinsz et al.'s. research, the current study will provide insight into the manner with which participants attempt to integrate base rate and case specific information and will further provide a comparison between case specific information that is incriminating or exonerating. In addition to providing information on what kind of cognitive models participants are using to integrate different sources of information, the findings will also speak to the consistency of the models used across different forms of evidence.

Participants

Two hundred and eleven students (172 female, 39 male) enrolled in an Introduction to Psychology course at Ryerson University participated in this study as a partial requirement for course credit. The study was advertised on the online participation program SONA, under the title "Can You Identify the Culprit?" Each participant received one course credit for their participation in the one hour study.

Materials

Participants read ten scenarios based on Hinsz et al.s (2008) modifications of the blue bus problem. Some of the scenarios were constructed for this study while others where modifications of some of the scenarios used by Hinsz et al. Five of the problems involved crime scenarios and the remaining five involved work negligence scenarios. For each of the problems, two parties are involved with a specific crime or work negligence scene, one at a high base rate (85%) and the other at a low base rate (15%) (see Appendix I and II for examples of problem

scenarios). For each scenario, after the crime or work negligence occurs, case specific eyewitness evidence is presented at varying degrees of accuracy (20, 40, 50, 60, 80% accuracy) that either incriminates or exonerates one of the two parties. The degree of accuracy of the eyewitness was determined through a diagnostic test that is presented as part of the scenario description and provides an accuracy rating for the witness on a task similar to the one relevant to the specific scenario. This diagnostic test is designed to present information on the witness' hit and false alarm rates.

Design

Every participant completed one study booklet that contained ten scenarios, five of which involved crimes and five of which involved work negligence. Within each subset of scenarios, each accuracy level of the eyewitness appeared only once so that there were five crime scenarios and five work negligence scenarios with witnesses at 20, 40, 50, 60, and 80% accuracy. A Latin square design was used to create and select five study booklets. This ensured that accuracy-scenario pairings and the presentation order of scenarios were counterbalanced across participants. The same five booklet orders were used for exonerating and incriminating witness scenarios. Participants were randomly assigned to read scenarios that included exonerating eyewitness evidence or incriminating eyewitness evidence. The study followed a 2 (condition: exonerating or incriminating case specific evidence) X 5 (accuracy of eyewitness: 20, 40, 50, 60, 80%) X 2 (subset: crime or work negligence) mixed factorial design with condition manipulated between subjects and accuracy and subset manipulated within subjects. Participants provided three responses for each scenario. The first was a judgement for the probability of guilt of the party identified by the eyewitness expressed as a percentage (0-

100%). The second was a confidence rating in their probability judgement given on a 9 point Likert-type scale ranging from “Not at all confident” to “Extremely confident” and the third was a categorical response to a 3 point multiple choice question asking whether they believed that a)given the eyewitness evidence the base rate information was irrelevant, b)given the base rate information the eyewitness information was irrelevant, or c)both the eyewitness and base rate information are equally relevant.

Procedure

Participants were tested in groups of up to four but were notified prior to beginning that even though they were completing the task in a group setting, it was an individual task and they should refrain from discussing the scenarios or their answers until the study was done. Signed consent was obtained from the participants on the day of their participation after a description of the task and experimental procedure was given. Participants were told that they would be taking part in a study that investigates how potential jury members use different forms of evidence in order to make a verdict decision. They were instructed to read the scenarios and imagine being a jury or committee member attempting to determine how likely a given party is responsible for a crime or act of work negligence given only the information presented in each scenario. Participants were asked to complete their responses to the scenarios in the order that the scenarios were presented in and when in groups, asked to wait upon completion of their booklet for all other participants to finish before moving on to the next step of the study. The researcher visited the study room at 10 minute intervals to address any questions and check on the progress of the group. When all participants were finished, the researcher collected all the study booklets and provided the participants with a verbal debriefing of the study. The

participants were informed that the aim of the study was to investigate the interaction between eyewitness evidence and base rate information especially in light of the effects of base rate neglect (which was explained to the participants) and the potentially different effects that different kinds of eyewitness information (whether it is exonerating or incriminating) may have. Participants were also asked about their reactions to the scenarios and any strategies they may have been using to address the task. Time was also given to address any questions or comments that the participants may have had. Participants were finally informed that their participation credit had been granted and thanked for their participation. They also received a written debriefing form outlining the purpose of the study and its hypotheses as well as contact information for the lab and researcher in case there were any future inquiries.

Results

Probability of guilt judgements and confidence ratings were subjected to a 2 (condition: extenuating or insinuating case specific evidence) X 5 (accuracy of eyewitness: 20, 40, 50, 60, 80%) X 2 (subset: crime or work negligence) repeated measures mixed model ANOVA.

Comparison of Scenarios and Participant Booklets:

An analysis of variance was conducted to make sure specific scenarios and presentation order were not significantly affecting probability of guilt judgements and confidence ratings. There were no significant differences between the ten presented scenarios on average probability estimates, F (9, 210) = 1.256, p = .256, $\eta_p^2 < .01$ or confidence ratings, F (9, 210) = 0.514, p = .865, $\eta_p^2 < .01$. This was true when the scenarios implied both innocence and guilt. Scenario booklets containing exonerating evidence showed no order effects for probability judgements, F (4, 105) = 1.16, p = .327, $\eta_p^2 < .01$ but did show a significant effect for confidence

ratings, F (4, 105) = 4.991, p = .001, η_p^2 = .02. Booklets containing scenarios with incriminating evidence showed order effects for probability judgements, F (4, 106) = 2.923, p = .020, η_p^2 = .01 and confidence ratings, F (4, 106) = 3.963, p = .003, η_p^2 = .02. Mean probability and confidence ratings can be seen for the scenarios and booklets in Tables 1 and 2 below, respectively.

Table 1. *Average Probability of Guilt and Average Confidence for Each Scenario*

		Alibi				Eyewitness			
	Scenario	**Avg. Prob**	**SD**	**Avg. Conf**	**SD**	**Avg. Prob**	**SD**	**Avg. Conf**	**SD**
Crime	**Hit & Run**	58.79	26.20	5.97	1.74	39.00	23.90	6.18	1.72
	Get Away Car	56.57	18.22	5.89	1.78	43.53	21.56	6.17	1.67
	Ambulance	60.81	24.39	6.11	1.78	38.51	25.09	6.28	1.92
	Handguns	59.93	23.58	5.79	1.89	44.92	25.93	6.24	1.73
	Football	58.29	23.09	5.88	1.70	37.63	18.35	6.03	1.82
	Group Mean	***58.88***	***23.10***	***5.93***	***1.78***	***40.72***	***22.97***	***6.18***	***1.77***
Work Negligence	**Art Store**	59.50	21.22	5.67	1.78	36.83	22.41	6.17	1.57
	Kitchen	54.84	24.76	5.95	1.68	44.65	24.66	6.09	1.79
	Glass	53.64	25.99	5.81	1.71	39.90	24.45	6.19	1.69
	Wine	61.97	22.05	6.10	1.78	40.63	23.94	6.14	1.81
	Zoo	59.03	23.27	5.70	1.65	33.37	20.20	6.16	1.70
	Group Mean	***57.80***	***23.46***	***5.85***	***1.72***	***39.08***	***23.13***	***6.15***	***1.71***
	Grand Mean	**58.34**	**23.28**	**5.89**	**1.75**	**39.90**	**23.05**	**6.17**	**1.74**

Table 2. *Average Probability of Guilt and Average Confidence by Study Booklet*

	Study Booklet	**Avg. Prob**	**SD**	**Avg. Conf**	**SD**
Alibi	A-Alibi1	59.24	24.45	6.04	1.46
	B-Alibi2	58.55	23.15	5.83	1.79
	C-Alibi3	58.77	25.39	5.49	2.03
	D-Alibi4	59.72	22.74	6.20	1.60
	E-Alibi5	55.24	20.85	5.90	1.73
	Group Mean	***58.3***	***23.31***	***5.89***	***1.72***
Eyewitness	F-Eye1	40.06	21.80	5.82	2.04
	G-Eye2	44.28	24.59	6.10	1.83
	H-Eye3	37.02	23.51	6.42	1.66
	I-Eye4	38.29	23.89	6.34	1.49
	J-Eye5	39.83	22.36	6.15	1.54
	Group Mean	***39.9***	***23.23***	***6.17***	***1.71***
	Grand Mean	**49.07**	**23.27**	**6.03**	**1.72**

The implications and magnitude of order effects will be further discussed later in the context of the limitations of the current study.

Tests of assumptions of ANOVA for probability of guilt estimates and confidence ratings:

Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of accuracy on probability judgements, $\chi^2(9) = 144.73$, $p < .001$, and confidence, $\chi^2(9) = 17.90$, $p < .001$. The assumption of sphericity was also violated for the effect of the subset by accuracy interaction on probability judgements, $\chi^2(9) = 33.58$, $p < .001$, and confidence, $\chi^2(9) = 35.76$, $p < .001$. Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .70$ and $\varepsilon = .96$ for the main effect of accuracy on probability judgements and confidence respectively. $\varepsilon = .93$ and $\varepsilon = .92$ for the main effect of the subset by accuracy interaction on probability judgements and confidence respectively). Results are reported with unadjusted degrees of freedom where tests of significance did not differ. All results are reported as significant at $p < .05$.

In order to account for violations of variance, the analyses were also tested with a log10 transformation of the data. Result patterns did not differ for the transformed data. Friedman's ANOVAs were also conducted as a precautionary measure and confirm the observed results from the ANOVAs.

Probability of guilt estimates and confidence ratings for subset, accuracy, and condition

The main effect for the subset of the scenario on judgements of probability of guilt approached significance, $F(1, 209) = 3.49$, $p = .063$, $\eta_p^2 = .02$ and was not significant for confidence in judgements, $F(1, 209) = 1.45$, $p = .230$, $\eta_p^2 = .01$. This implies that participants may have been using similar decision strategies to evaluate crime and work negligence

scenarios although some differences may be apparent. There was a significant main effect for the accuracy of the witness on judgements of probability of guilt, $F = (4, 836) = 43.47$, $p < .001$, $\eta_p^2 = .17$. Contrasts revealed significant linear, $F\ (1, 209) = 77.32$, $p < .001$, $\eta_p^2 = .27$, quadratic, $F = (1, 209) = 12.53$, $p < .001$, $\eta_p^2 = .06$, and cubic, $F\ (1, 209) = 16.23$, $p < .001$, $\eta_p^2 = .07$ trends indicating that the function of change for probability estimates varied as witness accuracy increased. The accuracy of the witness also showed a significant main effect on confidence in judgements, $F\ (4, 836) = 13.69$, $p < .001$, $\eta_p^2 = .06$. Contrasts revealed a significant quartic trend, $F\ (1, 209) = 17.68$, $p < .001$, $\eta_p^2 = .08$, indicating that confidence in judgements rises and falls several times across different levels of witness accuracy. Figure 1 below, highlights that confidence ratings were higher at 20%, 50%, and 80% levels, but dropped at 40% and 60% levels.

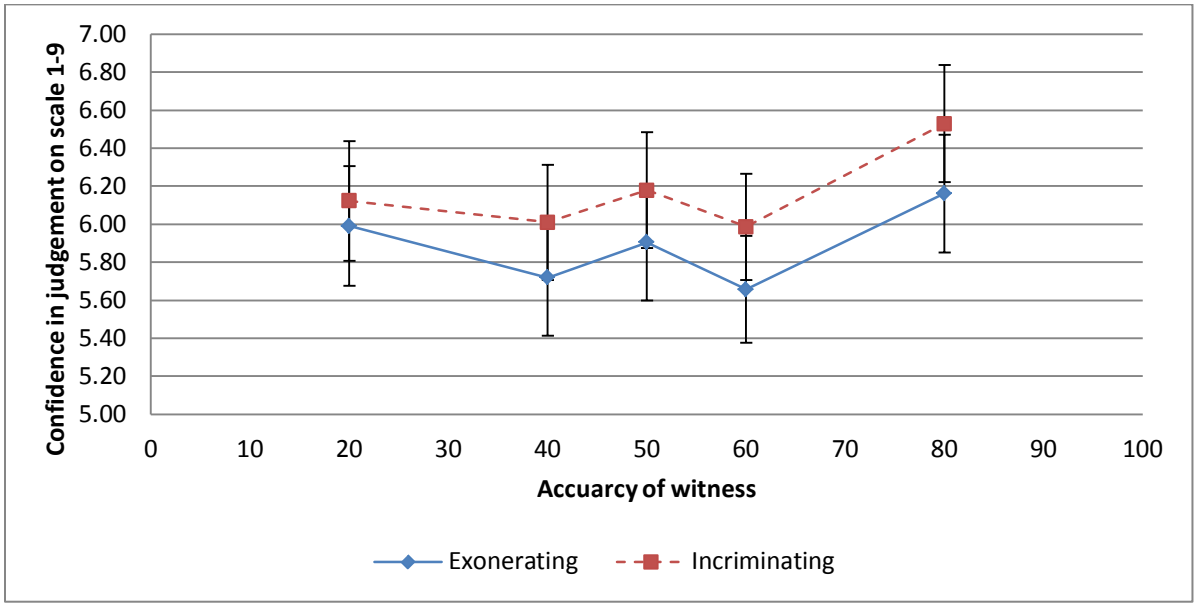

Figure 1. Axis range is truncated to highlight effect. Mean confidence rating for exonerating and incriminating scenarios across varying levels of accuracy of witness plotted with 95% confidence intervals.

There was a significant main effect of condition, $F\ (1, 209) = 111.22$, $p < .001$, $\eta_p^2 = .37$ on probability of guilt judgements indicating that participants who read exonerating evidence

gave higher probability of guilt estimates on average than participants who read incriminating evidence. This implies that the participants did use the base rates of implication in the exonerating and incriminating scenarios (85% and 15% respectively) as an anchor for their estimates. There was no significant main effect of condition on confidence in judgements, F (1, 209) = 2.16, p = .143, η_p^2 = .01.

There was no significant interaction between the subset of the scenarios (crime or work negligence) and condition (exonerating or incriminating eyewitness testimony) for probability judgements, F (1, 209) = .15, p = .701, η_p^2 = .01, or confidence in judgements, F (1, 209) = .34, p = .559, η_p^2 < .01. There was a significant interaction between the accuracy of the witness and condition for probability judgements, F (4, 836) = 66.41, p < .001, η_p^2 = .24 but not for confidence F (3.86, 586.58) = .64, p = .637, η_p^2 < .01. This indicates that the effect of increasing witness accuracy on the probability of guilt estimates produced by the participants differed when they encountered scenarios with exonerating or incriminating evidence. Contrasts revealed significant linear, F (1, 209) = 121.98, p < .001, η_p^2 = .37, quadratic, F (1, 209) = 23.72, p < .001, η_p^2 = .10 and cubic trends, F (1, 209) = 5.67, p = .020, η_p^2 = .03. Further, estimates at 80% accuracy were higher than those at 60% accuracy, F (1, 209) = 5.26, which were higher than those at 50%, F (1, 209) = 9.94, which were higher than those at 40%, F (1, 209) = 13.77, which were higher than those at 20%, F (1, 209) = 56.01. Looking at the interaction graph in Figure 2 below indicates that this interaction is largely mediated by changes in probability of guilt estimates for the incriminating group. Further analysis indicates that the main effect of accuracy on probability judgements is significant for the incriminating group, F (4, 420) = 126.49, p < .001, η_p^2 = .55 but not for the exonerating group, F (4, 416) = 1.44, p = .223, η_p^2 =

.01. Thus, increasing the accuracy of a witness produced a significant increase in estimates of probability of guilt when the witness provided incriminating testimony. Witnesses that provided exonerating testimony did not produce any significant decrease in estimates of probability of guilt.

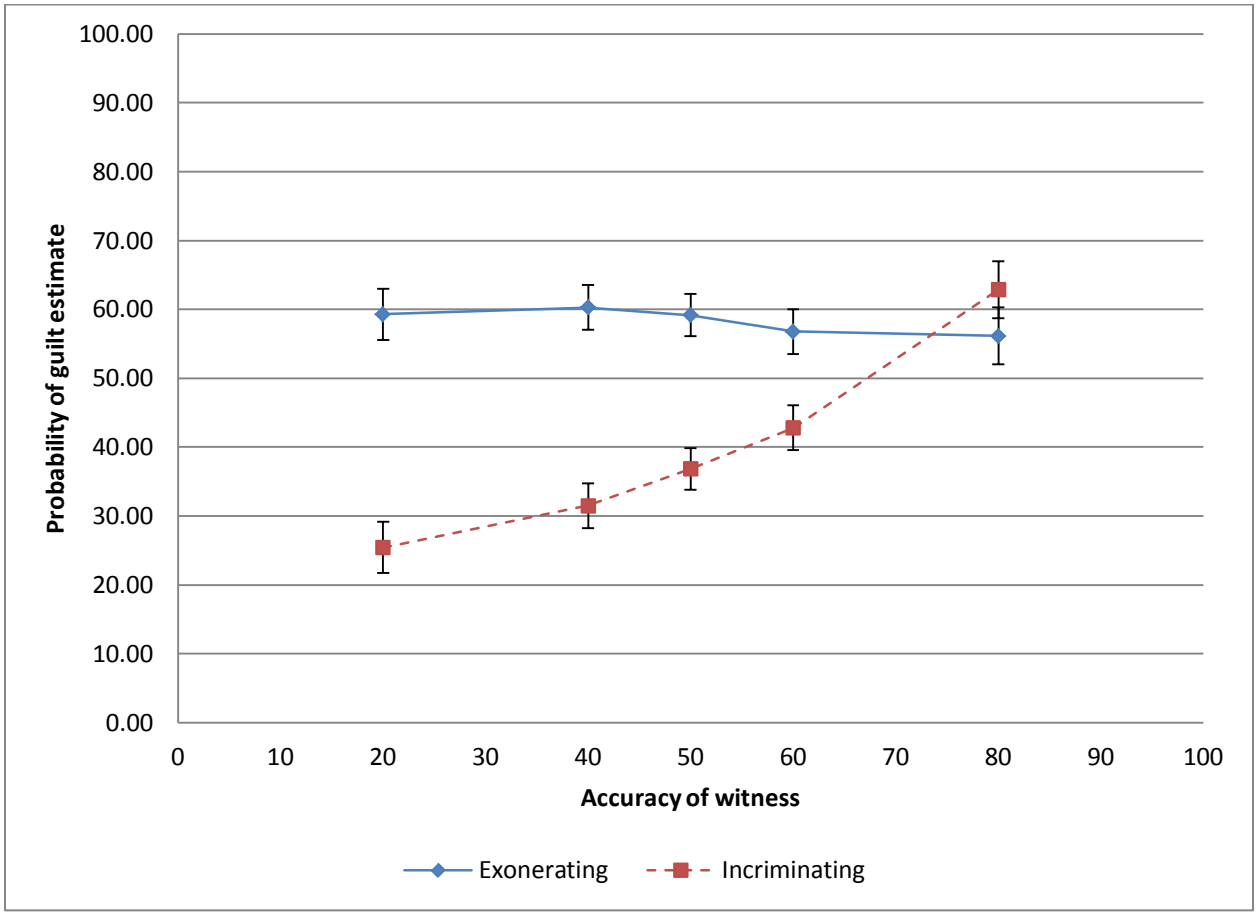

Figure 2. Mean probability of guilt judgement for exonerating and incriminating scenarios across different levels of witness accuracy plotted with 95% confidence intervals.

There was a significant interaction effect between subset and accuracy on probability judgements, F (4, 836) = 3.77, p = .005, η_p^2 = .02 and on confidence in judgements, F (4, 836) = 3.27, p = .011, η_p^2 = .02. Further analysis indicates that there is a significant difference between probability judgements for crime and work negligence scenes between 80% accurate witnesses and 50%, F (1, 209) = 6.83, 40%, F (1, 209) = 5.56, and 20%, F (1, 209) = 10.48 accurate

witnesses. The interaction graph in Figure 3 below indicates that while probability of guilt judgements are similar between 20% and 60% witness accuracy levels, there is a larger increase in probability estimates between 60% and 80% for crime scenarios.

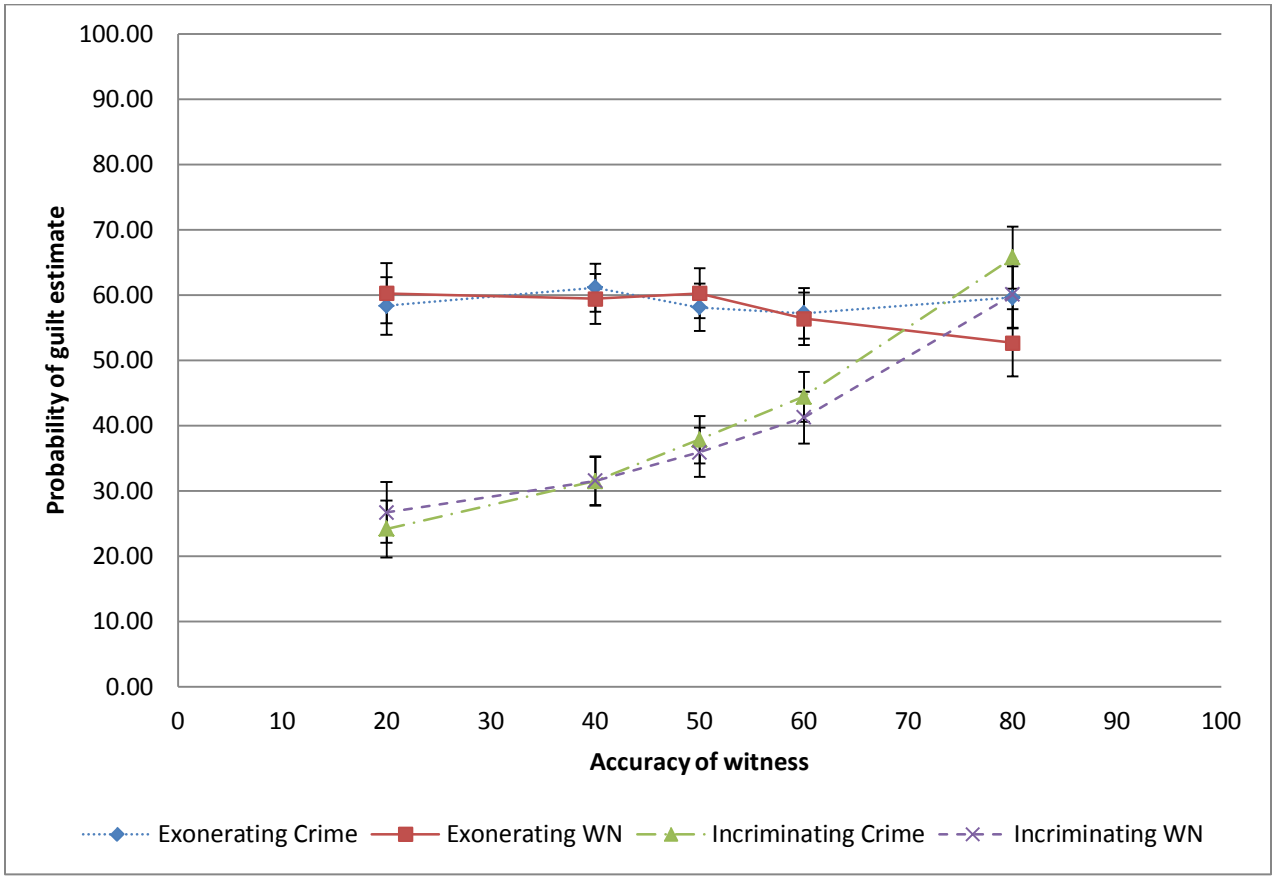

Figure 3. Mean probability of guilt judgement for incriminating and exonerating scenarios across crime and work negligence subsets plotted with 95% confidence intervals.

Contrasts for the effect of the subset by accuracy interaction indicate there are significant differences in confidence ratings between the 20%, 50% and 80% accuracy levels and levels at 40%, F (1, 209) = 5.02, and 60%, F (1, 209) = 11.39. These effects are more pronounced for the crime scenarios and can be seen in the interaction graph in Figure 4 below.

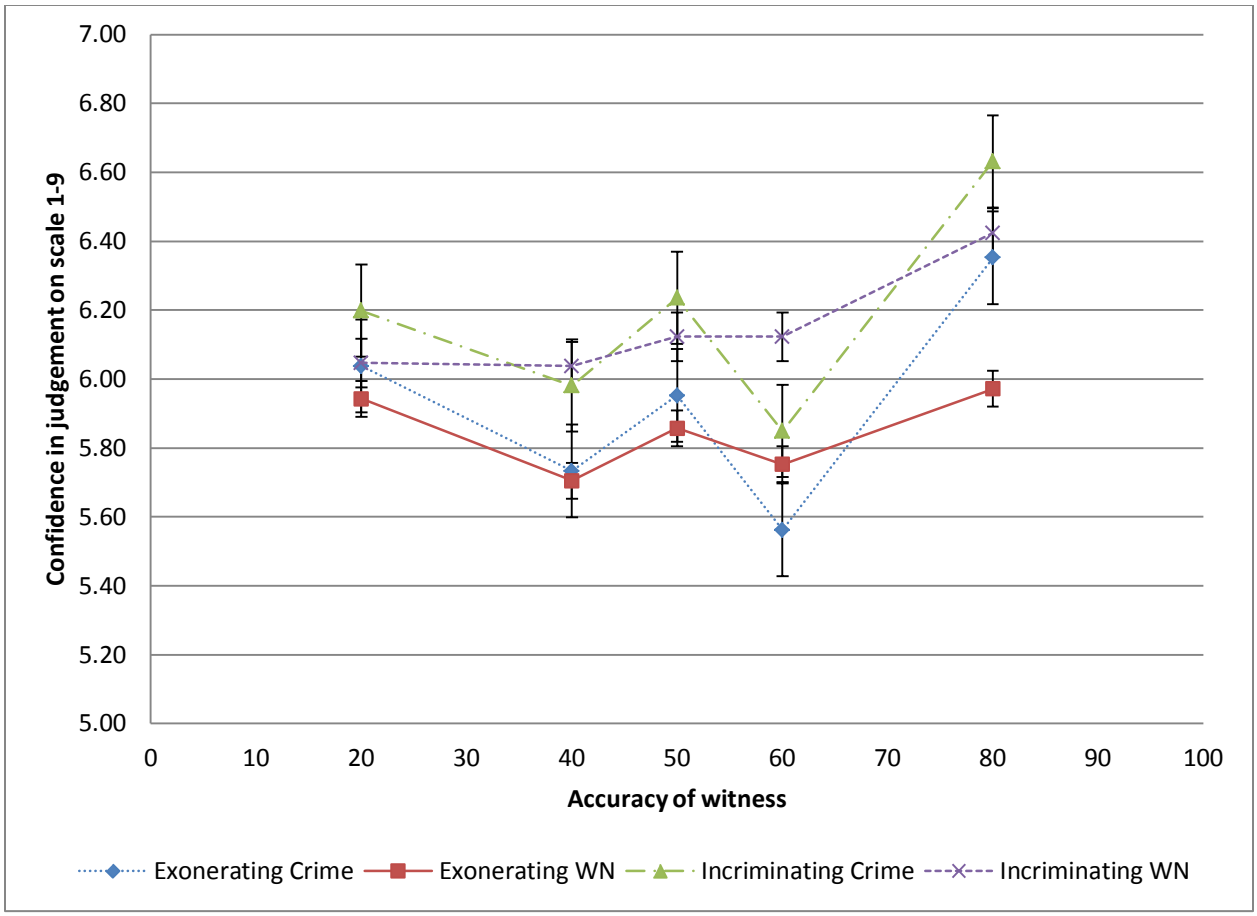

Figure 4. Axis range is truncated to highlight effect. Mean confidence rating for exonerating and incriminating scenarios for crime and work negligence scenes across varying levels of accuracy of witness plotted with 95% confidence intervals.

There was no significant three way interaction between subset, accuracy, and condition for probability of guilt judgements, F (4, 836) = 0.62, p = .648, $\eta_p^2 < .01$, or confidence in judgements, F (4, 836) = 0.18, p = .949, $\eta_p^2 < .01$.

Exploratory analysis for use and importance of base-rates and witness evidence

An exploratory analysis was conducted on participants' categorical responses in order to investigate trends in the reported usefulness and utilization of specific forms of information in the scenarios (i.e., the base-rates and witness evidence). After participant responses were recorded, a computed categorical variable was produced to indicate whether participant

responses were equal to the base-rate, equal to the witness accuracy in the given scenario, or indicative of some operation of mental math and estimation. Overall, participants indicated that they found the accuracy of the witness to be irrelevant in roughly 30% of their responses. Roughly 20% of the time, participants indicated that the base-rate information was irrelevant and on approximately 50% of their responses, they indicated that both pieces of information were equally relevant. A summary of relevance distributions across all levels of incriminating and exonerating witness accuracy can be seen in Table 3 below.

Table 3. *Relevance of Information for Probability Decision Across All Levels of Witness Accuracy (Acc = Accuracy, Br = Base-Rate)*

Incriminating	**Relevance Category**		
Witness Accuracy	**Acc irrelevant**	**BR irrelevant**	**Equal relevance**
20	42.92%	15.09%	41.98%
40	28.77%	21.23%	50.00%
50	24.53%	16.51%	58.96%
60	21.70%	20.28%	58.02%
80	15.57%	34.43%	50.00%
Grand Total	**26.70%**	**21.51%**	**51.79%**
Exonerating	**Relevance Category**		
Witness Accuracy	**Acc irrelevant**	**BR irrelevant**	**Equal relevance**
20	54.29%	14.29%	31.43%
40	44.76%	13.33%	41.90%
50	30.48%	16.19%	53.33%
60	17.62%	19.05%	63.33%
80	12.38%	30.48%	57.14%
Grand Total	**31.90%**	**18.67%**	**49.43%**

Analysis of the response category measure indicates that overall, participants provided probability of guilt estimates that were equivalent to the witness accuracy in a given scenario for approximately 20% of their responses. Their estimates were equal to the base-rate in approximately 15% of their responses and indicative of some form of mental math or estimation in approximately 65% of their responses.

A comparison of the relevance category and response category for participants indicates that their responses to the relevance of the information are not always indicative of their behaviour when making a probability estimate (i.e., they may indicate that the base-rate is irrelevant but still try to integrate it into an estimate). This is indicated by the higher percentage of responses representing mental math or estimation compared to the proportion of participants who reported that both pieces of information were equally relevant. Participant response categories were separately analysed for the incriminating and exonerating testimony groups. Table 4 below summarizes these findings.

Table 4. *Response Category for Probability Decision Across All Levels of Witness Accuracy for Incriminating and Exonerating Witnesses.*

Incriminating	**Response Category**		
Witness Accuracy	**EQ BR**	**EQ ACC**	**MENTAL MATH**
20	38.21%	16.51%	45.28%
40	26.42%	16.98%	56.60%
50	23.58%	34.91%	41.51%
60	17.92%	18.87%	63.21%
80	10.38%	30.66%	58.96%
Grand Total	**23.30%**	**23.58%**	**53.11%**
Exonerating	**Response Category**		
Witness Accuracy	**EQ BR**	**EQ ACC**	**MENTAL MATH**
20	9.05%	8.10%	82.86%
40	6.67%	11.43%	81.90%
50	5.71%	36.67%	57.62%
60	4.29%	13.81%	81.90%
80	4.76%	20.48%	74.76%
Grand Total	**6.10%**	**18.10%**	**75.81%**

The response category distributions show a dramatic difference in the proportion of responses indicative of mental math or estimation from the incriminating group (53.11%) compared to the exonerating group (75.81%). This indicates that participants in the exonerating group seemed more compelled to rely on some form of mental estimation in order to produce an estimated measure of culpability.

In summary, the order effects apparent in the study booklets imply that some biases could exist in evaluations of probability of guilt across different groups of participants in the study. While these order effects were more apparent for the incriminating scenarios, average probability of guilt ratings for the study booklets ranged between 55%-60% (5% range) and 37%-45% (8% range) for the exonerating and incriminating booklets respectively. Average confidence ratings ranged through less than one point on the nine point scale. While the order effects are statistically significant, the range of variance implies there may still be some value in comparing the results in an ANOVA.

The effects of subset appear to play a larger role for scenarios with exonerating evidence although this effect only approached significance. Accuracy of the witness on the other hand shows a significant main effect on probability judgements and confidence and seems to be moderated through increases in estimates of guilt as incriminating witnesses increase in accuracy. The significant difference between condition groups (incriminating vs exonerating) indicate that participants probably relied on the base-rate presentation of the suspects in the scenarios as an anchor for their probability judgements.

Discussion

Subset. The results of the study indicate that the subset of the scenario may potentially play a minimal role in participants' estimates of culpability in specific scenarios. Overall, participants gave similar probability of guilt estimates whether a scenario was based on a crime or a work negligence scene. This supports Devine's (2012) suggestions that decision models that apply to juror perceptions of guilt in criminal trials may also apply to jurors in civil trials, even when the response criteria (e.g., verdict, liability, etc.) may differ. Interestingly however,

distinctions between civil and criminal proceedings may be contingent on the nature of exonerating and incriminating evidence that is presented in both cases. The interaction between the subset of the scenario and the accuracy of the witness gives an interesting perspective on the effect of witness testimony that is perceived as highly accurate. In this study, highly accurate witnesses (80%) produced a stronger shift in estimates of probability of guilt in crime scenarios than they did in work negligence scenarios. Similarly, variations in confidence showed a higher sensitivity in crime scenarios than work negligence scenarios. The practical impact of these observations is difficult to interpret given the limited scope of the effect of the subset of the scenarios. These findings may however speak to the notion that while juror decision models may be similar in many cases, there could be deviations from this consistency in cases where evidence is perceived to be very strong or accurate. Further, participants' evaluations of crime and work negligence scenes did not differ when they encountered incriminating or exonerating witness information. Thus, estimates of culpability and confidence were consistent across different levels of the study design further implying that there is little difference in the way participants interpreted work negligence and crime scenes.

Accuracy. The accuracy of the witness had a significant effect on participant estimates of probability of guilt and on their confidence in those estimates. Generally, participants were more confident in their estimates of probability of guilt at high (80%), low (20%), and neutral (50%) levels of witness accuracy. Confidence ratings dropped at medial levels of witness accuracy (40% and 60%) indicating that participant confidence was higher when there was a perceived clear effect for the witness accuracy as opposed to when accuracy levels were more ambiguous. Responses at the 50% accuracy level were particularly interesting since this level of

accuracy is non-diagnostic but extremely useful to the decision process (i.e., it does not affect the mathematical computation but is very informative since it can essentially be ignored). Oddly, participants gave modal responses of around 50% culpability when the witness was 50% accurate and thus applied significant importance to this level though the attribution of significance may be somewhat inappropriate from a Bayesian perspective.

The significant accuracy by condition interaction further indicates that while participant estimates of probability of guilt increased steadily with the increasing accuracy of an incriminating witness, increasing the accuracy of an exonerating witness had no effect on estimates of probability of guilt. This is a potentially concerning finding as it implies that participants may be utilizing different mental models in their evaluation of the relevance and impact of inculpating and exculpating evidence. Specifically, these findings imply that exculpating evidence will have a much smaller effect on perceptions of guilt than inculpating evidence which may in turn influence verdict. This provides a partial explanation for other empirical findings which indicate that an incriminating eyewitness will generally trump any competing form of exculpating evidence (e.g., Olson & Wells, 2004). Though the two types of evidence were not in direct competition in the current study design (the manipulation was between subjects) the data gives a clear indication that exculpating eyewitness testimony was far less effective at changing perceptions of culpability than inculpating testimony.

One strong alternative explanation for this finding is that differences in the efficacy of the testimony may have been due to the original base rate implication of suspects in the incriminating and exonerating scenarios. While the distribution of the base-rates was planned on the assumptions of increasing external validity and avoiding floor and ceiling effects in

participant responses, a confound is presented in that suspects in the exonerating scenarios are always implied at an 85% base-rate while suspects in the incriminating scenarios are implied at a 15% base-rate. Thus, differences between the two groups may be due to challenges relating to the tendency to shift estimates from the given anchoring base-rate, rather than the actual efficacy of the evidence itself. Gilbert (1993) for example indicated the difficulty associated with unbelieving information that we encounter. If we can assume that information presented at a high base rate will be perceived as more believable then it follows that adjustments from such beliefs will also be more difficult to attain. Generally, research on adjustments from high and low anchors indicates that participants usually anchor at the higher end of a range of plausible values when adjusting from a high anchor and at the lower end of a plausible range when adjusting from low anchors (Chapman and Johnson, 2002). In other words, they try to stay as close to an anchor as possible. The data from the current study provide an interesting perspective in light of these predictions. While predictions at the lower end of witness accuracy seem to imply that participants may be using the base rates as an anchor for the probability of guilt of the suspect and give their estimates accordingly, shifts in the probability estimates as the accuracy of the witness increases imply a change in the anchoring strategy. Specifically, it is plausible that participants who read cases with incriminating testimony used the accuracy of the witness rather than the base-rate as an anchor for culpability and thus adjusted their estimates based on an increasing anchor, while participants who read cases with exonerating testimony consistently used the base-rate as an anchor and so showed little change in their estimates of culpability. The findings combined with the various theoretical interpretations

require further study to differentiate the effects of the base-rates from those of the evidence itself. This was partly achieved through follow up testing in studies 2 and 3 of this dissertation.

Condition. In addition to the interaction with accuracy, condition also showed a main effect on probability of guilt judgements indicating that participants who read scenarios with exonerating witnesses generally gave culpability estimates that were higher than those given by participants who read incriminating scenarios. As mentioned above, this implies that participants did rely on the base-rates to some extent in order to establish an anchor for probability of guilt judgements even though their strategies for belief adjustment do differ across the two conditions.

Relevance of information and response category. The exploratory analysis of participant responses regarding the relevance of the information and its relation to the accuracy and base-rate strongly implies different strategies of information integration across varying conditions. Table 3 above indicates a clear trend of change for the relevance category as witness accuracy increases. It can be seen from the table that participants more often indicated that witness testimony was irrelevant at low witness accuracy levels. The opposite trend can be seen for indications of the base-rate being irrelevant. In other words, participants more often found the base-rate irrelevant at higher witness accuracy levels. Thus, it appears that participants indicate an orientation toward using the base-rate information more at lower levels of witness accuracy, and using the witness information more at higher levels of witness accuracy. Comparing these given responses to the categorization of the observed results however indicates that participants actually engaged in some process of integration more often than they indicated that both the base-rate and witness accuracy where relevant for an

estimate of culpability. Further, participants' responses indicated a higher attempt to integrate information and compute a unique answer when they encountered exonerating witness testimony as opposed to when they encountered incriminating witness testimony. This further supports the notion that interpreting an appropriate anchor and adjustment may have been more challenging for participants who read exonerating witness testimony.

While participants who encountered incriminating testimony more often relied on just the base-rate or accuracy for their response, participants who encountered exonerating testimony attempted to integrate the information more often. This is consistent with the decision task in the study which required participants to indicate how probable it was that the suspect was *guilty*, even when they encountered exonerating evidence. Given the base-rate implications in the study and the nature of the task, the cognitive processing of the way exonerating information should affect probability judgements can be interpreted as slightly more complicated than that involved with incriminating information, especially in light of using the base rate and accuracy level as an anchor for probability judgements. In cases with incriminating information, the correlation between the accuracy rating and probability is direct. In other words as accuracy increases so do probability judgements. On the other hand, for exonerating information, this correlation is inversed; as accuracy increases, probability judgements decrease. In this sense, it may appear more obvious that an 80% accurate incriminating witness implies 80% guilt as opposed to an 80% accurate witness implying 50, 40, 20, or 10% guilt. It may be in this case that utilizing the exonerating evidence actually requires an effort at calculating odds of guilt beyond what is afforded by incriminating evidence, especially when the question asked requires an impression of guilt. Perhaps if the task required

participants to indicate the probability of innocence of the party in question this interaction may be reversed.

The results have interesting implications for the way jury members may integrate and use evidence of this nature in a real case. The interaction between accuracy and condition indicates that there seems to be a significant difference in the way we manage different forms of eyewitness testimony. Specifically, it does not appear that the exonerating evidence provides an equal but opposite effect to its counterpart in incriminating evidence. In other words, the cognitive models we use to evaluate evidence for innocence may be different than those we use to evaluate evidence for guilt and this can be further compounded by the manner in which we frame the verdict decision task (i.e., identifying guilt vs. identifying innocence). Interestingly, it does not seem in any case that participants actually completed the given Bayesian inference problems by using the appropriate parameters of the Bayesian solution. In fact, comparing the observed results to normative Bayesian solutions indicates that participants seem to follow a Bayesian adjustment for integrating incriminating testimony, though their response curve is a positive translation of the Bayesian norm. In other words, changes in participant estimates of guilt across increasing levels of witness accuracy seem to follow a Bayesian adjustment but the baseline for the adjustment (the base-rate in the normative solution) is shifted to a higher degree of guilt. For participants who encounter exonerating evidence, the function of change for estimates of guilt across different levels of witness accuracy is statistically non-existent and significantly different than the Bayesian norm, which is a mirror reflection of the function for incriminating testimony. Thus, if we assume the independence of the accuracy of the witness from the base-rate occurrence in the scenario in question, a Bayesian approach would predict

equal but opposite effects for incriminating and exonerating witness testimony. A comparison of the observed results to the Bayesian norms can be seen in Figure 5 below and a summary of the Bayesian reasoning and calculations can be seen in Appendix III.

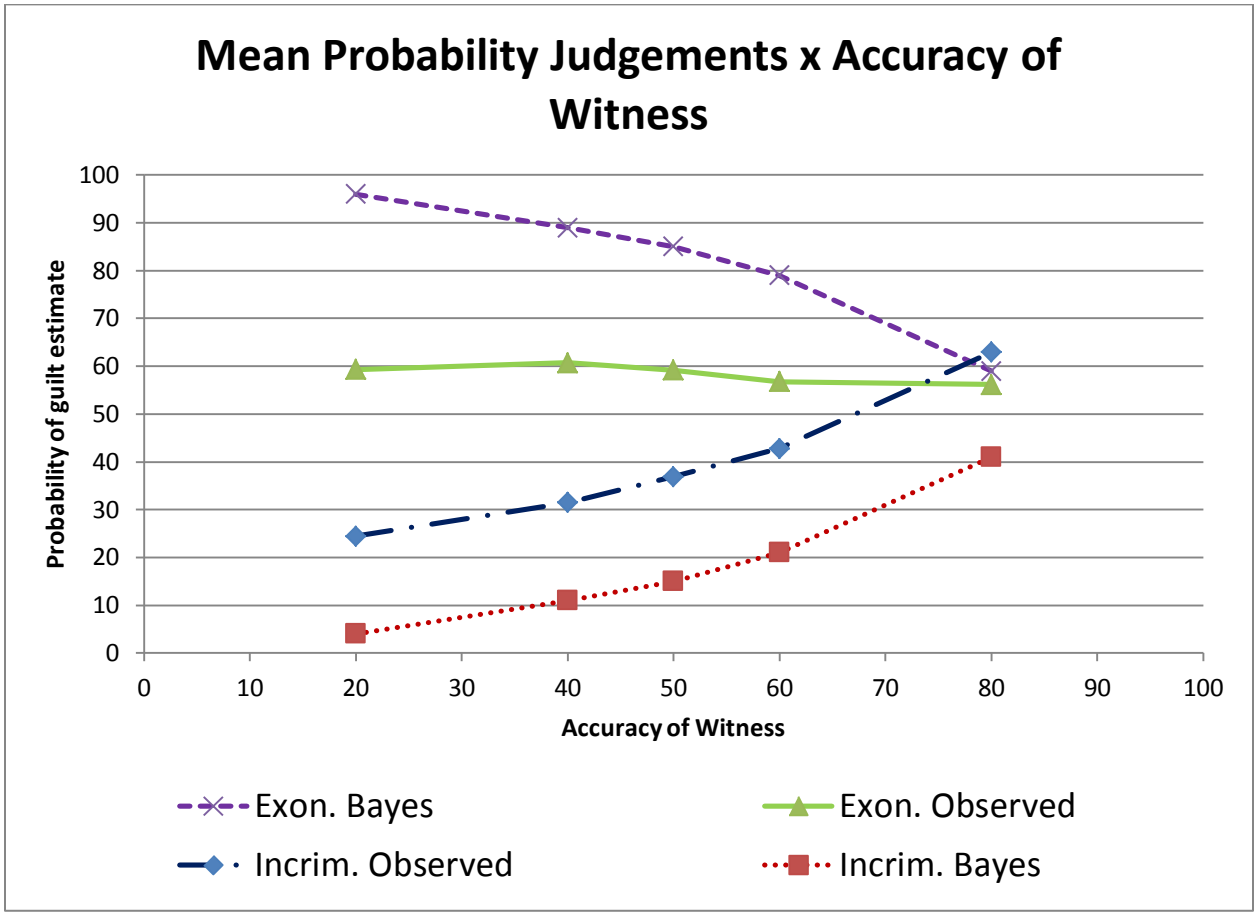

Figure 5. Bayesian norm estimates against observed data for mean probability of guilt judgements

Though the utility of this comparison is confounded by the reference class problem, the trends in the data imply that most participants did not rely on a complete Bayesian model to estimate probability of guilt. Particularly, most participants failed to recognize the diagnostic value of low accuracy witnesses in a Bayesian calculation. In cases with exonerating information for example, a 20% accurate witness actually should increase guilt beyond the level of the base rate because the binary nature of the inference problems essentially translates this low accuracy to an 80% accurate witness in the opposite direction. In other words, the assumption

that the information set is exhaustive of all possibilities necessarily implies that the absence of evidence in one direction implies the presence of evidence in the opposite direction, and vice versa. Further, the context of the comparative model (Bayesian in this case) is irrelevant to the argument relating to the consistency of juror decision models when evaluating incriminating and exonerating evidence. Specifically, the differences between the Bayesian comparisons and the observed results above indicate that there are clear inconsistencies in the mental processes used to estimate the utility of witness testimony when it implies guilt versus when it implies innocence.

The results also present an interesting insight into how the starting point of the decision for a probability judgement of guilt may affect the way jury members integrate and use different forms of evidence. On one hand, participants in the study did seem to utilize the base rate information in some cases or try to integrate it with the witness information which indicates that they did at least attempt to use all different forms of evidence presented. On the other hand, it also appears that starting with a high or low base rate had a strong effect on the way the participants used the evidence and attempted to make probability judgements. This raises concern over the effects that the natural dispositions towards guilt or innocence that jury members may possess prior to evaluating evidence may have on the way evidence is evaluated. Particularly, it is interesting to further investigate whether the effects of the evidence identified in this study will replicate for incriminating and exonerating scenarios at varying levels of base-rate implication. Studies 2 and 3 will aim to further explore this notion and address whether the accuracy by condition interaction was due to a difference in the effectiveness of the evidence or to a difficulty in shifting opinion at varying base rates.

Chapter 4: Studies 2 and 3 Methods, Results and Discussion

Studies 2 and 3 aim to further investigate the condition by accuracy interaction observed in study 1. In study 1, participants who read scenarios with exonerating witnesses gave significantly higher estimates for the probability of guilt of a suspect than participants who read scenarios with incriminating witnesses. Further, participants in the exonerating group show no statistical differences in their estimates of guilt as witness accuracy increased while participants in the incriminating group show consistent increases in estimates of guilt that directly correlate with witness accuracy. However, it was not clear whether these differences were due to the nature of the evidence or to differences in the base-rate implication of the suspect across the two study conditions. Therefore, studies 2 and 3 replicated the procedure from study 1 using only exonerating evidence (study 2) or incriminating evidence (study 3) that needs to be integrated with base-rate implications at low, neutral or high levels. These comparisons will give a more complete picture of the effect of different forms of witness accuracy at varying levels of suspect base-rate involvement and address the confound from the condition manipulation in study 1. If the interaction in study 1 was due to some property of exonerating witness testimony, then study 2 will show no effect for witness accuracy on estimates of guilt regardless of base-rate levels. If, however, a significant interaction is observed in study 2, this may be an indication that the framing of the decision task in combination with the high-base rate implication provide a better explanation for the observed interaction in study 1. Alternatively, seeing an inverse trend in study 3 for incriminating witness testimony will also indicate that framing and anchoring effects may be playing a bigger role influencing participant judgements than the efficacy of the evidence per se.

Participants

168 students enrolled in an Introduction to Psychology course at Ryerson University participated in studies 2 and 3 as a partial requirement for course credit. Eighty four students (73 female, 11 male) participated in study 2 and eighty four (71 female, 13 male) in study 3. The studies were advertised on the online participation program SONA, under the title "Can You Identify the Culprit?" (the same title as study 1) to discourage students who had already participated in study 1 from participating again. Each participant received one course credit for their participation in the study.

Materials

Studies 2 and 3 followed an identical design and procedure with the exception that study 2 used scenarios with exonerating witness testimony only and study 3 used scenarios with incriminating witness testimony only. Participants used the same set of scenarios and booklets used in study 1 however the implication of the suspect in question was set at a low base-rate (15%), a medium level base-rate (50%), or a high base-rate (85%) for different groups of participants in each study.

Design

The study design followed a similar format as study 1 with the exceptions mentioned above. Considering the minimal effect of subset that was observed in study 1, this category was collapsed for analysis to produce a 3 (Base-rate: 15%, 50%, 85%) X 5 (accuracy of eyewitness: 20, 40, 50, 60, 80%) mixed design. Base-rate was manipulated between subjects and accuracy was manipulated within subjects. Different groups of participants read scenarios where the suspect identified by the witness was implicated at a base-rate that was low (15%), medium-

level (50%), or high (85%). The dependent variables in studies 2 and 3 were identical to those in study 1. Participants were asked to give an estimate for the probability of guilt of the suspect in question along with confidence ratings in their judgements. They were also asked to indicated what information they believed was most relevant to their decision process.

Procedure

The Procedure for studies 2 and 3 was identical to that used for study 1.

Results

Probability of guilt judgements and confidence ratings were subjected to a 3 (base-rate: 15%, 50%, 85%) X 5 (accuracy of eyewitness: 20, 40, 50, 60, 80%) repeated measures mixed modal ANOVA. All results are reported as significant at $p < .05$.

Study 2 - Probability of guilt estimates and confidence ratings for base-rate and accuracy

Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of accuracy on probability judgement, $\chi^2(9) = 81.97$. Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .62$. Results are reported with unadjusted degrees of freedom where tests of significance did not differ. All results are reported as significant at $p < .05$.

There was no significant main effect for the accuracy of the witness on probability of guilt judgements, $F(4, 324) = 1.90$, $p = .110$, $\eta_p^2 = .02$. This implies that estimates of probability of guilt generally did not change across different levels of the accuracy of the witness. There was a main effect for accuracy of the witness on confidence in judgement ratings, $F(4, 324) = 6.77$, $p < .001$, $\eta_p^2 = .08$. Contrasts revealed a significant quartic trend, $F(1, 81) = 15.45$, $p < .001$, $\eta_p^2 = .16$ indicating that confidence in judgements rises and falls several times across

different levels of witness accuracy. Further analysis indicated that there was a significant difference in confidence ratings between witnesses who were 50% and 60% accurate, F (1, 82) = 16.82, $p < .001$, and witness who were 60% and 80% accurate, F (1, 82) = 26.81, $p < .001$. This can be seen in Figure 6 below.

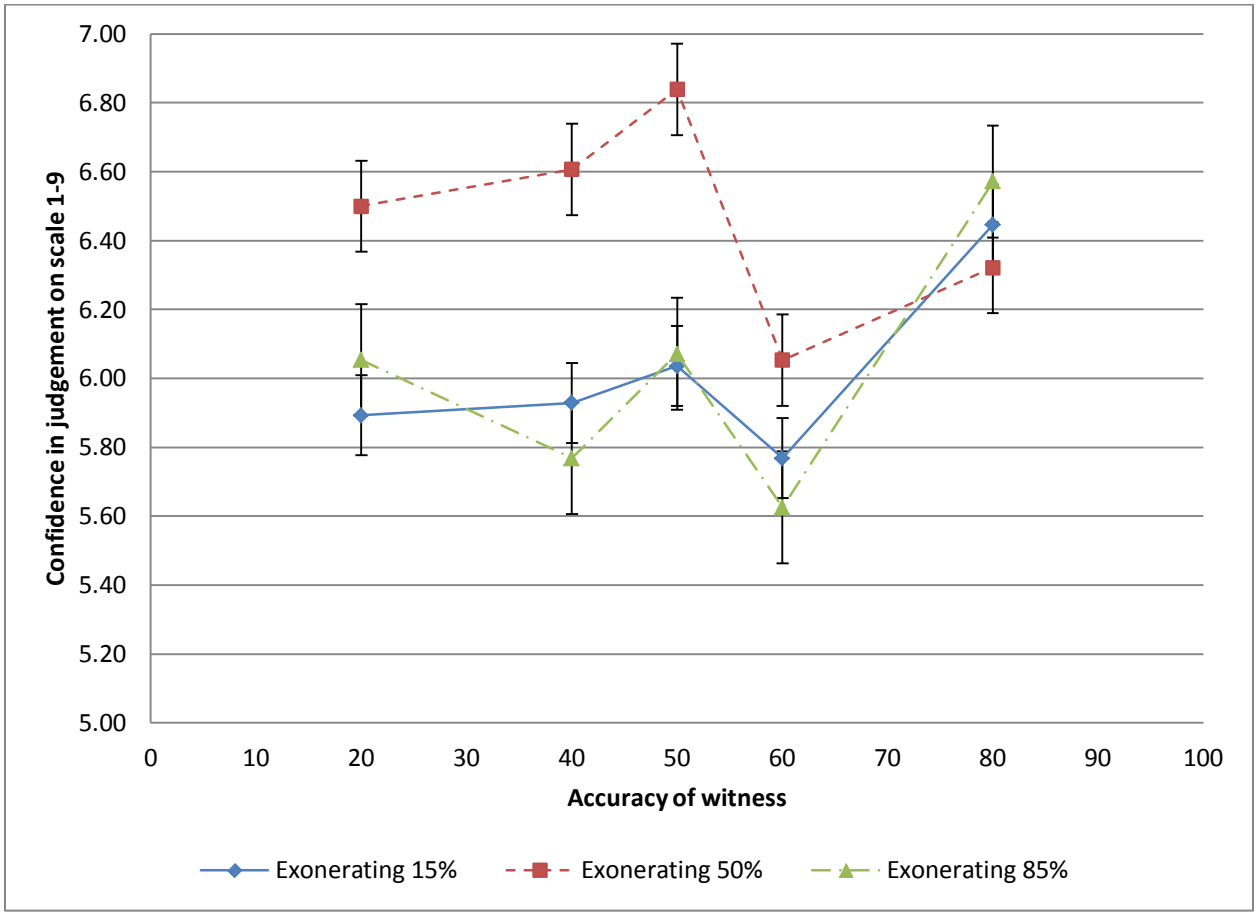

Figure 6. Axis truncated to highlight effect. Mean confidence rating across 15%, 50% and 85% base-rate implication plotted with 95% confidence intervals.

There was a significant main effect of base-rate condition on probability of guilt judgements, F (2, 81) = 35.73, $p < .001$, $\eta_p^2 = .47$ but not on confidence in judgements, F (2, 81) = 1.37, $p = .259$, $\eta_p^2 = .03$ indicating that participants again seem to be using the level of base rate implication as an anchor for decisions. Post hoc comparisons revealed that estimates at 15% base-rate (M = 35.06%, SE = 1.93) where lower than those at 50% (M = 47.54%, SE = 1.93)

which in turn were lower than those at 85% (*M* = 58.11%, SE = 1.93). Average estimates of guilt seem to shift in a linear fashion.

The effect of the accuracy by condition interaction on probability of guilt estimates approaches significance, F (4.94, 200.081) = 2.08, p = .070, η_p^2 = .05. Although the accuracy of the witness had no statistically significant effect on estimates of guilt regardless of base-rate implication, the trends in the data imply that further study could highlight differential strategies in belief adjustment depending on initial base-rate implications. For example, downward adjustments may be more likely at high base-rate implications while upward adjustments may be seen at lower base-rate implications. This can be seen in Figure 7 below.

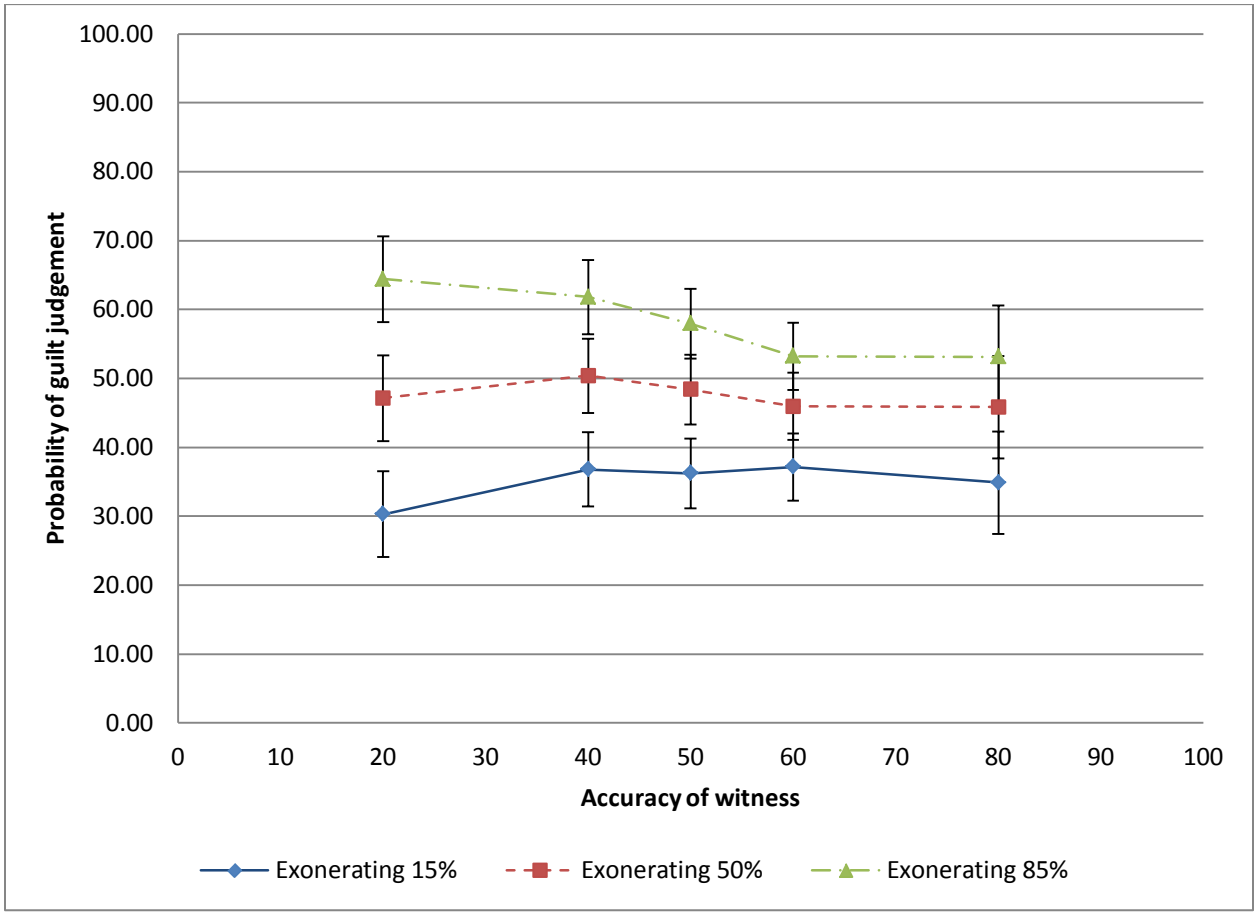

Figure 7. Mean probability of guilt estimates for 15%, 50% and 85% base-rate implication with exonerating witness plotted with 95% confidence intervals.

There was a significant effect of the accuracy by condition interaction on confidence in judgements, F (8, 324) = 2.29, p = .021, η_p^2 = .05. Further analysis indicated that values for confidence were higher when the witness was 80% accurate than at any other level of witness accuracy in the 15% and 85% base-rate conditions. In the 50% base rate condition, confidence ratings were generally higher than the other two groups but did not differ across different levels of witness accuracy. This can be seen in the interaction graph in Figure 6 above.

Study 2 - Exploratory analysis for use and importance of base-rates and witness evidence

An exploratory analysis was conducted on the participants' categorical responses regarding the relevance of the different forms of evidence presented. The correspondence between their given response and their observed data was also calculated. Overall, participants indicated that they found the accuracy of the witness to be irrelevant in roughly 35% of their responses. Roughly 15% of the time, participants indicated that the base-rate information was irrelevant and on approximately 50% of their responses, they indicated that both pieces of information were equally relevant. A summary of relevance distributions across all levels of witness accuracy can be seen in Table 5 below.

Table 5. *Relevance of Information for Probability Decision Across All Levels of Witness Accuracy and Base-Rates for Exonerating Witnesses*

Base-rate 15%	**Relevance Category**		
Witness Accuracy	**Acc irrelevant**	**BR irrelevant**	**Equal relevance**
20	55.36%	8.93%	35.71%
40	44.64%	14.29%	41.07%
50	41.07%	10.71%	48.21%
60	25.00%	14.29%	60.71%
80	16.07%	21.43%	62.50%
Group Mean	**36.43%**	**13.93%**	**49.64%**
Base-rate 50%	**Relevance Category**		
Witness Accuracy	**Acc irrelevant**	**BR irrelevant**	**Equal relevance**
20	57.14%	8.93%	33.93%
40	50.00%	7.14%	42.86%
50	37.50%	12.50%	50.00%
60	30.36%	10.71%	58.93%
80	28.57%	10.71%	60.71%
Group Mean	**40.71%**	**10.00%**	**49.29%**
Base-rate 85%	**Relevance Category**		
Witness Accuracy	**Acc irrelevant**	**BR irrelevant**	**Equal relevance**
20	57.14%	10.71%	32.14%
40	42.86%	17.86%	39.29%
50	21.43%	14.29%	64.29%
60	5.36%	23.21%	71.43%
80	8.93%	35.71%	55.36%
Group Mean	**27.14%**	**20.36%**	**52.50%**
Grand Total	**34.76%**	**14.76%**	**50.48%**

The data again indicates that participants show a shift towards increasingly believing that base-rate information is irrelevant as witness accuracy increases. Similarly, participants indicated that witness accuracy was increasingly irrelevant as witness accuracy decreased. Analysis of the response category measure indicates that participants provided probability of guilt estimates that were equivalent to the witness accuracy in a given scenario for approximately 14% of their responses. Their estimates were equal to the base-rate in approximately 27% of their responses and indicative of some form of mental math or

estimation in approximately 60% of their responses. A summary of response category distribution across all levels of witness accuracy can be seen in Table 6 below.

Table 6. *Response Category for Probability Decision Across All Levels of Witness Accuracy and Base-Rates for Exonerating Witnesses*.

Base-rate 15%	**Response Category**		
Witness Accuracy	**EQ ACC**	**EQ BR**	**MENTAL MATH**
20	14.29%	26.79%	58.93%
40	16.07%	21.43%	62.50%
50	26.79%	26.79%	46.43%
60	5.36%	12.50%	82.14%
80	10.71%	16.07%	73.21%
Group Mean	**14.64%**	**20.71%**	**64.64%**
Base-rate 50%	**Response Category**		
Witness Accuracy	**EQ ACC**	**EQ BR**	**MENTAL MATH**
20	8.93%	51.79%	39.29%
40	7.14%	51.79%	41.07%
50	32.14%	48.21%	19.64%
60	5.36%	41.07%	53.57%
80	7.14%	19.64%	73.21%
Group Mean	**12.14%**	**42.50%**	**45.36%**
Base-rate 85%	**Response Category**		
Witness Accuracy	**EQ ACC**	**EQ BR**	**MENTAL MATH**
20	8.93%	32.14%	58.93%
40	7.14%	19.64%	73.21%
50	42.86%	14.29%	42.86%
60	7.14%	12.50%	80.36%
80	3.57%	14.29%	82.14%
Group Mean	**13.93%**	**18.57%**	**67.50%**
Grand Total	**13.57%**	**27.26%**	**59.17%**

Once again, the data indicate that participants actually engage in some form of mental math or estimation more often than they report that both pieces of evidence are equally relevant.

Study 3 - Probability of guilt estimates and confidence ratings for base-rate and accuracy

Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of accuracy on probability judgement, $\chi^2(9) = 55.13$. Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .71$. Results are reported with unadjusted degrees of freedom where tests of significance did not differ. All results are reported as significant at $p < .05$.

There was a significant main effect for the accuracy of the witness on probability of guilt judgements, $F(4, 324) = 86.62$, $p < .001$, $\eta_p^2 = .52$. Contrasts revealed significant differences between all adjacent levels of witness accuracy indicating a consistent increase in probability of guilt judgements as witness accuracy increased. There was a significant main effect for accuracy of the witness on confidence in judgement ratings as well, $F(4, 324) = 7.35$, $p < .001$, $\eta_p^2 = .08$. Contrasts revealed a significant quartic trend, $F(1, 81) = 12.91$, indicating that confidence in judgements rises and falls several times across different levels of witness accuracy. Further analysis indicated that there were significant differences in confidence ratings at all adjacent levels of witness accuracy indicating a similar trend for confidence as that observed in study 1. Confidence ratings drop at the 40% and 60% witness accuracy levels and rise at the 20%, 50% and 80% levels producing a zigzag like pattern.

There was a significant main effect of base-rate condition on probability of guilt judgements, $F(2, 81) = 33.91$, $p < .001$, $\eta_p^2 = .46$ indicating once more that participants seem to be using the level of base-rate implication as an anchor for decisions. There was no significant main effect of base-rate on confidence in judgements, $F(2, 81) = 1.16$, $p = .320$, $\eta_p^2 = .03$. Post hoc comparisons revealed that probability estimates at 15% base-rate ($M = 42.42\%$, SE = 2.02)

where lower than those at 50% (*M* = 52.15%, SE = 2.02) which in turn were lower than those at 85% (*M* = 65.78%, SE = 2.02). Average estimates of guilt seem to shift in a linear fashion.

There was a significant effect of the accuracy by condition interaction on probability of guilt estimates, F (8, 324) = 6.07, $p < .001$, $\eta_p^2 = .13$. Contrasts revealed that the rate of increase for probability of guilt estimates varied across different base-rate conditions as witness accuracy increased. For the 15% base-rate condition, contrasts revealed significant differences at all adjacent levels of witness accuracy indicating a linear increase in judgements of probability of guilt. The 50% base-rate condition showed the same trend as the 15% condition. For the 85% base-rate condition, there was a significant increase in probability judgements between 20% and 40% accurate witnesses. Probability judgements were not statistically different at the 40%, 50%, and 60% levels but a significant increase in probability judgements was seen at the 80% witness accuracy level. While witness accuracy always had an effect on judgements of probability of guilt, the extent of this effect varied depending on the base-rate condition. Overall, it appears that the effect is similar in pattern, but the slope across accuracy decreases with increases in base-rate. This may be due to diminishing returns as base-rate implications approach ceiling effects. This can be seen in the interaction graph in Figure 8 below.

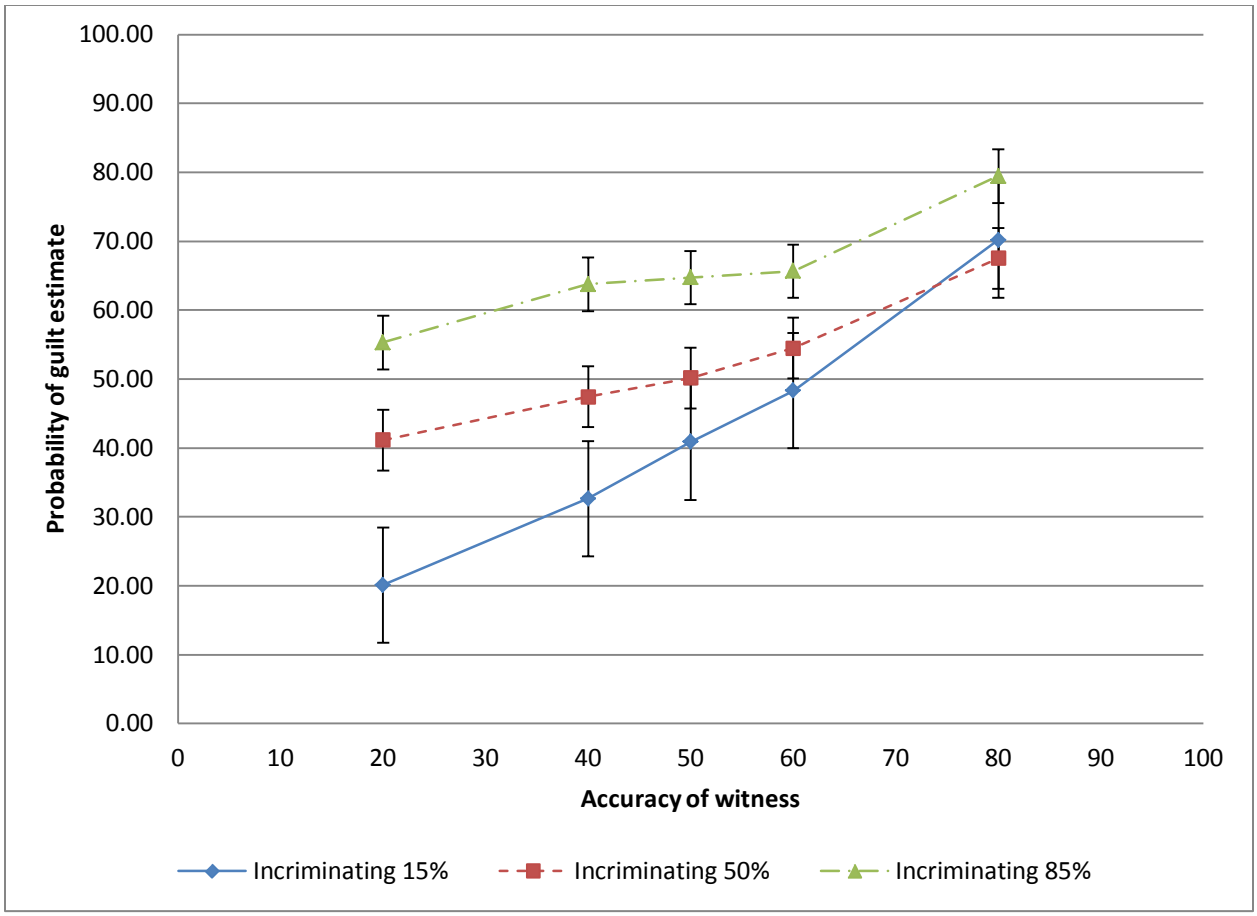

Figure 8. Mean probability of guilt estimates for 15%, 50% and 85% base-rate implication with incriminating witness plotted with 95% confidence intervals.

.

There was also a significant effect of the accuracy by condition interaction on confidence in judgements, F (8, 324) = 5.81, p < .001, η_p^2 = .13. Further analysis indicated that at a base-rate of 15%, confidence ratings show significant drops between 20% and 40% witness accuracy levels and between 50% and 60% accuracy levels. There was a significant increase in confidence between 60% and 80% accuracy levels. At the 50% base-rate level, there was a significant increase in confidence between 40% and 50% accuracy levels followed by a significant decrease between 50% and 60%. At the 85% base-rate level, confidence ratings showed a significant increase between 60% and 80% witness accuracy levels. The interaction graph can be seen in Figure 9 below.

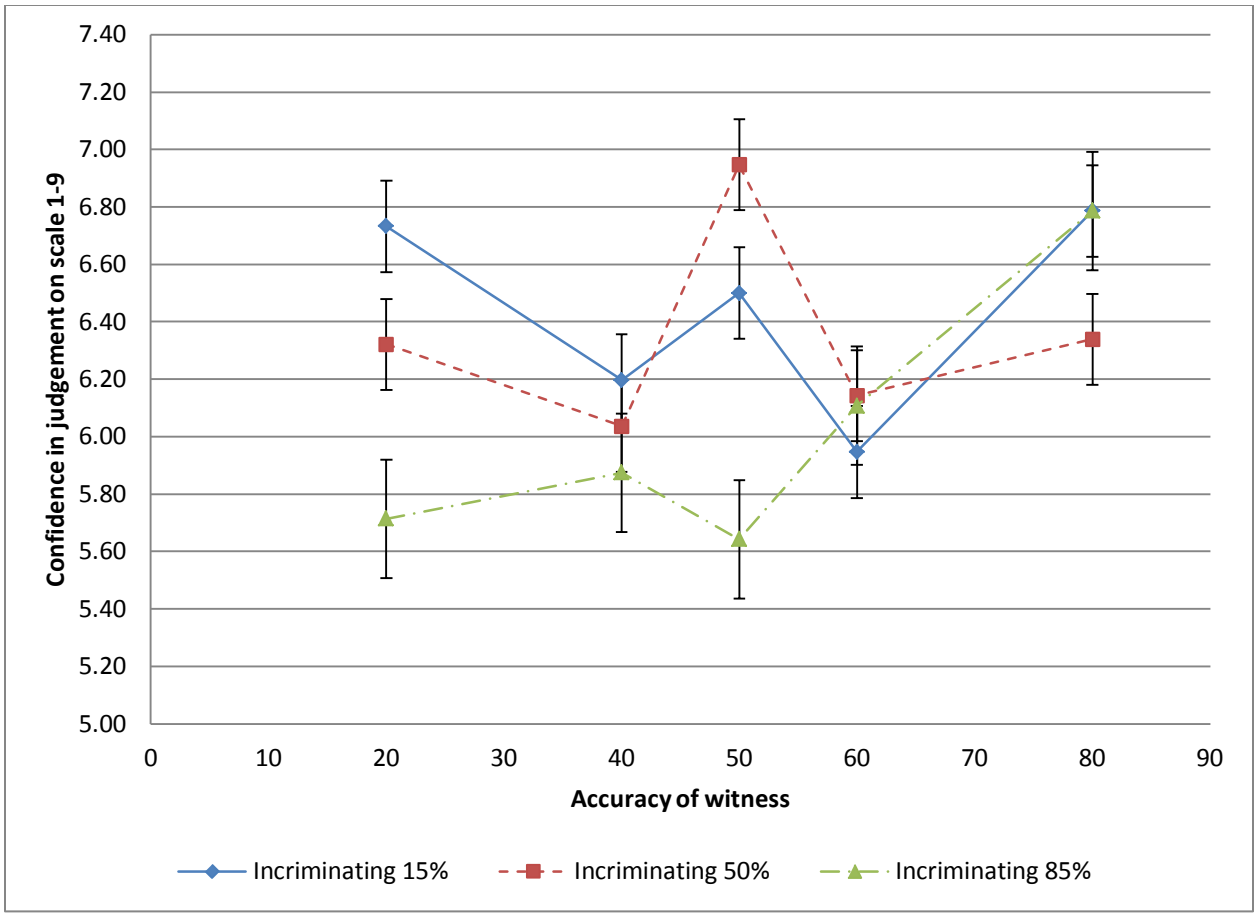

Figure 9. Axis truncated to highlight effect. Mean confidence rating across 15%, 50% and 85% base-rate implication with incriminating witness plotted with 95% confidence intervals.

Study 3 - Exploratory analysis for use and importance of base-rates and witness evidence

An exploratory analysis was conducted on the participants' categorical responses regarding the relevance of the different forms of evidence presented. The correspondence between their given response and their observed data was also calculated. Overall, participants indicated that they found the accuracy of the witness to be irrelevant in roughly 30% of their responses. Roughly 21% of the time, participants indicated that the base-rate information was irrelevant and on approximately 49% of their responses, they indicated that both pieces of information were equally relevant. A summary of relevance distributions across all levels of witness accuracy can be seen in Table 7 below.

Table 7. *Relevance Of Information for Probability Decision Across All Levels of Witness Accuracy and Base-Rates for Incriminating Witnesses*

Base-rate 15%	**Relevance Category**		
Witness Accuracy	**Acc irrelevant**	**BR irrelevant**	**Equal relevance**
20	53.57%	21.43%	25.00%
40	44.64%	19.64%	35.71%
50	35.71%	25.00%	39.29%
60	19.64%	21.43%	58.93%
80	10.71%	37.50%	51.79%
Group Mean	**32.86%**	**25.00%**	**42.14%**
Base-rate 50%	**Relevance Category**		
Witness Accuracy	**Acc irrelevant**	**BR irrelevant**	**Equal relevance**
20	62.50%	12.50%	25.00%
40	46.43%	14.29%	39.29%
50	28.57%	7.14%	64.29%
60	23.21%	14.29%	62.50%
80	12.50%	35.71%	51.79%
Group Mean	**34.64%**	**16.79%**	**48.57%**
Base-rate 85%	**Relevance Category**		
Witness Accuracy	**Acc irrelevant**	**BR irrelevant**	**Equal relevance**
20	42.86%	28.57%	28.57%
40	42.86%	17.86%	39.29%
50	21.43%	21.43%	57.14%
60	3.57%	23.21%	73.21%
80	5.36%	12.50%	82.14%
Group Mean	**23.21%**	**20.71%**	**56.07%**
Grand Total	**30.24%**	**20.83%**	**48.93%**

Once again, the data indicate a strong shift in reporting that the witness accuracy was irrelevant as witness accuracy decreased. The data also highlight the trend observed in studies 1 and 2 indicating that participants more often report that the base-rate information is irrelevant as witness accuracy increases. The data also indicate that participants once again seem to ignore the diagnosticity of low accuracy witnesses in the study scenarios as they more frequently report that witness testimony is irrelevant at lower witness accuracy levels. Taking a closer look at participant relevance responses across the base-rate conditions also reveals that participants most often indicate that both pieces of information are equally relevant in the 85%

base-rate group. In the highest incriminating category in the study design (base-rate = 85%, witness accuracy = 80%), approximately 82% of participants indicated that both pieces of information were equally relevant for the decision. In the corresponding highest exonerating (or lowest incriminating) category in study 2 (base-rate = 15%, witness accuracy = 80%), 62% of participants indicated that both pieces of information were equally relevant for the decision.

Analysis of the response category measure indicates that participants provided probability of guilt estimates that were equivalent to the witness accuracy in a given scenario for approximately 25% of their responses. Their estimates were equal to the base-rate in approximately 35% of their responses and indicative of some form of mental math or estimation in approximately 40% of their responses. In this case, participants actually seem to engage in mental math or estimation less frequently than they indicate that both pieces of information are equally relevant for an evaluation of culpability. Further analysis indicated that while the trends from study 1 and 2 were observed at the 15% base-rate condition, participants' observed responses were more frequently equal to the base-rate or accuracy of the scenario in the 50% and 85% base-rate conditions. Further, participants in these groups used mental math or estimation less frequently than they indicated that both pieces of information are equally relevant. A summary of the response category distribution across the three base rate levels can be seen in Table 8 below.

Table 8. *Response Category for Probability Decision Across All Levels of Witness Accuracy and Base-Rates for Incriminating Witnesses*

Base-rate 15%	**Response Category**		
Witness Accuracy	**EQ ACC**	**EQ BR**	**MENTAL MATH**
20	21.43%	48.21%	30.36%
40	19.64%	28.57%	51.79%
50	51.79%	26.79%	21.43%
60	23.21%	7.14%	69.64%
80	41.07%	5.36%	53.57%
Group Mean	**31.43%**	**23.21%**	**45.36%**
Base-rate 50%	**Response Category**		
Witness Accuracy	**EQ ACC**	**EQ BR**	**MENTAL MATH**
20	16.07%	57.14%	26.79%
40	14.29%	53.57%	32.14%
50	25.00%	67.86%	7.14%
60	28.57%	44.64%	26.79%
80	35.71%	21.43%	42.86%
Group Mean	**23.93%**	**48.93%**	**27.14%**
Base-rate 85%	**Response Category**		
Witness Accuracy	**EQ ACC**	**EQ BR**	**MENTAL MATH**
20	12.50%	35.71%	51.79%
40	10.71%	39.29%	50.00%
50	28.57%	32.14%	39.29%
60	26.79%	23.21%	50.00%
80	28.57%	30.36%	41.07%
Group Mean	**21.43%**	**32.14%**	**46.43%**
Grand Total	**25.60%**	**34.76%**	**39.64%**

Discussion

Accuracy. The findings from studies 2 and 3 strongly support the interpretation that the changes in probability of guilt judgements across different levels of witness accuracy observed in study 1 are due to the efficacy of the evidence rather than the discrepancy in base-rate implication. The results were replicated across the two studies with accuracy again showing no main effect for scenarios with exonerating evidence and a main effect for scenarios with incriminating evidence. This implies that participants are underutilizing the exonerating

testimony and evaluating its utility for the overall judgement of culpability in a different model than that used for evaluating scenarios with incriminating testimony. The effects for confidence also replicate in studies 2 and 3. Confidence levels are generally lower when incorporating testimony from witnesses who are 40% or 60% accurate indicating the added challenge of evaluating testimony without a clear anchor for accuracy. A comparison of mean probability judgements across studies 1, 2 and 3 can be seen in Figure 10 below.

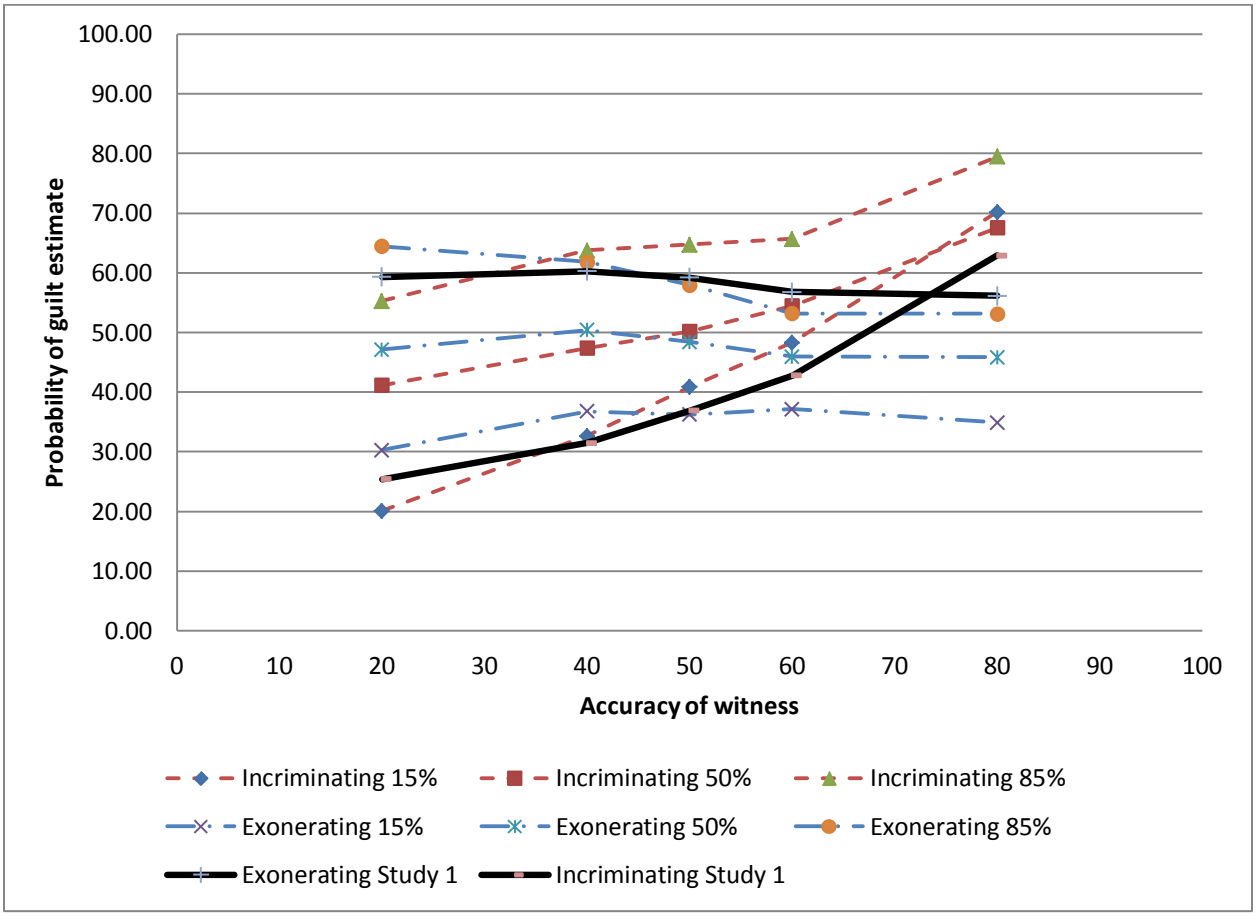

Figure 10. Comparison of mean probability of guilt judgements across studies 1, 2 and 3.

Interestingly, while participants are justified to show higher confidence at 50% witness accuracy levels (since the evidence is completely non-diagnostic and can be ignored) their observed response patterns do not indicate that their confidence stems from the fact that they ignore the witness testimony when evaluating culpability. In fact, in many cases 50% probability

of guilt was the modal observed response in situations where the base-rate and/or witness accuracy were reported at 50% as well. This finding was dramatically exaggerated when the scenarios included exonerating testimony as in these cases the witness accuracy cannot be used as an anchor for the judgement of culpability. This further exemplifies the participants' deviation from a standard stochastic approach to integrating the base-rate and witness information (at least a deviation from Bayesian norms) and highlights the strong anchoring effects of the information presented regardless of its actual diagnosticity.

Regarding the participants overall confidence levels, it is important to note that though the data does show significant changes in participant confidence, the scale of these changes is generally in the order of a fraction of a point on a nine point scale.

Condition. The main effect for condition on probability of guilt judgements in both study 2 and study 3 indicates that participants seem strongly inclined to rely on the reported base-rate for the scenario as an anchor for their decision. In other words, participants do not seem to demonstrate base-rate neglect in its purest sense. Rather, base-rates are integrated into their decision model but utilized outside of the parameters of a strict probabilistic model. One major challenge to interpreting the stochastic effect of the base-rate on probability of guilt estimates is the difficulty associated with identifying the reference class within which participants are integrating the presented information. As mentioned earlier, the parameters of the reference class have strong implications regarding the applicability of mathematical models and participants' approximation of these models. A follow up question or series of questions in the study aimed at identifying each individual participant's interpretation of the parameters of the

probabilistic scenarios would have been helpful towards better mapping observed and model-predicted results.

Interestingly, the average probability of guilt at each base-rate level was higher for incriminating scenarios than it was for exonerating scenarios. This further indicates that anchoring effects appear to be the result of integration between base-rates and the valence of the witness testimony. Although it is not clear whether participants are adjusting upwards for incriminating scenarios or downwards for exonerating scenarios, the observed effects of the accuracy and the accuracy by condition interaction indicate that it is more likely that incriminating scenarios contribute to an upward shift in base-line estimates of guilt. Either way, it is clear that even if the base-rate is acting as an anchor for probability of guilt judgements, the context of the scenario and testimony also have an effect on these judgements.

Comparing the interaction effects across studies 2 and 3 again strongly supports the notion that exonerating testimony has little effect on decreasing probability of guilt estimates while incriminating testimony shows strong effects on increasing probability of guilt estimates as witness accuracy increases. This was observed at low, medium and high levels of base-rate implication indicating that levels of the culpability estimate were mediated by the interaction of the testimony and the base-rate rather than just the base-rate level.

Relevance of information and response category:

Comparing the relevance and response categories for studies 2 and 3 with each other and with the results from study 1 again highlight that participants' observed responses exemplify some form of mental math or approximation more often than they report that both pieces of information are equally relevant. This also happens more frequently in study 2 with

the exonerating evidence than it does in study 3 with the incriminating evidence. Further, even when participants indicate that the eyewitness or base-rate information is irrelevant their response category does not always correspond to the appropriate counterpart of being either equal to the base-rate or equal to the accuracy. Thus, it is clear that participants often try to engage any source of perceived information that is available for the decision, even if it is not explicitly stated as a source of reference. Participants may be engaging elements of their own beliefs or other cues that were acquired from the scenario description in their expressed estimates of culpability. Deviations from responses that rely only on the base-rate or only on the accuracy level might also be the result of demand characteristics implicit in the study design and the desire to give an impression that an attempt at a mathematical calculation was being made. The consistency of response patterns across participants however indicates that there is likely some systematic influence on probability estimates. Even if participants were giving responses indicative of mental math simply to impress the researcher, elements of the study appear to have a consistent effect on the results of these simulated operations.

The relevance and response category analysis also yields some very interesting exploratory grounds for further study. For example, it was observed that participants more often reported that the eyewitness testimony was irrelevant when they encountered exonerating testimony than when they encountered incriminating testimony (35% vs 30% of participants). Although this analysis is exploratory, it speaks to research discussed earlier indicating that evidence for innocence, such as alibi testimony, is often ignored or underutilized in the face of incriminating evidence. Further, in the 85% base-rate condition for incriminating scenarios, an overwhelming proportion of participants indicate that both pieces of information

are equally relevant, especially at higher witness accuracy levels. This deviates from the general trend where participants often indicated that the witness accuracy was most relevant at high witness accuracy levels and implies a strong focus on any and all information that is indicative of guilt. This point is especially interesting when considering that the flipside of this observation for exonerating evidence did not show a similar effect. If we can imagine all the cells across the two study designs on a continuum of culpability, then scenarios where the suspect is implied at an 85% base-rate and an incriminating witness who is 80% accurate provide the highest likelihood of culpability. On the other hand, a scenario with a 15% base-rate implication and an exonerating witness that is 80% accurate provide the lowest likelihood of guilt. In the later cases however, participants often indicate that that base-rate is irrelevant and don't consider its utility as evidence for innocence.

These findings speak strongly to the effect the framing of the decision task has on participants' perceptions and utilization of evidence. The requirement to estimate "the probability that a suspect is guilty" could play an influential part in how participants orient to, evaluate, and utilize different forms of evidence. For example, a replication of studies 2 and 3 with the decision task to estimate "the probability that a suspect is innocent" may provide great insight into how framing affects the use of certain decision models for belief evaluation and updating.

Chapter 5: General Discussion and Conclusions

An important goal of any research on the decision models used in the juridical process is to increase the rate of accurate decisions and avoid the worst case of wrongful conviction but the traditional route to this goal has almost always started with the assumption that instinctual strategies will lead to more error than rational models. Considering the historical context of heuristics research, this perception does not come as a great surprise. The bulk of research in this area indicates that heuristics rely on reducing cognitive load by examining less information or modifying the way available information is weighted but the goal remains, like other strategies, to reach a decision that is not only faster but also more accurate (Gigerenzer and Gaissmaier, 2011). Further, heuristics are almost invariably relied upon in one way or another during the process of a decision. Even when attempting to utilize a probabilistic model, heuristics can have an effect on the reference classes selected which will ultimately affect decisions resulting from the model as well. There is no doubt that ideal (at least comparatively) strategies do exist but they exist for an ideal world where the boundaries of a decision task are clearly defined and known. In a sense then, we need to take a step back and consider what the word “ideal” could mean in the context of a verdict decision strategy and can start by building a better contextual understanding of the boundaries of the verdict decision. In other words, it is more important to pick the most appropriate decision model for the given context than to rely on traditional views of what a good or bad strategy may be. Further, it is also important to build this understanding with an eye for both internal and external factors that may trigger different decision strategies at different points in the process as these factors can have a great impact on policy making. For example, juror instructions could be designed to alert and educate jurors

about when they may be using sub-optimal strategies for information integration. Alternatively, more careful evaluation and understanding of the methods and order of evidence presentation will also clarify when specific decision strategies will be triggered and when potentially important information will be overlooked.

That being said, it is quite likely that a variety of strategies or models are used throughout the verdict decision-making process depending on the type, amount and context of the evidence and so it is also likely that a model that encompasses this decision task will not be purely logical, statistical or heuristic in nature but rather a combination of the three. In turn, building an understanding of such a model will help significantly in two ways. On one hand, such a model would give a solid ground for investigating related policy change that could lead to reductions in verdict decisions that contribute to wrongful conviction which will have great moral and monetary benefit for the legal system. On the other hand, even in situations where a decision strategy cannot be controlled, understanding which decision strategies are being used will provide insight into how and when an error in judgement may have occurred during the process of the verdict decision. Early detection of such errors can help tremendously in reducing the likelihood of a wrongful conviction as well and building a better platform for the appropriate utilization of juridical feedback. Thus, having a tool that can inform how more diagnostic verdict decisions could be made and also provide an early detection mechanism for error will be a great asset to the juridical process and the goal of insuring a fair and just trial.

A primary goal of this research project is to investigate the consistency of juror decision models between evidence that implies innocence and evidence that implies guilt. The utilization of models in general decision theory is a widely accepted finding and a large amount

of research also indicates that model thinking does often lead to better, more objective, and sometimes faster decisions (Kahneman, 2011; Lehrer, 2010). Thus, the utilization of models to interpret the process of jury decision making can be highly beneficial to both explaining how jurors actually go about making verdict decisions and to helping them optimize their decision process in order to reach a just and objective verdict, or at least have the ability to recognize when that is not possible.

To date, most efforts to model juror decision making have focused on describing the process jurors may go through while encountering and integrating different forms of evidence. This pursuit has been extremely helpful in delineating many of the factors that influence juror and jury verdict decisions including exposure to pre-trial publicity, the perceived strength of evidence, and the formation and coherence of story samples that help aggregate the presented evidence. However, little input is given on how this knowledge can help optimize juror decisions.

Part of this challenge is beyond the control of current research methods as certain integral aspects of the decision process remain outside the boundaries of any model. For example, a universal definition or legal standard of the concept of reasonable doubt has not yet been established, and it is still unclear how we could obtain an absolute (rather than comparative) measure of the veracity of evidence. However, understanding juror decision goals and the mental models that they use can help in this area. For example, research efforts focused on comparing functions of change in perceptions of culpability throughout a trial process rather than a single point measure of verdict or estimate of guilt may be helpful. Such a comparison could highlight both individual decision strategies and the range and scope of

subjective evaluations of reasonable doubt and its effect on the relationship between perceptions of guilt and verdict decisions. Though issues of this nature are often a matter of legal concern rather than a quality of any formal model of reasoning, efforts to expand our models and provide a temporal comparison of perceptions of guilt may allow advances in understanding when and why such issues may have influence on juror decisions and how to potentially mitigate them.

The current research project adds to this discussion by investigating the consistency of decision models (represented by the function describing shifts in probability of guilt estimates across different levels of witness accuracy) between incriminating and exonerating evidence. The data indicate a large discrepancy in the effects of incriminating and exonerating witness testimony that varies in its accuracy, when participants are asked to estimate the probability of guilt of a suspect. Exonerating testimony was ineffective at reducing estimates of guilt from baseline measures, even when it was described as highly accurate. It appears that participants in these cases strongly orient to the base-rate measure as an anchor for decisions and are not influenced much by varying witness testimony. On the other hand, incriminating testimony that increases in its presented accuracy produces a consistent rise in estimates of guilt. Furthermore, exonerating testimony fails to produce reductions in estimates of guilt regardless of whether the base-rate implication of the suspect is low, medium, or high while incriminating testimony produces increases in estimates of guilt at all levels of base-rate implication. The data present an interesting framework for understanding the consistency of decisions models used for belief updating and the evaluation of evidence and present some intriguing theoretical viewpoints as well.

The data partially support Devine's (2012) notion that the same decision models can be used to evaluate decisions in both criminal and civil trials though further research is needed in this field. In this study, the main effect of subset approached significance though shifts in decisions across witness accuracy levels were fairly consistent whether participants were reading work negligence or crime scenes. As noted earlier however, the effect of the context of the decision may be more apparent in the actual output of these decisions models. For example, the decision to act upon perceptions of guilt and convict may be more attractive in criminal trials while the willingness to accept exonerating evidence and acquit may be more present in civil trials. These are trends that were observed in the data but only approached significance.

The fact that participants more often engage in estimation rather than relying on any given piece of evidence when they encounter exonerating testimony also raises some interesting theoretical concerns. Particularly, while current juror decision models provide valuable insight into the effects of the external and internal factors of a trial on decisions, they make the fundamental assumption that these effects will behave consistently on exonerating and incriminating evidence. This is primarily due to the fundamental assumption that individual pieces of evidence hold an objective and true value of culpability. For example, many models assume that a peripheral appeal to the reliability of a witness will affect perceptions of the testimony presented by an exonerating witness the same way it would affect perceptions of testimony from an incriminating witness. The consistency of these effects however may be largely contingent on the framing of the decision task at hand; and the decision task at trial often orients the decision maker towards a search for culpability. Even the presumption of

innocence and the requirement to assume that a suspect is innocent until proven guilty support this framing of the decision task. In essence the presumption of innocence requires decision makers to pursue evidence for guilt that may convince them beyond a reasonable doubt that a suspect is guilty. In the absence of this evidence, a suspect is found not guilty. Yet, there is no incentive to pursue or highlight evidence that would lead a decision maker to believe beyond a reasonable doubt that a suspect is innocent. In fact, the proportion of "not guilty" verdicts in mock trial studies significantly decreases when participants are presented with a "not proven" verdict category (Hope, Greene, Memon, Gavisk, & Houston, 2008) highlighting the participants' strong motivation to search for proof of guilt.

One of the original intentions of the current research was to try to compare observed decision patterns to Bayesian norms in order to investigate the plausibility of approaching or modeling "optimal" juror decisions. In light of the findings of the research and my experience during the project, I now believe that a more relevant question to ask may rather be which of many decision models a juror is utilizing at any given point in their decision task. I believe that understanding how the decision task for the search of culpability is framed is integral to better understanding this challenge. Specifically, it is of utmost importance to identify the objectives of the juror as a decision maker before it is possible to investigate whether he/she is utilizing rational or optimal decision models. This is not a novel notion although it is one that is somewhat overlooked in many juror decision making models, be them probabilistic, heuristic, or story models. Even if we do assume the presence of an optimal decision model that accounts for the function of the evidence the framing of the decision task will still have a large impact on the function of the evidence in this model.

Let us assume for example that the function of evidence (incriminating or exonerating) can be represented by "f(E)" and the optimal decision function (for example the Bayesian solution) can be represented by "B". In this example, the observed results from a real decision maker (let's call this "V") can be represented as some deviation from the optimal function based on each individual juror's function of evidence. In other words, $V = B \pm f(E)$. In turn, the research discussed in this dissertation on evidence evaluation and belief updating indicates that the function of evidence, f(E), likely follows a power law. This is often represented in the form $f(E)=aE^x$. In general terms, this power law allows the function of evidence, f(E), to hold at least 3 distinct states. First, the function can be equal to zero if "a" is equal to zero. Second, the function can be a constant if "x" is equal to zero. Third, the function can be asymptotic if "a" and "x" are non-zero factors. In this case the function will show a curvilinear trend with diminishing returns at either extreme of the scale. In the case of a power law relating to evidence, "a" can be considered a representation of the baseline importance given to a specific piece of evidence. Thus, if a decision maker gives a certain piece of evidence no importance, its functional effect on their decision will be zero. Alternatively, if they give it high importance this will have a multiplicative effect on the functional utility of that evidence. "x" can be considered a representation of the effect of the evidence over time and in the context of other aspects of the decision task. If the effect of the evidence does not change due to its interaction with other aspects of the decision, then "E^x" is equal to one and f(E) = a. In other words, the evidence will have a constant effect on the decision regardless of its context. On the other hand, higher values of x are indicative of a stronger interaction of the evidence with the context of the decision. This includes the framing of the decision task, the type and amount of evidence

available, the personal bias of the decision maker and his/her peers, and the other factors described earlier in the literature review. This implies an interactive chain of influence between the probative value of the evidence of the case, the verdicts that jurors are considering, and the story models that support juror decision. Figure 11 below provides a visual representation of this interaction.

Figure 11. Evidence, stories, and verdict feedback loop. Evidence can influence story formation through peripheral and informational routes the same way that stories influence verdicts through peripheral and informational routes. Verdicts can create feedback loops that influence story models and story can create feedback loops that influence the selection and utilization of evidence.

Albert Einstein was quoted as saying: "If I had an hour to solve a problem, and my life depended upon the solution, I would spend the first 55 minutes determining how to frame the problem. For once I know the proper question to ask, I can solve the problem in less than 5 minutes". I believe it is very optimistic to think that we can address the challenge of helping jurors optimize their decision process and their decisions in a five minute period. However, better understanding the framing of the juror's decision task and its effect on the utilization of evidence can be an incredible step towards this goal that will aid in reducing the likelihood of wrongful conviction and increasing the overall efficiency of the legal decision process.

Appendix I. Incriminating Scenarios

Situation A

Andrew's moving company owns two trucks, one blue and one green. The blue truck is used for 85% of inner city orders while the green truck is used for 15% of inner city orders. On the night of August 7th, one of Andrew's trucks was involved in a hit-and-run accident. A witness testified in court that the color of the truck involved in the accident was green. In order to ascertain whether the witness could actually distinguish between blue and green trucks at night, the court subjected the witness to a test which consisted of showing the witness film clips of green and blue trucks under lighting conditions similar to the night in question. The witness was asked to identify the color of the cab in each film clip. The witness made correct identifications (20, 40, 50, 60, 80%) of the time, making an equal number of errors for each color truck.

1) What is the probability that the truck involved in the accident was green?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

   a) Given the witness' testimony, the information pertaining to the proportion of orders involving blue and green cabs is irrelevant for making this judgement.

   b) Given the information pertaining to the proportion of orders involving blue and green cabs, the witness' testimony is irrelevant for making this judgement.

   c) Both the information pertaining to the proportion of orders involving blue and green cabs and the witness' testimony are about equally relevant for making this judgement.

Situation B

A critical piece of evidence in a bank robbery case in Washington concerns the type of get-away car driven. Although the number of American and Japanese designed cars in Washington are equal, according to police records, the percentage of American cars used as get-away cars in robberies in Washington is 85% compared to 15% for Japanese cars. A witness to the bank robbery testified that the get-away car used by the robbers was a Japanese model. The witness argued that as a used car saleswoman, she is familiar with the styles of different types of cars. In order to test the witness' ability to distinguish between Japanese and American cars, the witness was asked to observe a number of Japanese and American cars at various angles. The saleswoman was then asked to state whether the car was of American or Japanese design. The witness was correct on (20, 40, 50, 60, 80%) of the trials, making an equal number of errors for each type of car.

1) What is the probability that the get-away car was Japanese?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of get-away cars used in robberies is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of get-away cars used in robberies, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of get-away cars used in robberies and the witness' testimony are about equally relevant for making this judgement.

Situation C

In Madison there are two ambulance companies, Mayday and Valley, serve the local hospitals. Even though there is an equal number of each type of ambulance in Madison, Mayday ambulances have 85% of the complaints by Madison residents about driving that nearly causes or causes accidents, and Valley ambulances receive 15% of the complaints. An accident that involved many cars occurred at an intersection near a hospital. One driver said that an ambulance passing through the intersection caused the accident. However, the drivers of the cars could not identify which type of ambulance passed through the intersection just before the accident. At the time of the accident, a blind man was waiting at the intersection. He said that he could distinguish between the sirens of the two ambulance companies. The blind man stated that the ambulance involved in the accident was a Valley ambulance. In order to test the accuracy of the blind man's hearing, the blind man listened to tape recordings of sirens produced by the ambulances of the two different companies. The blind man listened to a large number of tape recorded sirens and correctly identified (20, 40, 50, 60, 80%) of them, making an equal number of errors for each company's siren.

1) What is the probability that the ambulance involved the accident was a Valley ambulance? (Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

   a) Given the witness' testimony, the information pertaining to the proportion of complaints involving reckless driving is irrelevant for making this judgement.

   b) Given the information pertaining to the proportion of complaints involving reckless driving, the witness' testimony is irrelevant for making this judgement.

   c) Both the information pertaining to the proportion of complaints involving reckless driving and the witness' testimony are about equally relevant for making this judgement.

Situation D

In the city of Windsor, records indicate that there are equal numbers of .22 caliber rifles and handguns. However, of the guns used in drive-by shootings, 85% of the .22 caliber guns are handguns and 15% are .22 caliber rifles. In a recent drive-by shooting, a young child was shot and killed. The bullet removed from her body was a .22 caliber bullet. The police need to know whether the gun used was a rifle or a handgun. A neighbor said that he heard the shot, but did not see the car or the shooters. However, the neighbor said he was a gun enthusiast, and that he could tell whether the gunshot was from a .22 caliber rifle or a .22 caliber handgun. The neighbor reported that the shot he heard came from a rifle. In order to determine whether the neighbor could actually distinguish between the sounds of shots fired from .22 caliber rifles and handguns, the police conducted tests in which the neighbor was asked to identify gunshots of .22 caliber rifles and handguns by their sound. The neighbor was correct (20, 40, 50, 60, 80%) of the time, and made an equal number of errors for each type of gun.

1) What is the probability that the weapon used was a rifle?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of handguns and rifles used in drive-by shootings is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of handguns and rifles used in drive by-shootings, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of handguns and rifles used in drive by-shootings and the witness' testimony are about equally relevant for making this judgement.

Situation E

The fans of the Spurs and Wolves football clubs are known to have violent confrontations after games that often lead to property damage. Of the property damage reports received by the police for which a suspect is arrested, 85% are caused by Spurs fans while 15% are caused by Wolves fans. The night of a game between the two teams, a grocery store owner called 911 claiming that his outdoor display was destroyed by some people in football jerseys and several goods were stolen. Upon their arrival, the police were approached by a witness who said he saw the incident. Although both teams wear red Jerseys that look similar, the witness claimed that as an avid football fan he could identify the jerseys and indicated that the individuals who damaged the store display were wearing Wolves Jerseys. In order to test his ability to correctly identify the jerseys, the police had the witness view pictures of the two teams' jerseys from different angles. The witness was correct in identifying the jersey (20, 40, 50, 60, 80%) of the time making an equal number of errors for both team jerseys.

1) What is the probability that the suspect were Wolves fans?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of arrests relating to property damage is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of arrests relating to property damage, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of arrests relating to property damage and the witness' testimony are about equally relevant for making this judgement.

Situation F

Robin's arts supply store packages paints from two companies, Primary Arts and Michelangelo's, under the same label. Although the store sells an equal amount of both paints, an inspection of the store's inventory indicates that of all the bad paint (paints that suffers from rapid color decay) distributed, 85% was produces by Primary Arts while 15% was produced by Michelangelo's. Recently, a customer complained that paint he had purchased from Robin's store was no good and lead to negative reviews of paintings of his at a gallery he was participating in. The customer claimed that as an experienced painter, he could tell the difference between the two paints and stated that the paint he used was from Michelangelo's. In order to determine whether the painter could actually make such a distinction, Robin gave him a test in which he was asked to identify the producer of a sample of paint. He was given an equal amount of samples of each company's paint. The painter was correct on (20, 40, 50, 60, 80%) of the samples, and made an equal number of errors for each company's paint.

1) What is the probability that the bad paint was produced by Michelangelo's?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of bad paints produces by each company is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of bad paints produces by each company, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of bad paints produces by each company and the witness' testimony are about equally relevant for making this judgement.

Situation G

Robert and Jamie are cooks at a restaurant and are both responsible for preparing ingredients by inspecting, washing and cutting them for the days service. Robert does 85% of the preparation of vegetables while Jamie does the remaining 15%. On September 2nd, Gordon, the head chef, noticed a batch of spoiled vegetables in the preparation tray. Neither Robert nor Jamie recalls using the spoiled vegetables. Gordon, however, claims he can identify who cut the vegetables because of subtle differences in Robert and Jamie's cooking styles. Gordon claims that it was likely Jamie who cut the vegetables. As a result of his experience and training, Gordon was involved in the hiring process and training of both Robert and Jamie. During the training period he was successful at identifying what dishes were created by what cook (20, 40, 50, 60, 80%) of the time, making an equal number of errors for both cooks.

1) What is the probability that Jamie cut the spoiled vegetables?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of vegetables prepared by each cook is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of vegetables prepared by each cook, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of vegetables prepared by each cook and the witness' testimony are about equally relevant for making this judgement.

Situation H

Two glass companies, Sunlight and Clearview, supplied special, unbreakable glass windows to a construction company that was building a high-rise building. Although Sunlight and Clearview each supplied an equal number of the windows, 85% of the windows that broke during installation due to defects were supplied by Sunlight and the other 15% that broke due to defects were supplied by Clearview. Just moments after a workman finished putting in one of the windows, it shattered because of a defect. Although the windows made by each company are very similar, the workman who installed the window claimed that he could tell the difference between the windows of the two glass companies, and stated that the window which shattered was a Clearview window because he saw all the Clearview windows being moved to the floor he was working on for installation. In order to test whether the workman could actually make this distinction, the foreman presented the workman with a number of windows from both companies and asked him to identify which company had made which window. The workman was found to be correct (20, 40, 50, 60, 80%) of the time, and made an equal number of errors for each type of window.

1) What is the probability that the window that shattered was a Clearview window? (Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

   a) Given the witness' testimony, the information pertaining to the proportion of windows that broke due to a defect is irrelevant for making this judgement.

   b) Given the information pertaining to the proportion of windows that broke due to a defect, the witness' testimony is irrelevant for making this judgement.

   c) Both the information pertaining to the proportion of windows that broke due to a defect and the witness' testimony are about equally relevant for making this judgement.

Situation I

Both the Pablo Wine Company and the Roberto Wine Company make a rosé wine. However, the Pablo Company distributes both wines in bottles with the same label. Although the Pablo Company distributes equal amounts of both wines, an inspection of both company's quality control records indicates that of all the bad rosé wine (wine that could produce food poisoning) distributed, 85% was produced by the Pablo Company themselves and 15% by the Roberto Company. Recently, a wine connoisseur bought a bottle of Pablo's rosé wine, and after drinking it, contracted food poisoning and became very ill. The connoisseur said that she could tell the difference between the wines produced by the two different companies and that the bad wine had been produced by the Roberto Company. In order to determine whether the connoisseur could actually make such a distinction, the wine companies gave her a test in which she tasted an equal number of samples of each company's wine. The connoisseur was then asked to identify which company had produced each sample. The connoisseur was correct on (20, 40, 50, 60, 80%) of the samples, and made an equal number of errors for each company's wine.

1) What is the probability that the bad wine was produced by the Roberto company? (Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

   a) Given the witness' testimony, the information pertaining to the proportion of bad wine produced by each company is irrelevant for making this judgement.

   b) Given the information pertaining to the proportion of bad wine produced by each company, the witness' testimony is irrelevant for making this judgement.

   c) Both the information pertaining to the proportion of bad wine produced by each company and the witness' testimony are about equally relevant for making this judgement.

Situation J

The Pari Wildlife Park recently obtained two baboons of different subspecies to roam wild at the park: one of them is a brown colored baboon and the other a black colored baboon. Reported incidents of baboons biting park visitors indicate that 85% of the baboon bites are made by the brown colored baboon and 15% are by the black colored baboon. Recently, a park visitor was bitten by one of the baboons in the park. The park ranger needed to know which subspecies of baboon had bitten the visitor in order to check the baboons for rabies as quickly as possible. A witness identified the baboon which bit the visitor as the black baboon. In order to test the ability of the visitor to discriminate between the two subspecies of baboons, the ranger presented her with photographs of the different types of baboons and asked her to identify which subspecies of baboon was pictured in each photo. The visitor made correct identifications (20, 40, 50, 60, 80%) of the time, and made an equal number of errors for each of the two subspecies of baboon.

1) What is the probability that the baboon that bit the visitor was of the black subspecies? (Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

   a) Given the witness' testimony, the information pertaining to the proportion of biting incidents involving black or brown baboons is irrelevant for making this judgement.

   b) Given the information pertaining to the proportion of biting incidents involving black or brown baboons, the witness' testimony is irrelevant for making this judgement.

   c) Both the information pertaining to the proportion of biting incidents involving black or brown baboons and the witness' testimony are about equally relevant for making this judgement.

Appendix II. Exonerating Scenarios

Situation A

Andrew's moving company owns two trucks, one blue and one green. The blue truck is used for 85% of inner city orders while the green truck is used for 15% of inner city orders. On the night of August 7th, one of Andrew's trucks was involved in a hit-and-run accident. A customer visiting Andrew's garage to place an order testified that he saw a blue truck parked at the garage. In order to ascertain whether the witness could actually distinguish between blue and green trucks at night, the court subjected the witness to a test which consisted of showing the witness film clips of green and blue trucks under lighting conditions similar to the night in question. The witness was asked to identify the color of the cab in each film clip. The witness made correct identifications (20, 40, 50, 60, 80%) of the time, making an equal number of errors for each color truck.

1) What is the probability that the truck involved in the accident was blue?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

   a) Given the witness' testimony, the information pertaining to the proportion of orders involving blue and green cabs is irrelevant for making this judgement.

   b) Given the information pertaining to the proportion of orders involving blue and green cabs, the witness' testimony is irrelevant for making this judgement.

   c) Both the information pertaining to the proportion of orders involving blue and green cabs and the witness' testimony are about equally relevant for making this judgement.

Situation B

A critical piece of evidence in a bank robbery case in Washington concerns the type of get-away car driven. Although the number of American and Japanese designed cars in Washington are equal, according to police records, the percentage of American cars used as get-away cars in robberies in Washington is 85% compared to 15% for Japanese cars. A witness to the bank robbery testified that the get-away car used by the robbers was not an American Model. The witness argued that as a used car saleswoman, she is familiar with the styles of different types of cars. In order to test the witness' ability to distinguish between Japanese and American cars, the witness was asked to observe a number of Japanese and American cars at various angles. The saleswoman was then asked to state whether the car was of American or Japanese design. The witness was correct on (20, 40, 50, 60, 80%) of the trials, making an equal number of errors for each type of car.

1) What is the probability that the get-away car was American?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of get-away cars used in robberies is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of get-away cars used in robberies, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of orders involving blue and green cabs and the witness' testimony are about equally relevant for making this judgement.

Situation C

In Madison there are two ambulance companies, Mayday and Valley, serve the local hospitals. Even though there is an equal number of each type of ambulance in Madison, Mayday ambulances have 85% of the complaints by Madison residents about driving that nearly causes or causes accidents, and Valley ambulances receive 15% of the complaints. An accident that involved many cars occurred at an intersection near a hospital. One driver said that an ambulance passing through the intersection caused the accident. However, the drivers of the cars could not identify which type of ambulance passed through the intersection just before the accident. At the time of the accident, a blind man was waiting at the intersection. He said that he could distinguish between the sirens of the two ambulance companies. The blind man stated that the ambulance involved in the accident could not have been a Mayday ambulance. In order to test the accuracy of the blind man's hearing, the blind man listened to tape recordings of sirens produced by the ambulances of the two different companies. The blind man listened to a large number of tape recorded sirens and correctly identified (20, 40, 50, 60, 80%) of them, making an equal number of errors for each company's siren.

1) What is the probability that the ambulance involved the accident was Mayday ambulance? (Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

   a) Given the witness' testimony, the information pertaining to the proportion of complaints involving reckless driving is irrelevant for making this judgement.

   b) Given the information pertaining to the proportion of complaints involving reckless driving, the witness' testimony is irrelevant for making this judgement.

   c) Both the information pertaining to the proportion of complaints involving reckless driving and the witness' testimony are about equally relevant for making this judgement.

Situation D

In the city of Windsor, records indicate that there are equal numbers of .22 caliber rifles and handguns. However, of the guns used in drive-by shootings, 85% of the .22 caliber guns are handguns and 15% are .22 caliber rifles. In a recent drive-by shooting, a young child was shot and killed. The bullet removed from her body was a .22 caliber bullet. The police need to know whether the gun used was a rifle or a handgun. A neighbor said that he heard the shot, but did not see the car or the shooters. However, the neighbor said he was a gun enthusiast, and that he could tell whether the gunshot was from a .22 caliber rifle or a .22 caliber handgun. The neighbor reported that the shot he heard did not come from a handgun. In order to determine whether the neighbor could actually distinguish between the sounds of shots fired from .22 caliber rifles and handguns, the police conducted tests in which the neighbor was asked to identify gunshots of .22 caliber rifles and handguns by their sound. The neighbor was correct (20, 40, 50, 60, 80%) of the time, and made an equal number of errors for each type of gun.

1) What is the probability that the weapon used was a handgun?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of handguns and rifles used in drive-by shootings is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of handguns and rifles used in drive by-shootings, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of handguns and rifles used in drive by-shootings and the witness' testimony are about equally relevant for making this judgement.

Situation E

The fans of the Spurs and Wolves football clubs are known to have violent confrontations after games that often lead to property damage. Of the property damage reports received by the police for which a suspect is arrested, 85% are caused by Spurs fans while 15% are caused by Wolves fans. The night of a game between the two teams, a grocery store owner called 911 claiming that his outdoor display was destroyed by some people in football jerseys and several goods were stolen. Upon their arrival, the police were approached by a witness who said he saw the incident. Although both teams wear red Jerseys that look similar, the witness claimed that as an avid football fan he could identify the jerseys and indicated that there was no one in a Spurs Jersey in the area of the store at the time of incident. In order to test his ability to correctly identify the jerseys, the police had the witness view pictures of the two teams' jerseys from different angles. The witness was correct in identifying the jersey (20, 40, 50, 60, 80%) of the time making an equal number of errors for both team jerseys.

1) What is the probability that the suspect were Spurs fans?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of arrests relating to property damage is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of arrests relating to property damage, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of arrests relating to property damage and the witness' testimony are about equally relevant for making this judgement.

Situation F

Robin's arts supply store packages paints from two companies, Primary Arts and Michelangelo's, under the same label. Although the store sells an equal amount of both paints, an inspection of the store's inventory indicates that of all the bad paint (paints that suffers from rapid color decay) distributed, 85% was produces by Primary Arts while 15% was produced by Michelangelo's. Recently, a customer complained that paint he had purchased from Robin's store was no good and lead to negative reviews of paintings of his at a gallery he was participating in. The customer claimed that as an experienced painter, he could tell the difference between the two paints and stated that the paint he used was not from Primary Arts. In order to determine whether the painter could actually make such a distinction, Robin gave him a test in which he was asked to identify the producer of a sample of paint. He was given an equal amount of samples of each company's paint. The painter was correct on (20, 40, 50, 60, 80%) of the samples, and made an equal number of errors for each company's paint.

1) What is the probability that the bad paint was produced by Primary Arts?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of bad paints produces by each company is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of bad paints produces by each company, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of bad paints produces by each company and the witness' testimony are about equally relevant for making this judgement.

Situation G

Robert and Jamie are cooks at a restaurant and are both responsible for preparing ingredients by inspecting, washing and cutting them for the day's service. Robert does 85% of the preparation of vegetables while Jamie does the remaining 15%. On September 2nd, Gordon, the head chef, noticed a batch of spoiled vegetables in the preparation tray. Neither Robert nor Jamie recalls using the spoiled vegetables. Gordon, however, claims he can identify who cut the vegetables because of subtle differences in Robert and Jamie's cooking styles. Gordon claims that it likely was not Robert who cut the vegetables. As a result of his experience and training, Gordon was involved in the hiring process and training of both Robert and Jamie. During the training period he was successful at identifying what dishes were created by what cook (20, 40, 50, 60, 80%) of the time, making an equal number of errors for both cooks.

1) What is the probability that Robert cut the spoiled vegetables?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of vegetables prepared by each cook is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of vegetables prepared by each cook, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of vegetables prepared by each cook and the witness' testimony are about equally relevant for making this judgement.

Situation H

Two glass companies, Sunlight and Clearview, supplied special, unbreakable glass windows to a construction company that was building a high-rise building. Although Sunlight and Clearview each supplied an equal number of the windows, 85% of the windows that broke during installation due to defects were supplied by Sunlight and the other 15% that broke due to defects were supplied by Clearview. Just moments after a workman finished putting in one of the windows, it shattered because of a defect. Although the windows made by each company are very similar, the workman who installed the window claimed that he could tell the difference between the windows of the two glass companies, and stated that the window which shattered was not a Sunlight window because he saw all the Sunlight windows being moved to a different floor for installation there. In order to test whether the workman could actually make this distinction, the foreman presented the workman with a number of windows from both companies and asked him to identify which company had made which window. The workman was found to be correct (20, 40, 50, 60, 80%) of the time, and made an equal number of errors for each type of window.

1) What is the probability that the window that shattered was a Sunlight window? (Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

   a) Given the witness' testimony, the information pertaining to the proportion of windows that broke due to a defect is irrelevant for making this judgement.

   b) Given the information pertaining to the proportion of windows that broke due to a defect, the witness' testimony is irrelevant for making this judgement.

   c) Both the information pertaining to the proportion of windows that broke due to a defect and the witness' testimony are about equally relevant for making this judgement.

Situation I

Both the Pablo Wine Company and the Roberto Wine Company make a rosé wine. However, the Pablo Company distributes both wines in bottles with the same label. Although the Pablo Company distributes equal amounts of both wines, an inspection of both company's quality control records indicates that of all the bad rosé wine (wine that could produce food poisoning) distributed, 85% was produced by the Pablo Company themselves and 15% by the Roberto Company. Recently, a wine connoisseur bought a bottle of Pablo's rosé wine, and after drinking it, contracted food poisoning and became very ill. The connoisseur said that she could tell the difference between the wines produced by the two different companies and that the bad wine was not produced by the Pablo Company. In order to determine whether the connoisseur could actually make such a distinction, the wine companies gave her a test in which she tasted an equal number of samples of each company's wine. The connoisseur was then asked to identify which company had produced each sample. The connoisseur was correct on (20, 40, 50, 60, 80%) of the samples, and made an equal number of errors for each company's wine.

1) What is the probability that the bad wine was produced by the Pablo company?
(Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

a) Given the witness' testimony, the information pertaining to the proportion of bad wine produced by each company is irrelevant for making this judgement.

b) Given the information pertaining to the proportion of bad wine produced by each company, the witness' testimony is irrelevant for making this judgement.

c) Both the information pertaining to the proportion of bad wine produced by each company and the witness' testimony are about equally relevant for making this judgement.

Situation J

The Pari Wildlife Park recently obtained two baboons of different subspecies to roam wild at the park: one of them is a brown colored baboon and the other a black colored baboon. Reported incidents of baboons biting park visitors indicate that 85% of the baboon bites are made by the brown colored baboon and 15% are by the black colored baboon. Recently, a park visitor was bitten by one of the baboons in the park. The park ranger needed to know which subspecies of baboon had bitten the visitor in order to check the baboons for rabies as quickly as possible. A visitor indicated to the park ranger that it could not have been the brown baboon because he was observing her at a location in the park that was distant from where the biting incident occurred. In order to test the ability of the visitor to discriminate between the two subspecies of baboons, the ranger presented her with photographs of the different types of baboons and asked her to identify which subspecies of baboon was pictured in each photo. The visitor made correct identifications (20, 40, 50, 60, 80%) of the time, and made an equal number of errors for each of the two subspecies of baboon.

1) What is the probability that the baboon that bit the visitor was of the brown subspecies? (Write your answer as a percentage)

2) Select the number from the scale below that indicates how confident you are that your probability judgement above (#1) is justified

Not at all Confident	1	2 barely	3	4	5 moderately	6	7	8 highly	9	Extremely Confident

3) Indicate (by circling either a, b, or c below) which of the following three statements best represents the logic you used in making your probability judgement (#1) above.

   a) Given the witness' testimony, the information pertaining to the proportion of biting incidents involving black or brown baboons is irrelevant for making this judgement.

   b) Given the information pertaining to the proportion of biting incidents involving black or brown baboons, the witness' testimony is irrelevant for making this judgement.

   c) Both the information pertaining to the proportion of biting incidents involving black or brown baboons and the witness' testimony are about equally relevant for making this judgement.

Appendix III: Bayesian solutions for inference problems.

Frist, we need to examine the information that is available in the scenarios that could contribute to a Bayesian estimate of culpability. I will use the scenario A (Andrew's moving company) to demonstrate this process. Below is a summary of the scenario.

1) The company owns two trucks, one blue and one green.
2) The blue truck is implicated at a base-rate of 85% while the green truck is implicated at 15%.
3) In exonerating scenarios participants are asked what the likelihood of guilt for a hit and run accident is for a blue truck that is exonerated by an eyewitness.
4) In incriminating scenarios they are asked what the likelihood of guilt for a hit and run accident is for a green truck that is incriminated by an eyewitness.
5) Witness accuracy varies between 20%, 40%, 50%, 60% and 80%.
6) In a diagnostic test, the witness makes an equal number of errors for each color truck, implying that he is just as likely to say that a truck is green when it is actually blue, as he is to say that a truck is blue when it is actually green.

The base rates are related to a binary choice. So, it is only possible that the truck was either blue or green (i.e., when it is not blue, it must be green). Using this information and the information in the summary above, we can deduce that:

1) **P(A)** = The probability that the truck was blue/green = **0.85, 0.15 or 85%, 15%**
2) **P(~A)** = The probability that the truck was not blue / not green (i.e., the probability that the truck was green / blue) = **0.15, 0.85 or 15%, 85%**

*P(A) and P(~A) are complimentary probabilities

3) **P(B|A)** = The probability that a witness says the truck is blue when the truck is blue = the accuracy of the witness on the diagnostic task = **0.2, 0.4, 0.5, 0.6, 0.8 or 20%, 40%, 50%, 60% and 80%.**

From the accuracy rate, we can infer the error rate since the decision is binary. So,

4) **P(~B|A)** = The probability that witness says the truck is blue when the truck is green = 1 – P(B|A) = 1- (0.2, 0.4, 0.5, 0.6, or 0.8) = **0.8, 0.6, 0.5, 0.4, 0.2 or 80%, 60%, 50%, 40%, 80%.**

Since we know the witness makes an equal number of error for each color,

5) **P(B|~A)** = The probability that the witness says the truck is green when the truck is blue = P(B|A) = **0.8, 0.6, 0.5, 0.4, 0.2 or 80%, 60%, 50%, 40%, 80%.**

Unknown parameter:

6) **P(B)** = The probability that the witness says the truck is blue
7) **P(~B)** = The probability that the witness says the truck is not blue (i.e., the witness says the truck is green)

* It is important to note that in this solution, we assume that P(B) and P(~B) are independent of P(A) and P(~A).

In both incriminating and exonerating scenarios, the participant is asked to make a decision about the truck that is identified by the witness. Thus, the decision task for incriminating scenarios is to identify:

8) **P(A|B)** = The probability that the truck is green given that the witness says that the truck is green

In Bayesian probability this can be calculated with the following formulas:

9) P(A|B) = [P(B|A) x P(A)] / P(B)

10) P(B) = [P(B|A) x P(A)] + [P(B|~A) x P(~A)]

For exonerating scenarios, the decision task is to identify:

11) **P(A|~B)** = The probability that the truck is blue given that the witness says that the truck is not blue

In Bayesian probability this can be calculated with the following formulas:

12) P(A|~B) = [P(~B|A) x P(A)] / P(~B)

13) P(~B) = [P(~B|A) x P(A)] + [P(~B|~A) x P(~A)]

*P(~B|~A) is equivalent to witness accuracy (i.e., P(B|A)) since witness is equally likely to be correct for either color. In this case, "not blue" is equivalent to "green"

The summary table below shows the calculated Bayesian estimate for incriminating and exonerating witnesses at different levels of accuracy and base-rate implications of 15%, 50% and 85%:

Accuracy	P(A|B)	P(A|~B)
15% Base-rate implication		
20%	0.04	0.41
40%	0.11	0.21
50%	0.15	0.15
60%	0.21	0.11
80%	0.41	0.04
50% Base-rate implication		
20%	0.2	0.8
40%	0.4	0.6
50%	0.5	0.5
60%	0.6	0.4
80%	0.8	0.2
85% Base-rate implication		
20%	0.59	0.96
40%	0.79	0.89
50%	0.85	0.85
60%	0.89	0.79
80%	0.96	0.59

Appendix IV: Full ANOVA Tables

Table 10. Study 1 full ANOVA table.

Source	Measure		SS	df	MS	*F*	*p*	η_p^2	Observed Power[a]
Subset	Probability	Sphericity Assumed	979.449	1	979.449	3.489	.063	.016	.460
		Greenhouse-Geisser	979.449	1.000	979.449	3.489	.063	.016	.460
		Huynh-Feldt	979.449	1.000	979.449	3.489	.063	.016	.460
		Lower-bound	979.449	1.000	979.449	3.489	.063	.016	.460
	Confidence	Sphericity Assumed	1.602	1	1.602	1.446	.230	.007	.224
		Greenhouse-Geisser	1.602	1.000	1.602	1.446	.230	.007	.224
		Huynh-Feldt	1.602	1.000	1.602	1.446	.230	.007	.224
		Lower-bound	1.602	1.000	1.602	1.446	.230	.007	.224
Subset * Condition	Probability	Sphericity Assumed	41.576	1	41.576	.148	.701	.001	.067
		Greenhouse-Geisser	41.576	1.000	41.576	.148	.701	.001	.067
		Huynh-Feldt	41.576	1.000	41.576	.148	.701	.001	.067
		Lower-bound	41.576	1.000	41.576	.148	.701	.001	.067
	Confidence	Sphericity Assumed	.379	1	.379	.342	.559	.002	.090
		Greenhouse-Geisser	.379	1.000	.379	.342	.559	.002	.090
		Huynh-Feldt	.379	1.000	.379	.342	.559	.002	.090
		Lower-bound	.379	1.000	.379	.342	.559	.002	.090
Error(Subset)	Probability	Sphericity Assumed	58663.940	209	280.689				
		Greenhouse-Geisser	58663.940	209.000	280.689				
		Huynh-Feldt	58663.940	209.000	280.689				
		Lower-bound	58663.940	209.000	280.689				
	Confidence	Sphericity Assumed	231.427	209	1.107				
		Greenhouse-Geisser	231.427	209.000	1.107				
		Huynh-Feldt	231.427	209.000	1.107				
		Lower-bound	231.427	209.000	1.107				
Accuracy	Probability	Sphericity Assumed	70063.991	4	17515.998	43.469	.000	.172	1.000
		Greenhouse-Geisser	70063.991	2.807	24964.100	43.469	.000	.172	1.000

		Huynh-Feldt	70063.991	2.862	24478.086	43.469	.000	.172	1.000
		Lower-bound	70063.991	1.000	70063.991	43.469	.000	.172	1.000
	Confidence	Sphericity Assumed	72.162	4	18.041	13.685	.000	.061	1.000
		Greenhouse-Geisser	72.162	3.855	18.720	13.685	.000	.061	1.000
		Huynh-Feldt	72.162	3.955	18.245	13.685	.000	.061	1.000
		Lower-bound	72.162	1.000	72.162	13.685	.000	.061	.957
Accuracy * Condition	Probability	Sphericity Assumed	107038.017	4	26759.504	66.408	.000	.241	1.000
		Greenhouse-Geisser	107038.017	2.807	38138.104	66.408	.000	.241	1.000
		Huynh-Feldt	107038.017	2.862	37395.612	66.408	.000	.241	1.000
		Lower-bound	107038.017	1.000	107038.017	66.408	.000	.241	1.000
	Confidence	Sphericity Assumed	3.357	4	.839	.637	.637	.003	.210
		Greenhouse-Geisser	3.357	3.855	.871	.637	.631	.003	.206
		Huynh-Feldt	3.357	3.955	.849	.637	.635	.003	.208
		Lower-bound	3.357	1.000	3.357	.637	.426	.003	.125
Error(Accuracy)	Probability	Sphericity Assumed	336869.012	836	402.953				
		Greenhouse-Geisser	336869.012	586.577	574.296				
		Huynh-Feldt	336869.012	598.224	563.115				
		Lower-bound	336869.012	209.000	1611.813				
	Confidence	Sphericity Assumed	1102.057	836	1.318				
		Greenhouse-Geisser	1102.057	805.679	1.368				
		Huynh-Feldt	1102.057	826.636	1.333				
		Lower-bound	1102.057	209.000	5.273				
Subset * Accuracy	Probability	Sphericity Assumed	4240.735	4	1060.184	3.774	.005	.018	.891
		Greenhouse-Geisser	4240.735	3.721	1139.635	3.774	.006	.018	.874
		Huynh-Feldt	4240.735	3.815	1111.566	3.774	.005	.018	.880
		Lower-bound	4240.735	1.000	4240.735	3.774	.053	.018	.490
	Confidence	Sphericity Assumed	15.978	4	3.994	3.271	.011	.015	.837
		Greenhouse-Geisser	15.978	3.682	4.339	3.271	.014	.015	.813
		Huynh-Feldt	15.978	3.775	4.233	3.271	.013	.015	.820
		Lower-bound	15.978	1.000	15.978	3.271	.072	.015	.437

Subset * Accuracy * Condition	Probability	Sphericity Assumed	697.252	4	174.313	.621	.648	.003	.205
		Greenhouse-Geisser	697.252	3.721	187.376	.621	.636	.003	.198
		Huynh-Feldt	697.252	3.815	182.761	.621	.640	.003	.200
		Lower-bound	697.252	1.000	697.252	.621	.432	.003	.123
	Confidence	Sphericity Assumed	.878	4	.220	.180	.949	.001	.088
		Greenhouse-Geisser	.878	3.682	.238	.180	.939	.001	.087
		Huynh-Feldt	.878	3.775	.233	.180	.942	.001	.087
		Lower-bound	.878	1.000	.878	.180	.672	.001	.071
Error(Subset*Accuracy)	Probability	Sphericity Assumed	234833.445	836	280.901				
		Greenhouse-Geisser	234833.445	777.717	301.952				
		Huynh-Feldt	234833.445	797.356	294.515				
		Lower-bound	234833.445	209.000	1123.605				
	Confidence	Sphericity Assumed	1020.748	836	1.221				
		Greenhouse-Geisser	1020.748	769.610	1.326				
		Huynh-Feldt	1020.748	788.874	1.294				
		Lower-bound	1020.748	209.000	4.884				

a. Computed using alpha = .05

Table 11. Study 2 full ANOVA table.

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2	Observed Power[a]
Accuracy	Probability	Sphericity Assumed	1309.062	4	327.265	1.904	.110	.023	.573
		Greenhouse-Geisser	1309.062	2.470	529.955	1.904	.141	.023	.439
		Huynh-Feldt	1309.062	2.617	500.266	1.904	.138	.023	.453
		Lower-bound	1309.062	1.000	1309.062	1.904	.171	.023	.276
	Confidence	Sphericity Assumed	19.181	4	4.795	6.772	.000	.077	.993
		Greenhouse-Geisser	19.181	3.745	5.122	6.772	.000	.077	.990
		Huynh-Feldt	19.181	4.000	4.795	6.772	.000	.077	.993
		Lower-bound	19.181	1.000	19.181	6.772	.011	.077	.730
Accuracy * Condition	Probability	Sphericity Assumed	2858.920	8	357.365	2.079	.037	.049	.836
		Greenhouse-Geisser	2858.920	4.940	578.696	2.079	.070	.049	.679
		Huynh-Feldt	2858.920	5.233	546.276	2.079	.066	.049	.699
		Lower-bound	2858.920	2.000	1429.460	2.079	.132	.049	.416
	Confidence	Sphericity Assumed	12.987	8	1.623	2.292	.021	.054	.877
		Greenhouse-Geisser	12.987	7.490	1.734	2.292	.024	.054	.859
		Huynh-Feldt	12.987	8.000	1.623	2.292	.021	.054	.877
		Lower-bound	12.987	2.000	6.493	2.292	.108	.054	.453
Error(Accuracy)	Probability	Sphericity Assumed	55702.918	324	171.923				
		Greenhouse-Geisser	55702.918	200.081	278.401				
		Huynh-Feldt	55702.918	211.955	262.805				
		Lower-bound	55702.918	81.000	687.690				
	Confidence	Sphericity Assumed	229.432	324	.708				
		Greenhouse-Geisser	229.432	303.352	.756				
		Huynh-Feldt	229.432	324.000	.708				
		Lower-bound	229.432	81.000	2.832				

a. Computed using alpha = .05

Table 12. Study3 full ANOVA table.

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	η_p^2	Observed Power[a]
Accuracy	Probability	Sphericity Assumed	51390.206	4	12847.551	86.620	.000	.517	1.000
		Greenhouse-Geisser	51390.206	2.848	18046.781	86.620	.000	.517	1.000
		Huynh-Feldt	51390.206	3.035	16932.038	86.620	.000	.517	1.000
		Lower-bound	51390.206	1.000	51390.206	86.620	.000	.517	1.000
	Confidence	Sphericity Assumed	20.176	4	5.044	7.354	.000	.083	.996
		Greenhouse-Geisser	20.176	3.666	5.504	7.354	.000	.083	.994
		Huynh-Feldt	20.176	3.956	5.100	7.354	.000	.083	.996
		Lower-bound	20.176	1.000	20.176	7.354	.008	.083	.764
Accuracy * Condition	Probability	Sphericity Assumed	7199.501	8	899.938	6.068	.000	.130	1.000
		Greenhouse-Geisser	7199.501	5.695	1264.130	6.068	.000	.130	.998
		Huynh-Feldt	7199.501	6.070	1186.045	6.068	.000	.130	.999
		Lower-bound	7199.501	2.000	3599.751	6.068	.004	.130	.874
	Confidence	Sphericity Assumed	31.895	8	3.987	5.813	.000	.126	1.000
		Greenhouse-Geisser	31.895	7.332	4.350	5.813	.000	.126	.999
		Huynh-Feldt	31.895	7.912	4.031	5.813	.000	.126	1.000
		Lower-bound	31.895	2.000	15.948	5.813	.004	.126	.859
Error(Accuracy)	Probability	Sphericity Assumed	48055.693	324	148.320				
		Greenhouse-Geisser	48055.693	230.656	208.343				
		Huynh-Feldt	48055.693	245.842	195.474				
		Lower-bound	48055.693	81.000	593.280				
	Confidence	Sphericity Assumed	222.229	324	.686				
		Greenhouse-Geisser	222.229	296.929	.748				
		Huynh-Feldt	222.229	320.428	.694				
		Lower-bound	222.229	81.000	2.744				

a. Computed using alpha = .05

References

About Us-AIDWYC.org (2013). Retrieved October 16, 2013, from The Association in defence of the wrongly convicted (AIDWYC) web site: http://www.aidwyc.org/about/

Allen, R. J., & Pardo, M. S. (2007). *The problematic value of mathematical models of evidence* University of Chicago Press. doi:10.1086/508269

Allison, M., & Brimacombe, C. A. E. (2010). Alibi believability: The effect of prior convictions and judicial instructions. *Journal of Applied Social Psychology, 40*(5), 1054. doi:10.1111/j.1559-1816.2010.00610.x

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review, 78*(3), 171-206. doi:10.1037/h0030834

Ask, K., Rebelius, A., & Granhag, P. A. (2008). The 'elasticity' of criminal evidence: A moderator of investigator bias. *Applied Cognitive Psychology, 22*(9), 1245-1259. doi:10.1002/acp.1432

Aslan, A., Bäuml, K., & Grundgeiger, T. (2007). The role of inhibitory processes in part-list cuing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 335-341. doi:10.1037/0278-7393.33.2.335

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *The Behavioral and Brain Sciences, 30*(3), 241-254. doi:10.1017/S0140525X07001653

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgements. *Acta Psychologica, 44*(3), 211-233. doi:10.1016/0001-6918(80)90046-3

Bäuml, K., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition, 32*(4), 610-617. doi:10.3758/BF03195852

Bex, F. J., van Koppen, P. J., Prakken, H., & Verheij, B. (2010). A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law, 18*(2), 123-152. doi: 10.1007/s10506-010-9092-x

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *The American Journal of Psychology, 96*(1), 85-94.

Bornstein, B. H., & Greene, E. (2011). Jury decision making: Implications for and from psychology. *Current Directions in Psychological Science, 20*(1), 63-67. doi:10.1177/0963721410397282

Burke, T., Turtle, J., & Olson, E. (2006). Alibis in criminal investigations and trials. In M. Toglia, R.C.L. Lindsay, D. Ross, & J.D. Read (Eds.), *The Handbook of Eyewitness Psychology: Volume I. Memory for Events*. (pp. 157-174). Mahwah, NJ: Lawrence Erlbaum Associates.

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgement of belief and value. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgement* (pp. 120 –138). Cambridge, England: Cambridge University Press.

Culhane, S. E., & Hosch, H. M. (2004). An aiibi witness' influence on mock jurors' verdicts. *Journal of Applied Social Psychology, 34*(8), 1604-1616. doi: 10.1111/j.1559-1816.2004.tb02789.x

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior, 12*(1), 41-55. doi:10.1007/BF01064273

Dahl, L. C., Brimacombe, C. A. E., & Lindsay, D. S. (2009). Investigating investigators: How presentation order influences participant-investigators' interpretations of eyewitness identification and alibi evidence. *Law and Human Behavior, 33*(5), 368-380. doi:10.1007/s10979-008-9151-y

Davis, J. H., Kerr, N. L., Atkin, R. S., Holt, R., & Meek, D. (1975). The decision processes of 6- and 12-person mock juries assigned unanimous and two-thirds majority rules. *Journal of Personality and Social Psychology, 32*(1), 1-14. doi: http://dx.doi.org/10.1037/h0076849

Devine, D. J. (2012). *Jury decision making : The state of the science* (1st ed.). New York: NYU Press.

Dhami, M. K. (2008). On measuring quantitative interpretations of reasonable doubt. *Journal of Experimental Psychology.Applied, 14*(4), 353-363. doi:10.1037/a0013344

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology, 75*(1), 53-69. doi:10.1037/0022-3514.75.1.53

Garrett, B. L. (2008). Judging innocence. *Columbia Law Review, 108*(1), 55-142.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*(1), 451-482. doi:10.1146/annurev-psych-120709-145346

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221-233. doi:10.1037/0022-3514.65.2.221

Gilovich, T., & Griffin, D. (2002). Introduction – Heuristics and biases: Then and now. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgement* (pp. 120 –138). Cambridge, England: Cambridge University Press.

Givelber, D. (2005). Lost innocence: Speculation and data about the acquitted. *The American Criminal Law Review, 42*(4), 1167.

Govier, T. (2005). Back ground knowledge and inference to the best explanation. In T. Govier (Ed.), *A practical study of Argument* (pp.336-340). Belmont, USA: Wadsworth/Thompson Learning.

Hampikian, G., West, E., & Akselrod, O. (2011). The genetics of innocence: Analysis of 194 U.S. DNA exonerations. *Annual Review of Genomics and Human Genetics, 12*, 97-120. doi:10.1146/annurev-genom-082509-141715

Hannaford-Agor, P. L., Hans, V. P., Mott, N.L., & Munsterman, G.T. (September 2002). *Are Hung juries a Problem?* Williamsburg, VA: National Center for State Courts. Retrieved from: https://www.ncjrs.gov/pdffiles1/nij/grants/199372.pdf

Harris, A. J. L., & Hahn, U. (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1366-1373. doi:10.1037/a0016567

Hastie, R., Penrod, S.D., Pennington, N. (1983) *Inside the jury*. Cambridge: Harvard University Press.

Hatvany, N., & Strack, F. (1980). The impact of a discredited key witness. *Journal of Applied Social Psychology, 10*(6), 490-509. doi:10.1111/j.1559-1816.1980.tb00728.x

Hinsz, V. B., Scott Tindale, R., & Nagao, D. H. (2008). Accentuation of information processes and biases in group judgements integrating base-rate and case-specific information. *Journal of Experimental Social Psychology, 44*(1), 116-126. doi:10.1016/j.jesp.2007.02.013

Hinsz, V. B., Tindale, R. S., Nagao, D. H., Davis, J. H., & Robertson, B. A. (1988). The influence of the accuracy of individuating information on the use of base rate information in probability judgement. *Journal of Experimental Social Psychology, 24*(2), 127-145. doi:10.1016/0022-1031(88)90017-0

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*(1), 1-55. doi:10.1016/0010-0285(92)90002-J

Hogarth, R. M., & Einhorn, H. J. (1981). Behavioral decision theory: Processes of judgement and choice. *Journal of Accounting Research, 19*(1), 1-31.

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. doi:10.1007/s10979-007-9106-8

Kahneman, D. (2011). *Thinking, Fast and Slow*. Canada: Doubleday Canada

Kassin, S. M., Leo, R. A., Meissner, C. A., Richman, K. D., Colwell, L. H., Leach, A., & La Fon, D. (2007). Police interviewing and interrogation: A self-report survey of police practices and beliefs. *Law and Human Behavior, 31*(4), 381-400. doi:10.1007/s10979-006-9073-5

Koehler, J.J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*, 1-53

Koehler, J. J. (2002). When do courts think base rate statistics are relevant? *Jurimetrics Journal of Law, Science and Technology,42*(4), 373.

Koehler, J. J., & Macchi, L. (2004). Thinking about low-probability events: An exemplar-cuing theory. *Psychological Science, 15*(8), 540-546. doi:10.1111/j.0956-7976.2004.00716.x

Lecci, L., & Myers, B. (2008). Individual differences in attitudes relevant to juror decision making: Development and validation of the pretrial juror attitude questionnaire (PJAQ). *Journal of Applied Social Psychology, 38*(8), 2010-2038. doi:10.1111/j.1559-1816.2008.00378.x

Lehrer, J. (2010). *How We Decide*. New York: Houghton Mifflin Harcourt Publishing Company

Lieberman, J. D. (2011). The utility of scientific jury selection: Still murky after 30 years. *Current Directions in Psychological Science,20*(1), 48-52. doi:10.1177/0963721410396628

Lieberman, J. D. (2009). The psychology of the jury instruction process. In Lieberman, J. D., & Krauss, D. A. (Eds.), *Jury psychology: Social aspects of trial processes* (pp. 129-155). Farnham: Ashgate Publishing Ltd.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica, 40*(4), 287-298. doi:10.1016/0001-6918(76)90032-9

McAllister, H. A., & Bregman, N. J. (1989). Juror underutilization of eyewitness nonidentifications: A test of the disconfirmed expectancy explanation. *Journal of Applied Social Psychology, 19*(1), 20.

Olson, E. A. (2013). “You don't expect me to believe that, do you?” Expectations influence recall and belief of alibi information. *Journal of Applied Social Psychology, 43*(6), 1238-1247. doi:10.1111/jasp.12086

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

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. doi:10.1037/0022-3514.36.4.436

Pennington, N., & Hastie, R. (1991). A cognitive theory of juror decision making: The story model. *Cardozo Law Review, 13*(2-3), 519.

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. doi:10.1037/0022-3514.62.2.189

Penrod, S., & Hastie, R. (1979). Models of jury decision making: A critical review. *Psychological Bulletin, 86*(3), 462-492. doi:10.1037/0033-2909.86.3.462

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19*, 123-205. doi:10.1016/S0065-2601(08)60214-2

Poveda, T.G. (2001). Estimating wrongful convictions. *Justice Quarterly, 18*(3), 689-708. doi:10.1080/07418820100095061

Price, H. L., & Dahl, L. C. (2014). Order and strength matter for evaluation of alibi and eyewitness evidence. *Applied Cognitive Psychology, 28*(2), 143-150. doi:10.1002/acp.2983

Ruva, C. L., Guenther, C. C.,& Yarbrough, A (2011). Deciphering the effects of positive and negative PTP: Examining the roles of impression formation, emotion and predecisional distortion. *Criminal Justice and Behavior*, *38*, 511-534. doi: 10.1177/0093854811400823

Saks, M.J. & Marti, M.W. (1997). A meta-analysis of the effects of jury size. *Law & Human Behavior, 21*, 451-467.

Saunders, H.D. (2005). Quantifying Reasonable Doubt: A Proposed Solution to an Equal Protection Problem. ExpressO, Available at: http://works.bepress.com/harry_saunders/2

Seltzer, R. (2006). Scientific jury selection: Does it work? *Journal of Applied Social Psychology, 36*(10), 2417-2435. doi:10.1111/j.0021-9029.2006.00110.x

Shaw III, J.S., Bjork, R.A, & Handal, A. (1995). Retrieval induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin and Review, 2*(2), 249-253.

Stasser, G., & Davis, J. H. (1981). Group decision making and social influence: A social interaction sequence model. *Psychological Review, 88*(6), 523-551. doi:10.1037/0033-295X.88.6.523

Steblay, N., Hosch, H. M., Culhane, S. E., & McWethy, A. (2006). The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: A meta-analysis. *Law and Human Behavior, 30*(4), 469-492. doi:10.1007/s10979-006-9039-7

Storm, B. C., & White, H. A. (2010). ADHD and retrieval-induced forgetting: Evidence for a deficit in the inhibitory control of memory. *Memory, 18*(3), 265-271. doi:10.1080/09658210903547884

Thagard, P. (2004). Causal inference in legal decision making: Explanatory coherence vs. bayesian networks. *Applied Artificial Intelligence, 18*(3-4), 231-249. doi:10.1080/08839510490279861

Ministry of the Attorney General (2014). Report of the Kaufman Commission on Proceedings Involving Guy Paul Morin. Retrieved June 13, 2014 from Ministry of the Attorney General website: http://www.attorneygeneral.jus.gov.on.ca/english/about/pubs/morin/

The Innocence Project - About us (2014) Retrieved May 6, 2014, from The Innocence Project website: http://www.innocenceproject.org/about/

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. doi:10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1980). Causal Schemas in judgements under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (pp.49-72). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458. doi:10.1126/science.7455683

Wells, G. L. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology, 62*(5), 739-752. doi:10.1037/0022-3514.62.5.739

Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88*(3), 776-784. doi:10.1037/0033-2909.88.3.776

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22*(6), 603-647. doi:10.1023/A:1025750605807

Zumpano, L. V., Elder, H. W., & Baryla, E. A. (1996). Buying a house and the decision to use a real estate broker. *Journal of Real Estate Finance and Economics, 13*(2), 169-181. doi:10.1007/BF00154054