8-2013

Jurors' Subjective Certainty and Standards of Proof: The Role of Emotion and Severity of Charge in Subjective Probability Judgment

Yimoon Choi
University of Nebraska-Lincoln, yimoon@gmail.com

JURORS' SUBJECTIVE CERTAINTY AND STANDARDS OF PROOF: THE ROLE OF EMOTION AND SEVERITY OF CHARGE IN SUBJECTIVE PROBABILITY JUDGMENT

by

Yimoon Choi

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Psychology

Under the Supervision of Professor Richard L. Wiener

Lincoln, Nebraska

August, 2013

JURORS' SUBJECTIVE CERTAINTY AND STANDARDS OF PROOF: THE ROLE OF EMOTION AND SEVERITY OF CHARGE IN SUBJECTIVE PROBABILITY JUDGMENT

Yimoon Choi, Ph.D.

University of Nebraska, 2013

Advisor: Richard L. Wiener

Recent empirical research suggests that jurors struggle to understand and correctly apply the standard of proof. Many researchers have focused on methods to re-write jury instructions so that standards of proof are clearer and easier for jurors to understand. This dissertation suggests the fundamental cause of jurors' confusion concerning standards of proof is that jurors may use different decision processes (intuitive decision processing or systematic decision processing) and decision indices (objective probabilistic judgment or subjective confidence) depending upon their transient emotions or the seriousness of charge.

Study 1 assessed whether experiencing particular emotions (sadness or anger) could change mock jurors' decision styles and their application of the standard of proof. Study 2 examined whether the severity of the charge against the defendant and the order of questionnaire administration could influence mock jurors' fact processing and final verdicts. Results of study 1 showed that mock jurors induced to feel incidental sadness used rational decision processing, employed standards of proof, and made use of strength of evidence (objective probability information) as the law intended. In contrast, mock jurors induced to feel incidental anger relied on intuitive processing, and failed to make appropriate use of standards of proof and the strength of the evidence. Results of study 2 indicated that the order of questionnaire administration influenced mock jurors' decision styles. When mock jurors assessed case arguments before

reaching verdicts, they used rational decision processing and followed the order of questionnaires. In contrast, when mock jurors decided verdicts first, they relied on intuitive processing. However, contrary to expectations, the severity of charges did not influence mock jurors' decision styles.

The results of these studies suggest that jurors' misunderstanding of the standard of proof may arise from the different types of fact-finding processes they use. Trial characteristics (i.e., severity of charge) or incidental emotion might contribute varied styles of processing that characterizes jurors' decision-making. Implications of these results and possible future studies are discussed.

Acknowledgements

I would like to thank my advisor and doctoral committee chair, Dr. Richard Wiener, for all of the time and energy he devoted to mentoring me over the years. I would also like to thank to the members of my committee, professor of Law Craig Lawson, Dr. Brian Bornstein, and Dr. Sarah Gervais for their thoughtful feedback and support. I am extremely grateful to my friend, Michael Holtje, whose encouragement and help to finish this dissertation. I could not have gotten through this project without him.

Finally, special thank you to my family, whose encouragement and belief in me have helped me accomplish much in my life.

Table of Contents

Acknowledgements iv
Table of Contents v
Table of Figures vii
Table of Tables viii
Chapter 1: Introduction 1
Chapter 2: Standards of Proof in the Legal Literature 6
Standards of proof: A Brief History 6
Legal status of standards of proof 8
Chapter 3: Research Regarding Standards of Proof and Jury Decision Making 15
Empirical studies of standards of proof 15
Objective probability and subjective certainty 23
Models of jury decision making and standards of proof 31
Chapter 4. Dual Processing Models 40
Dual processing models and subjective probability estimation 45
What triggers type of processing 47
Motivation 47
Individual differences. 48
Emotion 49
Trial characteristics & jurors' heuristic processing 52
Severity of charge. 52
Question order. 53
Appropriate use of standards of proof 56
Chapter 5. Research Project & Hypotheses 57
Overview 57
Hypotheses 62
Chapter 6: Study 1 65
Overview 65
Method 65
Participants. 65
Procedure. 66
Results 70
Rational-experiential inventory (REI) scores. 70
Completion time 71
Emotion manipulation check. 71
Effects of emotion on certainty and control appraisals. 72
Verdict analyses. 73
Guilty versus not guilty verdicts 73
Standard of proof ratings (threshold probability for conviction) 76
Coherence shifts. 77
Guilty certainty ratings. 80
Mediation analyses. 80
Discussion 83

Chapter 7: Study 2 88
Method 89
Participants. 89
Procedure. 90
Results 91
Rational-experiential inventory (REI) scores. 92
Completion time 92
Seriousness manipulation check 92
Accuracy motivation 93
Verdict Analyses 93
Standard of proof rating (threshold probability for conviction). 97
Coherence Shifts 97
Guilty certainty ratings 99
Mediation analyses. 100
Discussion 101
Chapter 8: General Discussion 105
Summary of results 105
Implications for psychology 108
Implications for law 113
Strengths of this study 115
Limitations and Qualifications 116
Future Research 119
Final conclusions 120
References 122
Appendixes 174

Table of Figures

Figure 1. Conviction rate for emotion, standard of proof, and strength of evidence in study 1 148
Figure 2. Relationship between standardized score of experiential measure and predicted probability of conviction for each condition of strength of evidence (Study 1)..... 151
Figure 3. Average coherence shift for standard of proof, strength of evidence, and verdict in study 1. 155
Figure 4. Average coherence shift for standard of proof, strength of evidence, and verdict in study 1.. 156
Figure 5. Conviction rate for order of questionnaires, severity of charge, and strength of evidence in study 2. 166
Figure 6. Relationship between standardized score of experiential measure and predicted probability of conviction for each condition of strength of evidence (Study 2)..... 167

Table of Tables

Table 1. Participants' Demographic Characteristics in Study 1 140

Table 2. Emotional Manipulation Checks..... 142

Table 3. Emotional Manipulation Checks..... 143

Table 4. Conviction Rate by Conditions (Emotion, Standard of Proof, and Strength of evidence) 144

Table 5. Predictors of guilty/Not guilty Verdicts..... 145

Table 6. Predictors of Guilty/Not Guilty Verdicts by Emotion Condition 146

Table 7. Four Logistic Regressions on Prediction of Conviction in Study 1. 149

Table 8. Effects of Emotion, Standard of Proof, and Strength of evidence on Subjective Standard of Proof 152

Table 9. Means of Coherence Shifts by Conditions (Emotion, Standard of Proof, and Strength of evidence) and Verdict..... 153

Table 10. Effects of Emotion, Standard of Proof, and Strength of evidence on Coherence Shift 154

Table 11. Effects of Emotion, Standard of Proof, and Strength of evidence on Guilty Certainty 157

Table 12. The Indirect Effects of Emotion on coherence shift 158

Table 13. Participants' Demographic Characteristics in Study 2..... 159

Table 14. Manipulation Checks for Severity of Charge 161

Table 15. Conviction Rate for Theft Charge by Conditions (order of questionnaires, severity of charge, and Strength of evidence) 162

Table 16. Predictors of guilty/Not guilty Verdicts..... 163

Table 17. Two Logistic Regressions on Prediction of Conviction in Study 2 164

Table 18. Four Simple Logistic Regressions on Prediction of Conviction in Study 2 168

Table 19. Effects of Severity of charge, Order, and Strength of evidence on Subjective Standard of Proof 169

Table 20. Means of Coherence Shifts by Conditions (the order of questionnaires, severity of charge, and Strength of evidence) and Verdict 170

Table 21. Effects of Seriousness, Order, and Strength of evidence on Coherence Shift 171

Table 22. Effects of Seriousness, Order, and Strength of Evidence on Guilty Certainty 172

Table 23. The Indirect Effects of severity of charge on verdict and coherence shift through Accuracy motivation measures 173

Chapter 1: Introduction

Standard of proof is an important concept in the American legal system. Jurors should understand the meaning of the appropriate standard and apply that standard in their decisions. In *In Re Winship* (1970), the United States Supreme Court held that the standard "beyond a reasonable doubt" is a constitutional requirement in criminal trials. Unfortunately, the Court did not articulate a definition of this standard of proof. In *Holland v. United States* (1954), the Court held that judges should not "add confusion and controversy" by attempting to define the standard of proof or the presumption of innocence in jury instructions.

The Court later held that "the Constitution neither prohibits trial courts from defining "reasonable doubt" nor requires them to do so" (*Victor v. Nebraska*, 1994, p.1243). Some judges have assumed that the "reasonable doubt" standard is "self-explanatory," such that "laymen" can easily understand and apply the concept (*United States v. Taylor*, 1993). In *Victor* (1994), the Court admitted that the phrase "moral certainty" in a criminal jury instruction, as it applies to the notions of "beyond a reasonable doubt," is ambiguous. However, the majority concluded that the question is not whether there is a constitutional violation with the jury instruction itself, but whether such a violation reasonably influences the verdict. The justices speculated about what the jury might have thought based on the defective instructions because the risk of a vague definition could be just as problematic as violation of the *Winship* requirement. Ultimately, the Court concluded that the "beyond a reasonable doubt" standard is so fundamental in criminal cases that the failure to apply it appropriately will bring great injustice and will lead ordinary citizens to mistrust the justice system.

Unfortunately, however, a growing body of empirical evidence suggests that "beyond a reasonable doubt" is not self-explanatory. Moreover, such undefined legal concepts let jurors

modify, alter, or reject legal requirements based on their own "common sense" (i.e., to the extent that the standards do not fit to their folk-psychology understanding of the concepts; Finkel, 1995).

One problem with the "beyond a reasonable doubt" standard arises from the word "reasonable." The word "reasonable" refers to the degree of certainty required for convicting the defendant, but this degree of certainty is subject to interpretation (Horowitz & Kirkpatrick, 1996). Sometimes jurors' intuitive notions of fairness may change the "reasonableness" standard to reflect their own beliefs about the level of certainty necessary to convict in a criminal trial. For example, in one study Horowitz and Kirkpatrick (1996) gave five "reasonable doubt" instructions to eighty mock juries (each consisting of 6 mock jurors) and found that (1) all jury instructions produced unacceptably low levels of certainty (ranging from 52 percent to 82 percent) and (2) the instructions changed the jurors' self-reported reasonable doubt standards. In addition, the mock juries changed their standards of proof depending upon the strength of evidence presented. They adopted lower self-reported standards of proof when the trial evidence was weak and higher standards when the evidence was strong. The researchers interpreted these results to suggest that jurors might construe the phrase to mean that they may use their own common sense understanding of fairness to define "beyond a reasonable doubt." Such confusion might permit a jury to apply a subjective standard of proof for criminal conviction that is much lower than level *Winship* (1970) requires.

Other research has investigated methods to reduce juror confusion by modifying jury instructions, leading some researchers to propose quantified jury instructions (Kagehiro, 1990; Kagehiro & Stanton, 1985; Mosteller & Youtz, 1990; R. J. Simon & Mahan, 1970; Tillers & Gottfried, 2006). Legal commentators generally agree that standards of proof imply quantitative meanings (Lieberman & Krauss, 2009; Shapiro, 2008), and "the establishment of truth of alleged

facts' is a matter of probabilities, falling short of absolute certainty" (Twining, 1984). Moreover, using probability language is consistent with Anglo-American legal doctrine and fits well with "Western" thinking styles, as compared to "Eastern" thought patterns (R. Friedman & Yates, 2003; Yates, 2010). For example, Americans express certainty more frequently, and feel more comfortable when they do so in terms of probability (Buehler, Griffin, & Ross, 2002; Griffin & Tversky, 2002; Wrightsman, Greene, Nietzel, & Fortune, 2002).

Although some social scientists have shown that quantitative or numeric instructions can reduce mock jurors' confusion in understanding and applying standards of proof (e.g., Kagehiro, 1985), it remains unclear how jurors apply those standards in fact-finding. Possibly, jurors could interpret quantified standards of proof as an objective probability statement and use probabilistic thinking in fact-finding. This "objective probability" theory assumes that people can transform mathematical information into rational beliefs and that objective probability calculations should be *de facto* standards of reasonableness (Gigerenzer, 1994). Bayesian decision theory, algebraic weighing, and stochastic choice models all adopt this explanation of information processing (Hastie, 1993; Pennington & Hastie, 1986; Penrod & Hastie, 1979).

However, empirical studies have shown that mock jurors usually cannot calculate the correct objective probability based on Bayesian theory (Faigman & Baglioni, 1988; Thompson and Schumann, 1987), and even when they calculate it correctly, their verdicts do not always follow their calculated objective probabilities (Wells, 1992). Instead, jurors construe the standard of proof as a level of subjective probability based on the intensity of their beliefs or the coherence of the story (Engel & Glokner, 2013; Hastie, 1993; D. Simon, 2004). This explanation focuses more on psychological evidence than normative rules. This "subjective" theory assumes that jurors use psychological and heuristic processes rather than rational or logical processes.

One possible explanation for the disjunction between objective probability estimation and subjective judgment could be that jurors use different decision-making processes depending on the type of case they are evaluating. In other words, jurors' misunderstanding of various standards of proof may arise not only from ambiguous or archaic legal language, but also from different types of fact-finding processes they use. In some cases, jurors may use normative and systematic decision processing with objective probability statements. In other cases, though, jurors may use intuitive decision processing with subjective probabilities.

Dual process theories in social psychology assume that a perceiver's capacity for attending to and processing information is limited. Systematic processing occurs only when perceivers have enough cognitive capacity or when they are motivated. Systematic processing is slow, high-effort, rule based, and involves serial processing. Without motivation and cognitive capacity, perceivers rely on heuristic processing, a fast, low-effort, associative process with easily accessible clues such as feelings, peripheral cues, or stereotypes.

If dual processing theories accurately depict juror decision-making, researchers should be able to observe the rules jurors use when performing their decision tasks. When using analytic-rational reasoning, jurors evaluate evidence impartially and estimate the probability of culpability based on these evaluations. In other words, jurors fit the evidence to standards of proof using probability estimates as a threshold to a decision. Conversely, when using heuristic reasoning, jurors use subjective probability estimates to evaluate a case and focus less on the evidence or the standard of proof.

Specific emotions or severity of charge may facilitate or inhibit heuristic processing. According to cognitive appraisal theory and the appraisal tendency framework (Lerner & Keltner, 2001), specific emotions give rise to specific cognitive and motivational processes, and those

processes help explain the effects of each emotion upon judgment and decision-making. For example, angry people rely more on heuristics when processing social information (Bodenhausen, 1994). Moreover, perceivers of negative events who experience anger report greater certainty about what happened and who is blameworthy for those negative events. Conversely, perceivers of negative events who experience fear report greater uncertainty about what transpired and look for situational control of negative events (Lerner & Tiedens, 2006).

In addition to decision-makers' emotions, seriousness of charge may inhibit heuristic processing. Severe charges might motivate jurors to make judgments that are more accurate. Because the punishment for serious crimes is severe and jurors may know the possible penalties in serious cases, they may be more motivated than are jurors evaluating less serious crimes.

This research explores a hierarchical judgment model that posits jurors' emotions and the severity of charges influence their understanding of standards of proof, which in turn affect jurors' final verdicts. Jurors' misunderstanding of various standards of proof may arise not only from ambiguous legal language, but also from different types of fact-finding processes they use.

To understand these phenomena better, first, this dissertation summarizes the role of standards of proof in the legal literature and in scientific research regarding jury decision-making. Next, this dissertation addresses dual process models in psychology and the relationship between those models and standards of proof. The paper includes a discussion that describes how and why jurors use different psychological mechanisms to evaluate evidence. It also explains why a strictly legal model of standard of proof fails to explain variation in jurors' verdicts. The paper then discusses how emotions, severity of charge, individual differences, and inducing systematic process might change jurors' decision-making processes by altering their understanding of standards of proof. In the end, this dissertation presents two experiments that evaluated the

impact of these extra-legal factors on jurors' decision-making styles and their application of standards of proof.

Chapter 2: Standards of Proof in the Legal Literature

Standards of proof: A Brief History

What is sufficient proof for jurors to convict a defendant or find a defendant liable? This question has its historical roots in the early Middle Ages (Shapiro, 1991). During the twelfth century, trials relied on so-called "irrational proofs," such as trial by ordeal, battle, or compurgation (i.e., taking an oath and having a required number of other persons swear that they believe the oath). Such methods relied on the assumption that an omniscient God reveals truth. However, after the twelfth century, even medieval societies no longer accepted such irrational proofs as consistent with justice or the determination of truth. On the European continent, the Romano-canon inquisitorial system replaced irrational proofs, and in England, the jury trial did the same.

Until the fourteenth century, however, the verdicts of juries in England were "assertive" rather than evidentiary, meaning that jurors (usually men of the town or neighborhood) were "self-informing" (Shapiro, 1991, p.4). The court assumed that jurors knew the facts and would incorporate that knowledge in their verdicts. Juries engaged in fact-finding through their own common sense and common knowledge. It was not until the late fifteenth century, that jurors started to obtain new information from evidence introduced in court. Thus, by the time judges began to understand that jurors did not have enough knowledge to decide cases, jurors had already become passive viewers of trial facts (Groot, 1988, cited in Shapiro, 1991). Thus not until late in the fifteenth century, did courts articulate standards of proof.

Around the year 1700, judges and legal commentators started using Lockean theories of knowledge to define certainty and doubt, relying on a graduated probability scale (Shapiro, 1991). They separated "the probable" from "the unlikely" by looking to "rational belief" or "moral certainty." Moral certainty and rational belief were synonymous with "knowledge" (Shapiro, 1991, p.8). For Locke, moral certainty was the highest level of probability. It produced "so near to a certainty, that it governed our thoughts as absolutely as the most evident demonstration." Lower levels of probability produced "confidence," "confident belief," or "mere opinion" (Locke & Leyden, 2002, p. 281). By the late eighteenth century, judges began instructing juries in detailed Lockean terms of probability and degrees of certainty (Shapiro, 1991). "Satisfied conscience" and "moral certainty" implied the highest level of certainty (roughly equivalent to the modern legal requirement of "beyond reasonable doubt"). The first recorded use of "beyond reasonable doubt" appeared in the Boston Massacre trial of 1770 (Shapiro, 1991).

Originally, the purpose of "beyond a reasonable doubt" was to facilitate conviction, not to protect defendants from false conviction. Prosecutors introduced the "beyond reasonable doubt" threshold to make it easier for jurors to convict. The alternative "any doubt" test did not require that doubts should be reasonable. The prosecution indicated that if "evidence is not sufficient to convince beyond reasonable doubt" then the jury must acquit (Shapiro, 1991, p.22). Subsequently, criminal cases became subject to the "beyond a reasonable doubt" standard. Judges accepted this eighteenth century formulation with little criticism.

More recently, legal commentators borrowed concepts from mathematics and probability theory to define the highest level of certainty (Shapiro, 2009). In America, one school of legal scholars approached law from an empirical perspective emphasizing sociological, psychological,

and biological perspectives (Shapiro, 2009; Twining, 2007). Some of these scholars describe decision-making using a Bayesian model (Kaye, 2000). Bayesian theory, a normative decision making model, describes a process that mechanically evaluates evidence without emotion, stereotyping, or bias. Bayesian theorists argue that quantifying the standard of proof can provide a more effective way to describe its meaning to jurors (Allen & Lively, 2003; Clermont & Hall, 2008; Kaye, 1999). Simply put, because jurors can and should evaluate every element of fact with a probability judgment, quantifying the standard of proof can provide a clearer picture of the level of certainty required to convict a defendant in a criminal case or to find for a plaintiff in a civil case.

For centuries, trial judges have modified the form that jury instructions take to express the highest level of certainty, but their ultimate goal was unchanged. Judges made a consistent effort to "hold juries to the highest standard possible for day-to-day human affairs" (Shapiro, 2009, p. 279). Shapiro (2009) wonders whether the language of mathematics and probability theory could replace the inexact language currently used to describe standards of proof, which in turn should give jurors clearer guidance. This possibility leads to two related empirical questions. First, do probability and mathematical theories describe how jurors understand and apply standards of proof? Alternatively, is it possible that using probabilistic and mathematical language in jury instructions could facilitate jurors' conformity to those models? Before answering these questions, I will first discuss several criteria that define the optimal meaning of certain standards.

Legal status of standards of proof

This chapter describes three criteria that courts have either explicitly or implicitly endorsed: (a) standards of proof have objective or probability meaning rather than qualitative or

absolute certainty, (b) standards of proof should be distinguishable from each other, and (c) jurors should apply standards of proof consistently regardless of their own emotion, moral judgment, or seriousness of charge on the same case.

First, consider the option that standards of proof have objective and probability meaning. Since the 1780s, increasing participation by defense counsel in the U.S. was one force driving consistency across decisions and objective laws of evidence (Langbein, 2003). A comparative view may help illustrate the stark differences between common law and civil law. In civil law countries, fact-finding processes are subjective and qualitative. Trial judges collect enough evidence to reach their subjective certitude of guilt. Judges collect evidence and evaluate that evidence based on their discretion (*free evaluation of evidence*). Fact-finding is the strictly subjective impression in a judge's mind (*intime conviction*). In contrast, the fact-finding process in the U.S. and other common law countries is objective and quantitative (even though the fact-finder must reach a subjective certainty of guilt). The U.S. has "objective" evidentiary rules rather than subjective judicial discretion. Proof is an objective and probabilistic concept (Engel, 2009).

The concept of probability plays a central role in the Anglo-American rationalist tradition of adjudication and evidence (Friedman & Yates, 2003; Twining, 1994). In this tradition, establishment of factual propositions is generally a matter of probability rather than absolute certainty. Fact-finding does not presume a dichotomy between truth and falsity. Rather, it conceptualizes "facts" in terms of qualitative and graduated probability (e.g., whether a "fact" is more or less likely to be true). Twining (1994) maintains that in Western legal cultures "the establishment of the truth of alleged facts in adjudication is typically a matter of probabilities, falling short of absolute certainty" (p.73).

For its part, the United States Supreme Court has defined standards of proof with quasi-probabilistic meanings (Clermont, 1986). In *In re Winship*, Justice Harlan observed that it is an unattainable goal for the fact finder to acquire indisputably accurate knowledge of what actually happened. Instead, the fact finder can only acquire "a belief of what *probably* happened" (*In re Winship*, 1970, p. 370), and this judgment of probabilities should result from reasoning about evidence. A belief or decision is rational if it is in agreement with the knowledge acquired by a proper procedure. Rationality, therefore, is procedural. It depends on the quality of the process used to obtain the resulting belief. The purpose of procedural law is to impose maximum rationality (Koehler, 2001).

Secondly, standards of proof should be distinguishable from each other. Higher standards should convey higher thresholds. The purpose of the standard of proof is "to instruct the fact finder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions for particular type of adjudication" (*In re Winship*, 1970, p.370). In other words, "a task of the law is making the choice appropriate to the situation; the law may aim to minimize overall errors, to decrease dangers of deception or bias or to disfavor certain claims, or to avoid a special kind of error such as convicting the innocent" (Clermont, 1986, p. 1120). In short, the United States legal system recognizes that truth (and hence fact-finding) is a matter of probability, and the system should seek to optimize probabilistic standards of proof (R. D. Friedman, 1998).

To achieve those aims, U.S. law has settled on three graduated standards that differ in how likely a contested fact must appear before a fact-finder determines that the fact exists. Of course, different standards of proof apply to different types of cases. First, the lowest-threshold standard, "preponderance of evidence," also called "more-likely-than-not", is the common

standard in civil litigation. There is considerable debate regarding the practical meaning of this standard. Legally, the standard distributes "the risk of error in roughly equal fashion" between the plaintiff and the defendant (*Addington v. Texas*, 1979, p.423). Second, the law uses an intermediate standard, "clear and convincing evidence" for civil cases involving allegations of fraud or some other quasi-criminal wrongdoing by the defendant. The risk of error is mere loss of money or tarnished reputation, but there is no loss of liberty (confinement or death) at stake. The last standard, embodying the highest level of certainty, requires "proof beyond a reasonable doubt," and it is the threshold for verdicts in criminal cases. This standard protects defendants from false accusation. The Due Process Clause of the Constitution requires the state to prove the guilt of an accused individual beyond a reasonable doubt (*In re Winship*, 1970).

These three standards form a continuum of probabilities that triers of fact sometimes need to correctly understand and differentiate (Clermont, 2004). There is some evidence that legislators and judges presume that there are genuine differences between the three standards (Tiller & Gottfried, 2006). For example, under some circumstances, misuse of standards may be a violation of an individual's constitutional rights. In *re Winship* (1970), the Supreme Court held that when the state charges a juvenile with an act that would be a crime if committed by an adult, the state must prove every element of the offense beyond a reasonable doubt. Despite lawmakers' belief that there are genuine differences between standards of proof, however, the lines separating these standards are thin (*United States v. Feinberg*, 1944). Justice Hand astutely noted how differences between preponderance of evidence and beyond a reasonable doubt are clear but differences between clear convincing evidence with either of these two standards would be hard to recognize.

Lastly, jurors should not apply standards of proof differently based on their own emotion, moral judgment, or severity of charge. In *Cage* (1990) and *Victor* (1994), potential variability in jurors' application of standards of proof led to objections by the defendants. In *Cage* (1990), the United States Supreme Court disapproved use of the phrase "moral certainty" because it would lead jurors to reasonably believe that they can convict based on their moral standards or emotion. In *Victor* (1994), one defendant argued that the terms "moral certainty" and "moral evidence" drew jurors' attention to their own ethics or the morality of the crime, not the evidence presented, which might lead jurors to decide the case based on their own emotion and prejudice. Though the Supreme Court rejected the defendant's objection based on context review, the court confirmed that jurors' variability in interpretation of the criminal standard is not acceptable.

Given these legal rules, an interesting jurisprudential question is whether quantifying language can explain standards of proof. The English jurist Blackstone used a quantifiable standard long before social scientists suggested doing so. Blackstone asserted, "better that ten guilty persons escape than that one innocent suffer" (Blackstone, Sharswood, & Field, 1760, p.353). The objective of Blackstone's 10-to-1 ratio is to ensure that erroneous acquittals are more frequent and socially acceptable than erroneous convictions (Nagel, 1979). Other legal commentators agree that standards of proof have quantitative roots (i.e., Shapiro, 2008).

Some social scientists suggest that the stringency of a standard of proof is analogous to the stringency of an alpha level in statistics (i.e., Kerr et al., 1976). The alpha level in statistical analysis reflects researchers' attempts to control false positives. Likewise, the standard of proof in legal cases reflects society's interest to control wrongful convictions or erroneous liability. In the criminal context, the standard is highest because the state assigns more risk to an erroneous conviction than an erroneous acquittal. In the civil context, the standard is lower because

society's concern that a defendant may be erroneously liable is less severe. If jurors sometimes confuse the criminal standard with the civil standard, "guilty beyond a reasonable doubt" may not protect against wrongful convictions. Furthermore, an inappropriately high standard in civil cases, combined with a burden of proof that requires plaintiffs to prove every element of their claims, may make it difficult for deserving plaintiffs to succeed.

In addition, social scientists have shown that using probability language is consistent with Anglo-American legal doctrine and fits well with Western-thinking styles, at least as compared to Eastern thought patterns (R. Friedman & Yates, 2003; Yates, 2010). Americans express certainty more frequently, and they feel more comfortable when they express certainty in terms of probability (Buehler, Griffin, & Ross, 2002; Griffin & Tversky, 2002; Wrightsman, Greene, Nietzel, & Fortune, 2002). Some researchers suggest that conveying standards of proof as quantification terms may improve jurors' comprehension and application of those standards (Kagehiro, 1990).

However, American courts usually do not allow a quantified definition of the "beyond a reasonable doubt" standard. Occasionally in criminal trials, a prosecutor or a trial judge uses numbers and percentages to explain the reasonable doubt standard, but not always successfully. For example, in *People v. Ibarra* (2001), the appellate court disapproved a prosecutor's explanation of "beyond a reasonable doubt" by using of a bar graph with numeric values ranging from 100 percent certainty to lower levels (*People v. Ibarra*, 2001, as cited in Tiller et al., 2006). The prosecutor visualized the correct level of certainty as follows: Beyond a reasonable doubt was equal to 60 percent, beyond a shadow of a doubt to 70 percent, beyond all doubt to 80 percent, and absolute certainty as 90 percent. There have also been several attempts to explain beyond reasonable doubt with football-field analogies (*State v. Del Vecchio*, 191 Conn. 1983;

State v. Casey, 2004, as cited in Tiller et al., 2006), but appellate courts repeatedly have struck down these sports analogies. In *McCullough v. State* (1983), the trial judge described standards of proof using a scale from 0 to 10 to indicate relative certainty required for each standard. The judge described preponderance of evidence as 5 and reasonable doubt as 7.5. The Nevada Supreme Court struck down this description because the particular number the judge chose for reasonable doubt is not appropriate and reasonable doubt is "inherently qualitative" (p.75). Occasionally, though, an appellate court will allow puzzle analogies that refer to numeric values. The trial judge in *Rosa* (1996) successfully described proof beyond a reasonable doubt as "a 1,000 piece puzzle with only sixty pieces missing" (*Commonwealth v. Rosa*, 1996, as cited in Tillers et al., 2006).

Why do appellate judges disallow quantified jury instruction? One reason is the unquantifiability of degrees of belief (Tillers & Gottfried, 2006). Contrary to the views of many social scientists, judges doubt that anyone can objectively measure jurors' subjective confidence. If jurors cannot estimate their own subjective certainty, comparing their subjective certainty to quantified standards of proof would be impossible. Another objection is that there is no agreement on the threshold probability that courts should require to convict a defendant or to find a defendant liable. If "beyond a reasonable doubt" means 90 percent certainty, then jurors will convict 10 percent of innocent men. In *United States v. Hall* (1988), Judge Richard Posner pointed out that there is no consensus regarding the required threshold probability to convict defendants. He observed that trial judges and jurors might develop widely varying probabilities to describe the degree of certainty required by the reasonable doubt standard. For example, the level of certainty that defines "beyond a reasonable doubt" has been as low as seventy-six

percent (see *United States v. Fatico*, 1978) and in some cases, as low as 50 percent (see McCauliff, 1982).

These cautionary suggestions from the legal domain (i.e., that it is not possible to conclusively ascertain either jurors' subjective confidence or the appropriate objective legal threshold) highlight a point of divergence between the law and the social sciences. Most social scientists would likely respond to the suggestions by pointing out how the lack of an objective reference point can increase jurors' variability on both factors and that providing a reference point would likely decrease such variability. Nonetheless, as Tillers and Gottfried (2006) note, some judges maintain that jurors, not judges, should make these determinations.

In summary, some American commentators call for an infusion of objective and probabilistic meaning into standards of proof, and the law requires juries to distinguish different standards of proof. In addition, the court does not allow individuals' emotion or the morality of a crime to impact standards of proof. Nonetheless, judges remain skeptical about quantified standards of proof. It is not clear whether jurors' beliefs are accurately quantifiable, and it is not clear what level of certainty is required for each standard of proof.

Chapter 3: Research Regarding Standards of Proof and Jury Decision Making

Empirical studies of standards of proof

Considering the importance of standards of proof, it is not surprising that many empirical studies have investigated jury instructions' presentation of those standards. Most empirical studies have used an "application test," which measures jurors' abilities to understand the law as instructed when applying the law to a fact pattern (Severance & Loftus, 1982). This type of test approximates the task real jurors perform, and it can easily provide information about specific aspects of jury instructions that are not well understood by participants.

In some early studies, researchers measured lawyers' and mock jurors' understanding of standards of proof by simply asking them to replace verbal standards with probability statements. McCauliff (1982) asked federal judges to supply probability statements for nine terms of law, estimating the level of certainty that each conveyed. The average probability for "beyond a reasonable doubt" was 90.28 percent, for "clear and convincing evidence" was 74.99 percentage, and for "preponderance of the evidence" was 55.33 percent. Thus, judges displayed a reasonable range of probability estimates that distinguished the standards from each other.

Another line of studies focused on jurors understanding and application of the "beyond a reasonable doubt" standard under conditions when the definition of "beyond a reasonable doubt" varies significantly. These studies reasoned that if the jury instructions are effective and jurors understand standards correctly, higher standards should reduce conviction rates because higher standards increase jurors' threshold of conviction. In one mock jury study, Kerr et al. (1976) compared three definitions of reasonable doubt: (a) any doubt about the defendant's guilt qualified as a reasonable one (stringent criterion of guilt), (b) a reasonable doubt must be a substantial one (lax criterion), or (c) no definition of the concept (undefined). They gave these three definitions of reasonable doubt to individual or groups of mock jurors in a criminal case. They found a higher acquittal rate when using the most stringent definition (i.e., any doubt about the defendant's guilty qualified as a reasonable one) as compared to when no definition was used. Moreover, the effect of "reasonable doubt" definitions on verdicts was larger when juries initially split on the verdict.

In a seminal study, Horowitz and Kirkpatrick (1996) gave one of five jury instructions to six-person mock juries: (a) "firmly convinced" (b) "moral certainty" (c) does not waver or vacillate (d) real doubt (e) undefined. They presented the facts of a murder trial (Horowitz &

Seguin, 1986) and varied the evidence so that the case favored a conviction (strong evidence) or an acquittal (weak evidence). Mock jurors listened to audiotaped scripts, viewed slides of actors, and then deliberated in small groups to reach a verdict. Effective jury instructions should have induced jurors to be less likely to convict after receiving weak evidence (.50 certainty of guilty), compared to jurors who received strong evidence (.85 certainty of guilty). Results revealed that undefined or "does not waver" descriptions did not lead mock jurors to report higher guilt scores when the case was strong. In contrast, 'firmly convinced' definitions led mock jurors to report higher guilty scores after receiving strong evidence, compared to weak evidence. Specifically, when receiving strong evidence, there was no significant difference between guilty verdicts across jury instructions. However, when receiving weak evidence, only the "firmly convinced" definition significantly reduced guilty verdict. Furthermore, those applying the "firmly convinced" definition discussed the evidence and instructions in more detail than those using other definitions.

In addition, jurors set different standards corresponding to the strength of the evidence. Mock jurors' self-reported reasonable doubt standards showed that they set lower standards (63%) than the level that one would expect from courts and judges, that is about 90 percent as McCauliff (1982) reported. This finding supports the suggestion that jurors alter their standards of proof according to the likelihood of conviction. When jurors receive weak evidence, they set lower standards to decide guilty verdict. In contrast, when jurors receive strong evidence, they set higher standards to reach a guilty verdict. Interestingly, this low standard did not differ by deliberation (pre: 63% v. post: 64%). Only the "firmly convinced" instruction showed a significant increase in the self-reported standard of reasonable doubt after deliberation. The researchers interpreted the result to suggest that the juries with the "firmly convinced"

instruction probably identified the appropriate standard during deliberation. Together, these results suggest that, without appropriate guidance by jury instructions, jurors' verdicts likely do not reflect an appropriate level of certainty or an appropriate sensitivity to differences in evidentiary persuasiveness.

Relying on the same methodological assumption that decreases in conviction reflect increases in decision thresholds, some researchers have tested whether juries are able to distinguish different standards of proof presented in jury instructions. Though the law requires that jurors distinguish between the standards, empirical research indicates that jurors may have some difficulty doing so. For example, in a series of experiments (Kagehiro, 1990; Kagehiro & Stanton, 1985), Kagehiro and her colleagues provided undergraduate students a summary of a civil trial with legal definitions for one of three standards of proof: preponderance of evidence, clear and convincing evidence, or beyond a reasonable doubt. The particular wording of each instruction varied across the experiments. Kagehiro and her colleagues did not ask mock jurors to deliberate as a group because individual jurors should understand and apply standards of proof (*In re Winship*, 1970) and empirical evidence suggests that there is no significant discussion of the standard of proof in jury deliberation (Kagehiro, 1990). The researchers assumed that higher standards of proof should have increased protection for the defendant (as the law assumes). In other words, there would be fewer verdicts favoring plaintiffs as the standard of proof becomes higher. Surprisingly, though, only the quantified definitions reflected the hypothesized results (i.e., participants were more likely to find for the defendant under the "beyond a reasonable doubt" standard, as compared to the other lower standards). Across five studies with legal definitions, participants were no more likely to find for the defendant under the beyond a reasonable doubt standard than under the clear and convincing evidence. Conviction rates did

not differ when participants received the beyond a reasonable doubt standard, compared to when they received the preponderance of evidence standard. As a whole, these studies suggest that the legal definitions in common language failed to lead mock jurors to distinguish between standards.

Other empirical evidence suggests that legal decision makers may modify decision thresholds depending on potentially relevant case characteristics such as emotion or severity of the crime. In a mail survey, R. J. Simon (1969) asked a sample of state and federal trial court judges about the probability to declare a defendant guilty: "In arriving at a verdict for each offense, what would the probability that the defendant committed act have to be before you declared him guilty" (p. 110)? The list of crimes was murder, embezzlement, grand larceny, rape, petty larceny, and fraud. Simon assumed that judges would assign equal probability thresholds to all crimes regardless of severity. Instead, judges assigned higher thresholds to more severe crimes. In another study, Kerr (1978) investigated whether potential penalties may affect conviction thresholds. He asked participants to read a trial summary describing a case of first-degree murder, second degree murder, or manslaughter. Then, they read the penalty that the defendant would receive on conviction: a moderate (up to 5 years) or severe penalty (25 years to life or capital punishment). Participants were more likely to convict defendants accused of second-degree murder than first-degree murder. Kerr found that the conviction threshold was higher when the penalty was severe. If jurors modify the standard of proof depending on the severity of the crime or potential penalty, it is unclear whether severity of a crime only alters decision thresholds or whether it alters both decision thresholds and decision processes.

Another study tested the role of emotion on jurors' probability estimations and verdicts (Kassin & Garfield, 1991). Kassin and Garfield (1991) showed that crime scene videotapes influence jurors' decision thresholds. As part of a murder case, some mock jurors viewed a crime

scene videotape of close-ups of a bloodied victim that an assailant had stabbed to death and left lying in the street. Other jurors did not see that video. The mock jurors who viewed the crime scene videotapes established a significantly lower threshold for a guilty verdict than did those who did not view the videotape (77% v. 93%). However, there was no difference in conviction rate between conditions. This result indicates that the significantly lower threshold set by those who saw the bloody crimes scene did not result in more guilty verdicts. However, this study provides some evidence that emotions may influence self-reported decision thresholds but may not influence verdicts.

The general conclusions of these studies are the same. First, the phrase "beyond a reasonable doubt" is not self-explanatory, and its meaning varies depending on how the instructions define or describe the phrase. The required degree of certainty seems to vary by trial judges' definition of "reasonable doubt." Quite likely, some criminals have avoided punishment when jurors required too much certainty, and unwarranted convictions have occurred when jurors required too little certainty.

Second, empirical evidence indicates that sometimes jurors may have some difficulty distinguishing different standards. Although a legal distinction between standards could have important effects in a particular case, psychological evidence has demonstrated that different standard rarely have significant effects on decisions. Mock jurors are no more likely to find for a defendant under the "beyond a reasonable doubt" standard than under the "clear and convincing evidence" standard (Kagehiro, 1990).

Lastly, sometimes the required decision threshold may vary depending on characteristics of a case, such as emotional evidence or severity of crime. The severity of a criminal charge might affect juries' conviction thresholds in two particular ways. First, although the legal system

generally ignores jurors' speculation or simply admonishes the jury to refrain from considering the seriousness of a crime,[1] jurors likely speculate on possible consequences of their verdicts. If different jurors employ different thresholds in their own speculations, jurors may apply different standards to similar cases. Second, serious crimes positively correspond to emotional blameworthiness. If jurors confuse their emotions with subjective certainty of guilty, jurors' transient emotions could lead to an erroneous conviction or wrong liability. Notably, though, the studies described above do not fully uncover the psychological mechanisms underlying how jurors use jury instructions to understand standards of proof.

Other researchers have tried to fill this gap. After providing differently worded instructions describing the standards, researchers have used three methodological approaches to examine whether particular jury instructions adequately conveyed the appropriate threshold for deciding conviction or liability. First, mock jurors quantified the standard of proof (e.g., R. J. Simon & Mahan, 1970). Though this method helps to indicate whether jurors distinguish standards of proof and show which instructions provide more clear guidance, it cannot explain why jurors assigned specific probability to a standard. For example, it is unclear why jurors assigned 74.99 percent to "clear and convincing evidence" standard (McCauliff, 1982). Another problem is that many legal professionals assert that probabilities cannot encompass all factors in determining legal issues. Most judges do not allow quantified jury instructions because they believe that standards of proof are inherently qualitative (R. J. Simon, 1969). In addition, as Kassin and Garfield (1991) showed that changes in jurors' self-reported decision thresholds may not result in different conviction rates.

[1] Unlike civil cases in which the jury typically determines both liability and damage award, criminal jurors usually complete their jobs with a verdict on guilt. Jurors generally do not participate in criminal sentencing except in capital cases.

Second, some researchers relied on jurors' verdicts as a measure of jurors' use of standards of proof (e.g., Horowitz & Kirkpatrick, 1996). Horowitz and Kirkpatrick (1996) reasoned that if a "beyond a reasonable doubt" instruction was effective, jurors given weak evidence should have been less likely to convict a defendant than jurors given strong evidence. This method has many advantages over simple quantifying. The simulation looks more realistic, does not require any quantification, does not rely on self-reporting, and is more relevant to legal questions. However, this method does not explain the process by which different jury instructions increase or decrease jurors' decision thresholds. Moreover, this method cannot exclude the possibility that jurors might construe the evidence differently.

Another measurement of jury instruction efficacy is whether particular jury instructions lead mock jurors to follow the law. For example, Koch and Devine (1999) found that wording of reasonable doubt could interact with the availability of a lesser verdict option, and this interaction could influence verdicts. When defining reasonable doubt as being "firmly convinced," less severe crimes had no impact on conviction rates. However, when reasonable doubt was unexplained, mock jurors convicted more frequently when they received lesser charges. Despite this showing of jury instructions' influence, it remains unclear how mock jurors reach a verdict and how standards of proof influence that decision.

More significantly, although some studies suggest that case characteristics such as severity of charge (Kerr, 1978) or emotional evidence (Kassin & Garfield, 1991) can affect jurors' verdicts, empirical studies do not explicitly test the relationship between these case characteristics and jurors' understanding and application of standards of proof. If jurors interpret standards of proof differently as a function of case characteristics, courts and researchers cannot make standards of proof consistent simply by changing modifying jury instructions.

Furthermore, most empirical studies examining jury instructions and standards of proof are simply descriptive and lack psychological models to explain their findings. These studies only provide a description of jurors' decisions, not a psychological explanation for why or how such decisions occur. Undeniably, the influence of jury instructions is legally and practically significant; however, the underlying psychological mechanisms responsible for the effects of that wording remain unknown.

Objective probability and subjective certainty

In the following section, I will discuss the two probability judgment methods in the legal and empirical literature that show the disjunction of objective probability judgment and subjective certainty. For nearly fifty years, legal scholars have debated the use of overtly probabilistic evidence and methods (Koehler & Shaviro, 1990). These debates involve three general issues. The first issue concerns the appropriateness of allowing juries to base factual determinations on probabilities derived from base-rate evidence. The second issue concerns the use of subjective probabilities that a factual proposition is true. The third issue concerns the purposes of trials, verdict accuracy or imposing guilt or liability when facts warrant doing so. This discussion focuses on the nature and function of subjective probabilities because the purpose of this research project is to learn how jurors understand and apply standards of proof.

Accordingly, it is appropriate to begin with the observation that employing probability theory is useful to assess the combined significance of several subjective estimates that pertain to the same case. Almost in a tautological sense, all evidence is probabilistic because there is always a risk of error. The meanings of standards of proof are clearer under probability theory, compared to intuition. Probability theory assumes decision makers should rely on mathematical probability to decide the verdict. Standards of proof serve as decision thresholds, and numeric

values can describe the meanings of standards of proof. In contrast, the role of standards of proof is unclear under an intuition-based theory. With intuition, decision makers rely on individual value-choice, heuristic thinking, or empirical beliefs (Koehler & Shaviro, 1990). For example, jurors are more likely to recall vivid or salient events compared to less salient events even when the latter are more frequent. Under an intuition model, standards of proof take on a psychological meaning instead of a simple a mathematical threshold.

Furthermore, there are two ways to combine separate items of base-rate evidence into one probability judgment (Koehler & Shaviro, 1990). First, probability theory can facilitate combination of evidence. Base rates and subjective probabilities of evidence obey mathematical rules (Kaye, 1999). Decision makers who combine evidence based on mathematical rules are more consistent. Probabilistic and objective rules are rigorous techniques because they yield more accurate verdicts than other methods, such as unaided intuition. Second, intuition can be an alternative to probabilistic logic. While some commentators such as Kaye and Koehler generally support the use of base rate evidence and probability theory, other legal commentators, such as Brilmayer, Kornhauser, and Cohen, argue that mathematical calculation is unrealistic (Brilmayer & Kornhauser, 1961; Cohen, 1981). Not all evidence is convertible into probabilities, and there might be no feedback in real trials to calculate base-rates. Instead, intuitive or subjective probabilities may simply reflect laypersons' certainty to decide a verdict. Intuition, on the other hand, is internally coherent and preferable to probabilistic logic in the event of conflict.

As described later in mathematical models, probabilistic evidence is an important source of evidence for accurate verdicts. Although the law assumes a close relationship between the subjective certainty and objective probability that, a defendant committed an act and jurors' ensuing verdict, prior empirical research has shown that mock jurors often violate this

assumption. In a landmark study, Wells (1992) provided "the Blue Bus/Grey Bus problem" (Nesson, 1986, p.521) to several groups of participants. In this vignette, either a Blue or a Gray bus ran over Mrs. Prob's dog. One group received the base rate information that 80% of all buses belong to the Blue Bus Company and only 20% belong to Gray Bus. Most participants correctly estimated the base rate of 80%, but only 8.2% of participants decided against the Blue Bus Company. A second group received information designed to elicit a similar subjective probability that Blue Bus was culpable. In this condition, a Weigh Station attendant testified with 80% subjective confidence that the culprit was Blue Bus. Although the objective probability was identical in the two conditions, 67% of the participants ruled against Blue Bus in this condition. In short, "psychologically, there seems to be a difference between saying that there is an 80% chance that something is true and saying that something is true based on evidence that is 80% reliable" (Wells, 1992, p. 746).

Another example comes from a non-legal context. Windschitl and Wells (1998) asked participants to read a scenario in which they had 21 tickets for a lottery. In one condition, other players held 12, 13, 13, 14, and 15 tickets. In another condition, other players held 52, 6, 2, 2, and 5 tickets. In both conditions, participants' likelihood of winning was identical, but participants in the first scenario expressed greater certainty that they would win. Reviewing these studies, Windschitl and Young (1998) summarized that "it assumes that there can be a dissociation between a person's belief in the objective probability of an event and his/her more intuitive or 'gut-level' perception of certainty… Recent research… has uncovered related evidence for dissociations between gut-level perceptions of certainty and beliefs in objective likelihood" (p. 111).

Several studies attempt to parse this disjunction between gut-level perceptions of certainty and beliefs of objective likelihood (Niedermeier et al., 1999, Sykes & Johnson, 1999). Some researchers explain the blue bus/grey bus problem through an "ease of simulation" hypothesis (Niedermeier et al., 1999) or an analogous explanation based on a mental representation (Sykes & Johnson, 1999), while other researchers look to morality (Wasserman, 1991-1992) and still others suggest a different meaning of probability in law than in psychology (Nesson, 1986). These efforts suggest a possible relationship between simulation heuristics and the disjunction between subjective perception and objective likelihood. Neidermeier et al. (1999) proposed an ease-of-simulation hypothesis, postulating that jurors are reluctant to convict to the extent that they can simulate a situation in which the defendant is not guilty. For example, if a juror can easily imagine a grey bus killing the dog, the juror is less likely to rule against Blue Bus.

Because objective probability judgments may occur simultaneously with intuition, it may be difficult to differentiate between the two different processes. As Koehler and Shaviro (1990) argued, it can be difficult to separate jurors' probability judgments into probabilistic theory or intuition. However, social psychologists suggest the possibility that both intuitive decision processing and systematic decision processes may co-occur and interact with each other to exert either independent or interdependent effects on evaluation (e.g., Chaiken & Maheswaran, 1994). For example, German scholars Glockner and Engel (2013) describe an interesting method to separate objective probability judgment and intuitive probability judgment by manipulating manipulated base-rate information, which rational decision makers should use to calculate objective probabilities (Koehler & Shaviro, 1990). They assume that if mock jurors use systematic decision processing with probabilistic methods, mock jurors should be sensitive to

this information. In contrast, mock jurors using intuitive processing would not use probability information.

Simon and colleagues modified this method to investigate cognitive coherence models. Researchers have used their approach in judicial decision making studies (Holyoak & Simon, 1999), in studies of criminal cases (Simon, Snow, et al., 2008), to evaluate romantic relationships (Simon, Stenstrom, & Read, 2011), to understand lay judgments of judicial decision making (Simon & Scurich, 2011), and to investigate role-induced bias in court (Engel & Glockner, 2012). Using a summary vignette to represent a legal case, Simon, Snow and Read (2004) asked participants to make six factual assessments in a pretest exercise. For example, participants indicated whether, in a flower business, financial deals typically occur through cash payments. Then, participants read a new fact pattern which involved defendant Jason Wells, who was accused of stealing $5,200 from a company's safe. Participants played the role of arbiter. The evidence included six pieces of circumstantial, probabilistic evidence. Participants made a factual decision whether Jason Wells stole the money and evaluated each argument that the defense or prosecution presented in the fact pattern. For example, one piece of evidence showed that the defendant paid off his credit card debt with cash. The defendant argued that he received the money from his sister, but he could not prove it because she worked in a flower business where financial deals typically occur through cash payments. The researchers assumed that if participants use intuitive decision-making, they would distort their evaluation about identical factual statements (i.e., flower business transactions occurring via cash payments) following their verdict decisions in order to make their evaluations consistent with their verdicts. As the researchers predicted, participants demonstrated a strong distortion in order to maintain coherence in their judgments. Unfortunately, this study did not investigate objective probability

judgments, limiting any inferences that the authors could make about the divergence between subjective and objective probabilities. Nor did the researchers manipulate standards of proof. However, Simon and his colleagues' research provided an interesting method to study how jurors integrated evidence to reach verdicts.

Glockner and Engel (2013) extended the work of Simon and his colleagues to investigate the psychological role of standards of proof by manipulating those standards while controlling other legal information. For example, after reading the same vignette and evidence, one group of participants applied the clear and convincing evidence standard and another group applied the beyond a reasonable doubt standard. Moreover, to separate subjective probability from objective probability, the researchers also manipulated the defendants' probability of culpability. The researchers provided several pieces of statistical evidence, such as the likelihood that eyewitnesses were correct (80% to 99%). Similarly, the researchers manipulated the number of workers who had the ability to open the safe from which money was allegedly stolen (from 18 persons to 4 persons). The researchers showed that if participants applied Bayesian reasoning to these pieces of evidence, the likelihood of mock jurors finding the defendant guilty would be 50%, 95%, or 99%.

The researchers included several measurements to separate intuitive processing from systematic processing. First, they used Simon and colleagues' cognitive coherence index, which reflected the average change in six factual assessments (which participants made before and after the verdict decision). Second, the researchers asked participants to complete a 13-item questionnaire that directly queried whether individuals applied a mathematical strategy based on deliberate calculation, a simple rule of thumb, or cognitive consistent interpretations of the evidence. For example, one question measuring heuristic processing asked whether participants'

agreed with the statement, "I counted the number of arguments for and against the defendant and decide solely on the basis of this amount." Another question asked whether they agreed with the statement, "My decision was based on an objective consideration of the information." Participants answered each question on a scale ranging from strongly disagree (-5) to strongly agree (+5).

In the first experiment, Glockner and Engel (2013) provided German undergraduate students with a modified version of Simon and colleagues' fact pattern (Simon, 2004; Simon, Snow, & Read, 2004) in which a company accused an employee of stealing money from the company safe. The case included six pro-guilty pieces of evidence and six contra-guilty pieces of evidence, each with some form of probabilistic information. The authors manipulated the probability of defendant guilt from medium (55%) to high (95%) by changing the probabilistic values of the pieces of evidence. One group of mock jury participants received a criminal scenario with the "beyond a reasonable doubt" standard, as articulated by the U.S. Ninth Circuit Court of Appeals. Another group received an arbitration scenario with the Ninth Circuit's "preponderance of evidence" standard. (Although, this experimental design confounded type of case – criminal/civil – with standard of proof – reasonable doubt/preponderance –the results are nonetheless interesting and worth considering.) Participants reported a liability or guilt verdict, the probability that the accused had stolen the money, the necessary level of confidence for convicting the accused, and the probative value of each piece of evidence. The participants also completed measures of their processing style (rational strategy, heuristic, or consistent maximizing).

In a second similar experiment, the researchers paired both standards of proof instructions (preponderance of evidence & beyond a reasonable doubt) with two different

decision frames (criminal & civil). In other words, they added two legally inappropriate scenarios (i.e. preponderance of evidence in a criminal case decision and beyond a reasonable doubt in a civil case) to the appropriate scenarios (i.e., preponderance of the evidenced in a civil case and beyond reasonable doubt in a criminal case) in order to test their hypotheses. That is, to separate the effects of the standard of proof from the type of case, participants received a criminal standard of proof in the civil case and the civil standard of proof in a criminal case. Finally, the last experiment tested the effect of strength of evidence on verdicts by manipulating strength of case, as either a strong case (99% likelihood of guilt) or a slightly weaker case (95% of guilty).

The researchers found that varying the standard of proof influenced the conviction rate in the hypothesized direction. Participants reported higher conviction rates when using the preponderance of evidence standard, as compared to the reasonable doubt standard. However, participants did not use probabilistic information rationally. Conviction rates were insensitive to increased probability of guilty. Participants indicated no differences in conviction rates between the 99% and 95% likelihood of being guilty conditions. Participants self-reported that they used consistency-maximization to examine the evidence and that they did not use deliberate decision-making or simple heuristics. Perhaps most interestingly, participants whose understanding of standards of proof relied only on translated jury instructions were unable to differentiate adequately between the beyond a reasonable doubt standard and the preponderance of evidence standard. They reported that "beyond a reasonable doubt" requires 85 percent certainty and "preponderance of the evidence" requires 75 percent certainty, a difference, which although statistically significant was not close enough to what we would expect on an objective probability scale. Furthermore, participants' reported that the 85 percent threshold for "beyond a

reasonable doubt" was significantly lower than 90 percent, and their reported 75 percent threshold for "preponderance of the evidence" was significantly higher than 50 percent certainty.

Although these results help illuminate our understanding, there are some problems. First, in the second study, participants received a criminal case with a preponderance of evidence standard. While experimentally plausible, it is legally unrealistic for anyone who already knows that criminal cases require the beyond a reasonable doubt standard. Of course, participants were German undergraduate students who might not be familiar with U.S. standards of proof, particularly since Germany is a civil law country that does not articulate standards of proof (Clermont, 2008). Second, the researchers did not include individual-difference variables that might influence participants' engagement in the process. Empirical studies suggest that some people tend to enjoy using rational decision strategies more than intuitive decision strategies, while others tend to rely on intuitive strategies more frequently than rational strategies (e.g., Epstein, 1994). Third, the researchers assumed that legal decision makers use one decision making process regardless of characteristics of a case, such as severity of charge. Social psychology's dual process models suggest that people might process information in different ways depending on the type of decision, severity of crime, or emotional influences. Finally, the study measured decision styles relying on self-reports. Empirical evidence suggests that people may not be able to accurately assess and report their cognitive processes (Nisbett & Wilson, 1977).

Models of jury decision making and standards of proof

Some legal psychologists have tried to theorize how jurors use standards of proof in their decision making. Any model must account for the fact that jurors listen to conflicting evidence from multiple sources and then synthesize that information to reach decisions. To integrate this

information, jurors may create a story (Penrod & Hastie, 1979) or they may rely on probabilistic estimates based solely on the evidence at trial (Schum, 2009). In either case, the law presumes that jurors arrive at a subjective estimate of guilt and then compare that estimate with a specified threshold, which is the standard of proof as articulated in jury instructions. If a juror's subjective estimate of guilt exceeds the standard of proof, then the decision should be to convict the defendant (Ostrom, Werner, & Saks, 1978).

The two main theories of how jurors accomplish this task are the mathematical and the story models of jury decision making (Groscup & Tallon, 2009). This section will briefly review each model and explain how each incorporates the standard of proof. The mathematical model assumes that jurors rely heavily on objective probability judgments that are derivable from probabilities, base rates, algebraic processes, and a stochastic process (Hastie, 1993). This model possesses the advantage of compatibility with legal rules and formal expression (Devine, Clayton, Dunford, Seying, & Pryce, 2001). The basic idea underlying the mathematical model is that, for each piece of evidence, jurors independently assign a weight on a single dimension, such as certainty of guilty, and decide a verdict by averaging the weighted values of all available pieces of evidence (Kaplan & Kemmerick, 1974).This model assumes that standards of proof specify decision rules regarding jurors' subjective probability-assessment that the facts warrant imposing liability. For example, this model interprets the preponderance of evidence as "return a verdict for plaintiff if the probability is greater than ½ that the facts that plaintiff need to prevail are as plaintiff alleges (Kaye, 1999, p.3)." The threshold probability for 'beyond a reasonable doubt' is much higher.

The difference between the civil and criminal standards comes from "the command to maximize expected utility or minimize expected loss" (Kaye, 1999, p. 3). For example, the

decision threshold is no more than a function of two error costs – the loss associated with a false verdict for the plaintiff and the loss associated with a false verdict for defendant. When these two losses are equal, such as in civil cases, the "more probable than not" standard minimizes the expected loss. In contrast, if the loss of false conviction is higher, such as in criminal cases, a higher threshold probability minimizes expected loss. Following Blackstone's assertion, if it is ten times worse to convict an innocent person than to acquit a guilty person, then the threshold probability is 10/1, or 91 percent.

How might the severity of a charge influence this theory? In mathematical approaches, jurors engage in a series of mental calculations. They weight the relevancy and strength of each independent piece of evidence and aggregate the resulting score. Then, they compare this score with the standard of proof. Severity of charge is not relevant to deciding verdicts. In addition, this theory does not expect the severity of a charge to influence jurors' evaluations of evidence. The standard of proof is only applicable after jurors evaluate the evidence and estimate the subjective likelihood of the defendant's liability or guilt. Similarly, the theory does not expect other extralegal factors, such as emotion, to affect the meaning of the standard of proof.

There are two central problems with mathematical models as they relate to standards of proof. First, there is ambiguity in the meanings of "proof by preponderance of the evidence" and "proof beyond a reasonable doubt." Legal language does not convey clear thresholds for these standards (contrary to quantified standards). Even if judicial instructions were more accurate, there is no agreement on the threshold probability that jurors should require in order to convict a defendant or to find a defendant liable (Allen & Leiter, 2001). Second, the model does not account for extra-legal factors that may influence base rate estimations and verdicts. Jurors are human beings with emotions, morals, and prejudices, and these extra-legal factors can influence

the outcomes of cases (Villejoubert, Almond, & Alison, 2009). Recall that mock jurors who viewed a videotape of a murder scene applied lower conviction standards than those who saw no video or a video depicting an unrelated crime scene (Kassin & Garfield, 1991). Furthermore, even if jurors estimate likelihoods based on Bayesian theory, they are unlikely to decide verdicts following mathematical calculations. In Wells' (2001) influential study, even though participants could calculate the likelihood of liability accurately, they hesitated to decide for the plaintiff.

The most widely adopted approach to juror decision making is the story model (Devine et al., 2001), which portrays jurors as story writers who integrate evidence into coherent stories (Pennington & Hastie, 1986). The story model is an explanation-based model of juror decision making. The strength of this model is its completeness. Jurors create one or more stories based upon the trial evidence, prior knowledge, personal expectation, and emotional reactions to the evidence (Pennington & Hastie, 1992). They then select the best story based upon the "certainty principle," in which the fact finder constructs a story that satisfies their required level of certainty (Pennington & Hastie, 1992, p.190). Three factors determine the jurors' certainty in their stories: coverage of the evidence, coherence of the story, and uniqueness of the story. Coverage of the story concerns whether the story accounts for most of the evidence. Coherence of the story pertains to whether the story is internally consistent and compatible with the fact finder's world knowledge. Uniqueness of the story concerns the absence of plausible alternative stories.

As with mathematical models, the story model assumes that standards of proof work as thresholds. Applying a goodness-of-fit test, jurors select a story that best fits the evidence, and if the certainty of the story exceeds the standard of proof, the verdict will be "guilty." However, if certainty falls below the threshold, the verdict will be "not guilty" (in accordance with the presumption of innocence). Similar to a mathematical model, jurors' estimated values for the

"beyond a reasonable doubt" and "presumption of innocence" standards do not change as a function of the verdict (Hastie, 1991). The role of the standards of proof is unidirectional, from confidence to verdict. Contrary to relying on objective probability, as mathematical models presume, the story model assumes jurors reach verdicts based upon subjective certainty (Engel, 2009). In the story model, the goodness of fit is somewhat subjective because evaluations about coherence, uniqueness, and coverage are more difficult to quantify than the base rate information in mathematical models.

Emotion contributes more to decision making in the story model than in the mathematical model (Hastie, 2001). According to the model, the story that the juror constructs is the central determinant of the decision. Most emotion-related effects will pertain to the characteristics of jurors' story. There are three ways that emotion might influence jurors' story constructions (Hastie, 2001). In the initial stage of construction, emotions can distort initial story creation. Jurors might use incidental or appraised emotions as cues to construct their initial story. For example, jurors' emotional states can bias retrieval of information from their memories. An angry juror might exaggerate the egregiousness of a defendant's alleged conduct or the severity of an injury. Second, emotions can fill in the details of a story. When a juror tries to infer the intentions or goals of an actor, emotions can bias the juror's reasoning by ascribing motivations associated with the juror's current emotional state. If a juror is angry or fearful, that juror may attribute morally reprehensible motives to an actor. Thirdly, in later stages of story construction, a juror's emotion can influence the verdict. For example, sympathy toward one party might influence a juror's threshold for the acceptance of a story.

Closely related to the story model of juror decision making, coherence-based models suggest that requiring a higher standard of proof makes jurors more confident of their decisions.

Contrary to the mathematical and story models, coherence-based models do not predict that higher standards of proof will lead to lower conviction rates. According to coherence-based models, certainty flows from the degree of coherence that people find in complex decision tasks (Simon, 2004). During a trial, jurors' mental representations of the evidence changes, gradually shifting toward a state of coherence that supports one of the possible decisions (e.g., guilty or not guilty verdict). Information supporting a verdict is overestimated, and conflicting information is underestimated (Simon, 2004). This process reflects cognitive consistency and a "coherence shift" that results in a subjective probability that exceeds the certainty required by the standard of proof (Simon, Snow, & Read, 2004).

Coherence-based reasoning shares many features with heuristic processing. Generally, empirical findings support the notion that consistency-maximizing processes are general, automatic mechanisms in human cognition (Read, Simon, & Stenstrom, 2010; Simon, 2004; Glockner, 2008). These processes do not require a trigger for initiation and involve little cognitive effort. They help people make sense of information by actively organizing it. Consistency-maximizing processes increase consistency in mental representations automatically. Because these processes are automatic without awareness, they cannot be easily reduced or changed consciously (Bargh & Chartrand, 1999). People are not aware of the underlying mental processes, but they get a consistent mental representation (Glockner, 2008; Simon, 2004). Based on this reasoning, Simon (2004) and Glockner (2008) argue that consistency-maximizing processes play on operative role in decision-making, rather than in simple post-decisional-rationalization.

Coherence-based reasoning and heuristic processing are more alike than they are different, but there are important distinctions. Heuristic research deals with specific and narrowly

defined judgment tasks, and heuristic research does not purport to describe cohesive and general cognitive processing. Conversely, cognitive coherence models focus on underlying cognitive mechanisms that drive complex processing (Simon, 2004). Coherence models provide some interesting insights regarding standards of proof. First, heightened standards of proof may not guarantee higher protection from erroneous decisions. From a legal perspective, the failure of both mathematical and story models are worrisome, because the implications of the alternative coherence model is troublesome for the due process principle (D. Simon, 2004). Coherence-shifts have the effect of undercutting the protection promised by a heightened standard of proof. In addition, these models illustrate how jurors can distort or change reference points. Such intuitive decision-making can have some serious disadvantages. For example, jurors may not evaluate the diagnostic value of evidence correctly; they may ignore admissible evidence; or they could use their feelings as information (Schwarz, Gilovich, Griffin, & Kahneman, 2002). If these models are accurate, jurors likely rely on their subjective confidence of their own simulations to make verdict decisions, and jurors are less sensitive to probative value of evidence. Simply put, even if jurors understand heighten meaning of criminal standards, it is difficult for them to apply these standards of proof to their verdict in a consistent manner.

In coherence models, emotions may also have an influential role. According to Thagard (2003), when people make judgments, their decisions come not only from what they believe but also from their emotional assessment. In addition, inferences about what people believe merge with the production of feelings about people, things, and situations. On this theory, most propositions and concepts have an emotional valence, which derives from evaluative features such as likeability, desirability, or similar kinds of positive or negative attitudes. For example,

the positive valence when individuals win a race influences their mental processing regarding the race, such that people purportedly distort objective criteria (Simon, 2004).

Although the two theories are similar, Simon (2004) argued that the coherence model corrects an important limitation of the story model. The story model proposes that the representation of evidence bears a distinct narrative structure. Stories center on a narrative that captures the intentional and causal scheme of a defendant's behavior. However, not all legal cases reflect the primacy of intention, goals, or motivation. For example, human intentionality is of little relevance to negligence cases where the core issues concern an actor's failure to appreciate risk and a subsequent damage assessment. Still, one might think of the coherence model as a story model with a consistency process that Hastie did not, but perhaps should have, included. Unlike the mathematical model, the story model is much more malleable and would be easy to amend to suggest that the interpretation of a certainty threshold (e.g., how jurors interpret the meaning of beyond a reasonable doubt or preponderance of the evidence) could fluctuate depending upon the type of story that they made up.

In sum, mathematical models describe how a juror ought to decide a case based on evidence. Base rate information should help jurors determine the subjective probability of an event. Standards of proof serve as thresholds for decisions (i.e., verdict or liability). If they take this approach, jurors should distinguish standards and understand the required level of certainty articulated in a judicial instruction. Emotion is unrelated and undesirable if it influences jurors' decision making. While the severity of the charge might influence jurors' expected utility calculation, the fact finding role should help the jurors from changing their calculation processes.

In contrast to mathematical models, story models and coherence models describe the psychology of how jurors decide cases. Jurors construct plausible stories rather than calculate

probabilities. Standards of proof are decision thresholds, but the discrimination between standards of proof is not as clear as in mathematical models. In these latter models, heightened standards of proof sometimes cannot guarantee higher decision thresholds because jurors simply increase subjective certainty by re-evaluating the evidence. These models assume jurors might use their emotions broadly and that the severity of the charge might influence decisions and story construction. As noted above, seriousness of charge may inhibit heuristic processing or motivate jurors to make accurate judgments. Because the punishment for serious crimes is severe and jurors may know the possible penalties in serious cases, they may be more motivated than jurors evaluating less serious crimes.

This dissertation proposes that one explanation of the findings described above may be jurors' transient emotions or the severity of the charge in the case. Most jury-decision research assumes a general model of decision-making and examines how different standards of proof influence verdicts. Researchers hypothesize that the role of standards of proof in jury decision making is universal, regardless of the case characteristics. However, it is possible that the characteristics of crimes researchers employ could contribute to some of the differences between the models described above. The scenarios used in coherence and story model research are more emotional and involve severe crimes. In fact, investigations into the story model and the coherence model present murder trials (D. Simon, Snow, & Read, 2004; Pennington & Hastie, 1986), hit-and-run cases (Pennington & Hastie, 1992), and theft cases (Engel & Glockner, 2008). In contrast, researchers studying Bayesian models have used DNA evidence, traffic violations (Kaplan & Kemmerick, 1974), or murder.

Although some empirical studies noted the role of emotion (Kassin & Garfield, 1991) or seriousness of charge (Kerr, 1978; Freedman, Krismer, MacDonald, & Cunningham, 1994) on

jurors' verdicts or decision thresholds, it remains unclear whether they affect jurors' processing of evidence, such as probability information, and jurors' application of standards of proof. In one set of studies Simon, Snow, and Read (2004) manipulated severity of charge comparing a simple theft charge to a theft resulting in an individual's death. Since murder is morally blameworthy, it likely has an emotional effect on decision makers. The researchers found no significant effect for severity of charge (theft vs. murder) on conviction rates, but they found coherence shifts changed the meaning of the evidence with both charges. They did not investigate the role of probability information and standard of proof on coherence-shifts. Before investigating how jurors' emotion or the severity of the charge might impact jurors' decision-making processes concerning probability information, it is necessary to first review dual process models in social psychology, because those models may help answer this question.

Chapter 4. Dual Processing Models

Dual process models in social psychology

This dissertation will investigate whether social psychology's dual process models can help explain jurors' differential use of standards of proof. First, I will review the development of dual process models that distinguish between two processing styles. Then, I will review the dual process explanation of subjective probability estimation. Finally, I will review what factors lead people to engage in one of these processes. Psychologists and philosophers argue that there are two distinct types of processing that guide cognition, reasoning, and decision-making. The American psychologist and philosopher William James (1890; cited in Frankish & Evans, 1999) asserted that there are two qualitatively different mechanisms of information processing: associative reasoning and true reasoning. He asserted that true reasoning is more useful than associative reasoning because associative thought is "only reproductive." In other words, objects

of associative thought are all elements of or abstractions from past experience. True reasoning, by contrast, is "productive" because it can address novel data. "Reasoning helps us out of unprecedented situations" (p.330). James (1907/1991) named his dual-process theory a "pluralistic monism" (p.13). Empirical efforts to examine this duality of mind were frustrated because researchers lacked adequate methods. However, in the 1970s cognitive psychology made it possible to explore the influence of unintentional components of human reasoning, distinct from the intentional components (Devine & Monteith, 1999).

In the 1970s, cognitive psychology made it theoretically and methodologically possible to independently explore the influence of intentional (conscious) and unintentional (unconscious) components of human thought and behavior. In a series of experiments, Schneider and Shiffrin (1977) distinguished between automatic and controlled processes and proposed a two-process theory of human information processing in detection, search, and attention phenomena. Automatic processing relies on activation of long-term memory. Controlled processing is a temporary activation of sequential elements that need attention that has limited capacity and is under the control of perceivers. These two processing modes are qualitatively different activities. To demonstrate the differences, the researchers conducted a serious of complex studies using both reaction time and accuracy measures. In one experiment, participants looked at an array of stimuli while searching for a particular target stimulus (e.g., the letter "D") as rapidly as possible. Participants practiced this task many times, and as expected, they spent more time trying to identify target stimuli when there were more distracter stimuli. However, that effect decreased over time. After a period of practice, finding the target became an automatic process, whereby the number of distracter stimuli (4, 9, or 16) showed no effect on participants' speed in detecting targets. Shiffrin and Schneider (1977) theorized that participants began the task by devoting

attention to each stimulus separately, but with practice, attention became parallel and automatic. Participants could attend to multiple stimuli simultaneously. This distinction between automatic and controlled processing in attention and the methods used to dissociate the separate systems provided an approach for a highly influential theory of automaticity in social judgment (e.g., Bargh, 1994; Devine, 1989; Higgins, Rholes, & Hones, 1977).

Another important finding from cognitive psychology, concerns multiple memory systems. In the 1960s, Reber postulated an implicit learning model involving the "acquisition of knowledge that takes place largely independently of conscious attempts to learn and largely in the absence of explicit knowledge about what was acquired" (Reber, 1993, p.5). To demonstrate implicit learning, Reber invented the artificial grammar-learning paradigm, which demonstrated the unconscious acquisition of knowledge and the use of that knowledge without consciously activating it. In the first stage of Reber's experiments, participants received letter strings to memorize. The strings were the product of a set of complex rules (artificial grammars) unknown to participants. In the second stage, participants learned of the existence of the complex rules, and the researchers asked them to report whether new strings did or did not conform to those rules. Even though participants could not articulate the rules, they recognized correct strings at a rate better than chance. Reber interpreted this result to mean that participants had implicitly extracted rules and argued that they did so without conscious effort or awareness. He coined the term "cognitive unconscious" (Reber, 1989, p.230) to refer to the idea that a large number of cognitive processes can occur outside of consciousness.

This notion of implicit learning separates implicit and explicit knowledge. Reber not only provided evidence for distinct implicit and explicit learning processes, but he also provided a theoretical foundation for the generic dual system theory of cognition. His work was a major

influence on dual system theory, particularly regarding the development of dual process accounts of deductive reasoning (Evans, 2009). Reber (1993) made a number of claims about the nature of these implicit learning and memory systems; including an argument, that implicit function had low variability across individuals and was independent of general intelligence.

Another important influence from cognitive psychology is the model of associative networks in memory. For decades, cognitive psychologists used priming experiments to explore the structure of knowledge (e.g., Tulving, Schacter, & Stark, 1982) because priming associative networks isolates automaticity and implicit memory (see Schacter, 1987). Cognitive and social psychologists have primed networks by asking participants to complete unrelated prior tasks, which include content that activates the relevant unconscious knowledge structures, such as implicit stereotypes and attitudes. In the general paradigm, researchers observe subsequent behavior to determine if the primes influence performance on the main task of social judgment or perception. For example, Bargh, Chen, and Burrows (1996) primed participants with stereotype activating cues of elderly people (e.g., forgetful, wrinkles) and found that these participants walked more slowly through a hallway than did those primed with neutral words. Adopting this method, social psychologists were able to demonstrate the existence of social memories, stereotypes, and attitudes without explicit measurements. Another method favored by social psychologists is to measure correlations between social judgments and individual differences in thinking style, such as need for cognition (Cacioppo, & Petty, 1986) or the Rational-Experiential Inventory (Pacini, & Epstein, 1999).

Dual process theories of social cognition originated in the 1980s and have shaped and guided much of the research in persuasion (e.g., Chaiken, 1987), prejudice and stereotyping (e.g., Devine, 1989), impression formation (e.g., Brewer, 1998), and dispositional attribution (e.g.,

Trope, 1986). These dual process theories of social cognition come under many labels, but they have a shared premise: thinking proceeds in one of two modes. Intuition is automatic, effortless, associative, holistic, and less accurate. Rational thought is conscious, effortful, rule-based, analytic, and more accurate.

Perhaps, the most well-known of these theories is the heuristic-systematic model (HSM: Chaiken, 1980; Chen & Chaiken, 1999), which is highly applicable to jury decision-making (Groscup & Tallon, 2009). The heuristic-systematic model assumes two qualitatively different information-processing modes. The first mode, systematic processing, is relatively slow, controlled, and effortful. It is analytic and as such includes a comprehensive treatment of judgment-relevant information. Systematic processing produces judgments that are responsive to the actual content of information. The second mode, heuristic processing, is relatively quick, automatic, and nearly effortless. Heuristic processing entails the activation and application of simple and easily enacted judgmental rules or "heuristics." Thus, heuristic processing results in judgments that rely on easily processed and generally relevant cues. For example, the heuristic "experts are usually correct" can lead people to accept the validity of a message based simply on the message source (e.g., an "expert"), rather than by independently assessing the support for that message (Chaiken et al., 1989).

Chaiken (1989) introduced two principles to determine whether a decision maker engages in heuristic processing or systematic processing. The least-effort principle follows from the observation that it is impossible and undesirable for individuals to devote all cognitive resources to one task (Moskowitz, 2005). Therefore, they must allocate their limited cognitive resources in the most efficient way and fast manner. Fiske and Taylor (1984) coined the phrase "cognitive miser" to describe this process. Rather than evaluating information rationally and objectively, the

cognitive miser assigns new information to categories that are easy to process. These categories come from prior information, preexisting scripts, and other knowledge structures that are stored in memory.

The second principle emerges from individuals' natural "accuracy motivation" (Chaiken, 1980). Chaiken argues that people assess the validity of persuasive messages in a manner that is consistent with their own beliefs about the importance of collecting valid and accurate information. Given that decision makers have both limited cognitive resources and accuracy motivation, they must compromise to function under these competing psychological demands. Chaiken and colleagues extended their research to specify the psychological conditions that trigger either systematic or heuristic modes of processing based on the discrepancy between participants' desired level of certainty and their actual certainty. The resulting "sufficiency principle" (Chen & Chaiken, 1999, p.74) asserts that individuals are sometimes motivated to exert additional cognitive effort to reach a certain level of judgmental confidence. They balance their preference for maximizing cognitive efficiency with accuracy motivation. This balancing point is the sufficiency threshold. If the certainty threshold is high and heuristic processing does not close the gap between actual and desired confidence, people will engage in systematic processing. Systematic processing requires more cognitive effort than heuristic processing, similar to controlled processing in cognitive psychology literature (Posner & Snyder, 1975 cited in Chaiken et al., 1989). Alternatively, if the gap between actual and desired confidence is acceptable, individuals may default to heuristic processing, which makes minimal cognitive demands on the perceiver. Chaiken and her colleagues conducted studies in which they manipulated motivation and observed behavior to indirectly measure the sufficiency threshold.

Dual processing models and subjective probability estimation

Dual processing models might explain jurors' different use of probability information. Kirkpatrick and Epstein (1992) demonstrated that people who estimate objective probability (i.e., calibrate their subjective probability estimates to an objective criterion) do so differently when in a rational processing mode (i.e., relying on controlled analytic thought), as compared to when in an experiential processing mode (i.e., relying on emotional and heuristic thought). To explain this finding, Epstein and colleagues proposed the Cognitive-Experiential Self-Theory (CEST: Epstein, 1994; Pacini & Epstein, 1999). CEST is slightly different from the Heuristic-systematic model, but the basic assumptions are similar. From the CEST perspective, people adapt to their environment through two different information-processing systems: the experiential system and the rational system. The rational system is effortful, analytical, and logical. The experiential system is the default processing mode. It is automatic and is dependent upon affect. Rational processing cannot occur unless there is a direct appeal to the rational system that emphasizes the importance of effortful, logical processing.

Experimental research supports the CEST model. For example, Kirkpatrick and Epstein (1992) presented participants the opportunity to win money by drawing a red jellybean from one of two bowls. One bowl had one red jellybean among 20 beans, and the other bowl contained 10 red jellybeans among 200 beans. Although both bowls contained 5% red beans, more than two thirds of participants chose to draw from larger number of beans (*i.e.*, 10 red jellybeans in 200 beans). Moreover, most participants were willing to spend money for an opportunity to select their choice, rather than accept a random choice. They commented that they knew their behavior was irrational, but they felt they had a better chance when the absolute number of red jellybeans was greater. Kirkpatrick and Epstein (1992) proposed CEST's concrete and experiential principles to explain these results. In the experiential system, individuals encode information in

the form of concrete representations. Because the larger absolute number is more concrete than the ratio information, people prefer the larger number lottery, even though they know the objective probability between two lotteries is identical. Moreover, in the experiential system, people use experiences like this to construct a schema that includes emotionally significant information.

What triggers type of processing

What factors influence an individual's motivation or ability to process systematically? The following section will review the likely candidates: motivation, individual differences, and emotion.

Motivation. Early research by Chaiken and colleagues focused on personal involvement that can motivate individuals to make judgments that are more accurate. In one study, Chaiken (1980) had participants read a persuasive message from a likable or unlikable target who presented either six or two arguments about one of two topics. In the high involvement condition, participants learned that they would be discussing that message with others at a future time. In the low involvement condition, participants anticipated discussing a different topic. Participants in the high involvement condition showed more opinion change when they received six arguments, as compared to two arguments. However, in the low involvement condition, the likability of the communicator, not the number of arguments, predicted opinion changes. The researcher interpreted these results to conclude that the impact of communicator likability reflects participants' use of a simple judgment heuristic, namely that likeable communicators are reliable. Chaiken (1980) concluded that in the high involvement condition participants made use of systematic processing of judgment-relevant information (e.g., the number of pieces of

confirming information) and participants in the low involvement conditions made use of heuristic processing (in the form of the rule: likeable communicators are reliable).

Individual differences. Another motivational factor known to influence the systematic processing of persuasive argument is individuals' need for cognition (NFC, Cacioppo & Petty, 1982). NFC refers to the tendency for individuals to vary the extent to which they engage in or enjoy effortful cognitive activities. Some individuals prefer saving their cognitive effort, whereas other individuals engage and enjoy cognitive challenging activities. People low in NFC may use cognitive heuristics more frequently than those high in NFC (Chaiken et al., 1988).

Elaborating on the NFC construct, Epstein and colleagues developed a self-report instrument, the Rational-Experiential Inventory (REI: Epstein, Pacini, Denes-Raj, & Heier, 1996), which measures individual differences in rational and experiential thinking styles. The REI is composed of 10 items within four REI subscales, which researchers have shown to predict self-reported gambling in real life (Denes-Raj & Epstein, 1994), heuristic responses to vignettes (Denes-Raj & Epstein, 1994), and depression (Pacini, Muir, & Epstein, 1998). The four subscales are rational ability, rational engagement, experiential ability, and experiential engagement. Rational ability refers to a high level of ability to think logically. Rational engagement refers to individuals' reliance on an analytic, logical manner of thought. Experiential ability refers to individuals' intuitive impressions and feelings. Experiential engagement refers to individuals' reliance on feelings and intuitions in making decisions. Epstein et al. (1996) found that the relation between REI and various vignettes suggested that people with high experiential scores responded more heuristically and individuals with high rationality scores responded less heuristically.

One might expect that individuals who rely on the rational system would not also rely on the experiential system. However, empirical evidence suggests that the rational system and experiential system are uncorrelated. In other words, individuals high in NFC make variable use of both their rational and experiential systems (Epstein, 1996).

Emotion. A growing body of research has demonstrated that transient emotions, often referred to as mood, can affect judgments in dual processing models. Some moods or emotions (e.g., anger, disgust, happiness) are typically associated with a greater sense of subjective certainty, while others (e.g., hope, anxiety, fear, some forms of sadness) are typically associated with uncertainty. Furthermore, the more certain people feel, the less likely they are to use systematic processing. They are confident enough without relying on the careful deliberations that result from systematic processing (Tiedens & Linton, 2001). In the criminal legal context, some crimes can make jurors angry and likely to feel more certain even before they see any evidence (Bandes, 2009; Sunstein, 2002; Georges, Wiener, and Keller, 2013).

If a case triggers a high certainty emotion (e.g. anger) or automatically invokes high levels of moral culpability, jurors might begin their deliberations with enough certainty to rely on heuristic or experiential processing to exclusion of rational thought. In these cases, jurors may ignore the objective probability of evidence but instead exhibit a coherence shift succumbing to the need for cognitive consistency and accept lower standards of proof. Conversely, if experiencing an uncertainty emotion (e.g. sadness), jurors are likely to seek more confidence in their judgments through rational or systematic processing. Systematic or rational processing should lead jurors to focus more on probability information derived from evidence and less on their feelings of cognitive consistency (Tiedens & Linton, 2001). As a result, jurors should avoid

a coherence shift and instead focus on understanding the legal standards of proof and probability information in the evidence.

For example, Bodenhausen, Sheppard, and Kramer (1994) conducted three experiments to investigate anger's impact on culpability judgments. In the first study, they induced college students to experience neutral, sad, or angry states by asking them to write autobiographical statements regarding neutral, sad, or angry events they previously experienced. Participants then acted as members of a judicial review board evaluating either an assault or an academic dishonesty case. To measure heuristic processing, the authors manipulated the group to which the defendant belonged. For the assault case, the defendant was White or Hispanic and for the academic misconduct case, the defendant was a nondescript student or an athlete. The researchers assumed that people hold stereotypes of Hispanic defendants in criminal cases and stereotypes of athletes in academic misconduct cases. Results revealed that the stereotype cue manipulation did not affect neutral or sad participants' judgments but angry participants relied more on heuristic, stereotype based judgments. They viewed the Hispanic assailant as more guilty and the athlete as more dishonest.

Semmler and Brewer (2002) similarly examined whether emotions induce decision makers to engage in deeper cognitive processing. The researchers manipulated emotions by inducing participants to feel sad or neutral emotions through witness statements, which did or did not provide vivid details about harm in a mock criminal trial. Then, one group of participants received consistent testimony and the other group received inconsistent testimony. The researchers hypothesized that sadness would induce participant to use systematic processing, so that sad participants would report more testimonial inconsistencies. Results suggested that sad moods resulted in more accurate reporting of testimonial inconsistencies than did neutral moods.

However, emotion induction is a difficult psychological process to manipulate, especially in the context of long and complicated task materials like criminal cases. To induce perceivers to feel a desired mood state, many researchers ask participants to describe in writing recent incidents that made them feel sad, fear, anger, or happy (C. A. Smith & Ellsworth, 1985). Another method is to give all participants the same scenario and then ask them what features of the case made them feel angry or sad (Lerner, & Keltner, 2001). An additional problem arises when using emotion induction for online studies because in web-based experiments, researchers have less control over participants and their environments, which can result in smaller effects than similar experiments conducted in a lab (Birnbaum, 2004). Another problem is that technical and environmental variance is higher in online studies because participants' hardware and software differ widely. On the other hand, online studies can be beneficial because participants usually perform tasks in familiar environments and demand effects may be lower.

While researchers have developed many emotion induction techniques for face-to-face experiments, few exist for online use (Goritz & Moser, 2006; Goritz, 2007, Verheyen & Goritz, 2009). Nonetheless, Goritz and colleagues have shown autobiographical recall, mood-suggestive photography (Goritz & Moser, 2006) cartoons, jokes, emotion-laden words, emotive texts (Goritz, 2007), and plain texts (Verheyen & Goritz, 2009) to successfully induce emotion online. They found that online experiments can induce negative mood more effectively than positive mood, and mood induction by plain texts is more promising than by photography. Based on these findings, negative emotion seems more suitable for online studies. It follows that negative emotions with different levels of invoked certainty, such as anger and sadness, are most suitable for this project, which intends to test the coherence shift as a way of understanding how jurors understand different standards of proof.

Trial characteristics & jurors' heuristic processing

Severity of charge. Rational decision makers are mindful of the consequences of their decisions. When a charge is seriousness, there is a high likelihood that a victim suffered a severe injury and that the defendant will, in turn, suffer a severe punishment or pay high damages. Some empirical studies support the notion that jurors apply standards of proof differently depending upon severity of charge. Based on previous research on dual process models, a possible severe penalty for conviction should motivate jurors to be accurate (Chaiken, 1992) and use systematic processing to reach criminal or civil verdicts. Because the consequences of more severe criminal charges are greater than less severe charges, jurors are more likely to be motivated to process accurately and therefore systematically when the charges are more severe. Empirical evidence suggests that jurors tend to be less susceptible to heuristic evidence when they decide crimes that are more serious.

For example, Rind, Jaeger, and Strohmetz (1995) presented low, intermediate, and high seriousness crimes (vandalism, arson, or murder) to undergraduate mock jurors. Half of the participants encountered ambiguous evidence, and the other half encountered inadmissible evidence (i.e., an illegal wiretap on the defendant). The researchers found that mock jurors were less likely to rely on inadmissible evidence when the crime was serious. Interestingly, crime-seriousness predicted guilty ratings when evidence was ambiguous, but not when inadmissible evidence was included. Essentially, participants used inadmissible evidence for non-serious crimes but not for serious crimes.

Similarly, jurors might interpret standards of proof to set higher thresholds for more severe crimes or penalties. In a classic study, R. J. Simon (1969) asked a sample of state and federal trial court judges about the probability of finding guilty defendants accused of murder,

embezzlement, grand larceny, rape, petty larceny, and fraud. Judges assigned higher thresholds to more severe crimes as opposed to less sever crimes. Furthermore, Kerr (1978) asked participants to read a trial summary describing a case of first-degree murder, second degree murder, or manslaughter. Then, they read the penalty that the defendant would receive on conviction: a moderate (up to 5 years) or severe penalty (25 years to life or capital punishment). Participants were more likely to convict defendants accused of second-degree murder than first-degree murder. In addition, participants' conviction threshold was higher when the penalty was severe, as compared to mild.

Although the literature suggests that actual judges set higher thresholds for more severe crimes and mock jurors consider more evidence (Freedman, Krismer, MacDonald, & Cunningham, 1994), it remains unclear whether jurors will seek higher probabilistic judgments or alternatively rely more on subjective certainty and follow a cognitive coherence model of judgment. That is, jurors faced with more severe charges may consider information more carefully, as the law requires. Alternatively, they may rely on higher subjective confidence, based on increased cognitive coherence processing. The latter should result in jurors distorting evidence to fit their higher confidence levels.

Question order. In 2000, Kahneman proposed that heuristic responses flow from the impression or "natural assessment" that people automatically draw about the objects of perception and the objects of thought (Kahneman, 2000, p. 701). Recent research by Inbar, Cone, and Gilovich (2010) indicates that the characteristics of a task can trigger systematic processing and rational thought. These authors collected data to show that "task-cuing" influences individuals' selection of processing style. In the first study, researchers showed to one group of participants 25 choice dilemmas, ranging from the trivial ("selecting an entrée") to the

consequential ("choosing a college to attend"). Then, the student participants rated whether they used "intuition" or "rational analysis." A separate group of students indicated the extent to which they could objectively quantify each dilemma. The results showed that when students rated a choice as objectively quantifiable, participants in the first group were more likely to select a rational approach to the dilemma.

In a second study, participants evaluated 40 consumer products ranging from quite complex (e.g., "computer," "car") to moderately complex (e.g., "curtains," "dress") to relatively simple (e.g., "brush," "toothpaste"). When encountering more complex choices, participants reported they were using reason more than intuition. A third study was similar to Epstein's (1991) earlier work. Participants selected one of two jars. One jar contained one blue and nine yellow pieces. Another jar contained nine blue and ninety-one yellow pieces. If they drew a blue piece from the selected jar, participants could receive a freshly baked cookie. Participants' only task was selecting one of the two jars. In a non-sequential, holistic condition, they would win if they drew a blue piece from a selected jar. However, in the sequential condition, they needed to win a series of tasks to receive the cookie. That is even if they drew a blue piece from the selected jar, they needed to also solve an anagram word task and Chinese Tangram in order to receive cookies. Researchers assumed that participants who use rational decision processes should choose the jar that contained one blue and nine yellow pieces because, regardless of condition, selecting that jar increased the probability to win the prize. However, in the holistic condition, participants selected the jar that included more pieces. In the sequential condition, participants selected the jar with ten total pieces. Inbar and colleagues interpreted these results to conclude that the decision process with sequential tasks induced participants to use more serial-analytic processing so that they used systematic or rational thought to select the optimal task.

Thus, "task cueing" not only influences people's thoughts about proper decision strategies, but it also determines how they resolve conflicts between intuition and reason. Inbar and colleagues' third experiment is similar to Semmler and Brewer's (2002) study that showed how mock jurors who referred to a flow chart while evaluating a case scenario comprehended and applied judicial instructions better than those who used only traditional instructions. Wiener and colleagues (2006) similarly demonstrated that, even in the highly emotional context of capital murder trials, jurors provided with flow chart instructions are better able to understand the complicated penalty scheme conveyed in jury instructions. Furthermore, there is a possibility that flow chart instructions not only aid understanding of the law but also can change the process that jurors use to make decisions.

In this dissertation, I examined whether the order of questionnaires might also induce mock jurors to follow either deliberate processing or heuristic processing. The main dependent variables in many legal decision-making tasks are verdicts, measures of subjective certainty, and measures of the required level of certainty based on standards of proof (i.e., Simon, Snow, & Read, 2004). Notably, the order of these dependent variables is almost always the same. Verdict is the first dependent variable. Other measurements then investigate the decision-making processes used to reach a verdict. This is a reasonable sequence of measurements since the main interest of legal decision-making studies are often the verdict. However, according to the task cuing account (Inbar, et al., 2010), the sequence of these measurements might induce intuitive information processing, such as a coherence shift. If participants are first asked to offer a verdict (verdict-first questionnaires), then they may use their verdicts as an anchor and rate other measurements to be coherent with their verdicts. In contrast, if participants first rate their subjective probability of guilt (probability-first questionnaires) and then rate standards of proof

and guilty verdict, they may use a more rational process. If a flow chart can guide mock jurors' decision processes (Wiener, et al., 2004), the sequence of questionnaires may also guide their decision processes.

Appropriate use of standards of proof

As described above, there are two kinds of probability judgments: objective probability and intuitive-subjective certainty. Objective probability consists of a numerical estimation of objective likelihood calculated by normative or mathematical rules, such as Bayesian calculation. Subjective probability certainty is an intuitive estimation of likelihood that people make without recourse to an objective mathematical scale (see Windschitl & Wells, 1996). Subjective certainty often reflects objective probability, but some research methodology can separate the two (Wells, 1992). The law is silent concerning which type of probability (objective or subjective) triers of fact should use. In fact, judges following the rules of evidence are not likely to accept a trial arguments based solely upon statistics and objective probability judgments (Kaye, 1999; Koehler, 2001). Ironically, judges may ask jurors to update judgments based upon new information (in a manner that Bayesian statisticians would recognize), but then ask jurors to decide cases based on their subjective likelihood judgments (Koehler, 2001).

This dissertation does not argue that one of these probability judgments is correct. The purpose of this research is not to identify a "correct" or "accurate" judgment method. Instead, this dissertation focuses on discovering how jurors use two different probability judgments and standards of proof. When jurors engage in systematic processing, they should be more likely to consider objective probability information. They should be more likely to use standards of proof, as the law requires because the law assumes jurors are rational decision makers. However, when jurors engage in heuristic processing, that is, they use associative and intuitive processing, such

as the cognitive coherence process; they may not use probability information correctly, at least not in the way that the law intends them to use that information.

Chapter 5. Research Project & Hypotheses

Overview

Theory. The law uses different standards of proof to explain to jurors which levels of certainty are appropriate for various criminal and civil charges. The criminal standard ("beyond a reasonable doubt") is the highest level of certainty, and civil standards ("clear and convincing evidence" or "preponderance of evidence") suggest lower and graduated levels of certainty. The law requires and assumes that jurors will make rational decisions when applying these standards of proof. Jurors ought not to rely on other, non-rational factors such as emotions. Moreover, some legal commentators and social scientists assume that jurors can apply standards of proof consistently, regardless of the types of cases jurors evaluate.

In contrast to these assumptions, social psychology's dual process theories suggest that jurors might use two distinctly different processing styles. When jurors hold expectations about a case or when they are not highly motivated, jurors may employ heuristic processing leading them to fall victim to cognitive consistency demands and the coherence shift. Conversely, when jurors are motivated to make accurate judgments or when they have no available or accessible heuristics, they are more likely to use systematic processing. Moreover, emotions or severity of charges can trigger either heuristic or systematic processing.

While they are engaged in heuristic processing, jurors are most likely to use subjective probability to guide their decisions. This type of probability judgment distorts the meaning of standards of proof. In contrast, systematic processing allows jurors to use objective probability, which increases the likelihood that they will differentiate between different standards of proof.

Therefore, the goal of these experiments is to explore a hierarchical judgment model in which jurors' transient emotions and the severity of charges influence their understanding of standards of proof, which ultimately shapes their final verdicts. Study 1 investigated whether, and under what conditions, mock jurors use subjective probability to make judgments. Study 2 investigated whether the severity of the charge influences mock jurors' use of subjective probabilities.

Overview of experiments 1 and 2. The purpose of study 1 was to test whether experienced emotion could change mock jurors' decision styles and their use of the standard of proof. To control the influence of specific emotions, this study manipulated emotions instead of measuring naturally occurring emotions. Study 1 invoked the emotions of sadness and anger because both are negative emotions that jurors might experience in real courtrooms and there is a history of social psychological studies to compare processing under anger and sadness. Both anger and sadness are similar in negative valence, though anger is associated with high levels of certainty and individual control (Smith & Ellsworth, 1985). In contrast, sadness is associated with uncertainty and situational control. According to Tiedens and Linton (2001), when people feel sad, they are uncertain but when they feel anger, they are certain. Therefore, if people experience sadness and uncertainty, they will process information more systematically, and focus more on probability information. They will do this in order to gain a sufficient level of certainty (Chen & Chaiken, 1999). On the other hand, if they experience anger and certainty, they will feel a stronger sense of certainty initially and process more heuristically. Moreover, negative emotions are better suited for internet studies, especially those that involve criminal and civil wrongdoings, which naturally induce negative valence rather than positive valence emotions.

Another purpose of study 1 was to test the psychological mechanism that guides the use of standards of proof. A substantial number of experimental jury decision making studies have

investigated this topic, but only a few have focused on the role of probability judgments and the applicable standards of proof (Engel & Glockner, 2013; D. Simon, Snow, & Read, 2004). Rational (mathematical) models assume higher standards of proof increase decisions thresholds only and should not influence assessments of evidence. In contrast, cognitive coherence models assume higher standards of proof simply increase subjective certainty through cognitive shift. No experiments have examined how transient emotions or the severity of a charge may facilitate heuristic processing and variable applications of standards of proof. Researchers have only recently begun to examine the effects of emotion in legal decision making (Feigenson, 2009; Wiener, Bornstein, & Voss, 2006; Georges & Wiener, 2013; Wiener, Gervais, Allen, & Marquez, 2013).

To manipulate standards of proof while controlling evidence and procedures, both experiments presented the decision scenarios as part of an arbitration process, for several reasons. First, there is no objective and clear standard in arbitration practice. Employee theft is unauthorized taking, control, or transfer of money or property belonging to the employer or to a fellow employee. Among labor arbitrators, employee theft is one of the few offenses for which summary discharge is appropriate. Discharge by theft is tantamount to industrial capital punishment and is severely punitive. Therefore, the employer carries a heavy burden in supporting such action. In cases involving employee theft, some arbitrators require clear and convincing evidence of an employee's guilt, while other hold that proof must be beyond a reasonable doubt (Scheinman, 1977).

Second, in both study 1 and study 2, I used the same scenario ('Jason Wells') that Simon and colleagues presented to participants. Simon and colleagues have used the arbitration scenario repeatedly (D. Simon, Snow, & Read, 2004). They report robust results in which decision

makers follow the procedure quite well. Engel and Glockner replicated Simon and colleagues' results (Engel & Glockner, 2013).

The purpose of study 2 was to test whether severity of charge can influence mock jurors' decision styles. Instead of manipulating emotion and standards of proof, study 2 manipulated the type of charge (theft v. murder) and the order of questionnaires (verdict first v. questionnaires first). Contrary to study 1, study 2 used only a criminal case and standard of proof (i.e., beyond a reasonable doubt). The manipulation of standard of proof was inappropriate because murder was a criminal charge more serious than theft, and unlike theft, murder could not plausibly be subject to arbitration. While study 1 investigated the role of standard of proof directly by manipulating standards, study 2 assessed the role of the criminal standard of proof indirectly, through verdicts.

Before completing the main portions of study 1 and study 2, participants completed the Rational-Experiential Inventory questionnaire assessing their individual differences in rational and experiential thinking styles (REI: Epstein, Pacini, Denes-Raj, & Heier, 1996; Epstein, 1998). For the main portion of study 1, participants completed an emotion manipulation, following which they completed a short appraisal questionnaire and an emotion manipulation check. Then, participants read four apparently unrelated vignettes. Following each vignette, participants read two or four inferences about some of the facts that made up the vignette and rated their agreement with each inference. Next, they read a written summary of the Jason Wells case, which involves an individual accused of stealing money from the company safe.

Participants played the role of an arbiter whose task is to decide culpability in a disciplinary proceeding that the employer brought against Jason Wells. Participants received some general background information about Jason, followed by the evidence and arguments presented by lawyers from both parties. Participants read pieces of evidence indicative of

medium (55%) or high (99%) probability of culpability, a manipulation of the strength of the evidence. The pieces of evidence were unrelated to each other and their likelihoods were numerically defined. Bayesian computation with the numeric values of evidence defined the "posterior likelihood of guilt."[2] Then, participants decided whether Jason Wells stole the money, using one of two standards of proof: preponderance of evidence or beyond a reasonable doubt. Next, they reassessed the inferences offered in the pretest using almost identical questions. Finally, the participants answered a questionnaire about how they reached their decisions. The resulting design was a 2 (emotion: anger v. sadness) x 2 (standard of proof: preponderance of evidence v. beyond a reasonable doubt) x 2 (Strength of evidence: medium, 55% v. high, 99%) between subjects design.

For the main portion of study 2, participants similarly read four apparently unrelated vignettes, each followed by either two or four inferences about the facts in the vignettes. Participants rated their agreement with each inference. Then, they read a written trial summary of the same Jason Wells case involving in which the state charges Jason Wells with employer theft. Half of the participants received a version with only a theft charge, and the other half received a version in which the culprit brutally murdered a guard during the theft. Then, participants played the role of a juror whose task was to reach a verdict for the criminal charges. Participants received some general background information about Jason, following the evidence and arguments presented by lawyers from both parties. Participants again read pieces of evidence indicative of medium (55%) or high (99%) probability of culpability, again a manipulation of the strength of the evidence. As in study 1, the pieces of evidence were unrelated to each other and

[2] The posterior likelihood of guilty will be calculated by this equation (Glockner & Engel, 2013):

$$p(g|ev_1 \cap ev_2 \cap ev_3 \cap ev_4 \cap ev_5)=\frac{p(ev_1|g)*p(ev_2|g)*p(ev_3|g)*p(ev_4|g)*p(ev_5|g)}{p(ev_1|g)*p(ev_2|g)*p(ev_3|g)*p(ev_4|g)*p(ev_5|g)+p(ev_1|i)*p(ev_2|i)*p(ev_3|i)*p(ev_4|i)*p(ev_5|i)}$$

Where g stands for "guilty", i stands for "innocent", and ev_i stands for the quantifiable pieces of evidence.

numerically defined and as in study 1, with Bayesian computation such that the numeric values of evidence produced the "posterior likelihood of guilt." Next, half of the participants decided a verdict regarding whether Jason Wells stole the money, applying the beyond a reasonable doubt standard. Then, they answered almost identical questions as those in the pretest. The other half of participants assessed the pre-test questions first and then decided a verdict. Finally, the participants answered a questionnaire about how they reached their decisions. The resulting design was a 2 (type of case: theft v. murder) x 2 (probability of culpability: Medium v. High) x 2 (order of questionnaires: verdict- first v. Probability- first) between subjects design.

The rational experiential inventory was included as a covariate because substantial research has shown how individual differences in personality, processing style, and cognitive ability may affect juror performance (Epstein, 2003; Denes-Raj & Epstein, 1994). Several inventories measure individual differences but because the REI includes the need for cognition and the focus of this research is on the decision making style this study used the REI.

Hypotheses

Study 1 investigated whether emotions and higher standards of proof can influence jurors' use of either systematic or intuitive processing.

H1: There should be an interaction of emotion and standards of proof on verdicts. Among participants who experience anger, differences in the standards of proof should have a smaller effect on verdicts (because participants should use intuitive processing when experiencing anger). In contrast, among participants who experience sadness, differences in the standards of proof should have a stronger effect on verdicts (because participants should use systematic processing when experiencing sadness).

H2: There should be an interaction of emotion and standards of proof on jurors' re-evaluation of evidence. Participants who experience anger should demonstrate more coherent re-evaluation of evidence when those participants are required to apply a higher standard of proof (i.e., beyond a reasonable doubt), as a result of using intuitive processing. Conversely, participants who experience sadness should demonstrate less coherent re-evaluation of evidence when required to apply a higher standard of proof (i.e., beyond a reasonable doubt), as a result of using systematic processing.

H3: There should be an interaction of emotion and the strength of evidence on verdict. Among participants who experience anger, strength of evidence should be less influential on participants' verdicts. In contrast, among participants who experience sadness, strength of evidence should be more influential on participants' verdicts.

Study 2 will investigate whether the severity of charge and the order of questionnaires can influence jurors' use of systematic or intuitive processing.

H4: There should a main effect for severity of charge on verdict. Participants faced with a less serious charge should be more likely to convict than participants faced with a charge that is more serious. Because the consequence (liability or penalty) following less serious charges are less severe, participants confronted with less severe charges should be less motivated to decide cases accurately. Participants should maintain lower required levels of guilty and should be more likely to convict. In contrast, participants who decide severe charges should be more motivated to decide cases accurately. Participants should maintain higher required levels of conviction and should be less likely to convict.

H5: There should be a main effect for the severity of the charge on jurors' re-evaluation of evidence. Participants faced with less serious charges should demonstrate more coherent re-

evaluation of evidence. Because the consequences (liability or penalty) following less serious charges are less severe, participants confronted with a less severe charge should be less motivated to decide cases accurately. Participants should rely on heuristic processing and show more coherence shift toward their decisions. In contrast, participants who decide a more severe charge should demonstrate less coherent re-evaluation of evidence. The severe charge should motivate participants to decide cases more accurately. Participants experiencing such accuracy motivation should use systematic processing,

H6: There should be a main effect for order of questionnaires on the re-evaluation of evidence. Participants who first decide the verdict should be more likely to re-evaluate evidence to increase coherence with their verdict. Similar to the Inbar et al. (2010) third experiment, when participants decide the verdict first in the current experiment, doing so should cue intuitive decision processes. These participants should seek subjective certainty through a coherence-shift rather than accurate evaluation of evidence. In contrast, participants who first re-evaluate evidence should be less likely to increase coherence to their verdict. When participants assess evidence and then decide a verdict, doing so should cue the rational choice system so that those participants should seek accurate evaluations of evidence rather than subjective certainty.

H7: There should be an interaction between severity of charge and strength of evidence on verdict. Participants faced with less serious charges should show lowered accuracy motivation, more heuristic thinking and be less likely to rely on strength of evidence because less severe penalties should induce participants to use intuitive processing. In contrast, participants faced with severe charges should be more likely to rely on strength of evidence because greater penalties should increase accuracy motivation.

H8: There should be an interaction between the order of questionnaires and strength of evidence on verdict. Participants who first decide the verdict first should be less likely to use strength of evidence in verdict decisions. Based on the task-cueing literature, participants who decide the verdict first should rely more on their intuitive processing and thus be less sensitive to strength of evidence. In contrast, participants who first assess the evidence should be more likely to use the strength of evidence in verdict decisions.

Chapter 6: Study 1

Overview

The experimental design was a 2 (Emotion: Anger v. Sad) x 2 (Standards of proof: preponderance of evidence v. beyond a reasonable doubt) x 2 (strength of evidence: 55%, 99%) between subjects with one individual-difference measure (REI).

Method

Participants. I recruited participants from the Amazon Mechanical Turk website. This site posted recruitment advertisements to individuals who agreed to participate in web-based research studies. Participants received a three-dollar gift certificate to Amazon.com for their participation.

Participants included 73 males and 108 females from across the United States (and one participant did not indicate his or her gender). The mean age was 34.32 years old, ranging from 18 to 64 years old. The majority of participants indicated that they were Caucasian (83.5%). 10 participants indicated they were Asian American (5.5 %), 8 African American (4.4 %), 6 Hispanic (3.3 %), 1 Latin American (0.5 %), 2 Native American (1.1 %), and 3 indicated that their ethnicity was “other”(1.6%). One participant had less than a high school education (0.6%). 52 indicated a high school level education (28.6%), 34 had an associate’s degree (18.7 %), 70

had a bachelor's degree (38.5 %), 20 had a master's degree (11.0 %), and 5 had a professional degree (2.7 %). The majority of participants were employed full time (N = 83, 45.6 %), though 47 were employed part-time (25.8 %) and 52 were unemployed (28.6 %).

Data cleaning. To be jury-eligible in most states, citizens must be at least 18 years old and have no felony convictions. I removed one participant who was not a U.S. citizen and four other participants who indicated that they were convicted felons without civil rights.

In the third part of the study, three participants realized that the emotion manipulation was related to the juror decision making task; however, none indicated that they knew how or why emotions could affect decision making. Therefore, no participants were excluded based on their responses to the suspicion-check questions.

The final data set consisted of 72 jury-eligible men, 104 jury-eligible women, and one jury-eligible participant who did not specify his or her gender. These participants were randomly assigned to one of the eight experimental conditions. Table 1 displays participants' demographic characteristics after data cleaning.

Procedure.

When participants selected this study from Mechanical Turk and finished the Rational-Experiential Inventory (Appendix A), they received email invitations to participate in the main study, which included a URL address to gain access to the website. Participants who chose to participate completed a standard informed consent form (Appendix B). They learned that they would participate in three short unrelated studies, one concerning emotional memories, one assessing social issues, and one concerning a hypothetical legal case.

Participants then completed an emotion manipulation survey that asked them to fill out a self-generated memory task (Appendix C). All participants wrote three autobiographical sketches

about events that had happened to them in the past (Tiedens & Linton, 2001). The instructions asked participants in the angry condition to "remember, relieve, and vividly recall" three negative events that made them angry. Participants in the sad condition did the same for three sad memories. Although the emotion manipulation occurred before the second study, prior research has shown that the effect of certainty appraisal emotions carry over and influence unrelated judgments and decisions in subsequent tasks (Lerner & Tiedens, 2006).

After completing the emotion manipulation, all participants completed a shortened version of Smith and Ellsworth's (1985) emotion manipulation check (Appendix D) and the appraisal questionnaire (Appendix E). The appraisal questionnaire includes six 9-point Likert scaled questions designed to assess individuals' understanding, certainty, and predictions of "what would happen next" at a time when those individuals experienced the manipulated emotion (Tiedens & Linton, 2001).

Once participants completed the emotion manipulation check, they moved on to a "separate" webpage that purported to host the ostensible "second study." First, participants read eight unrelated vignettes involving a variety of social situations (Appendix F). Simon and colleagues (Holyoak and Simon, 1999; Simon, 2004; Simon, Snow et al. 2004) constructed these vignettes and used them in their research on jury decision making. One or two factual questions about the vignette and 11-point response scales ranging from -5 (strongly disagree) to +5 (strongly agree) followed each vignette. For example, following one of the vignettes, participants rated their agreement with the following statement: "Wendy's identification makes it likely that it was Dale who left the flowers on Jessica's desk." Participants responded to 10 questions of this type, five of which were about facts discussed in the vignette, and five were general belief

related to the vignette. These ratings were the basis for calculating the participants' coherence shift index.

Once participants completed the "second study," to reinforce the emotion manipulation, they wrote another statement about the emotion that they described about in the first memory study (Appendix G). Then, participants completed the "third study" which was a fully-crossed 2 (standard of proof: preponderance of evidence v. beyond a reasonable doubt) x 2 (probability of culpability: 50% v. 90%) between subjects design. The website randomly assigned participants to one of four conditions.

Participants received slightly modified versions of a complex legal case. The scenario was the one that D. Simon and his colleagues constructed and used in prior research (Holyoak & D. Simon, 1999; D. Simon, 2004; D. Simon, Snow, et al. 2004) and which Engel and Glokner (2008) modified for their research. The case appears as a mediated arbitration in a disciplinary proceeding that Big Buildings Construction Company brought against Jason (Appendix H). The vignette describes a fact pattern in an embezzlement theft case. The civil case version left all evidence the same but altered the standard of proof. The standard of proof in the criminal case was beyond a reasonable doubt and it was preponderance of the evidence in the civil case.

In the vignette, the company accuses one of its employees of having stolen money from the company safe. Following a neutral background description, there were six pieces of information favoring conviction and six pieces of information favoring exoneration. For example, incriminating facts were a) the number of persons who knew the access code to the safe, b) the confidence level of an eyewitness who identified the employee, and c) the relative frequency of ownership of the type of car that witnesses observed at the scene of the crime, which was the

type of car the accused employee owned. The central exonerating fact was that other witnesses saw the accused employee at a distant location a short time after the estimated time of theft.

Following Engel and Glockner (2008)'s study, Jason's likelihood of guilt varied according to a numeric value altering the strength of evidence. The researchers manipulated the number of persons who knew the access code (medium: 18 persons, high: 8 persons), the self-reported confidence level of the eyewitness (medium: 80% certain, high: 95% certain), and the relative frequency of the type of car witnessed at the scene of the crime (medium: 6%, high: 0.1%). Aggregation of these numeric probabilities produced two different total posterior probabilities of culpability calculated by Bayesian theorem: 55% or 99% (Appendix I).

Each participant then acted as an arbitrator whose task was to decide the issue, either in favor of Jason Wells or in favor of the company. After reading the case-summary and reviewing the evidence, participants received one of two standards of proof, either "beyond a reasonable doubt" or "preponderance of the evidence." Both standards exactly reproduced the official model jury instructions of the Ninth Circuit Court of Appeals (Appendix J).

Participants then made a finding of fact (i.e., "yes, it was Jason who stole the money from the company's safe versus no it was not Jason who stole the money from the company's safe"). This finding of fact did not include a verdict on the case (to avoid possible confounds). Simon and colleagues as well as Glockner used this approach. Participants rated their subjective confidence of their factual finding on an 11-point scale ranging from 0 (completely uncertain) to 10 (completely certain) (Appendix K).

Then, participants estimated the probability that the accused person stole the money. They also specified the required level of probability necessary for finding against Jason under the given standard of proof. Next, as in the pretest, participants responded to 12 questions that

parallel the questions asked during the pretest: six questions concerned facts of the case and six questions concerned background beliefs (Appendix L). There were twelve questions in the "second study" (pretest) and the "third study" (posttest). To avoid suspicion, the researcher added six dummy questions in the "second study." The questions are identical in both "studies." Six questions concerned exoneration and six questions concerned incrimination. For example, in the second study, participants read eight vignettes and then rated exonerating or incriminating statements. One exonerating statement asks whether, "In the flower business, financial deals are typically done in cash." In the third study, the defendant testifies that he paid his debts using money he received from his sister-in-law for his work in her flower store, but he could not show bank documents for this transaction because in the flower business, financial transactions typically occur in cash. Finally, participants completed a demographic questionnaire (Appendix M) and read a debriefing statement (Appendix N).

Results

Rational-experiential inventory (REI) scores. I calculated REI scores for all participants on both the rational and experiential scales, each ranging from 0 to 100. The mean score on the rational scale for participants was 78.40 (SD = 10.67; minimum= 46; maximum= 98) with a median of 79. The mean score on the experiential scale for participants was 68.11 (SD = 15.71; minimum= 21; maximum= 100) with a median of 70. The scales, both together and separately, were found to be reliable (REI scale overall, $\alpha = .92$; for the rational scale, $\alpha = .90$; for the experiential scale, $\alpha = .96$). As previous studies indicated (Epstein et al., 1996), these scores suggest that rational and experiential processing are independent, $r(174) = .025$, $p = .74$. In subsequent analyses, I used standardized scores for both scales to increase stability and interpretability of the result.

Completion time. The average completion time for study 1 was 40 minutes and 53 seconds, with a standard deviation of 11 minutes and 39 seconds. Minimum completion time was 19 minutes and 55 seconds, and maximum completion time was 1 hour 08 minutes 25 seconds.

Emotion manipulation check. To test whether the emotion manipulation was effective, I divided self-reported emotions after the emotion manipulation into three groups: anger-related emotions (angry, hostile, irritable, scornful, and loathing), sadness-related emotions (sad, blue, downhearted, alone, and lonely), and other emotions (happy, disgusted, fearful, surprised, nervous, alert, proud, and excited). See Table 2 for the results of difference tests on each emotion manipulation check. Furthermore, the average scores for each emotional group shows two groups demonstrating strong internal consistency reliability (for anger related emotions, $\alpha = .94$; for sadness related emotions, $\alpha = .90$).

I conducted two separate t-tests (with either averaged anger-related emotions or averaged sadness-related emotions as the dependent variable). The emotion manipulation was the independent variable for both tests. There was a significant effect of the emotion manipulation on anger-related emotions, $t(174) = 4.87$, $p < .001$, $d = .74$. Participants in the anger condition demonstrated significantly higher ratings of anger-related emotions ($M = 3.99$, $SD = 2.38$) compared to those in the sadness condition ($M = 2.44$, $SD = 1.79$). There was also a significant effect of the emotion manipulation on sadness related emotions, $t(174) = 3.61$, $p < .001$, $d = .55$. Participants in the sadness condition had higher scores on sadness-related emotions ($M = 4.30$, $SD = 2.10$) than participants in anger condition ($M = 3.20$, $SD = 1.93$). For unrelated emotions, participants in the anger condition reported more disgusted, alert, proud, and excited emotions than participants in the sadness condition.

These results suggest that participants experienced a variety of negative emotions after completing the emotion manipulations. Though the mean of anger-related emotions was highly correlated with the mean of sadness-related emotions, $r(175) = .29$, $p < .001$, the fact that participants experienced other emotions does not mean that the emotion manipulation was ineffective. Rather, the primary emotion manipulation checks show that the manipulations were effective.

Effects of emotion on certainty and control appraisals. The three certainty measures demonstrated poor reliability (internal consistency $\alpha = .68$), as did the three control measures (internal consistency $\alpha = .62$). Therefore, all subsequent analyses used these measures separately rather than as averaged together into one scale.

Two multivariate analysis variance (MANOVA) tested whether manipulated emotion affected participants' control and certainty appraisals, as cognitive appraisal theory predicts. The first MANOVA included the emotion manipulation as the independent variable and the three control appraisal measures as dependent variables. This analysis examined whether emotion impacted participant's control appraisals. For the control measures, higher scores indicated individual control, and lower scores indicated situational control. There was a significant multivariate effect for emotion on control appraisal measures, $F(5,166) = 7.207$, *Wilks*' $\lambda = .741$, $p < .01$, $\eta_p^2 = .259$.

Follow-up t-tests examined the effects of manipulated emotion on each of the three items measuring control appraisal (see Table 3). For the three items measuring control appraisals, participants in the anger condition demonstrated significantly higher ratings of individual control (as opposed to situational control) than did participants in the sadness condition ($p < .001$).

For the certainty measures, higher scores indicated greater certainty, and lower scores indicated greater uncertainty. There was a significant multivariate effect for emotion manipulation on the control measures, $F(3, 173) = 3.769$, *Wilks'* $\lambda = .939$, $p < .01$, $\eta_p^2 = .167$. Post-hoc tests using independent sample t-tests examined the effects of manipulated emotion on each of the three items measuring certainty appraisal (see Table 3).

One certainty measure (i.e., *In the events that you described on the previous pages, how well did you understand what was happening in those situations?*) showed no differences among the manipulated emotions. Participants in the anger condition did not demonstrated significantly different ratings of event certainty compared to those participants in the sad condition. The second certainty measure (*how uncertain were you about what would happen in these situations?*) showed no significant effect for the emotion manipulation. Participants in the anger condition did not demonstrate no significantly different certainty than participants in the sad condition. Only on the third certainty measure (*how well could you typically predict what was going to happen next?*) did participants demonstrate the hypothesized result. Participants in the anger condition indicated significantly greater certainty than did participants in the sadness condition. Overall, as cognitive appraisal theory predicts, participants in the anger condition reported higher individual control and some increased certainty than did participants in the sadness condition (see Table 3).

Verdict analyses.

Guilty versus not guilty verdicts. The experimental design was a 2 (emotion: anger v. sadness) x 2 (standard of proof: preponderance of evidence v. beyond a reasonable doubt) x 2 (strength of evidence: 55% v. 99%) between subjects design.

Overall, 59.9 percent of total participants (n = 177) found the defendant guilty. In the preponderance of evidence condition (n = 89), 69.7 percent of participants found the defendant

guilty. In the beyond a reasonable doubt condition (n = 88), 50 percent of participants found the defendant guilty. In the anger condition (n = 89), 60.7 percent of participants found the defendant guilty. In the sad condition (n = 88), 59.1 percent of participants found the defendant guilty (See Table 4 for conviction rates according to emotion, standard of proof, and strength of evidence).

A forced entry logistic regression tested the main effects and interactions of emotion, standard of proof, and strength of evidence on verdict. Guilty/not guilty verdicts served as the dependent variable. The predictor variables included manipulated emotion, standard of proof, strength of evidence, both subscales of rational experiential inventory, and several two way interactions and one three way interaction stemming from the hypotheses. Standardized subscales of rational experiential inventory served as covariates. The full model was significant, $x^2(15, N = 177) = 28.917$, $p < .016$, *Nagelkerke* $R^2 = .20$. It correctly predicted 84.9% ($n = 90$) of guilty verdicts and 39.4% ($n = 28$) of the not guilty verdicts, for an overall accuracy of 66.7% ($n = 117$). Table 5 shows the logistic regression coefficient, Wald test, and odds ratio for each of the predictors. Although there were no significant main effects, several two way interactions (i.e., emotion by standard, rationality score by emotion, experiential score by strength of evidence) and one three way interaction (strength of evidence by emotion by standard) significantly predicted verdicts.

First, there was the significant two way interaction between emotion and standard of proof on verdict, $B = -2.38$, *Wald* $\chi^2(1) = 5.08$, $p = .02$, such that the impact of standards of proof depended on manipulated emotion. To interpret this two-way interaction, two separate simple logistic regression analyses examined the relation between standard of proof and emotion.

The first simple logistic regression used only data from the angry condition and tested the simple effect of standard of proof. Standard of proof, strength of evidence, rationality score,

experiential score, and the relevant two way interaction terms served as predictors. Verdict was the dependent variable. This model was not statistically insignificant, $x^2(9, N = 89)$ = 9.96, p = .35, *Nagelkerke* R^2 = .14, and correctly predicted 40.0% (n = 14) of not guilty verdicts and 81.5% (n = 44) of guilty verdicts. Table 6 shows regression coefficients, Wald test, and odds ratio for each of the predictors. There was no main effect of standard of proof, such that changing the standard of proof did not impact verdicts reached by angry participants.

The second simple logistic regression used only the data from the sadness condition to help facilitate interpretations of interactions. Standard of proof, strength of evidence, rationality score, experiential score, and the relevant two way interaction terms served as predictors. Verdict was the dependent variable. The model was statistically significant, $x^2(9, N = 88)$ = 32.37, p < .001, *Nagelkerke* R^2 = .42. It correctly predicted 58.3% (n = 21) of not guilty verdicts and 90.4% (n = 47) of guilty verdicts. Table 6 shows the regression coefficient, Wald test, and odds ratio for each of the predictors. There was a significant main effect for standard of proof, such that sad participants applying the beyond a reasonable doubt standard showed a significantly lower conviction rate than did participants who applied the preponderance of evidence standard (45% v. 72%).

Next, there was a significant three way interaction of emotion by standard of proof by strength of evidence, B = 3.44, *Wald* $\chi^2(1)$ = 5.71, p = .02. Conviction rates for this interaction are described in Figure 1. To evaluate this three way interaction, I conducted four binary logistic regressions. Table 7 displays the results.

A main effect for strength of evidence was statistically significant when sad participants applied the beyond a reasonable doubt standard, B = 2.31, *Wald* $\chi^2(1)$ = 5.190, p = .02. Within these conditions participants who evaluated medium strength evidence were less likely to find

the defendant guilty (26% of participants) than were participants who evaluated strong evidence (66% of participants).

Another interesting result of primary logistic regression was the interaction of experiential score by strength of evidence. To interpret this interaction, I calculated the predicted probability based on logistic regression. I conducted two separate simple regression analyses using experiential score as the predictor and predicted probability of a guilty-verdict as the dependent variable. Figure 2 depicts this interaction. When the evidence was weak, the predicted probability of conviction increased as participants' experiential score increased, $b = .46$, $p < .001$. In contrast, when the evidence was strong, predicted probability of conviction decreased as participants' experiential score increased, $b = -36$, $p < .001$. In sum, when participants received medium-strength evidence, they were more likely to find the defendant guilty as they relied more on their intuition. However, when evaluating strong evidence, participants did not differ in their use of their intuition.

Standard of proof ratings (threshold probability for conviction). A 2 (emotion: anger, sad) x 2 (strength of evidence: moderate, high) x 2 (standard: preponderance of evidence, beyond a reasonable doubt) ANCOVA with both rational experiential inventory subscales as covariates used the "level of probability necessary for conviction" as the dependent variable. Reponses to the standard of proof measure could range from 1 to 100 percent.

There were significant main effects for the standard of proof and the rationality scale. The results are displayed in Table 8. Participants using the beyond a reasonable doubt standard reported a higher level of certainty necessary to find the defendant guilty (estimated marginal mean = 84.29, $SE = 2.09$) than did participants using the preponderance of evidence standard (estimated marginal mean = 70.70, $SE = 2.09$).

A linear regression tested the effect of rationality scores on participants' subjective threshold of conviction by using emotion, standard of proof, strength of evidence, rationality score, and experiential score as predictors and using subjective threshold of conviction as the dependent variable, $B_{\text{emotion}} = .51$, $p = .86$, $B_{\text{standard}} = 13.61$, $p < .001$, $B_{\text{strength of evidence}} = -.04$, $p = .99$, $B_{\text{rationality}} = 3.76$, $p = .02$, $B_{\text{experiential}} = 2.30$, $p = .11$. As participants' rationality scale increased by 1 standard deviation, the threshold for conviction increased by 3.76 percentiles.

Coherence shifts. To analyze coherence shifts, I first calculated for each participant the average reevaluation scores of pro-guilty and contra-guilty evidence by subtracting pre-test from post-test valuations of the same evidence. Participants were divided into those who reported guilty verdicts and those who reported not guilty verdicts. The coherence shift index ranged from -11 to 11. Positive values indicated that participants attached greater weight to evidence coherent with their verdict and devalued evidence that contrasted with their verdict decision. Negative values indicated that participants devalued evidence supporting their decision and added weight to evidence that was incoherent with their verdict decision (see Table 9).A 2 (verdict: not guilty v. guilty) x 2 (emotion: anger v. sad) x 2 (strength of evidence: moderate v. high) x 2 (standard: preponderance of evidence v. beyond a reasonable doubt) ANCOVA with both rational experiential inventory subscales as covariates, and coherence shift as the dependent variable, tested the hypothesized main effects and interactions. Table 10 displays the results.

First, there was a significant main effect for emotion, $F(1, 159) = 5.20$, $p = .02$, $\eta_p^2 = .03$. Participants in the anger condition showed more coherence shift (estimated marginal mean = .86, $SE = .10$) than participants in the sad condition (estimated marginal mean = .51, $SE = .47$).

Second, there was a significant main effect for verdict, $F(1,159) = 18.01$, $p < .001$, $\eta_p^2 = .10$. Those who found the defendant guilty showed higher coherence shift (estimated marginal

mean = 1.01, *SE* = .10) than did participants who found the defendant not guilty (estimated marginal mean = .36, *SE* = .12).

Furthermore, there were two significant three-way interactions. The first was the interaction between emotion, standard of proof, and verdict, $F(1, 159) = 4.85$, p = .03, η_p^2 =.03. The average coherence shift for this interaction is described in Figure 3. Post-hoc tests using the least significance difference (LSD) method revealed that those in the anger condition who found a guilty verdict showed higher coherence shift when they used the beyond a reasonable doubt standard (*M* = 1.43, *SE* = .19, $p < .05$) then when they used the preponderance of evidence standard (*M* = .90, *SE* = .17). However, sad participants who found a guilty verdict did not show significant differences in coherence shift when they used the beyond a reasonable doubt standard (*M* = .59, *SE* = .23) as compared to when they used the preponderance of evidence standard (*M* = 1.11, *SE* = .17).

For participants who acquitted the defendant, those in the anger condition showed no difference in cognitive coherence when they used the beyond a reasonable doubt standard (*M* = .44, *SE* = .21) than when they used the preponderance of evidence standard (*M* = .67, *SE* = .25). Participants in the sad condition also showed no difference in coherence shift when they used the beyond a reasonable doubt standard (*M* = .20, *SE* = .21) than when they used the preponderance of evidence standard (*M* = .13, *SE* =.28).

There was another significant three way interaction between standard of proof, strength of evidence, and verdict, $F(1, 159) = 7.03$, $p < .01$, $\eta_p^2 = .04$. Average coherence shift for this interaction is described in Figure 4. Based on post hoc pairwise comparisons using LSD ($p < .05$), there were two significant simple effects of strength of evidence. Among those participants applying the preponderance of evidence standard, those who decided that the

defendant was not guilty showed significantly higher coherence shift when they received strong evidence (M = .79, SE = .27) than when they received weak evidence (M = .14, SE = .25). Also, among those participants applying the beyond a reasonable doubt standard, those who decided that the defendant was guilty showed significant higher coherence shift when they received weak evidence (M = 1.25, SE = .19) than when they received strong evidence (M = .78, SE = .24).

In contrast, among those participants who applied the preponderance of evidence standard, those who decided that the defendant was guilty showed no difference in coherence shift when they received strong evidence (M = 1.01, SE = .17) than when they received weak evidence (M = 1.00, SE = .17). Likewise, among those participants applying the beyond a reasonable doubt standard, those who decided that the defendant was not guilty showed no difference in coherence shift when they received strong evidence (M = .12, SE = .23) than when they received weak evidence (M = .52, SE = .20).

In sum, coherence shift significantly increased as the strength of evidence increased in two conditions: not guilty verdict derived under the preponderance of evidence standard and guilty verdict derived under the beyond a reasonable doubt standard.

Relationship between verdict certainty ratings and coherence shift index. The results suggest that high levels of confidence are an indicator of constraint satisfaction processing. The evidence supporting participants' decisions dominated the remainder of the evidence. The greater the shift in cognitive coherence, the more confident participants were in their verdicts, $r(177) = .337$, $p < .001$. As cognitive coherence increased, certainty in verdicts increased. Moreover, the relationship between coherence shift and confidence in one's verdict varied by verdicts, *Fisher's Z* = 1.744, $p = .05$. Among those who reached guilty verdicts, a higher coherence shift

was associated with higher verdict certainty, $r(106) = .40$, $p < .001$. This was not true for those who reached not guilty verdicts, $r(71) = .15$, $p = .21$.

Guilty certainty ratings. I created a verdict-certainty variable by coding guilty verdicts as 1 and not guilty verdicts as -1. I then multiplied each participant's coded guilty/not guilty verdict by their certainty in that verdict. This produced verdict certainty scores that ranged from -11 to 11, with -11 representing those participants who were very certain in a not guilty verdict, and 11 representing those participants who were very certain in a guilty verdict.

A 2 (emotion: anger v. sad) x 2 (strength of evidence: moderate v. high) x 2 (standard: preponderance of evidence v. beyond a reasonable doubt) ANCOVA with both rational experiential inventory subscales as covariates tested the hypothesized main effects and interactions on the verdict certainty dependent variable. Table 11 displays the results.

There was a significant main effect for standard of proof, $F(1,167) = 6.00$, $p = .015$, $\eta_p^2 = .035$, such that participants who used the beyond a reasonable doubt standard were less certain in a guilty verdict (estimated marginal mean = 1.03, *SE* = .80) than participants who used the preponderance of evidence standard (estimated marginal mean = 3.81, *SE* = .79). There was also a nearly significant main effect for strength of evidence, $F(1,167) = 3.34$, $p = .07$, $\eta_p^2 = .02$, such that participants who received high strength of evidence were more certain in their guilty verdicts (estimated marginal mean = 3.45, *SE* = .79) than were participants who received medium strength of evidence (estimated marginal mean = 1.40, *SE* = .79).

Mediation analyses. Recall that, according to cognitive appraisal theory, specific emotions give rise to different patterns of cognitive appraisals. People feeling angry are likely to experience high levels of certainty and attribute high levels of person-centered control (contrary to situational control) when judging the cause of events. In contrast, people feeling fear or

sadness are likely to experience low levels of certainty and high levels of situational control. Dual processing models in social psychology further suggest that people with high certainty are not motivated to increase their certainty, and they are satisfied using heuristics. In contrast, people with low certainty tend to process information more systematically to increase their levels of certainty (Tiedens & Linton, 2001).

Therefore, angry mock jurors might be more likely to re-evaluate evidence coherent to their decisions, whereas sad mock jurors might be less likely to re-evaluate evidence coherent to their verdicts. Thus, I hypothesized that manipulated anger would produce higher levels of certainty and higher levels of person centered control, which in turn would produce more coherence shift. Unfortunately, there was no need to conduct a mediation analysis on verdict because the results did not confirm the main effect of emotion on verdict. Instead, mediation analyses tested the indirect effect of emotion on coherence shift.

There are two ways to analyze this mediation effect: Barron and Kenny (1986)'s approach, which includes a Sobel test (1982) between regression weights or Preacher and Hayes (2008)'s bootstrapping approach. The traditional method of examining whether measured factors mediate the relationship between a predictor and outcome is the Barron and Kenny (1986) approach. Baron and Kenny discussed four steps in examining mediation. The first step is to show that a predictor is correlated with the outcome. This step establishes that an effect that may be mediated ("path C"). The second step is to show that the predictor is correlated with the mediator. In the current experiment, there should be a positive association of manipulated anger with certainty and personal control ("path A"). The third step is to show that the mediator affects the outcome variable. In the current experiment, there should be a positive association between

certainty (or personal control) and coherence shift ("path B"). The fourth step entails testing whether the mediating variable attenuates the effect of the predictor on the outcome.

Barron and Kenny's four steps are not preferred because those steps produce zero and nonzero coefficients, rather than direct regression weights with unbiased tests of statistical significance. There is also a weakness in that very large coefficients can be non-significant with small sample sizes, but small coefficients could be statistically significant with large sample sizes. That is, the Barron and Kenny approach is low in statistical power.

Recent mediation analyses focus more on the indirect effect than the first step (direct effect) in Barron and Kenny's (1986) approach (Frazier, Tix, and Barron, 2004). One reason is that there is a possibility that the direction of the direct effect could be opposite of the indirect effect. In such a case, the total effect is likely to be very small because the direct and indirect effects will tend to cancel out each other.

There are two common ways to test the significance of indirect effects: a Sobel (1982) test and bootstrapping. MacKinnon, Lockwood, Hoffman, West, & Sheets (2002) recommend the Sobel test. This test is a specialized t-test that provides a method to determine whether there is a reduction in the independent variable's effect when the mediator is included in a regression model. However, the Sobel test has two problems. The first problem is a low power issue. The Sobel test requires large sample sizes. With small sample sizes, the Sobel test will not produce an accurate p-value (Preacher and Hayes, 2008). Another problem is that the Sobel test assumes the symmetric distribution of variables.

Recently, researchers have favored a mediation analysis using bootstrapping (Preacher and Hayes, 2008) over the Sobel test (Shrout & Bolger, 2002). Because bootstrapping is a non-parametric method based on resampling, it does not require an assumption of normality, and it

allows multivariate tests of simultaneous mediations. Unlike Barron and Kenney's method, unbiased confidence intervals, *p* values, and standard errors result.

I used Preacher and Hayes' SPSS macros to calculate a matrix of indirect effect coefficients of emotion (0- Anger, 1-Sad) on participants' coherence shifts through three control appraisal measures with 10,000 bootstraps. Table 12 lists the indirect effects and 95 percentile confidence intervals. The confidence intervals of two control appraisal measures do not contain zero, suggesting that these measures mediate the effect of emotion on verdict. The other control appraisal measure did not mediate the effect of emotion on coherence shift.

Next, three certainty appraisal measures were included as mediators. Bootstrapping produced a matrix of coefficients for indirect effect, standard error, and 95th percentile confidence intervals. Table 12 lists the indirect effects and 95th percentile confidence intervals. Unfortunately, none of the three measures mediated the effect of emotion on coherence shift.

These results suggest that sadness had a significant indirect effect on coherence shift through changes in respondents' control appraisal. Sad participants showed less personal control, and they were less likely to find the defendant guilty and/or re-evaluate evidence toward their verdicts, as compared to angry participants. However, there was no significant indirect effect of emotion on either verdict or coherence shift when certainty appraisal measures were used as mediating variables because there was no effect of the emotion manipulation on certainty appraisal measures, as described above.

Discussion

Study 1 attempted to examine whether experienced emotion could impact mock jurors' decision style and their use of the standard of proof. It was assumed that sad mock jurors would use systematic processing, demonstrate sensitivity to changing strength of evidence and

standards of proof, but would not increase coherence shift (as compared to angry mock jurors). The results of study 1 provided some support for these hypotheses.

The first prediction was that sad mock jurors would use systematic processing and that differences in standards of proof would have a stronger effect on verdicts. In contrast, I predicted that when mock jurors experience anger, they would use intuitive processing, and altering the standard of proof would have a weak effect on verdicts (*Hypothesis 1*). As expected, participants who experienced sadness were less likely to find the defendant guilty when they applied higher standards of proof (as compared to lower standards of proof). By contrast, among participants in the anger condition, verdicts did not vary by standard of proof.

Furthermore, based on the Appraisal-Tendency Framework and dual process theories in social psychology, I predicted that angry emotions would increase participants' certainty appraisal and eventually induce participants to use intuitive processing (i.e., coherence shift) compared with sadness emotions. However, mediation analyses showed very little evidence of certainty appraisal mediating the effects of emotion on coherence shift.

One possible explanation could be that participants in the sadness condition might have experienced the same level of certainty as the participants in the anger condition. Tiedens and Linton (2001) demonstrated that sadness falls more toward the middle of the certainty spectrum compared to fear. Therefore, people who feel sad may experience certainty sometimes and uncertainty other times. Future studies should include more emotions (e.g., fear) to further examine the influence of certainty appraisals on coherence shifts.

Instead, mediation analyses using control appraisal measures found that the effect of emotion on coherence shift was mediated by individual control (as opposed to situation control). Consistent with the appraisal tendency view (Tiendens & Linton, 2001), participants in the anger

condition perceived negative events as controlled by others and participants in the sad condition perceived negative events as controlled by the situation. Further, when participants perceived negative events to imply individual control, they re-evaluated evidence consistent with their verdicts. This result is consistent with a previous study that angry decision makers are more influenced by heuristic cues than sad decision makers (Bodenhausen et al., 1994).

A second hypothesis predicted that, because participants who experience anger would rely on intuitive processing, when these participants are required to apply a higher standard of proof, they would increase their subjective certainty by increasing coherent re-evaluation of evidence. Conversely, participants who experience sadness would rely on systematic processing, and when they are required to apply a higher standard of proof, they would not rely on coherent re-evaluation of evidence (at least as compared to angry participants) (*Hypothesis 2*).

Interestingly, study 1 found that angry mock jurors increased their certainty by using coherence shift when asked to apply a higher standard of proof, but sad mock jurors did not use coherence shift across standards of proof. However, this result occurred only when mock jurors found the defendant guilty. No interaction occurred for mock jurors who decided the defendant was not guilty.

One possible explanation for why guilty decisions might require more confidence than acquittal decision may be that mock jurors who reach a guilty verdict require more information processing (either intuitive or systematic) to bolster their confidence in the guilty verdict. Therefore, the effect of strength of evidence and standard of proof on coherence shift should occur only when mock jurors reach a guilty verdict. This result is consistent with previous studies (Glockner & Engel, 2013; Simon et al., 2004), which found that guilty verdicts necessitate more comprehensive coherence shifts than not guilty verdicts.

The third hypothesis posited that, because participants in the sad condition would use systematic processing, their verdicts would be more susceptible to variations in the strength of evidence. In contrast, because participants in the anger condition would use intuitive processing, strength of evidence would influence their verdicts less (*Hypothesis 3*).

The data partially supported this prediction. The results indicated that strength of evidence influenced verdicts only when sad participants applied the beyond a reasonable doubt standard. The fact that even in the weak evidence condition, the probative value of evidence was higher than 50 percent might explain this finding. Thus, even rational mock jurors would not be sensitive to variations in the strength of evidence when using the preponderance of evidence standard.

One other interesting finding concerns participants' self-reported threshold for conviction. Although participants applying the beyond a reasonable doubt standard reported significantly higher thresholds for conviction than did participants applying the preponderance of evidence standard, the difference between these standards was smaller than desired. Legal professionals (e.g., federal judges) estimated the level of certainty as 90 percent for the beyond a reasonable doubt standard, and 51 percent for the preponderance of evidence standard (McCauliff, 1985; Weinstein & Dewbury, 2006). However, in study 1, the mean for the preponderance of evidence standard was 70 percent, and the mean for the beyond a reasonable doubt standard was 85 percent. Essentially, participants underestimated the threshold of the beyond a reasonable doubt standard and overestimated the threshold of the preponderance of evidence standard.

This result is similar to that of Glockner and Engel's (2013) study. They asked German students to estimate the respective thresholds for conviction, and German students reported results of 75 percent in the preponderance of evidence condition and 85 percent in the beyond a

reasonable doubt condition. Of course, German students might be less familiar with American standards of proof because German law has no articulated civil standards of proof (Clermont, 2002). However, in this study, the participants were U.S. jury-eligible citizens who should be more familiar with standards of proof. Unfortunately, the participants in this study did not indicate their familiarity with standards of proof ahead of time. Nonetheless, the difference between the preponderance of evidence and the beyond a reasonable doubt standard was smaller than anticipated under the law.

Regarding individual differences, participants' rationality scores and experiential scores did not predict the amount of coherence shift or participants' ultimate verdicts. These results are inconsistent with previous studies using the rational experiential inventory, which showed that the experiential scale highly correlates with heuristic decisions (e.g., Pacini & Epstein, 1999). However, the current study is different because instead of using a general heuristic like Pacini and Epstein, it explored coherent shifts, which are more similar to cognitive consistency models than to heuristic processing. Future research should compare the two types of decision tasks with respect to rationality and experiential processing styles.

However, the current study did find that the effect of strength of evidence depends on participants' experiential score. When evidence was strong enough to convict the defendant, conviction rates decreased as experiential scores increased. In contrast, when the evidence was not strong enough to convict the defendant, conviction rates increased as experiential score increased. In other words, the difference in strength of evidence attenuated as mock jurors used the experiential system more.

The results of study 1 replicated and extended the previous work by Glockner and Engel (2013). Their previous study found that the beyond a reasonable doubt instruction could reduce

the conviction rate, as compared to the preponderance of evidence instruction. The current study found that sad mock jurors used the standard of proof as reported in the previous study. However, for angry mock jurors, the beyond a reasonable doubt instruction did not reduce conviction rates, as compared to the preponderance of evidence instruction.

In addition, Glockner and Engel (2013) did not find that mock jurors simply increased their confidence in their verdicts by using coherence shift when asked to apply the beyond a reasonable doubt standard. Those authors argued that the beyond a reasonable doubt jury instruction might work in a desired way. The current study replicates the previous results among sad mock jurors. In addition, however, this study further found that angry mock jurors increased their certainty by using coherence shift to meet higher standards of proof, likely through their reliance on intuitive processing.

Another interesting finding is that previous researchers (Simon et al., 2004; Simon & Scurich, 2011, Glockner & Engel, 2013) suggest that coherence shift is robust. The current study found that emotion (i.e., anger) can increase or decrease coherence shift. Mediation analyses indicated that individual control mediated the effect of emotion on coherence shift. Angry mock jurors attributed blame to other individuals and they re-evaluated evidence toward their verdicts.

Overall, these results support the impact of emotion on mock jurors' decision styles. When mock jurors are induced to feel sad, they use rational decision processing, and standards of proof and strength of evidence operate as intended by the legal system. Sad mock jurors also demonstrate less coherence shifts. In contrast, when mock jurors are induced to feel angry, they seem to rely on intuitive processing. Their verdicts were not sensitive to varying standards of proof or varying strength of evidence information.

Chapter 7: Study 2

The basic procedure and materials for Study 2 were the same as in study 1, though with the following changes. First, half of the participants received the theft version of the Jason materials (similar to the version used in study 1), with the criminal standard provided. The other half of participants received a version in which the culprit brutally murdered a guard during the theft, and as a result, the charges were theft and first-degree murder. Second, half of the participants received verdict questionnaires first, and the other half received assessment questionnaires first. Third, the case was a criminal trial instead of an arbitration case, and the participants acted as jurors rather than arbitrators. The experimental design was a fullycrossed 2 (severity of charge: theft v. murder) x 2 (strength of evidence: 55% v. 99%) x 2 (the order of questionnaires: assessing arguments first v. verdict first) between subject design with all participants completing the REI.

Method

Participants. I recruited participants from the Amazon Mechanical Turk website. This site posted a recruitment advertisement to individuals who agreed to participate in web-based research studies. Participants received a three-dollar gift certificate to Amazon.com for their participation.

Participants included 87 males and 100 females from across the United States. The mean age was 36.06 years, ranging from 19 to 67 years old. The majority of participants indicated that they were Caucasian (*N* = 157, 84%). Ten participants indicated that they were Asian American (5.3 %), nine were African American (4.8 %), eight were Hispanic (4.3 %), and three indicated that their ethnicity was "other" (1.6%). Two participants had less than a high school education (1.1%). 79 indicated they had a high school level of education (42.2%). 34 had an associate's degree (18.2 %), 52 a bachelor's degree (27.8 %), 17 a master's degree (9.1 %), and 2 had a

professional degree (1.1 %). One male participant did not indicate his level of education. Eighty-four participants were employed full time (44.9%), 53 were employed part time (28.3 %), and 49 were unemployed (26.2 %). One female participant did not indicate her employment status.

To be jury-eligible in most states, citizens must be at least 18 years old and without felony convictions. I removed six participants from the data set because they were not U.S. citizens and four participants indicated that they were felons without civil rights. I also removed six participants who had already completed study 1. The final data set consisted of 78 jury eligible men and 97 jury eligible women randomly assigned to one of the eight experimental conditions. Table 13 depicts participants' demographic characteristics after data cleaning.

Procedure. When participants selected this study from Mechanical Turk and finished the Rational-Experiential Inventory (Appendix A), they received email invitations to participate, which included a URL address to gain access to the website. Participants who chose to participate completed a standard informed consent form (Appendix O). They learned that they would participate in two short unrelated studies, one assessing social issues, and the other concerning a hypothetical legal case.

In the "first study," participants read eight unrelated vignettes involving a variety of social situations (Appendix F). One or two factual questions about the vignette and 11-point response scales ranging from -5 (strongly disagree) to +5 (strongly agree) followed each vignette. For example, following one of the vignettes, participants rated their agreement with the following statement: "Wendy's identification makes it likely that it was Dale who left the flowers on Jessica's desk." Participants responded to 10 questions of this type, five of which were about facts discussed in the vignette, and five were general belief related to the vignette.

These ratings were the basis for calculating participants' coherence shift index (see Appendix L, note 2).

After they completed the "first study," participants completed the "second study," which was a fully crossed 2 (seriousness of crime: theft v. murder) x 2 (strength of evidence: medium: 55% v. high: 99%) x 2 (the order of questionnaires: probability assessment first v. verdict first) between subject design. The website randomly assigned participants to one of eight conditions. The website instructed participants to assume the role of a juror in a criminal case whose task it is to decide whether Jason is guilty of a crime (Appendix P). Half of the participants received the theft version of the case, and the other half received the murder version. The two versions were essentially identical, with the exception that in the murder case the culprit brutally beat the guard to death (Appendix Q). Before deciding a verdict, participants received Simon and colleagues' (2001) jury instructions articulating the "beyond a reasonable doubt" standard (Appendix R).

Half of the participants received assessment-first questionnaires; the other half received verdict-first questionnaires (Appendix S). In the assessment-first condition, participants rated the defense and prosecution arguments using a scale ranging from -5 (strongly disagree) to +5 (strongly agree). Then, they offered a verdict. In verdict-first condition, participants offered a verdict, and then they rated the defense and prosecution arguments. The participants next answered an accuracy motivation questionnaire (Appendix T). Two items asked participants the degree to which they were motivated to decide accurately and a third item asked about the mock jurors' expectations regarding possible penalties. Finally, participants completed the same demographic questions (Appendix M) and read debriefing statement (Appendix U).

Results

Rational-experiential inventory (REI) scores. For each participant, I calculated REI scores, for both the rational and experiential scales. Participants' mean score on the rational scale was 76.58 (*SD* = 12.27; minimum= 42; maximum= 100) with a median of 78. Participants' mean score on the experiential scale was 67.46 (*SD* = 13.22; minimum= 21; maximum= 100) with a median of 70. The scales, both together and separately, were reliable (REI scale overall, $\alpha = .93$; for the rational scale, $\alpha = .90$; for the experiential scale, $\alpha = .95$). As previous studies indicate (Epstein et al., 1996), these scores suggest that rational and experiential processing are independent, $r(175) = .009$, $p = .91$. In subsequent analyses, I used standardized scores for both scales to increase the interpretability of the results.

Completion time. The average completion time for study 2 was 30 minutes and 29 seconds, with a standard deviation of 9 minutes and 57 seconds. Minimum completion time was 22 minutes and 34 seconds, and maximum completion time was 43 minutes 12 seconds.

Seriousness manipulation check. A multivariate analysis of variance (MANOVA) with five seriousness measures as dependent variables and manipulated charge as the independent variable tested the effect of charge on seriousness measures. There were significant effects of the seriousness manipulation, $F(5,154) = 747.656$, *Wilks*' $\lambda = .040$, $p < .001$, $\eta_p^2 = .960$. As follow-up *t*-tests showed (see Table 14), all five seriousness measures showed that significant difference between when participants faced a less serious charge and a more serious charge. I calculated the Cronbach's alpha reliability for these five items. This scale was reliable ($\alpha = .86$). In subsequent analyses, I used the combined score of these five measures to increase stability and interpretability. I also performed a t-test between more serious and less serious crime manipulation on this scale and found it to be significant, $t(173) = 31.79$, $p < .001$, $d = 4.80$, such

that participants scored significantly higher for murder (M = 10.55, SD = .66) than for theft (M = 5.66, SD = 1.28).

Accuracy motivation. A multivariate analysis of variance (MANOVA) with five accuracy motivation items as dependent variables and with manipulated charge as the dependent variable tested the effect of charge on accuracy manipulation checks. There were significant effects for the seriousness manipulation, $F(5,165) = 2.985$, *Wilks*' $\lambda = .917$, $p = .013$, $\eta_p^2 = .083$. Independent sample t-tests with each of the relevant self-reported accuracy motivation measures as dependent variables and with manipulated seriousness as the independent variable tested the effect of the seriousness manipulations (see Table 14). There was a significant effect of severity of charge only on the first motivation measure (*if you reached the wrong verdict, how much would you regret your decision?*). There was no significant effect of severity of charge on the other measures (*how hard did you try to judge the verdict accurately*? *how hard did you try to judge fairly? were you motivated to make your judgments accurately? Was it more important to reach a fair judgment or an accurate judgment?*). I calculated the Cronbach's alpha for these five items, and the scale was not reliable (α = .48).

Verdict Analyses

Guilty versus not guilty verdicts. The experimental design was a fully crossed 2 (severity of charge: theft v. murder) x 2 (strength of evidence: 55% v. 99%) x 2 (the order of questionnaires: assessing arguments first v. verdict first) between subject design. Participants in the verdict-first condition initially decided their verdict (guilty or not guilty) and then they rated percentile scores of guilt, threshold of conviction, and their confidence. Next, they assessed the strength of the arguments in the case. Participants in the probability assessment-first condition completed these tasks in the reverse order, first assessing the arguments, then rating percentile

score of guilt, threshold of conviction, and confidence. Only after completing all these measures did they provide verdicts.

Logistic regression tested the effects of severity of charge, order of questionnaires, and strength of evidence on verdict. Verdict for the theft charge was the dependent variable for both the less severe charge condition and the more severe charge condition. In the less severe charge condition, participants received the theft version of the case and decided only a theft verdict. In the more severe charge condition, participants received the murder version and decided both the theft charge and the murder charge. Only two participants found the defendant guilty of theft but not guilty of murder. Therefore, the verdict for murder was highly associated with the verdict for theft, $r(89) = .96$, $p < .001$, making it sensible to use participants' verdict for the theft charge as the only dependent variable in this analysis.

Overall, 42.9 percent of total participants (n = 175) found the defendant guilty. In the less severe crime condition (n = 86), 50.0 percent of participants found the defendant guilty. In the more severe crime condition (n = 89), 43.8 percent of participants found the defendant guilty of theft, 41.6 percent of participants found the defendant guilty of murder. Conviction rates for theft charge by conditions are illustrated in Table 15.

A forced entry logistic regression used guilty/not-guilty verdicts as the dependent variable. The predictor variables included severity of charge (theft v. murder), strength of evidence (medium v. high), and the order of questionnaires (verdict first v. assessing arguments first). Both subscales of rational experiential inventory, relevant two-way interactions, and one three-way interaction (all stemming from the hypotheses) were also included as predictors. I used standardized subscales of the rational experiential inventory to increase the interpretability of the resulting Beta weights.

The model was significant, $x^2(15, \mathrm{N} = 93)$ = 31.51, *p* = .008, *Nagelkerke* R^2 = .22, and correctly predicted 60.0% (*N* = 50) of guilty verdicts and 63.4% (*N* = 59) of the not guilty verdicts, for an overall accuracy of 62.3% (*N* = 109). There were a number of statistically significant predictors of guilt. Table 16 shows the logistic regression coefficient, Wald test, and odds ratio for each of the predictors. Severity of charge, three of the two-way interactions (severity of charge by order of questionnaires, severity of charge by strength of evidence, and rational score by order of questionnaires), and one three-way interaction (strength of evidence by the order of questionnaires by severity of charge) showed significant effects.

First, when holding all other variables constant, participants who decided a more serious charge were less likely to find the defendant guilty (41.6% of participants) compared to those who faced a less serious charge (50.0% of participants). There was no significant main effect for strength of evidence (Medium: 41% v. High: 58%) or for the order of questionnaires (Verdict first: 46.5% v. assessing arguments first: 45.3%).

Second, there was a significant two-way interaction between severity of charge and the order of questionnaires, as well as a marginally significant interaction between severity of charge and strength of evidence ($p < .07$). Table 16 displays the results. Two separate logistic regressions examined these interactions. Both regressions used verdicts as the dependent variable, and the order of questionnaires, strength of evidence, rationality score, experiential score, and two-way interactions as predictors. Table 17 shows the results. In a forced entry logistic regression using data only for theft condition, the model was not significant, $x^2(9, N = 86) = 13.15$, $p = .16$, *Nagelkerke* $R^2 = .19$, detecting no statistically significant dependence between verdict and any predictors. A second logistic regression using only the data from the murder condition was statistically significant, $x^2(9, N = 89) = 24.07$, $p = .004$, *Nagelkerke* $R^2 = .32$ and

correctly predicted 68.0% (n = 34) of the not guilty verdicts and 56.4% (*n* = 22) of guilty verdicts, for an overall accuracy of 62.9% (*n* = 56). The order of questionnaires was significant, $B = -2.87$, $Wald\ \chi^2(1) = 8.15$, $p = .004$, such that participants who assessed arguments before making a verdict decision were significantly less likely to find the defendant guilty (34.9% of participants) than participants who made a verdict decision before assessing arguments (47.8% of participants).

There was a significant three-way interaction between severity of charge, the order of questionnaires, and strength of evidence, $B = 3.41$, $Wald\ \chi^2(1) = 5.91$, $p = .02$. The results are displayed in Figure 5. To interpret this complicated three-way interaction, I performed four binary logistic regressions. Table 18 shows the results. The four separate logistic regression (2: severity of charge x 2: the order of questionnaires) included strength of evidence, rationality score, and experiential score as predictors with verdict as the dependent variable.

A main effect for strength of evidence was statistically significant only when participants faced a murder charge and assessed arguments before choosing a verdict, $B = 1.72$, $Wald\ \chi^2(1) = .4.52$, $p = .03$. Stronger evidence increased conviction rates (52% of participants) more so than weaker evidence (15% of participants).

As in study 1, there was no main effect or interaction for rationality score on verdict. However, there was a significant interaction between experiential score and strength of evidence. To interpret this interaction, I calculated predicted probability based on logistic regression. Two separate and simple regression analyses used experiential score as the predictor and predicted probability of guilty as the dependent variable. Figure 6 depicts this interaction. When the evidence was weak, the predicted probability of guilt increased as participants experiential score increased, $b = .53$, $p < .001$. In contrast, when the evidence was strong, the predicted probability

of guilt decreased as participants' experiential score increased, $b = -.46$, $p < .001$. This result replicates the result of study 1. Experiential score led to reduced effects of the strength of evidence.

Standard of proof rating (threshold probability for conviction). A 2 (seriousness of crime: theft v. murder) x 2 (strength of evidence: 55% v. 99%) x 2 (the order of questionnaires: probability assessment first v. verdict first) ANCOVA with both rational experiential inventory subscales as covariates used the level of probability necessary for convicting as the dependent variable. Reponses to the standard of proof measure could range from 1 to 100 percent. There was no significant predictor or interaction (See Table 19).

Coherence Shifts. As in study 1, I calculated average reevaluation scores for pro-guilty and contra-guilty evidence by subtracting pre-test from post-test valuations of the same evidence. Participants were divided into those who convicted and those who acquitted (based on the verdict they rendered). The coherence shift index ranged from -11 to 11. Positive values indicated that participants attached greater weight to arguments coherent with their verdict decision and devalued evidence contrary to their verdict decision. Negative values indicated that participants devalued evidence supporting their verdict decision and added weight to arguments incoherent with their verdict decision. Table 20 displays the means for coherence shift across conditions and verdicts.

A 2 (verdict: not guilty v. guilty) x 2 (seriousness of crime: theft v. murder) x 2 (strength of evidence: 55% v. 99%) x 2 (order of questionnaires: assessing arguments first v. verdict first) ANCOVA tested the hypothesized main effects and interactions. The ANCOVA included both rational experiential inventory subscales as covariates and the coherence shift as the dependent variable. Table 21 displays the *F*-value, degree of freedom, *p*-value, and partial η^2.

There was a significant main effect for order of questionnaires, $F(1, 156) = 4.22$, $p = .042$, partial $\eta_p^2 = .03$, such that participants who first offered a verdict showed more coherence shift (estimated marginal mean = .91, *SE* = .11) than participants who first assessed arguments (estimated marginal mean = .57, *SE* = .12).

There was also a significant main effect for verdict, $F(1,156) = 40.67$, $p < .001$, $\eta_p^2 = .21$. Those who convicted showed a higher coherence shift (estimated marginal mean = 1.28, SE = .13) than did those who acquitted (estimated marginal mean = .21, SE = .11).

There was a nearly significant main effect for strength of evidence, $F(1, 156) = 3.76$, $p = .052$, $\eta_p^2 = .02$. Participants who evaluated weak evidence (estimated marginal mean = .90, SE = .12) showed more coherence shift than did participants who evaluated strong evidence (estimated marginal mean = .20, SE = .11).

Furthermore, there was a significant two way interaction between seriousness of charge and strength of evidence, $F(1, 156) = 3.85$, $p = .05$, $\eta_p^2 = .02$. Pairwise comparisons using the least significance difference (LSD) method revealed ($p < .05$) a significant difference in coherence shift between participants who received different strength of evidence but only in the theft condition. Participants showed more coherence shift when evaluating weak evidence (estimated marginal mean = 1.07, *SE* = .17) than when evaluating strong evidence (estimated marginal mean = .42, *SE* = .16).

In the murder condition, however, there was no significant effect for strength of evidence on coherence shift. Participants who received weak evidence showed no significantly difference in coherence shift (estimated marginal mean = .74, *SE* = .20) than participants who received strong evidence (estimated marginal mean = .74, *SE* = .15).

There was another significant two way interaction between strength of evidence and verdict, $F(1, 156) = 5.30$, $p = .02$, $\eta_p^2 = .03$. Pairwise comparisons using the least significance difference (LSD) method revealed ($p < .05$) that participants who found the defendant guilty increased their coherence shift more when they received weak evidence (estimated marginal mean = 1.57, *SE* = .20) than when they received stronger evidence (estimated marginal mean = .98, *SE* = .15). In contrast, participants who found the defendant not guilty did not show significant differences in coherence shift between the weak evidence condition (estimated marginal mean = .20, *SE* = .15) and the strong evidence condition (estimated marginal mean = .24, *SE* = .16). There was no significant three-way interaction.

Relationship between verdict certainty ratings and coherence shift index. High levels of confidence are an indicator of constraint satisfaction processing because they are the natural consequence of the spreading apart of the subsets of evidence (Simon, 2004). Evidence supporting a chosen decision dominates the remaining evidence. As expected, the greater participants' cognitive coherence shift, the more confident participants were in their verdicts, $r(175) = .47$, $p < .001$.

Guilty certainty ratings

I created a verdict-certainty variable by coding guilty verdicts as 1 and not guilty verdicts as -1. I then multiplied each participant's coded guilty/not-guilty verdict by their certainty in that verdict. This produced verdict certainty scores that ranged from -11 to 11, with -11 representing those participants who were very certain in a not guilty verdict, and 11 representing those participants who were very certain in a guilty verdict.

A 2 (type of case: theft v. murder) x 2 (strength of evidence: Medium v. High) x 2 (order of questionnaires: verdict-first v. probability-first) ANCOVA with both rational experiential

inventory subscales as covariates used verdict certainty as the dependent variable. As can be seen in Table 22, there was only one significant effect: a three way interaction between severity of charge, order of questionnaires, and strength of evidence, $F(1,165) = 5.90$, $p = .02$, $\eta_p^2 = .03$. Pairwise comparisons using the LSD method ($p < .05$) revealed that the strength of evidence was significant only when participants evaluated a murder charge and decided on a verdict first (Med M = 1.99 v. high M = 1.21). This replicates the verdict findings for the logistic regression.

Mediation analyses.

Hypothesis 4 predicted that there should be a main effect for the severity of charge on verdict. Because the consequences (liability and penalty) following more serious charges are severe, participants faced with a more severe charge should be more motivated to decide cases accurately. To test this prediction, I conducted a mediation analysis using Preacher and Hayes' (2008) bootstrapping approach with 10000 bootstraps. The Preacher and Hayes' SPSS macro calculated a matrix of indirect effects of severity of charges (0- theft, 1- murder) on the verdict (0- not-guilty, 1- guilty) through five accuracy motivation measures. The hypothesized causal chain was that accuracy motivation measures would mediate the effect of different charges (theft or murder) on verdict. Unfortunately, there were no significant indirect effects of the severity of charges on verdicts through accuracy motivation measures (see Table 21).

Hypothesis 5 predicted that there should be a main effect for severity of charge on coherence shift. The consequences of a severe charge would increase participants' motivation to make accurate decisions. This increased motivation would decrease coherence shift. To test this causal chain, I analyzed the indirect effects of severity of charge on coherence shift through accuracy motivation. However, there were no significant indirect effects of the severity of charges on coherence shift through accuracy motivation measures (see Table 21).

Discussion

The purpose of study 2 was to examine whether variations in the severity of charges can influence mock jurors' decision styles. I predicted that, when the charge is more serious, participants would be motivated to reach accurate decisions. To do so, they would use systematic processing. When decision makers rely on systematic processing, I expected that the influence of strength of evidence would be strong and the coherence shift would attenuate. Conversely, when the charge is less serious, participants would lack the motivation to make an accurate decision, and they would rely on intuitive processing. As a result, the influence of strength of evidence would be weaker and coherence shift would intensify.

The first prediction was that participants who evaluated a less serious charge would be more likely to convict than participants who evaluated a more serious charge (*Hypothesis 4*). As expected, the results show that the severity of charges significantly impacted participants' verdicts. Participants who evaluated a less serious charge were more likely to find the defendant guilty than were participants who evaluated a more serious charge. This result is consistent with the *severity leniency hypothesis* (Kerr, 1978; R.J. Simon, 1969), which asserts that jurors (and juries) become less likely to convict a person as the penalty associated with conviction grows.

Furthermore, based on the *severity leniency hypothesis* and dual processing models, I predicted that the severity of the charge would increase participants' motivations for accurate judgments. In addition, increased motivations resulting from a more serious charge could decrease the conviction rates. Mediation analyses tested whether accuracy motivations mediate the effects of severity of charge on verdict, such that after controlling accuracy motivation, the relationship between severity of charge and verdict should attenuate. Unfortunately, there was no need for the mediation analyses because there were no differences in self-reported accuracy

motivations between different charges. One explanation for this result might be the problems associated with self-report data. Participants might have reported accuracy motivations because the instructions explicitly told them to decide accurately.

My second prediction was that participants who evaluated a less serious charge would use intuitive processing and show more coherence shift. By contrast, participants who evaluated a more serious charge would use systematic processing and show less coherence shift (*Hypothesis 5*).

Unfortunately, there were no direct effects of severity of charge on coherence shift. However, a two-way interaction between severity of charge and strength of evidence showed that the effect of severity of charge depend upon the strength of evidence. In the theft condition, participants showed higher coherence shifts when they received weak evidence than when they received strong evidence. In contrast, in the murder case, participants did not show any differences in coherence shift across the strength of evidence. This result suggests that the severity of charge does not simply impact coherence shift. Instead, participants evaluating a less severe charge increased their coherence shifts but only when the evidence was weak.

Another prediction was that participants who assessed arguments first would show a reduced coherence shift as compared to participants who first decided on a verdict (*Hypothesis 6*). The *task cueing* hypothesis suggests that the characteristics (i.e., sequence) of tasks can trigger systematic processing or intuitive processing (Inbar, Cone, &Gilovich, 2010). I predicted that by varying the order of questionnaires, I could induce participants to use systematic processing or intuitive processing. Examination of the main effect supports this prediction. Participants who first assessed arguments showed less coherence shift than participants who decided the verdict

first. The result suggests that the order of the questionnaires could guide participants thought processing and reduce intuitive processing.

Furthermore, cognitive dissonance (Festinger, 1957, 1964) predicts that re-evaluation of evidence would not begin before formation of verdicts. However, the order of questionnaires could not remove coherence shifts completely. This result indicates that coherence shift can occur even before participants make a decision.

The predicted two-way interaction between severity of charge and strength of evidence on verdict suggested that participants who evaluate a less serious charge would be less sensitive to the strength of evidence than those who evaluated a more severe charge (*Hypothesis 7*). Unfortunately, the results of simple logistic regressions suggest the opposite effect. The impact of strength of evidence was significant for participants who evaluated the theft charge. For the murder charge, the impact of strength of evidence was not significant.

This result might be explained by the effect of the order of questionnaires. The significant three-way interaction between severity of charge, strength of evidence, and order of questionnaires suggests that the effect of strength of evidence was significant only when participants evaluated a murder charge and evaluated evidence first. The effect of order of questionnaires might be stronger than the effect of severity of charge.

Another predicted two-way interaction between the order of the questionnaires and strength of evidence on verdict was that participants who first decided a verdict would be less sensitive to variations in the strength of evidence, as compared to those who assessed arguments first (*Hypothesis 8*). The results support this prediction, such that participants' verdicts in the arguments-first condition were more sensitive to variations in the strength of evidence than were participants in the verdict-first condition.

For individual differences, the results reflect almost the exact same results as study 1. Rationality score and experiential score did not predict the amount of coherence shift or verdicts. As in study 1, the current study found that the effect of strength of evidence is related to participants' experiential score. When the evidence was strong enough to convict the defendant, the conviction rate decreased as participants' experiential scores increased. In contrast, when the evidence was not strong enough, the conviction rate increased as participants' experiential scores increased. In other word, the differences that otherwise resulted from variations in the strength of evidence were attenuated as mock jurors relied more on their experiential systems.

The finding that the severity of charges impact participants' re-evaluation of evidence is inconsistent with a previous study (Simon et al., 2004, study 2). The researchers found no effect of the severity of charge on verdict. However, they did not investigate the interaction between the types of charges with coherence shift.

In addition, the results of study 2 partially replicated and extended Simon and colleagues' (2001, study 1) study which found that a coherence shift occurred before participants made decisions. However, they did not find that an increased coherence shift after participants made verdicts. Those authors argued that a coherence shift might occur before participants reached verdicts. In contrast, the current study replicated the result that a coherence shift occurred even before participants reached verdicts. In addition, the current study found that coherence shift increased after participants made decisions. One possible explanation is that Simon et al. used a civil case (libel) but I used criminal case (theft or murder). Presumption of innocence in criminal case could be a significant difference at the beginning of participants' decision making.

One limit to this research was the ambiguity in the definition of "severity of charge." Severity of charge is difficult to define because there are many possible dimensions in criminal

charges and this experiment used only two types of crime (theft and murder). I assumed that jurors see murder as a more severe charge than theft. Future studies should include a broader range of crimes in the investigation. Future research should investigate more types of crime that vary in their seriousness.

Overall, these results strongly support the impact of question order on mock jurors' decision styles. When mock jurors assessed arguments and estimated the probability of guilt before they made verdicts, they used more rational decision processing. In contrast, when mock jurors decided verdicts first, they seemed to rely more on intuitive processing.

However, these results only partially supported the notion that the severity of charges influence mock jurors' decision styles. The severity of charges reduced conviction rates but did not impact coherence shift. The only significant interaction was that a coherence shift reduced the effect of strength of evidence on verdicts when the charge was less serious (i.e., theft). In contrast, the coherence shift did not reduce the effect of evidence on verdicts when the charge was more serious. It is clear that the way in which severity of charges influences jury decision making is complicated and likely involves more than altering the use of objective and subjective probabilities.

Chapter 8: General Discussion

Summary of results

The present research set out to explore a hierarchical judgment model in which jurors' transient emotions and the severity of the charges influence their decision styles, final verdicts, and re-evaluation of evidence. It also sought to examine whether the order of questionnaires impact mock jurors' use of systematic information processing or intuitive information processing.

Study 1 aimed to test whether experiencing particular emotions could change mock jurors' decision styles and their application of the standard of proof. *Dual processing theories* and *the appraisal tendency framework* predict that if people experience sadness and uncertainty, they will process information more systematically. On the other hand, if they experience anger, they will feel a stronger sense of certainty and process more intuitively.

The results indicate that sad participants were sensitive to variations in the standard of proof, such that the *beyond a reasonable doubt* standard decreased conviction rates compared to the *preponderance with evidence* standard. In contrast, angry participants were less sensitive to changes in the standard of proof. There was no difference in conviction rates between these two standards among angry participants (supporting *Hypothesis 1*).

Another finding from Study 1 is that when participants were required to apply a higher standard of proof, those experiencing anger increased their subjective certainty by increasing coherent re-evaluation of evidence. Conversely, those in the sad condition did not manifest coherent re-evaluation of evidence when applying a heightened standard of proof (supporting *Hypothesis 2*).

Lastly, variations in the strength of evidence influenced verdicts only when sad participants applied the beyond a reasonable doubt standard. In contrast, variations in the strength of evidence affected participants less in the anger condition (partially supporting *Hypothesis 3*).

Study 2 aimed to test whether the severity of the charge against the defendant can influence mock jurors' fact processing and verdicts. *Dual processing theories* and *the severity leniency theory* suggest that mock jurors faced with severe charges would be more motivated to decide cases more accurately than would mock jurors faced with less severe charges. I also

predicted that the order of questionnaires could work as a "task cue" and induce mock jurors' decision style towards either systematic processing or intuitive processing.

Study 2 produced several interesting results. First, severe charges significantly reduced the conviction rate (supporting *Hypothesis 4*). This result is consistent with the *severity leniency hypothesis* (Kerr, 1978; R.J. Simon, 1969), which asserts that jurors (and juries) become less likely to convict a person as the penalty associated with conviction increases.

Second, there was an interaction between severity of charge and strength of evidence on coherence shift (supporting *Hypothesis 5*). Participants in the theft condition increased coherence shifts to compensate for weak evidence. In contrast, participants who evaluated a murder charge did not show any differences in coherence shifts across variations in the strength of evidence. This result suggests that participants who evaluated a less severe charge increased their coherence shifts when the evidence was weak, perhaps attempting to justify their decisions

Third, only participants in the less severe condition (i.e., theft) were sensitive to variations in the strength of the evidence when deciding verdicts (contrary to *Hypothesis 7*). This result might reflect the already low conviction rate for murder charges, even with strong evidence. Participants might have concluded that even strong evidence was insufficient to convict on the murder charge.

Fourth, the finding of a change in coherence shift fit well with expectations from *task cueing*, which predicted that participants' decision processes would follow the sequence of tasks (supporting *Hypothesis 6*). Participants who assessed arguments and probability of guilt, and who rated threshold of guilt before they made verdicts (as the mathematical model described), showed less coherence shift compared to when they made verdicts before evaluating evidence. The finding that coherence shifts increase when the mock jurors reached decisions without first

evaluating the threshold questions again suggests a strategy in which participants were attempting to justify their decisions when they were less certain of their accuracy.

Fifth, participants' verdicts in the assessing-arguments-first condition were more sensitive to variations in the strength of evidence than were participants in the verdict-first condition (*Hypothesis 8*).

Overall, these findings support the theories that jurors use two distinctly different processing methods depending on the characteristics of their cases, such as emotion, severity of charge, and question order. Specifically, when participants were angry, faced a less serious charge (i.e., theft), or answered verdict first, they relied on intuitive processing (coherence based reasoning). They were less sensitive to variations in the strength of evidence, and they showed increased levels of coherence shift. When participants were angry, higher standards of proof did not decrease conviction rates.

In contrast, when participants were sad, faced a more serious charge (i.e., murder), or assessed arguments first, they relied on systematic processing. They were more sensitive to variations in the strength of evidence, and they showed less coherence shift. In addition, higher standards of proof decreased conviction rate.

Instructing the mock jurors with different standards of proof may have had a significant effect in the desired direction when participants relied on systematic processing. However, these standards did not demonstrate a significant effect when participants relied on intuitive processing. Therefore, even good jury instructions may not be sufficient to influence jurors in the direction that the law intends because the jurors might use different decision styles that vary by contextual factors such as emotion, severity of the charge, and the order of the judgments that they make.

Implications for psychology

Emotion. The effects of transient emotions on mock jurors' fact-finding and verdicts were mostly in line with previous findings regarding the impact of emotions on judgment and decision-making (Bodenhausen, 1994; Lerner & Keltner, 2001). Sad participants were more likely to rely on systematic information processing, and they were sensitive to changes in the standard of proof and strength of evidence (*Hypothesis 1, 3*). The size of the coherence shift did not reduce the effect of standards of proof (*Hypothesis 2*).

In contrast, angry participants were more likely to rely on intuitive information processing. They were less sensitive to changes in the standard of proof and strength of evidence. The size and direction of the coherence shift reduced the effect of standards of proof, such that when angry mock jurors applied a higher standard of proof, they simply increased their certainty by using a coherence shift rather than carefully assessing the evidence.

Contrary to *Hypothesis 2*, there was no mediation effect of certainty on the size of and direction of the coherence shift. One possible explanation could be that participants in the sadness condition might have experienced similar levels of certainty to the participants in the anger condition. Tiedens and Linton (2001) demonstrated that sadness falls more toward the middle of the certainty spectrum while fear falls closer to the uncertainty end. Therefore, people who feel sad may not experience enough uncertainty to trigger systematic processing.

However, mediation analyses provided evidence that individual control mediated the effect of emotion on coherence shifts. Anger increased individual control appraisals, such that mock jurors feel that other people are blameworthy for the events that anger them. Such cognitive appraisal increased the coherence shift toward their decision. This result is consistent with previous findings that the biggest difference between angry and sad participants is the control appraisal (Tiedens & Linton, 2001).

This result is consistent with Thagard's (2006) hypothesis that hot cognitions play an important role in coherence reasoning suggesting that people make judgments by not only coming to a rational conclusion but also by making emotional assessments. For example, a juror who likes the defendant would evaluate evidence sympathetically in favor of their affinity to that defendant. The current study demonstrated that transient and unrelated emotions triggered by mock jurors' personal events before the trial can induce them to rely on systematic information processing or intuitive information processing. Eventually, such emotions influence their decision processing and re-evaluation of evidence.

The current research complements Glockner and Engel's (2013) work and extends it with the use of an emotion manipulation in which participants wrote down personal experiences that made them angry or sad. Glockner and Engel found that the standard, "beyond a reasonable doubt" decreased conviction rates compared to the standard, "preponderance of evidence." They argued that the current jury instructions work as the law intends them to work. However, the current study demonstrated that the instruction to only convict when the evidence is "beyond a reasonable doubt" reduced conviction rates only when the jurors were sad. Angry mock jurors recognized the difference in standards of proof, but conviction rates did not decrease. Instead, coherence shift increased to meet a higher standard of proof.

Severity of charge. The result of study 2 is consistent with the *severity leniency hypothesis* (Kerr, 1978; R.J. Simon, 1969), which asserts that jurors (and juries) become less likely to convict a person as the penalty associated with conviction grows.

I further hypothesized that the severity of the charge would increase participants' motivations for accurate judgments and induce mock jurors to use systematic processing. Unfortunately, the experimental results did not provide evidence of the hypothesized mediation

effect of an accuracy motivation. This lack of evidence might come from the measurements of accuracy motivation. Most studies in dual processing theories have used a manipulation of accuracy motivation (e.g., Chaiken, 1984) rather than relying on the participants' self-report.

Nonetheless, I found some evidence that participants who evaluated a more serious charge relied on systematic processing. The results support the prediction that participants evaluating theft charges simply increased their confidence when re-evaluating evidence, in order to compensate for weak evidence. In contrast, the participants evaluating murder charges did not show a difference in coherence shift when analyzing the strength of the evidence. In other words, increased coherence shifts lead to a reduced sensitivity for differences in the strength of evidence only when the charge was theft (but not if it was murder). These results are consistent with a previous study (Rind et al., 1995) which shows that more serious charges induces jurors to follow the law (i.e., ignore inadmissible evidence).

Question order. The finding of a change in coherence shift fit well with expectations from *task cueing*, which predicted that participants' decision processes would follow the sequence of tasks (*Hypothesis 6*). Participants who assessed argument and probability of guilt and rated threshold of guilt before they made verdicts showed less coherence shift than those who made verdicts before evaluating evidence.

The findings of coherence shift in different sequences of questionnaires is especially significant for research on the processes of decision making, particularly in light of cognitive dissonance theory's assertion that re-evaluation does not begin before the formation of commitment (Festinger, 1957, 1964). The results of study 2 provide some evidence that coherence shift occurs after participants offer verdicts, as cognitive dissonance predicted (Festinger, 1964).

However, changing the order of the questionnaires did not completely reduce coherence shift. Svenson's (1999) model of differentiation and consolidation closely parallels this finding. Differentiation and consolidation theory assumes that people make decisions to achieve specific goals. The decision processes aim to make one alternative sufficiently superior to other alternatives. The apparent superiority of the preferred alternative results from the application of decision rules and restructuring processing. Svenson calls this "differentiation." The decision-maker gradually differentiates one alternative from the others until the degree of differentiation is sufficient for a decision. Once they make decisions, "consolidation" occurs to make their choices more reasonable. The central assumption of this theory is that sufficient restructuring of information protects the decision-makers from internal (e.g., change of values) and external (e.g., poor outcome) threats. According to differentiation and consolidation theory, coherence shifts can occur through pre- and post-decision processes. The current study supports this differentiation and consolidation theory because coherence shifts occurred before participants made decisions but changing the order of questionnaires reduced the participants' coherence shift.

The present findings are inconsistent with Simon and his colleagues' findings (Simon, Pham, & Holyoak, 2001, study 1) in which they found no significant effect of changing the order of questionnaires. One explanation for the disparate results might be the difference in case scenarios between the previous study and the current research. The previous study used a civil case (*libel*) and a civil standard of proof (*preponderance of evidence*). In contrast, the current research used a criminal case (*theft* and/or *murder*) and a criminal standard of proof (*beyond a reasonable doubt*). In a criminal case, the law explicitly requires jurors to begin evaluating evidence with the assumption of innocence. There are several other differences between civil and criminal cases but one possible explanation for this inconsistent result could be the decision

options for jurors. Following the differentiation and consolidation theory, jurors might try to compare the two sides in civil cases to see which party (plaintiff or defendant) is more reasonable. However, for a criminal case, they would have to compare guilty/not guilty options. They might consider how much evidence increases their certainty in their judgments. Furthermore, jurors do not need to explain their reasoning for a not guilty verdict.

The results of study 2 suggest that the sequence of questionnaires can influence decision styles. Usually, researchers have used verdict-first and verdict confidence measures because verdict is the main concern of research and the rendering of a verdict is the only role of an actual juror. This sequence is convenient for researchers because quantitative dependent variables are preferred for statistical analysis. The current study suggests that this sequence influences jurors' decision processing. Studies that use juror simulation should be more cautious of the sequence of outcome variables and the interpretation of the results of outcomes.

Implications for law

These results complicate our understanding of the jurors' application of standards of proof and provide a possible explanation of jurors' confusion about those instructions. Jurors may fail to follow certain legal standards because their decision-making processes run counter to the approach that the law endorses. For some types of crimes, jurors may feel they know best how to make a decision and will therefore use their own intuition rather than the rational decision process outlined in the law. They may be most likely to do this when they feel anger or react with generic prejudice to the type of crime under consideration. Judges will not easily recognize the different decision processes that jurors use, and as a result, standards of proof may remain confusing to jurors.

Second, as expected, the threshold of conviction measured by participants' self-report was higher with the *beyond a reasonable doubt* instruction than the *preponderance of evidence* instruction. However, there are some additional issues regarding the threshold of conviction. The difference of threshold between preponderance of evidence and beyond a reasonable doubt was surprisingly small, as previous studies have shown (R. J. Simon & Mahan, 1970; Glockner & Engel, 2013). Legal professionals (e.g., Weinstein & Dewsbury, 2006) argued that the minimum threshold of guilt under *beyond a reasonable doubt* is as high as the 90 percent range (i.e., 95 %). In addition, federal judges estimated the probability for "beyond a reasonable doubt" was 90.28 percent and "preponderance of evidence" was 55.33 percent (McCauliff, 1982). Whereas, in the current study, the average probability of guilty under *beyond a reasonable doubt* was 84.81 percent and under *preponderance of evidence* was 70.13 percent. Hence, contrary to legal experts, participants lowered the threshold for *beyond a reasonable doubt* (study 1 and study 2), and heightened it for *preponderance of evidence* (study 1).

Third, this study instructed the jurors that they must be *firmly convinced* to meet the *beyond a reasonable doubt* standard. Previous studies (Horowitz & Kirkpatrick, 1996) have found that this instruction gives better information than other types of jury instructions. However, this research found that even good jury instructions might not give clear meaning to jurors because they might use different decision styles triggered by trial context, such as emotion, severity of the charge, and the order of the judgments that they make.

Changing the order of interrogatories may be useful when instructing actual jurors. Like the flow chart suggested by prior researchers (Wiener et al., 2004), changing of the sequence of questions and issues could guide jurors in summarizing evidence, understanding the meaning of the standard of proof, and reaching verdicts as law requires. Usually, the jury instruction is the

only way to guide jurors to follow the law. However, giving simple questions following the sequence of legal questions might help jurors to understand and apply the law in the desired way.

Strengths of this study

This research has several strengths. First, participants were jury-eligible citizen from various communities from across the United States. The previous study (Glockner & Engel, 2013) that examined this issue used German undergraduate students as participants, which could be problematic for several reasons. There is no *preponderance of evidence* standard in German law. With some exceptions, a reasonable doubt standard applies to private, criminal, and public law issues (Koriott, 1998). Therefore, German participants might be unfamiliar with the *preponderance of evidence* standard. Undergraduates are generally between the ages of 18 to 22 years old and are not representative of the average juror. Empirical evidence suggests that students sometimes make different legal decisions than community-representative jurors (Wiener, Krauss, & Lieberman, 2011). By using a sample of adults from across the country, the sample is more similar to actual jurors than studies using undergraduate participants.

Second, vignettes and judgment tasks used in this study were largely similar to previous studies (Simon et al., 2004, Simon &Scurich, 2011; Glockner & Engel, 2013). Therefore, the results of this study can be compared to and extend previous findings.

Third, previous studies did not control for individual differences on fact-finding processing and judgment. In this study, I included participants' individual differences in decision-making by using the rational experiential inventory (Pacini & Epstein, 1999). Although the only significant effect of individual differences was that experiential scores reduced the impact of culpability on verdicts, controlling for individual differences produced a model that better fit the data.

Finally, previous studies (Engel & Glockner, 2013, Simon et al., 2004) did not report the interactions and only focused on the main effect for type of charge (Simon et al., 2004, study 1) or the main effect for change in standards of proof (Engel & Glockner, 2013). However, this study focused on interactions, which produced interesting results. In addition, this research provided some evidence that mock jurors use different information processing approaches depending upon context and extra-legal factors.

Limitations and Qualifications

There are, of course, several caveats and qualifications. First, even though others have used these materials in previous studies (e.g., Glockner & Engel, 2013; Simon et al., 2004, 2011), the current study is not equivalent to an actual trial. For example, participants did not see oral arguments between the defense and the prosecution, and both study 1 and study 2 lasted less than 50 minutes. Furthermore, jurors did not deliberate in the current study. Although jurors should understand standards of proof by themselves, they may gain greater incite when they deliberate as a group. It is likely that jurors in a real trial would demonstrate different patterns of decision-making.

Second, participants finished both studies over the internet at their own pace. Although the survey took precautions to ensure that participants took their time and paid attention to the studies, and several validation questions (e.g., *While watching the television, have you ever had a fatal heart attack?*) were included, it is still possible that some participants did not carefully read the trial scenario and decided arbitrarily.

Third, the current study used two crimes that are comparable in terms of physical damage and cruelty to make the perceived difference between them significant. Additional research should test which dimensions of seriousness perception can affect mock jurors' decisions.

Multidimensional scaling could be useful to determine how people perceive the specific elements of the seriousness of crime.

Fourth, the severity of charge did not increase accuracy motivation based on the manipulation checks, which showed participants. The severity manipulation checks did show that participants did view murder as a more serious crime than theft. However, the severity of charge had no overall significant difference on the motivation accuracy measures. Therefore, it is possible that manipulating the type of charge (murder vs. theft) had little effect on accuracy motivation. Real jurors might feel responsibility about the fate of an actual defendant but the mock jurors in this study might have failed to show the accuracy motivation differences because they were not making judgments about a real defendant. Another possibility is that this lack of effect is a result of the ineffectiveness of the explicit accuracy motivation measures. These measures are high social desirability so that the respondents may have all claimed to be motivated to be accurate to look good and to fulfill the researcher's expectation. If this is the case, some other indirect measure of accuracy motivation might have shown better results. Future studies should manipulate the accuracy motivation more directly than simply providing different charges.

Fifth, this research by itself could not control for the ambiguity of the definition of severity of a charge. A previous study (e.g., Vidmar, 1972) manipulated only severity of the charge for only one crime (i.e., homicide). The current study did not separate the severity of the charge from the type of crime in order to maximize the difference between the two conditions. Instead, it relied on a seriousness manipulation check to measure participants' perception of seriousness. Future research should separate the type of crime from the seriousness of the charge.

Sixth, there is a possible confounding effect because the number of verdict(s) that the participants rendered was different in the murder and theft conditions. In the less severe charge condition, participants answered only a theft verdict. In contrast, in the more severe condition, participants decided both the theft charge and the murder charge. Although two charges are more severe than one charge, it is not clear exactly what severity of charge actually meant to the participants. While this may be either considered an internal validity problem of a labelling problem (construct validity), it certainly contributes to the fuzziness of the definition of charge severity. Charge severity is confused more generally in the literature and researchers need to develop a better way of manipulating that variable. This confusion may have contributed to the lack of significant results for the severity of charge in this study. That is, the threshold for theft condition could be same with the threshold for murder condition because participants answered theft charges in both condition. Future studies should seek better definitions for the severity of charge factor.

Seventh, one goal of this research was to investigate whether individual differences were a possible mechanism to explain coherence shifts. Unfortunately, the experimental results did not find any relationship between cognitive processing styles measured by rational experiential inventory and the size or direction of coherence shifts. Although rationality scores had a marginal effect on coherence shift in study 1, study 2 failed to replicate that finding.

This lack of a relationship might indicate that rational experiential inventory did not measure individual differences that contributed to participants' decision style (i.e., coherence shift). This research did not provide any evidence whether or not individual differences of the decision maker mediate coherence shifts. Future research should investigate alternative individual difference measures to search for coherence shift mediation. As Simon (2004)

suggested, tolerance for inconsistency might explain coherence based reasoning among jurors. Future research should measure or manipulate individuals' tolerance for inconsistency.

Finally, this experiment did not provide a clear answer as to why certainty appraisals were unrelated to jurors' decision styles because it included only two emotions: anger and sadness. That is, it did not include a neutral emotion control group. Future studies should include control groups and a broader range of emotions (e.g., fear and surprise), which show differences in certainty appraisals to better assess the influence of certainty appraisals on coherence shifts.

This study manipulated incidental sadness and anger to isolate emotions in decision-making. Although incidental emotions allow for experimental control, they may influence juror decision making less than do integral emotions. One way to control this effect would be to manipulate or measure integral emotions. Integral emotions are emotions provoked by attributes of a target (Lerner, Han, & Keltner, 2007). This task-generated emotion may be more intense because it comes from one's judgment of a target in question.

Future Research

The focus of this study was to test whether mock jurors' decision-making processes and judgments change as a function of experienced emotion, severity of the charge, or the order of questionnaires. However, in real trials, mock jurors deliberate with other jurors whose opinions can influence decision processing. One possible explanation of the differences between mock jurors and a trial jury is accountability. People make different decisions when they expect to rationalize their decision to others, such as other jurors or a judge (Lerner &Tetlock, 1999). Accountability sometimes attenuates cognitive biases, and causes people to think more carefully and focus on the task. In contrast, sometimes (if other's opinions are well known) accountability

can increase cognitive biases. A future study should manipulate accountability to study jurors' use of subjective and objective probabilities.

Study 1 and study 2 employed a modified version of Ninth Circuit pattern jury instructions, which include the expression, "firmly convinced." However, as study 2 showed, different tasks could induce jurors to use different decision styles. A quantification instruction instead of, or in conjunction with, the typical standard of proof instruction might trigger systematic processing. Previous studies have shown that quantitative or numeric based instructions can reduce mock jurors' confusion in understanding and applying standards of proof (e.g., Kagehiro, 1985).

Finally, future research should examine cultural differences in fact-finding processing. In an international and comparative legal perspective, the American legal system selects objective, serial, and analytic legal decision models. Contrarily, many European and East Asian countries select a more subjective and holistic approach to the fact-finding process. The current study provides some evidence that mock jurors process facts differently based on emotion, the severity of the charge, and the order of questionnaires. Future studies could investigate these differences across cultures. If people follow different decision-making processes based on their culture of origin, simply implanting a standard of proof may not work as the law intends.

Final conclusions

The results of these studies suggest that jurors' misunderstanding of standards of proof may arise from the different types of fact-finding processes they use. Trial characteristics (i.e., severity of charge) or incidental emotion might contribute to differentiating jurors' decision-making. While it may be possible to measure and even control these characteristics in the research environment, it is difficult to separate or control these properties in actual trials.

Therefore, simple word changes might not reduce mock jurors' misunderstandings about standards of proof.

This research contributes the growing body of literature addressing the important issue of whether mock jurors use two different methods for processing information. There is still too little evidence to reach a conclusion regarding whether jurors use different processing modes. However, the results of this research suggest that jurors may use different processing styles that shift according to incidental emotion, severity of charge, and the order of questionnaires. The meaning of the standard of proof may also differ according to the individual's processing style. Additional research will allow us to better understand jurors' fact-finding processing and how they use different standards of proof.

References

Addington v. Texas, 441 U.S. 418 (1979)

Allen, R. J., & Lively, S. (2003). Burdens of persuasion in civil cases: Algorithms v. Explanations. *Michigan State Law Review,* 2003, 893-944.

Bandes, S. A. (2009). Emotions in context: exploring the interaction between emotions and legal institutions: repellent crimes and rational deliberation: emotion and the death penalty. *Vermont Law Review, 33,* 489-518.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology,* 71, 230-244.

Birnbaum, M.H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology, 55,* 803–832.

Blackstone, S., Sharswood, G., & Field, B. (1760). *Commentaries on the laws of England: in four books*: Childs & Peterson.

Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgment: The differential impact of anger and sadness. *European Journal of Social Psychology, 24*(1), 45-62.

Brilmayer, L. & Kornhauser, L. (1961). Review: Quantitative methods and legal decisions, *The University of Chicago Law Review, 46,* 116-153.

Buehler, R., Griffin, D., & Ross, M. (2002). Inside the planning fallacy: The causes and consequences of optimistic time predictions. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment.* (pp. 250-270). Cambridge: Cambridge University Press.

Cage v. Louisiana, 498 U.S. 39 (1990)

Cacioppo, J. T., Petty, R. E., Kao, C. F., & Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology*, 51(5), 1032-1043.

Chaiken, S., Axsom, D., Liberman, A., & Wilson, D. (1992). *Heuristic processing of persuasive messages; Chronic and temporary sources of rule accessibility*. Unpublished manuscript, New York University.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended Thought.* (pp. 212-252). New York, NY: Guilford Press.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument, ambiguity, and task importance on attitude judgment, *Journal of Personality and Social Psychology, 66*(3), 460-473.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752-766.

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

Chen, S., Shechter, D., & Chaiken, S. (1996). Getting at the truth or getting along: Accuracy-versus impression-motivated heuristic and systematic processing. *Journal of Personality and Social Psychology, 71(2),* 262-275.

Clermont, K. (1986). Procedure's magical number three psychological bases for standards of decision. *Cornell Law Review, 72,* 1115-1156.

Clermont, K. (2004). Standards of proof in Japan and the United States. *Cornell International Law Journal, 37*(2), 4-21.

Clermont, K., & Hall, M. (2008). Standards of proof revisited. *Vermont Law Review, 33,* 469-488.

Clermont, K., & Sherwin, E. (2002). A comparative view of standards of proof. *The American Journal of Comparative Law, 50*(2), 243-275.

Cockburn (1988). Twelve silly men? The trial jury at assizes 1560-1670. In Cockburn and Green (eds.), *A retrospective on the criminal trial jury: 1200-1800* (pp. 158-81). London.

Cohen, L. J. (1981). Subjective probability and the paradox of the gate-crasher, *Arizona State Law Journal, 1981,* 627-634.

Commonwealth v. Rosa, 422 Mass. 18, 661 N.E.2d 56 (1996).

Damasio, A., Everitt, B., & Bishop, D. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex [and discussion]. *Philosophical transactions: Biological sciences, 351*(1346), 1413-1420.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology, 66,* 819-829.

Devine, D. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(2), 5-18.

Devine, D., Clayton, L., Dunford, B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law, 7*(3), 622-727.

Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 339-360). New York: Guilford Press.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without attention effect. Science, 311, 1005–1007.

Eagly, A. H., & Chaiken, S. (1993*). The psychology of attitudes.* New York: Harcourt Brace Jovanovich.

Engel, C. (2009). Emotions in context: Exploring the interaction between emotions and legal institutions: preponderance of the evidence versus intime conviction: a behavioral perspective on a conflict between american and continental european law. *Vermont Law Review*, 33, 435-467.

Engel, C., & Glockner, A. (2012). Role-induced bias in court: An experimental analysis. *Journal of Behavioral Decision Making*, Published online in Wiley Online Library.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49, 709-424.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, 71, 390-405.

Eskine, L. J., Kacinik, N. A., & Prinz, J. J. (2011). A bad taste in the mouth. *Psychological Science, 22(3),* 295-299.

Evans, J. (2009). How many dual-process theories do we need? One, two, or many? In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond.* (pp. 33-54). New York, NY: Oxford University Press.

Faigman, D. L., & Baglioni, A. J., Jr. (1988). Bayes' theorem in the trial process: Instructing jurors on the value of statistical evidence. *Law and Human Behavior,* 12(1), 1-17.

Feigenson, N. (2009). Emotional influences on judgments of legal blame: How they happen, whether they should, and what to do about it. In B. Bornstein & R. Wiener (Eds.), *Emotion and the Law* (pp. 45-96).

Finkel, N. J. (1995). *Commonsense justice: jurors' notions of the law* (Vol. 400). Cambridge, MA: Harvard University Press.

Finkel, N. J., & Groscup, J. L. (1997). When mistakes happen: Commonsense rules of culpability. *Psychology, Public Policy, and Law, 3*(1), 65-125.

Fisher, W. W.(2004). *Promises to keep: Technology, law, and the future of entertainment*. Stanford, CA: Stanford University Press

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117*(1), 39-66.

Frankish, K., & Evans, J. (2009). The duality of mind: An historical perspective. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond.* (pp. 1-29). New York, NY: Oxford University Press.

Freedman, J., Krismer, K., MacDonald, J., & Cunningham, J. (1994). Severity of penalty, seriousness of the charge, and mock jurors' verdicts. *Law and Human Behavior, 18*(2), 189-202.

Friedman, R. D. (1998). Review: Anchors and flotsam: Is evidence law "adrift"? *The Yale Law Journal, 107*(6), 1921-1967.

Friedman, R. D., & Yates, J. (2003). The triangle of culture, inference, and litigation system. *Law, Probability and Risk, 2*(2), 137-150.

Georges, L. C., Wiener, R. L., & Keller, S. R. (2013). The angry juror: sentencing decisions in first-degree murder. *Applied Cognitive Psychology*, *27*(2), 156-166

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability*. (pp. 129-161). Oxford, England: John Wiley & Sons.

Glockner, A., & Engel, C.(2013). Can we trust intuitive jurors? Standards of proof and the probative value of evidence in coherence based reasoning, *Journal of Empirical Legal Studies*, *10*(2), 230-252.

Göritz, A.S. (2007). The induction of mood via the WWW. *Motivation and Emotion*, *31*, 35–47.

Göritz, A.S., & Moser, K. (2006). Web-based mood induction. *Cognition and Emotion, 20*, 887–896.

Griffin, D., & Tversky, A. (2002). The weighing of evidence and the determinants of confidence. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment.* (pp. 230-249). New York, NY: Cambridge University Press.

Groscup, J., & Tallon, J. (2009). Theoretical models of jury decision-making. In J. Lieberman & D. Krauss (Eds.), *Jury psychology: social aspects of trial processes: psychology in the*

courtroom, Vol. 1: Psychology of the Courtroom (pp. 41-66). Burlington, VT: Ashgate Publishing Company.

Haidt, J . (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review. 108*, 814-834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*, 998-1002.

Hans, V., Kaye, D., Dann, B., Farley, E., & Albertson, S. (2011). Science in the jury box: jurors' comprehension of mitochondrial DNA evidence. *Law and Human Behavior, 35(1)*, 60-71.

Hastie, R. (1993). Algebraic models of juror decision processes *Inside the juror: The psychology of juror decision making.* (pp. 84-115). New York, NY: Cambridge University Press.

Hastie, R (2001). Emotions in juror decisions. *Brooklyn Law Review, 66*, 991-1009.

Higgins, E. T., & Brendl, C. M. (1995). Accessibility and applicability: some "activation rules" influencing judgment. *Journal of Experimental Social Psychology, 31*(3), 218-243.

Higgins, E. T, Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*, 141-154.

Holland v. United States, 348 U.S. 121 (1954).

Holyoak, K.J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General, 128*, 3–31.

Horowitz, I. A., & ForsterLee, L. (2001). The effects of note-taking and trial transcript access on mock jury decisions in a complex civil trial. *Law and Human Behavior, 25*(4), 373-391.

Horowitz, I. A., ForsterLee, L., & Brolly, I. (1996). Effects of trial complexity on decision making. *Journal of Applied Psychology, 81*(6), 757-768.

Horowitz, I. A., & Seguin, D. G. (1986). The effects of bifurcation and death qualification on assignment of penalty in capital crimes. *Journal of Applied Social Psychology, 16*(2), 165-185.

Horowitz, I. A., & Kirkpatrick, L. C. (1996). A concept in search of a definition: The effects of reasonable doubt instructions on certainty of guilt standards and jury verdicts. *Law and Human Behavior, 20*(6), 655-670.

Horowitz, I. A., & Seguin, D. G. (1986) The effects of bifurcation and death qualification on assignment of penalty in capital crimes. *Journal of Applied Social Psychology*, *16*(2), 165-185.

In Re Winship, 397 U.S. 358 (1970).

Inbar, Y., Cone, J., & Gilovich, T. (2010). People's intuitions about intuitive insight and intuitive choice. *Journal of Personality and Social Psychology, 99*(2), 232-247

Kagehiro, D. K. (1990). Defining the standard of proof in jury instructions. *Psychological Science, 1*(3), 194-200.

Kagehiro, D. K., & Stanton, W. C. (1985). Legal vs. quantified definitions of standards of proof. *Law and Human Behavior, 9*(2), 159-178.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697-720.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In Gilovich, T., Griffin, D., Kahneman, D. *Heuristics and Biases: The Psychology of Intuitive Judgmen.* (pp. 49-81). Cambridge: Cambridge University Press.

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

Kassin, S. M., & Garfield, D. A. (1991). Blood and guts: General and trial-specific effects of videotaped crime scenes on mock jurors. *Journal of Applied Social Psychology, 21*(18), 1459-1472.

Kaye, D. (1999). Clarifying the burden of persuasion: what Bayesian decision rules do and do not do. *International Journal of Evidence & Proof, 3*(1), 1-28.

Kaye, D. (2000). Bayes, burdens and base rates. *International Journal of Evidence & Proof, 4*, 260-267.

Kerr, N. L. (1978). Severity of prescribed penalty and mock jurors' verdicts. *Journal of Personality and Social Psychology, 36*(12), 1431-1442.

Kerr, N. L., Atkin, R. S., Stasser, G., Meek, D., Holt, R. W., & Davis, J. H. (1976). Guilt beyond a reasonable doubt: Effects of concept definition and assigned decision rule on the judgments of mock jurors. *Journal of Personality and Social Psychology, 34*(2), 282-294.

Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology, 63*(4), 534-544.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, 3(3), 172-179.

Koch, C., & Devine, D. (1999). Effects of reasonable doubt definition and inclusion of a lesser charge on jury verdicts. *Law and Human Behavior, 23*(6), 653-674.

Koehler, J. J., & Shaviro, D. N. (1990). Verdical verdicts: increasing verdict accuracy through the use of overtly probabilistic evidence and method, *Cornell Law Review, 75*, 247-279.

Koehler, J. J. (2001). When do courts think base rates statistics are relevant? *Jurimetrics, 42*, 373-402.

Krauss, D. A., Lieberman, J. G., & Olson, J. (2004). The effects of rational and experiential information processing of expert testimony in death penalty cases. *Behavioral Sciences and the Law*, *22*, 801–822.

Krauss, D. A., McCabe, J. G., & Lieberman, J. D. (2012). Dangerously misunderstood: Representative jurors' reactions to expert testimony on future dangerousness in a sexually violent predator trial. *Psychology, Public Policy, and Law, 18*(1), 18-49.

Langbein, J. (2003). *The origins of adversary criminal trial*. Oxford: Oxford University Press.

Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science*, *14*(2), 144-150.

Lerner, J. S., Han, S., & Ketlner, D. (2007). Feelings and consumer decision making: Extending the appraisal-tendency framework. *Journal of Consumer Psychology, 17(3)*, 184-187.

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81*(1), 146-159.

Lerner, J. S., Tiedens, L. Z. (2006). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of Behavioral Decision Making, 19*, 115-137.

Levine, M. (1998). Do standards of proof affect decision making in child protection investigations? *Law and Human Behavior, 22*(3), 341-347.

Lieberman, J. D., & Krauss, D. A. (2009). The psychology of the jury instruction process *Jury psychology: Social aspects of trial processes: Psychology in the courtroom, Vol. 1*. (pp. 129-155). Burlington, VT: Ashgate Publishing Company.

Locke, J., & Leyden, W. (2002). *John Locke: Essays on the law of nature : the Latin text with a translation, introduction, and notes ; together with transcripts of Locke's shorthand in his journal for 1676*. Oxford: Oxford University Press.

McCauliff, C. (1982). Burdens of proof: Degrees of belief, quanta of evidence, or constitutional guarantees. *Vanderbilt law review, 35*, 1293.

Montgomery, J. (1998). The criminal standard of proof. *New Law Journal, 148*, 582-585.

Moskowitz, G. B. (1996). The mediational effects of attributions and information processing in minority social influence. *British Journal of Social Psychology, 35*(1), 47-66.

Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science, 5*(1), 2-12.

Nagel, S. (1979). Bringing the values of jurors in line with the law. *Judicature, 63*, 189-195.

Napier, J., & Tyler, T. (2008). Does moral conviction really override concerns about procedural justice? A reexamination of the value protection model. *Social Justice Research, 21(4)*, 509-528.

Nesson, C. (1986). Agent Orange meets the Blue Bus: Factfinding at the frontier of knowledge. *Boston University Law Review, 66*, 521-540.

Niedermeier, K. E., Kerr, N. L., & Messé, L. A. (1999). Jurors' use of naked statistical evidence: Exploring bases and implications of the Wells effect. *Journal of Personality and Social Psychology, 76*(4), 533-542.

Ornelas v. United States, 517 US 690 (1996).

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.

Pacini, R., & Epstein, S. (1999). The relation of rational and Experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology,* 76, 972-987.

Pacini, R., Muir, F., & Epstein, S. (1998). Depressive realism from the perspective of cognitive-experiential self-theory. *Journal of Personality and Social Psychology, 74*(4), 1056-1068.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51*(2), 242-258.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*(2), 189-206.

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

People v. Ibarra, 2001 WL 1330296 (2001)

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219-235.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York, NY: Oxford University Press.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(5), 888-896.

Rind, B., Jaeger, M., & Strohmetz, D. B. (1995). Effect of crime seriousness on simulated jurors' use of inadmissible evidence, *Journal of Social Psychology, 135*(4), 417-424.

Rozin, P., Haidt, J., & McCauley, C. R. (2008). Disgust. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions, 3rd ed*. (pp. 757-776). New York, NY: Guilford Press

Sargent v. Massachusetts Accident Co., 307 Mass. 246 , 250 (1940)

Schacter, D. L. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 501-518.

Scheinman, M. F. (1977). *Evidence and Proof in Arbitration*. Ithaca, NY: Cornell University Press.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin, 34(8),* 1096-1109.

Schum, D. (2009). A science of evidence: contributions from law and probability. *Law, Probability and Risk, 8*(3), 197-231.

Schwarz, N., Gilovich, T., Griffin, D., & Kahneman, D. (2002). Feelings as information: Moods influence judgments and processing strategies, *Heuristics and biases: The psychology of intuitive judgment.* (pp. 534-547). New York, NY: Cambridge University Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

Semmler, C., & Brewer, N. (2002). Effects of mood and emotion on juror processing and judgments. *Behavioral Sciences & the law, 20(4),* 423-436.

Severance, L. J., & Loftus, E. F. (1982). Improving the ability of jurors to comprehend and apply criminal jury instructions. *Law and Society Review, 17*, 153-197.

Shapiro, B. (1991). " Beyond reasonable doubt" and" probable cause": *historical perspectives on the Anglo-American law of evidence*. Berkely and Los Angeles, California: University of California Press.

Shapiro, B. (2008). The beyond reasonable doubt doctrine: 'moral comfort' or standard of proof? *Law and Humanities, 2*(2), 149-173.

Shapiro, B. (2009). Changing language, unchanging standard: from satisfied conscience to moral certainty and beyond reasonable doubt. *Cardozo Journal of Internationall and Comparative Law, 17*, 261-280.

Simon, D. (2004). A third view of the black box: Cognitive coherence in legal decision making. *University of Chicago Law Review, 71*, 511-586.

Simon, D., & Scurich, N. (2011). Lay Judgments of Judicial Decision Making. *Journal of Empirical Legal Studies, 8*(4), 709-727.

Simon, D., Snow, C. J., & Read, S. J. (2004). The Redux of Cognitive Consistency Theories: Evidence Judgments by Constraint Satisfaction. *Journal of Personality and Social Psychology, 86*(6), 814-837.

Simon, D., Stenstrom, D., & Read, S. J. (2011). Blending hot and cold cognitions by constraint satisfaction.

Simon, R. J. (1969). Judges'translations of burdens of proof into statements of probability. *Trial Lawyer's Guide, 13*, 103-114.

Simon, R. J., & Mahan, L. (1970). Quantifying Burdens of Proof-A View from the Bench, the Jury, and the Classroom. *Law & Society Review, 5*, 319-330.

Skeem, J. L., & Golding, S. L. (2001). Describing jurors' personal conceptions of insanity and their relationship to case judgments. *Psychology, Public Policy, and Law, 7*(3), 561-621.

Skitka, L. J. (2002). Do the means justify the ends, or do the ends justify the means? A test of the value protection model of justice. *Personality and Social Psychology Bulletin, 28*, 588–597.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*(4), 813-838.

Smith v. Rapid Transit, Inc. 58 N.E.2d 754 (1945)

Smith, E. R., & DeCoster, J. (2000). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to. *Personality & Social Psychology Review, 4*(2), 108.

Smith, V. L. (1991). Prototypes in the courtroom: Lay representations of legal concepts. *Journal of Personality and Social Psychology, 61*(6), 857-872.

Smith, V. L. (1993). When prior knowledge and law collide. *Law and Human Behavior, 17*(5), 507-536.

Smith, V. L., & Studebaker, C. A. (1993). Prototypes of crimes operate as expectancies for typical information. Paper presented at the Fifth Annual Convention of the American Psychological Society, Chicago, IL.

Srull, *T. K.*, & Wyer, *R. S. (*1989*).* Person memory and judgment. *Psychological Review* , 96(1), 58–83.

Stalans, L. J. (1993). Citizens' crime stereotypes, biased recall, and punishment preferences in abstract cases: The educative role of interpersonal sources. *Law and Human Behavior, 17*(4), 451-470.

Stalans, L. J., & Diamond, S. S. (1990). Formation and change in lay evaluations of criminal sentencing. *Law and Human Behavior, 14*(3), 199-214.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8(3),* 220-247.

Sunstein, C. R. (2002). Probability neglect: Emotions, worst cases, and law. *The Yale Law Journal, 112*, 61-107.

Sunstein, C. R. (2005). moral heuristics. *Behavioral and Brain Sciences, 28*(4), 531-542.

Svenson, O. (1999). Differentiation and consolidation theory: decision making processes before and after a choice. In P. Juslin & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (pp. 175-197). Mahwah, NJ: Erlbaum.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Thagard, P. (2003). Why wasn't O.J. convicted? Emotional coherence in legal inference. Cognition & Emotion, 17 (3), 361-383.

Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials. Law and Human Behavior, 11(3), 167-187.

Tiedens, L. Z., & Linton, S. (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *Journal of Personality and Social Psychology, 81*(6), 973-988.

Tillers, P., & Gottfried, J. (2006). Case comment-United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond A Reasonable Doubt Is Unquantifiable? *Law, Probability & Risk, 5*(2), 135-157.

Tulving, E., Schacter, D.L., & Stark, H.A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(4), 336-342.

Twining, W. (1984). Some scepticism about some scepticisms. *Journal of Law and Society, 11*(2), 137-171.

Twining, W. (1994). *Rethinking Evidence: Exploratory Essays*. Northwestern University Press.

Twining, W. (2007). *General jurisprudence*. Cambridge: University College London.

United States v. Fatico, 458 F. Supp. 388 (1978)

United States v. Feinberg, 140 F.2d 592, 594 (2d Cir. 1944)

United States v. Hall, 854 F.2d 1036 (1988)

United States v. Taylor, 508 U.S. 333 (1993)

Verheyen, C., & Göritz, A.S. (2009). Plain texts as an online mood-induction procedure. *Social Psychology, 40*(1), 6-15.

Victor v. Nebraska, 511 U.S. 15 (1994)

Vidmar, N. (1997). Generic prejudice and the presumption of guilt in sex abuse trials. *Law and Human Behavior, 21*(1), 5-25.

Vidmar, N. (2003). When all of us are victims: Juror prejudice and "terrorist" trials. *Chicago-Kent Law Review, 78*, 1143–1178.

Villejoubert, G., Almond, L., & Alison, L. (2009). Interpreting claims in offender profiles: the role of probability phrases, base-rates and perceived dangerousness. *Applied Cognitive Psychology, 23*(1), 36-54.

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

Wiener, R., Gervais, S. J., Allen, J., & Marquez, A. (2013). In the eye of the beholder: The role of sexual objectification and affective forecasting on sexual harassment judgments. *Psychology, Public Policy, and the Law*, *19(2)*, 206-221.

Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied, 2*(4), 343-364.

Windschitl, P. D., & Wells, G. L. (1998). The alternative-outcomes effect. *Organizational Behavior and Human Decision Processes, 75*, 1411-1423.

Wiener, R. L., Arnot, L., Winter, R., & Redmond, B. (2006). Generic prejudice in the law: Sexual assault and homicide. *Basic and Applied Social Psychology, 28*(2), 145 - 155.

Wiener, R. L., Bornstein, B., & Voss, A. (2006). Emotion and the law: A framework for inquiry. *Law and Human Behavior, 30*(2), 231-248.

Wiener, R. L., Krauss, D. A., & Lieberman, J. D. (2011). Mock jury research: where do we go from here? [Review]. *Behavioral Sciences and the Law, 29*(3), 467-479.

Wiener, R. L., Richmond, T. L., Seib, H. M., Rauch, S. M., & Hackney, A. A. (2002). The psychology of telling murder stories: do we think in scripts, exemplars, or prototypes? *Behavioral Sciences & the Law, 20*(1-2), 119-139.

Windschitl, P. D. (2002). Judging the accuracy of a likelihood judgment: the case of smoking risk. *Journal of Behavioral Decision Making, 15*(1), 19-35.

Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied, 2*(4), 343-364.

Wrightsman, L. S., Greene, E., Nietzel, M. T., & Fortune, W. H. (2002). Jury Trials II: Concerns and Reforms (5th ed., pp. 408-441). Belmont, CA: Wadsworth Group.

Yates, J. F. (1998). Conceptualizing, explaining, and improving accuracy: Process models of probability judgment. *Cognitive Studies, 5*(4), 49-64.

Yates, J. F. (2010). Culture and probability judgment. *Social and Personality Psychology Compass, 4*(3), 174-188.

Tables

Table 1. Participants' Demographic Characteristics in Study 1

Age				
	18-20	8 (4.5%)	41-50	20 (11.2%)
	21-30	84 (47.4%)	51-60	19 (10.7%)
	31-40	38 (21.4%)	61-64	8 (4.5%)
Gender				
	Male	72 (40.9%)	Female	104 (59.1%)
Ethnic Origin				
	Caucasian	148 (83.6%)	Native	2 (1.1%)
	Asian	10 (5.6%)	Latin	1 (.6%)
	African	7 (4.0%)	Other	3 (1.7%)
	Hispanic	6 (3.4%)		
Education				
	Less than high school	1 (.6%)	Bachelor's degree	67 (37.9%)
	High school	51 (28.8%)	Master's degree	20 (11.3%)
	Associate Degree	33 (18.6%)	Profession degree	5 (2.8%)

	Yes	**No**
Are you registered to vote?	162 (91.5%)	15 (8.5%)
Do you have a valid driver license?	160 (90.4%)	16 (9.0%)
Have you served as a juror?		
State-Civil	9 (5.1%)	168 (94.9%)
State-Criminal	12 (6.8%)	165 (93.2%)
Federal-Civil	4 (2.3%)	172 (97.2%)
Federal-Criminal	1 (.6%)	176 (99.4%)
Current working status		
Full time	83 (46.9%)	
Part time	45 (25.4%)	
Unemployed	49 (27.7%)	

Residing State							
AR	3 (1.7%)	IN	4 (2.3%)	NC	9 (5.1%)	SC	1 (.6%)
AZ	5 (2.8%)	KS	4 (2.3%)	NH	1 (.6%)	TN	5 (2.8%)
CA	16 (9.0%)	KY	5 (2.8%)	NJ	1 (.6%)	TX	7 (4.0%)
CO	4 (2.3%)	LA	4 (2.3%)	NM	1 (.6%)	UT	1 (.6%)
CT	3 (1.7%)	MA	7 (4.0%)	NV	2 (1.1%)	VA	7 (4.0%)
DC	1 (.6%)	MD	4 (2.3%)	NY	15 (8.5%)	WA	5 (2.8%)
FL	14 (7.9%)	ME	2 (1.1%)	OH	5 (2.8%)	WI	2 (1.1%)
GA	4 (2.3%)	MI	2 (1.1%)	OK	3 (1.7%)	WV	1 (.6%)
IA	3 (1.7%)	MN	3 (1.7%)	OR	2 (1.1%)	Missing	4 (2.3%)
ID	1 (.6%)	MO	4 (2.3%)	PA	6 (3.4%)		
IL	4 (2.3%)	MS	1 (.6%)	RI	1 (.6%)		

Table 2. Emotional Manipulation Checks

	Anger		Sad		*t* (175)		*95% CI*		*Cohen's*
	Mean	SD	Mean	SD		*p*	*LL*	*UL*	*d*
Anger related emotions									
Angry	4.89	2.35	2.88	2.26	5.81	.000**	1.33	2.70	0.88
Hostile	3.53	2.60	1.86	1.61	5.12	.000**	1.03	2.31	0.78
Irritable	4.06	2.55	2.83	2.39	3.29	.001*	0.49	1.97	0.50
Scornful	3.80	2.74	2.24	2.10	4.23	.000**	0.83	2.28	0.64
Loathing	3.84	2.91	2.40	2.13	3.75	.000**	0.68	2.20	0.57
Sadness related emotions									
Sad	3.83	2.23	5.65	2.43	-5.18	.000**	-2.51	-1.12	-0.78
Blue	2.89	2.08	4.28	2.48	-4.03	.000**	-2.07	-0.71	-0.61
Down -hearted	3.30	2.19	4.67	2.59	-3.79	.000**	-2.09	-0.66	-0.58
Alone	3.20	2.55	3.63	2.47	-1.13	.258	-1.19	0.32	-0.17
Lonely	2.78	2.45	3.40	2.47	-1.67	.098	-1.36	0.12	-0.25
Other emotions									
Happy	3.66	2.04	3.24	1.86	1.44	.151	-0.16	1.00	0.22
Disgusted	4.70	2.70	2.83	2.32	4.94	.000**	1.13	2.62	0.75
Fearful	2.11	1.76	2.82	2.40	-2.22	.028*	-1.33	-0.08	-0.34
Surprised	1.77	1.42	1.60	1.17	.87	.386	-0.22	0.56	0.13
Nervous	2.11	1.77	2.43	2.09	-1.07	.285	-0.89	0.26	-0.16
Alert	6.63	2.15	5.56	2.11	3.32	.001**	0.43	1.70	0.50
Proud	3.18	2.39	2.47	2.24	2.05	.042*	0.03	1.40	0.31
Excited	2.63	2.02	1.86	1.60	2.80	.006**	0.22	1.31	0.42

Note. *$p < .05$ **$p < .01$

Table 3. Emotional Manipulation Checks

	Anger		Sad		t (175)	p	95% CI		d
	Mean	SD	Mean	SD			LL	UL	
Control appraisal measures									
to what extent did you typically feel that someone other than yourself had the ability to influence what was happening?	7.30	1.89	5.93	2.73	3.85	.000**	.665	2.06	.58
to what extent did you typically feel that someone else was to blame for what was happening in the situation?	7.74	1.59	5.08	3.04	7.31	.000**	1.94	3.38	1.10
to what extent were the events beyond anyone's control?(reversed)	6.72	2.33	4.69	2.83	5.21	.000**	1.26	2.79	.79
Certainty appraisal measures									
how well did you understand what was happening in those situations?	7.45	1.62	7.25	1.98	.73	.46	-.34	.74	.11
how uncertain were you about what would happen in these situations?(reversed)	5.17	2.35	4.52	2.77	1.67	.10	-.12	1.41	025
how well could you typically predict what was going to happen next?	5.46	2.37	4.27	2.37	3.33	.001**	.49	1.89	.50

Note. *$p < .05$, **$p < .01$

Table 4. Conviction Rate by Conditions (Emotion, Standard of Proof, and Strength of evidence)

	Emotion					
	Anger		Sad		Total	
Strength of Evidence	PoE	BRD	PoE	BRD	PoE	BRD
Medium	60.9% (23)	57.1% (21)	72.7% (22)	26.1% (23)	66.6% (45)	40.9% (44)
High	72.7% (22)	52.2% (23)	72.7% (22)	66.6% (21)	72.7% (44)	59.1% (44)
Total	66.6% (45)	54.5% (44)	72.7% (44)	45.5% (44)	69.7% (89)	50.0% (88)

Note. The percentile is conviction rate for each cell. PoE indicates 'preponderance of evidence' and BRD indicates 'beyond a reasonable doubt' standard. Numbers of total participants for each condition are in parentheses.

Table 5. Predictors of guilty/Not guilty Verdicts

Variable	*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*	*95% CI*	
							LL	*UL*
Emotion	1.07	.77	1.95	1	.16	2.92	.65	13.20
Standard of proof	-.19	.66	.09	1	.77	.82	.23	3.00
strength of evidence	.78	.68	1.33	1	.25	2.19	.58	8.28
Rationality score	.46	.37	1.53	1	.22	1.58	.77	3.25
Experiential score	.45	.38	1.37	1	.24	1.56	.74	3.30
Emotion								
x Standard	-2.38	1.06	5.08	1	.02*	.09	.12	.73
x strength of evidence	-1.37	1.05	1.70	1	.19	.25	.03	1.99
Standard x Strength of	-.91	.94	.95	1	.33	.40	.06	2.52
evidence								
Emotion x Standard x	3.44	1.44	5.71	1	.02*	31.15	1.85	523.3
strength of evidence								
Rationality score x								
Emotion	-0.66	0.38	2.97	1	0.09	0.52	.24	1.096
Standard of proof	0.03	0.37	0.01	1	0.94	1.03	.50	2.12
strength of evidence	-0.26	0.38	0.47	1	0.49	0.77	.36	1.63
Experiential score x				1				
Emotion	0.16	0.39	0.17	1	0.68	1.18	.55	2.50
Standard of proof	0.29	0.38	0.57	1	0.45	1.33	.63	2.79
strength of evidence	-1.06	0.39	7.19	1	0.01**	0.35	.16	.75

Note. *p < .05, **p < .01

Table 6. Predictors of Guilty/Not Guilty Verdicts by Emotion Condition

Variable	*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*	*95%CI*	
							LL	*UL*
Anger Condition								
Standard of proof	-0.50	0.69	0.53	1	0.47	0.60	0.16	2.35
Strength of evidence	0.37	0.70	0.28	1	0.60	1.45	0.37	5.64
Rationality score	0.93	0.48	3.72	1	0.05*	2.54	0.99	6.55
Experiential score	-0.19	0.52	0.13	1	0.72	0.83	0.30	2.31
Standard x	-0.38	0.96	0.16	1	0.69	0.68	0.10	4.44
Strength of evidence								
Rationality score x								
Standard of proof	-0.21	0.54	0.15	1	0.70	0.81	0.28	2.33
Strength of evidence	-1.17	0.55	4.54	1	0.03*	0.31	0.11	0.91
Experiential score x								
Standard of proof	0.65	0.52	1.55	1	0.21	1.91	0.69	5.30
Strength of evidence	-0.17	0.53	0.10	1	0.75	0.85	0.30	2.38
Sadness Condition								
Standard of proof	-3.98	1.34	8.83	1	0.00**	0.02	0.00	0.26
Strength of evidence	-1.61	1.20	1.79	1	0.18	0.20	0.02	2.11
Rationality score	-1.23	0.61	4.13	1	0.04*	0.29	0.09	0.96
Experiential score	1.45	0.59	6.08	1	0.01**	4.25	1.35	13.45
Standard x	3.66	1.52	5.79	1	0.02*	38.99	1.97	770.61
Strength of evidence								

Rationality score x								
Standard of proof	0.37	0.58	0.39	1	0.53	1.44	0.46	4.52
Strength of evidence	0.95	0.65	2.09	1	0.15	2.57	0.71	9.28
Experiential score x								
Standard of proof	0.25	0.70	0.13	1	0.72	1.28	0.33	5.06
Strength of evidence	-2.29	0.78	8.68	1	0.00**	0.10	0.02	0.46

Note. *p < .05, **p < .01

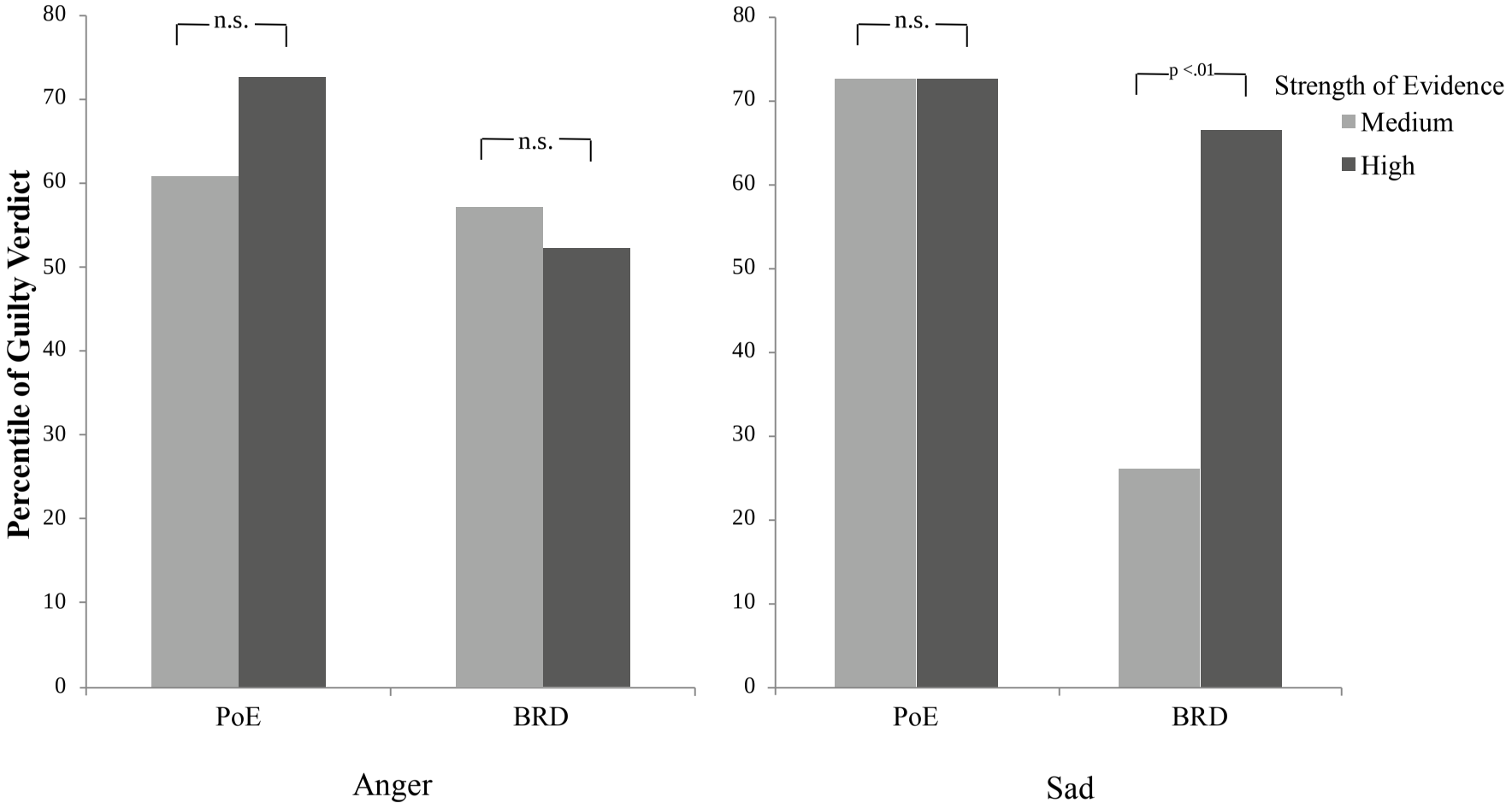

Figure 1. Conviction rate for emotion, standard of proof, and strength of evidence in study 1. PoE and BRD indicate 'preponderance of evidence' and 'beyond a reasonable doubt' instruction.

Table 7. Four Logistic Regressions on Prediction of Conviction in Study 1.

Effects	*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*	*95%CI*	
							LL	*UL*
Anger – PoE Condition								
Strength of evidence	.38	.74	.27	1.00	.60	1.47	.34	6.24
Rationality score	1.30	.68	3.67	1.00	.06	3.68	.97	13.96
Experiential score	.06	.74	.01	1.00	.93	1.06	.25	4.51
Rationality score x Strength of evidence	-1.81	.91	3.97	1.00	.05	.16	.03	.97
Experiential score x Strength of evidence	-.50	.91	.29	1.00	.59	.61	.10	3.65
Anger – BRD condition								
Strength of evidence	-.06	.63	.01	1.00	.93	.95	.27	3.27
Rationality score	.45	.51	.77	1.00	.38	1.56	.58	4.22
Experiential score	.31	.48	.40	1.00	.53	1.36	.53	3.52
Rationality score x Strength of evidence	-.66	.73	.80	1.00	.37	.52	.12	2.19
Experiential score x Strength of evidence	-.01	.65	.00	1.00	.99	.99	.28	3.57
Sad – PoE condition								
Strength of evidence	-1.26	1.10	1.29	1.00	.26	.28	.03	2.48
Rationality score	-.88	.67	1.74	1.00	.19	.42	.11	1.53
Experiential score	1.19	.57	4.36	1.00	.04*	3.29	1.08	10.06

Rationality score x Strength of evidence	.51	.83	.38	1.00	.54	1.66	.33	8.35
Experiential score x Strength of evidence	-1.94	.97	4.02	1.00	.05	.14	.02	.96
Sad – BRD condition								
Strength of evidence	2.31	1.01	5.19	1.00	.02*	10.09	1.38	73.66
Rationality score	-1.27	.78	2.65	1.00	.10	.28	.06	1.30
Experiential score	2.29	1.32	2.99	1.00	.08	9.88	.74	132.46
Rationality score x Strength of evidence	1.55	1.02	2.30	1.00	.13	4.69	.64	34.66
Experiential score x Strength of evidence	-2.93	1.45	4.05	1.00	.04*	.05	.00	.93

Variable “BRD” indicate “Beyond a reasonable doubt standard” and “POE” represents “Preponderance of evidence”
*$p < .05$

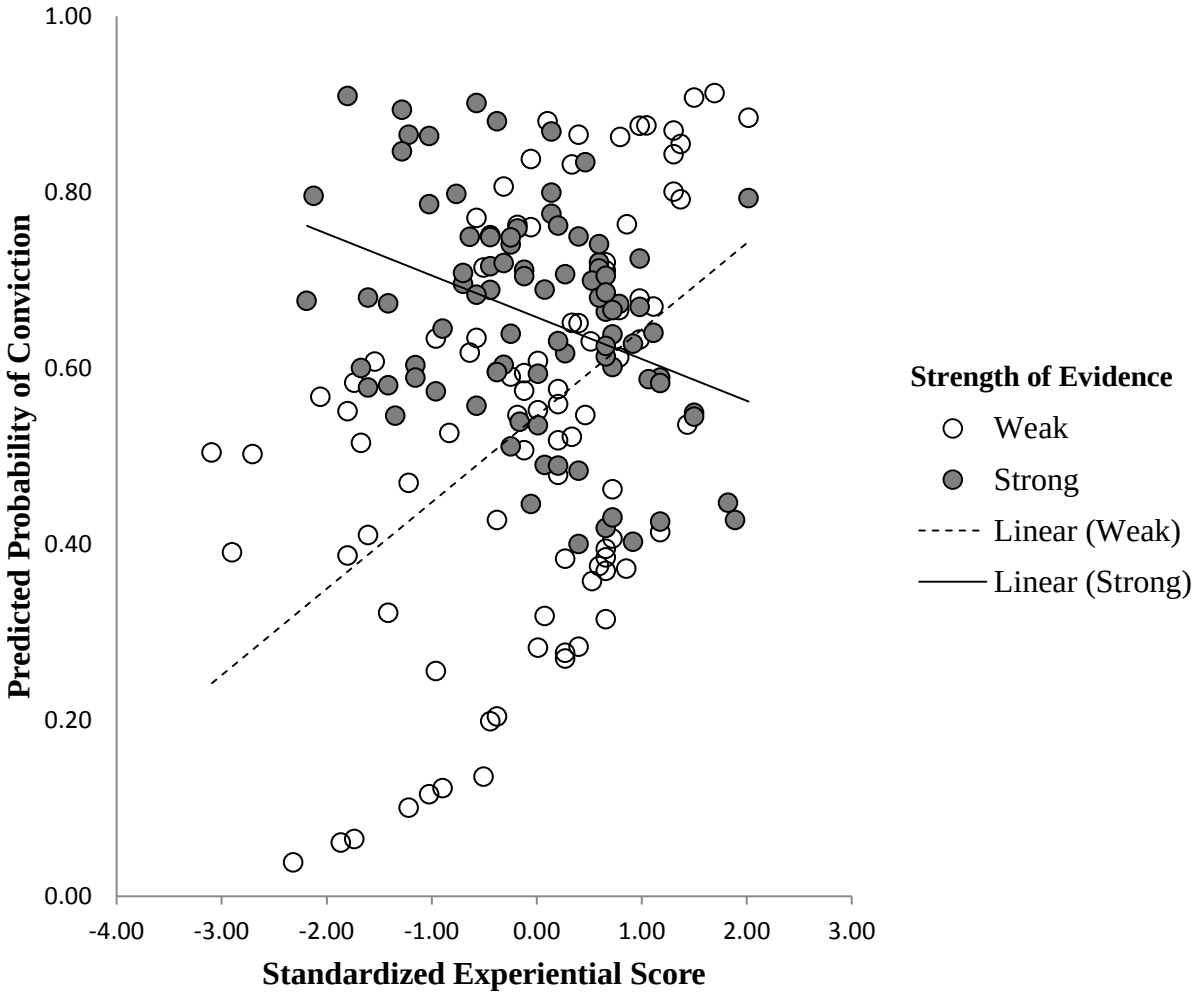

Figure 2. Relationship between standardized score of experiential measure and predicted probability of conviction for each condition of strength of evidence

Table 8. Effects of Emotion, Standard of Proof, and Strength of evidence on Subjective Standard of Proof

Effects	F	df	p	partial η_p^2
Emotion	.012	1	.912	.000
Standard	20.992	1	.000**	.113
Strength of evidence	.000	1	.994	.000
Rationality	6.096	1	.015*	.036
Experiential	1.738	1	.189	.010
Emotion * Standard	.352	1	.554	.002
Emotion * Strength of evidence	.878	1	.350	.005
Standard * Strength of evidence	2.032	1	.156	.012
Emotion * Standard * Strength of evidence	.008	1	.928	.000
Error		164		

Note. *p < .05, **p < .01

Table 9. Means of Coherence Shifts by Conditions (Emotion, Standard of Proof, and Strength of evidence) and Verdict

	Guilty				Not Guilty			
	Anger		Sad		Anger		Sad	
Strength of Evidence	PoE	BRD	PoE	BRD	PoE	BRD	PoE	BRD
Medium	1.02 (14)	1.23 (12)	1.02 (16)	.34 (6)	-.04 (9)	.36 (9)	.07 (6)	.75 (17)
High	.76 (16)	1.68 (12)	1.14 (16)	.87 (14)	1.28 (6)	.54 (11)	.25 (6)	-.29 (7)
Total	.88 (30)	1.45 (24)	1.08 (32)	.71 (20)	.49 (15)	.46 (20)	.16 (12)	.44 (24)

Note. PoE indicates 'preponderance of evidence' and BRD indicates 'beyond a reasonable doubt' standard. Total numbers of participants in each condition are in parentheses. 'Guilty' indicate participants who found the defendant guilty. 'Not Guilty' indicates participants who found the defendant not guilty.

Table 10. Effects of Emotion, Standard of Proof, and Strength of evidence on Coherence Shift

Effects	*F*	*df*	*p*	η_p^2
Emotion	5.202	1	.024*	.032
Standard	.060	1	.806	.000
Strength of evidence	1.925	1	.167	.012
Rationality	3.422	1	.066	.021
Experiential	.230	1	.632	.001
Verdict	18.008	1	.000**	.102
Verdict				
x Emotion	.054	1	.817	.000
x Standard	.072	1	.789	.000
x Strength of evidence	.026	1	.871	.000
Emotion				
x Standard	1.461	1	.229	.009
x Strength of evidence	2.209	1	.139	.014
Standard * Strength of evidence	1.356	1	.246	.008
Emotion * Standard * Strength of evidence	.365	1	.547	.002
Emotion * Standard * Verdict	4.853	1	.029*	.030
Emotion * Strength of evidence * Verdict	3.411	1	.067	.021
Standard * Strength of evidence * Verdict	7.025	1	.009**	.042
Emotion * Standard * Strength of evidence * Verdict	.024	1	.878	.000
Error		159		

Note. * $p < .05$, ** $p < .01$

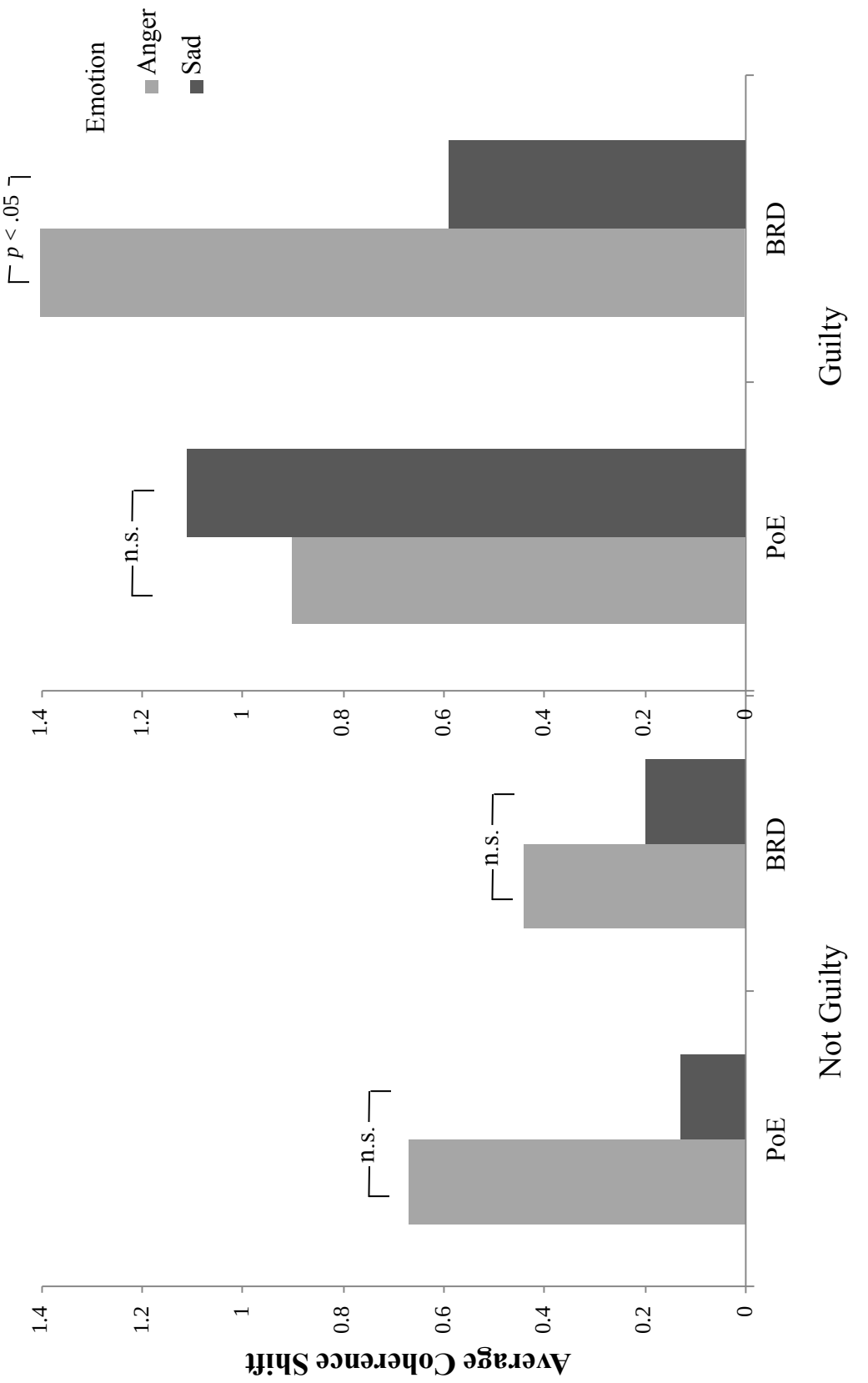

Figure 3. Average coherence shift for standard of proof, strength of evidence, and verdict in study 1. PoE and BRD indicates 'preponderance of evidence' and 'beyond a reasonable doubt' instruction.

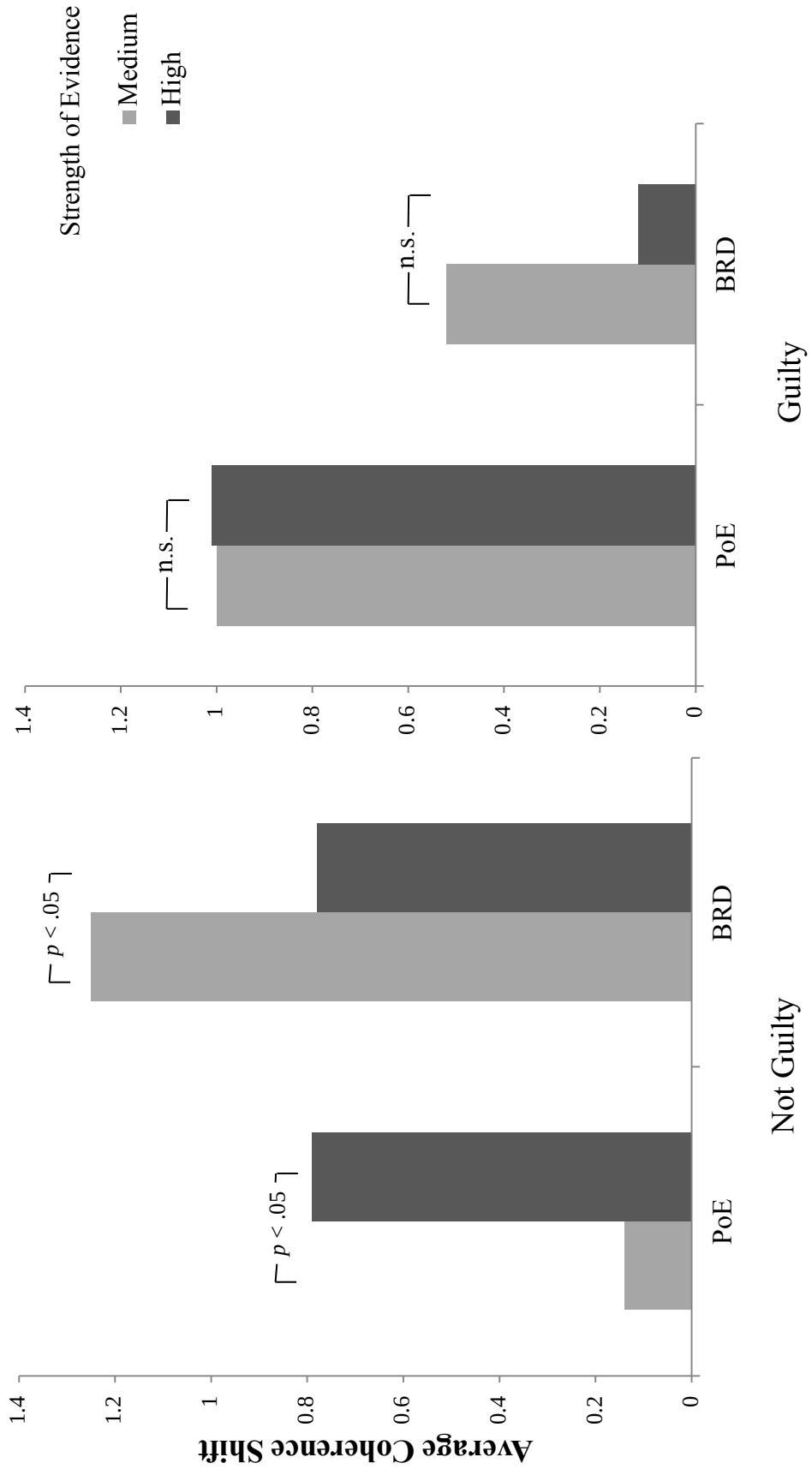

Figure 4. Average coherence shift for standard of proof, strength of evidence, and verdict in study 1. PoE and BRD indicates 'preponderance of evidence' and 'beyond a reasonable doubt' instruction.

Table 11. Effects of Emotion, Standard of Proof, and Strength of evidence on Guilty Certainty

Source	F	df	Sig.	Partial Eta Squared
Emotion	.048	1	.827	.000
Standard	5.998	1	.015*	.035
Strength of evidence	3.344	1	.069	.020
Rationality	.296	1	.587	.002
Experiential	1.352	1	.247	.008
Emotion * Standard	1.824	1	.179	.011
Emotion * Strength of evidence	1.457	1	.229	.009
Standard * Strength of evidence	.768	1	.382	.005
Emotion * Standard * Strength of evidence	3.259	1	.073	.019

Note. * $p < .05$

Table 12. The Indirect Effects of Emotion on coherence shift

Mediation Source	Data Indirect effect	Boot Indirect effect	Standard error	95% confidence interval LL	UL
Control appraisal measures					
to what extent did you typically feel that someone other than yourself had the ability to influence what was happening?	-.08	-.08*	.05	-.21	-.004
to what extent did you typically feel that someone else was to blame for what was happening in the situation?	-.18	-.19*	.09	-.37	-.009
to what extent were the events beyond anyone's control?	.05	.05	.06	-.08	.18
Certainty appraisal measures					
how well did you understand what was happening in those situations?	.01	.01	.02	-.01	.09
how uncertain were you about what would happen in these situations?	-.01	-.01	.02	-.07	.03
how well could you typically predict what was going to happen next?	-.05	-.05	.05	-.18	.02

Note. The data coefficient came from the regression weight and the boot coefficient from the bootstrapping.
*$p < .05$

Table 13. Participants' Demographic Characteristics in Study 2

Age				
	18-20	7 (4.0%)	41-50	26 (14.9%)
	21-30	73 (41.7%)	51-60	22 (12.6%)
	31-40	38 (21.7%)	61-64	9 (5.1%)
Gender				
	Male	78 (44.6%)	Female	97 (55.4%)
Ethnic Origin				
	Caucasian	146 (83.4%)	Hispanic	8 (4.6%)
	Asian	10 (5.7%)	Other	3 (1.7%)
	African	8 (4.6%)		
Education				
	Less than high school	2 (1.1%)	Bachelor's degree	50 (82.6%)
	High school	74 (42.3%)	Master's degree	15 (8.6%)
	Associate Degree	31 (17.7%)	Profession degree	2 (1.1%)

	Yes	**No**	Missing
Are you registered to vote?	156 (89.1%)	19 (10.9%)	
Do you have a valid driver license?	157 (89.7%)	15 (8.6%)	3 (1.7%)
Have you served as a juror?			
State-Civil	14 (8.0%)	160 (91.4%)	1 (.6%)
State-Criminal	10 (5.7%)	163 (93.1%)	2 (1.1%)
Federal-Civil	1 (.6%)	172 (98.3%)	2 (1.1%)
Federal-Criminal	1 (.6%)	171 (97.7%)	3 (1.7%)
Current working status			
Full time	78 (44.6%)	Unemployed	48 (27.4%)

	Part time		48 (27.4%)	Missing		1 (.6%)	
Residing State							
	AL	1 (0.6%)	IN	2 (1.1%)	MS	1 (0.6%)	SC	3 (1.7%)
	AR	2 (1.1%)	KS	5 (2.9%)	MT	1 (0.6%)	TN	2 (1.1%)
	AZ	3 (1.7%)	KY	4 (2.3%)	NC	6 (3.4%)	TX	9 (5.1%)
	CA	27 (15.4%)	LA	2 (1.1%)	NH	2 (1.1%)	UT	2 (1.1%)
	CO	4 (2.3%)	MA	4 (2.3%)	NJ	1 (0.6%)	VA	3 (1.7%)
	CT	1 (0.6%)	MD	5 (2.9%)	NV	1 (0.6%)	VT	1 (0.6%)
	FL	13 (7.4%)	ME	3 (1.7%)	NY	10 (5.7%)	WA	1 (0.6%)
	GA	5 (2.9%)	MI	6 (3.4%)	OH	5 (2.9%)	WI	4 (2.3%)
	IA	1 (0.6%)	MN	8 (4.6%)	OR	2 (1.1%)	Missing	3 (1.7%)
	IL	11 (6.3%)	MO	1 (0.6%)	PA	10 (5.7%)		

Table 14. Manipulation Checks for Severity of Charge

	theft		Murder		*t* (173)	*p*	95% CI		*d*
	Mean	SD	Mean	SD			LL	UL	
Seriousness manipulation check									
How seriousness was this crime?	6.45	1.81	10.52	.83	-19.22	.00**	-.49	-3.65	-2.93
if convicted, how morally blameworthy was the defendant?	8.99	2.12	10.47	1.26	-5.65	.00**	-2.00	-.97	.86
How much damage, both monetary damage and bodily harm, did this crime cause?	5.56	1.81	10.54	.81	-23.59	.00**	-5.40	-4.56	-3.58
How much violent was this crime?	1.32	1.15	10.69	.74	-61.42	.00**	-9.67	-9.07	-9.74
Which of the following best describes the penalty you would likely assign to the defendant	1.70	.69	4.23	2.39	-9.45	.00**	-3.06	-2.00	-6.30
Accuracy motivation check									
If you reached the wrong verdict, how much would you regret your decision?	8.13	3.01	9.54	2.08	-3.62	.00**	-2.18	-.64	-.55
How hard did you try to judge the verdict accurately?	10.13	1.51	9.92	1.49	.92	.36	-.24	.66	.14
How hard did you try to judge fairly?	10.33	1.28	10.15	1.51	.85	.39	-.24	.60	.13
Were you motivated to make your judgments accurately?	10.23	1.26	9.98	1.37	1.28	.20	-.14	.65	.19
Was it more important to reach a fair judgment or an accurate judgment?	7.80	2.87	7.90	3.00	-.22	.83	-.98	.78	-.03

Note. *$p < .05$, **$p < .01$

Table 15. Conviction Rate for Theft Charge by Conditions (order of questionnaires, severity of charge, and Strength of evidence)

	Order of Questionnaires					
	Verdict First		Probability assessment first		Total	
Strength of Evidence	Less Severe	More Severe	Less Severe	More Severe	Less Severe	More Severe
Medium	30.0% (20)	60.9% (23)	47.6% (21)	15.0% (20)	39.0% (41)	39.5% (43)
High	56.5% (23)	34.8% (23)	63.6% (22)	52.2% (23)	60.0% (45)	43.5% (46)
Total	44.2% (43)	47.8% (46)	55.8% (43)	34.9% (43)	50.0% (86)	41.6% (89)

Note. The percentile is conviction rate for each cell. Numbers of total participants are in parentheses.

Table 16. Predictors of guilty/Not guilty Verdicts

Variable	*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*	*95% CI*	
							LL	*UL*
Severity of charge	1.41	0.70	4.06	1	0.04*	4.09	1.03	16.12
Order	0.91	0.71	1.63	1	0.20	2.49	0.63	10.16
Strength of evidence	1.20	0.70	2.91	1	0.09	3.31	0.85	13.25
Rationality score	-0.62	0.38	2.72	1	0.10	0.54	0.26	1.15
Experiential score	0.42	0.37	1.30	1	0.25	1.52	0.75	3.13
Severity of charge								
x Order	-3.24	1.10	8.69	1	0.00**	0.04	0.00	0.34
x Strength of evidence	-1.74	0.95	3.37	1	0.06	0.18	0.02	0.76
Order x Strength of evidence	-0.54	0.96	0.31	1	0.58	0.58	0.09	3.70
Severity of charge x Order x Strength of evidence	2.98	1.40	4.56	1	0.03*	19.76	1.94	468.63
Rationality score x								
Severity of charge	0.36	0.39	0.85	1	0.36	1.43	0.52	2.28
Order	0.47	0.38	1.58	1	0.21	1.60	0.17	0.75
Strength of evidence	0.12	0.38	0.09	1	0.76	1.12	0.73	2.98
Experiential score x								
Severity of charge	0.31	0.36	0.76	1	0.39	1.37	0.63	2.90
Order	0.39	0.36	1.14	1	0.29	1.47	0.66	2.71
Strength of evidence	-1.11	0.38	8.61	1	0.00**	0.33	0.78	3.38

Note. * $p < .05$, ** $p < .01$

Table 17. Two Logistic Regressions on Prediction of Conviction in Study 2

Variable	*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*
Theft Condition						
Order	1.00	0.75	1.78	1.00	0.18	2.73
Strength of evidence	1.39	0.74	3.56	1.00	0.06	4.01
Rationality score	-0.25	0.47	0.28	1.00	0.60	0.78
Experiential score	0.93	0.54	2.99	1.00	0.08	2.54
Order x Strength of evidence	-0.63	0.99	0.41	1.00	0.52	0.53
Rationality score x						
Order	-0.03	0.49	0.00	1.00	0.95	0.97
Strength of evidence	-0.04	0.50	0.01	1.00	0.93	0.96
Experiential score x						
Order	-0.42	0.56	0.56	1.00	0.46	0.66
Strength of evidence	-1.21	0.55	4.82	1.00	0.03*	0.30
Murder Condition						
Order	-2.93	1.01	8.47	1.00	0.00*	0.05
Strength of evidence	-0.41	0.66	0.39	1.00	0.53	0.66
Rationality score	-0.49	0.40	1.55	1.00	0.21	0.61
Experiential score	0.58	0.46	1.58	1.00	0.21	1.78
Order x Strength of evidence	2.64	1.13	5.41	1.00	0.02	13.98
Rationality score x						

Order	1.48	0.83	3.19	1.00	0.07	4.38
Strength of evidence	-0.04	0.76	0.00	1.00	0.96	0.96
Experiential score x						
Order	1.29	0.61	4.43	1.00	0.04	3.63
Strength of evidence	-1.46	0.64	5.24	1.00	0.02*	0.23

*$p < .05$

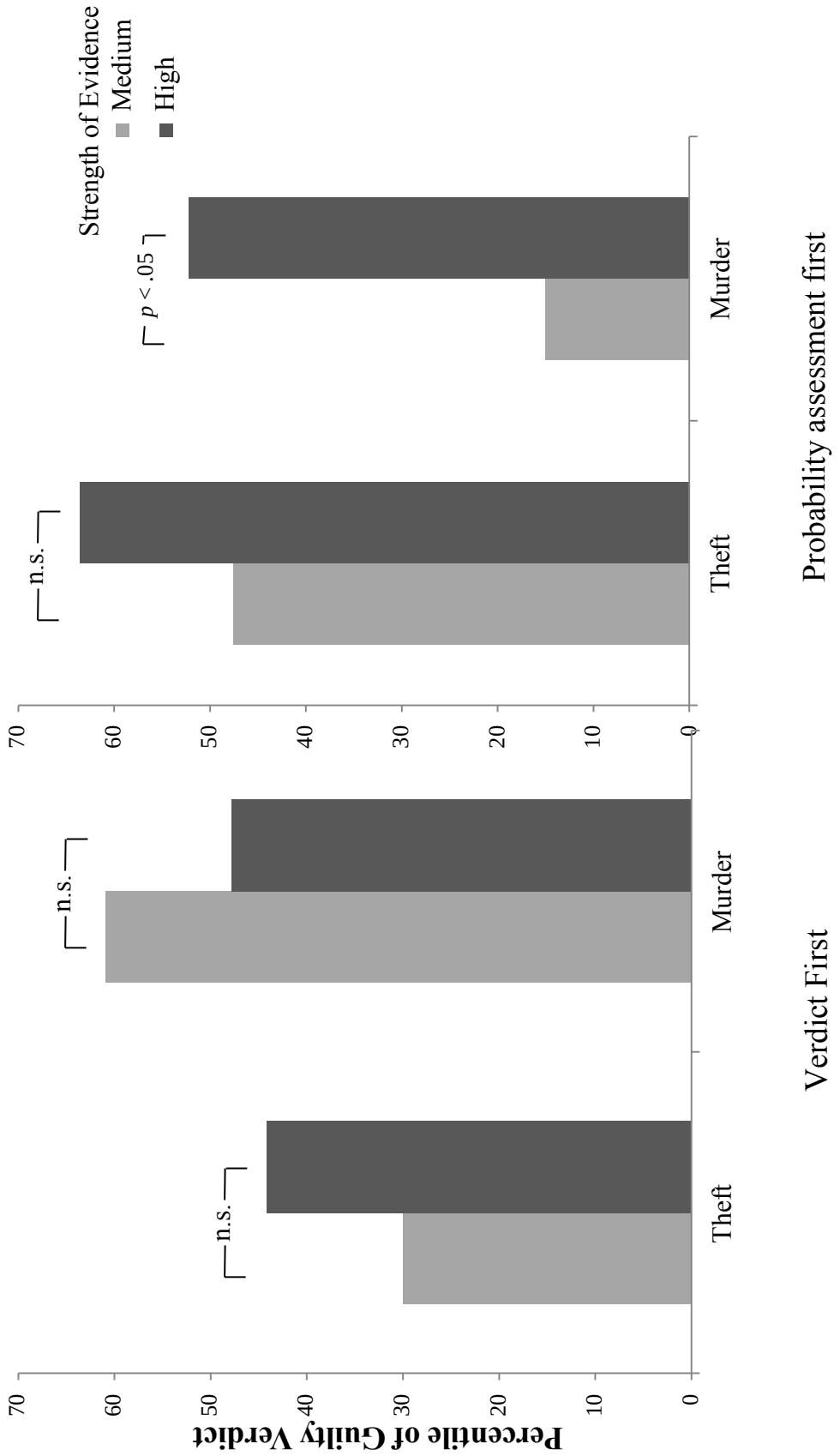

Figure 5. Conviction rate for order of questionnaires, severity of charge, and strength of evidence

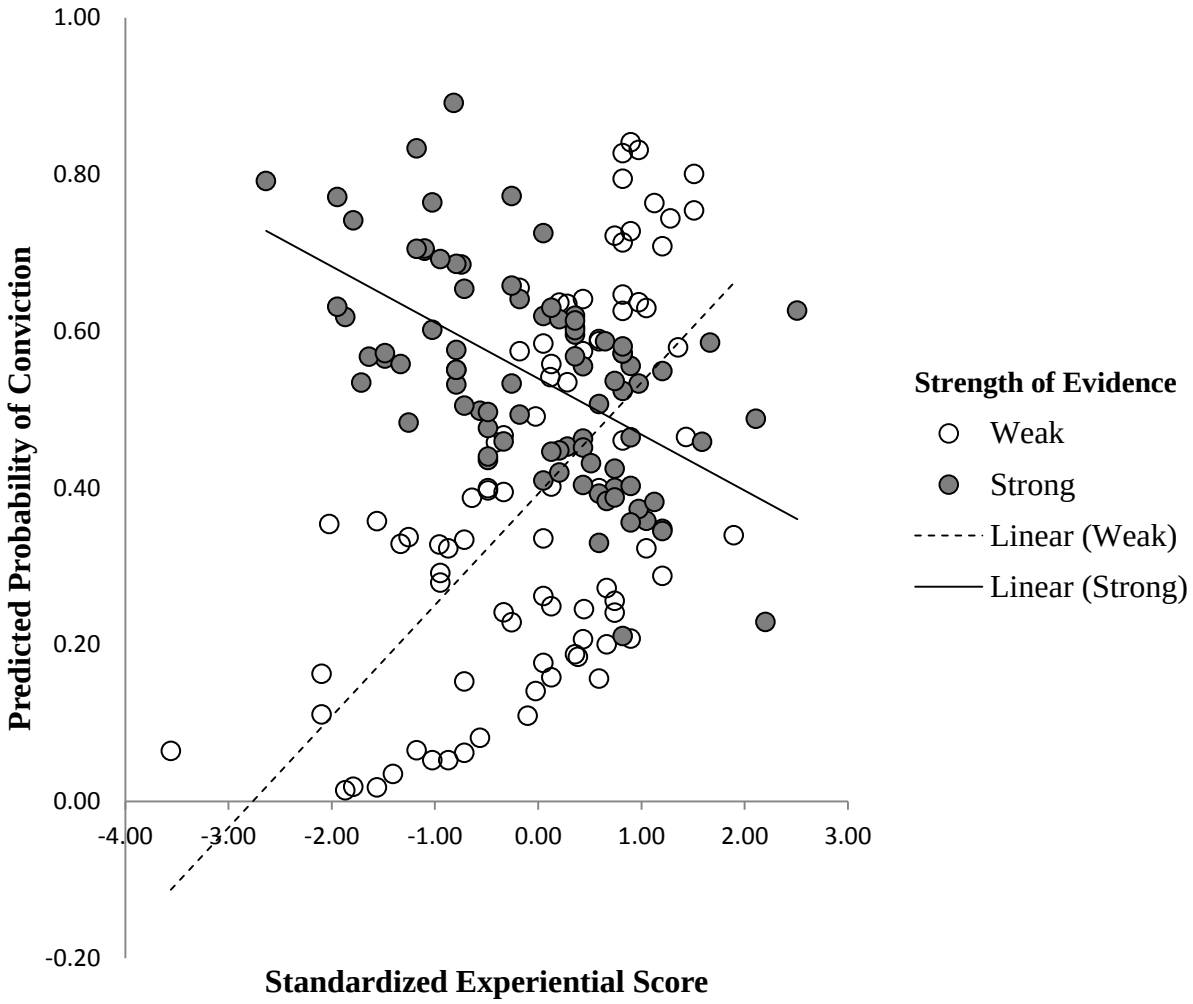

Figure 6. Relationship between standardized score of experiential measure and predicted probability of conviction for each condition of strength of evidence

Table 18. Four Simple Logistic Regressions on Prediction of Conviction in Study 2

Charge	Order		*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*	*95% CI*	
									LL	*UL*
theft	verdict first	Strength of evidence	1.22	.67	3.29	1.00	.07	3.40	.91	12.76
		Rationality score	-.23	.29	.63	1.00	.43	.79	.44	1.41
		Experiential score	.20	.36	.30	1.00	.58	1.22	.60	2.47
	question first	Strength of evidence	.79	.65	1.47	1.00	.23	2.19	.62	7.82
		Rationality score	-.31	.33	.93	1.00	.34	.73	.39	1.39
		Experiential score	.04	.30	.02	1.00	.90	1.04	.58	1.86
murder	verdict first	Strength of evidence	-.61	.61	.99	1.00	.32	.54	.16	1.81
		Rationality score	-.46	.34	1.88	1.00	.17	.63	.32	1.22
		Experiential score	-.16	.33	.22	1.00	.64	.86	.45	1.64
	question first	Strength of evidence	1.72	.81	4.52	1.00	.03*	5.60	1.14	27.45
		Rationality score	1.03	.67	2.40	1.00	.12	2.81	.76	10.39
		Experiential score	.66	.41	2.62	1.00	.11	1.93	.87	4.29

*$p < .05$

Table 19. Effects of Severity of charge, Order, and Strength of evidence on Subjective Standard of Proof

Effects	*F*	*df*	*p*	Partial η^2
Severity of charge	.012	1	.912	.000
Order	.041	1	.840	.000
Strength of evidence	.475	1	.491	.003
Rationality	.001	1	.977	.000
Experiential	.026	1	.872	.000
Severity of charge *Order	1.643	1	.202	.010
Severity of charge *Strength of evidence	.027	1	.870	.000
Order * Strength of evidence	.473	1	.493	.003
Severity of charge *Order * Strength of evidence	.753	1	.387	.005
Error		156		

Note. No effect was statistically significant.

Table 20. Means of Coherence Shifts by Conditions (the order of questionnaires, severity of charge, and Strength of evidence) and Verdict

	Guilty				Not Guilty			
	Verdict First		Probability Assessment First		Verdict First		Probability Assessment First	
Strength of Evidence	Theft	Murder	Theft	Murder	Theft	Murder	Theft	Murder
Medium	2.37 (6)	1.61 (14)	1.37 (10)	.94 (3)	.43 (14)	.18 (8)	.03 (11)	.15 (17)
High	1.01 (13)	1.27 (10)	.66 (14)	.81 (12)	-.18 (10)	.56 (13)	.25 (8)	.30 (11)
Total	1.43 (19)	1.47 (24)	.95 (24)	.83 (15)	.18 (24)	.43 (21)	.12 (19)	.21 (28)

Note. Verdict first indicates participants made verdicts before they evaluate evidence. Probability Assessment first indicates participants evaluate evidence before they made verdicts. Total numbers of participants in each condition are in parentheses. ‘Guilty’ indicate participants who found the defendant guilty. ‘Not Guilty’ indicates participants who found the defendant not guilty.

Table 21. Effects of Seriousness, Order, and Strength of evidence on Coherence Shift

Effects	*F*	*df*	*p*	Partial η_p^2
Severity of charge	0.00	1.00	0.95	0.00
Order	4.22	1.00	0.04*	0.03
Strength of evidence	3.78	1.00	0.05*	0.02
Rationality	0.49	1.00	0.48	0.00
Experiential	2.12	1.00	0.15	0.01
Verdict	41.57	1.00	0.00**	0.21
Verdict				
x Severity of charge	1.10	1.00	0.30	0.01
x Order	2.58	1.00	0.11	0.02
x Strength of evidence	5.41	1.00	0.02*	0.03
Severity of charge				
x Order	0.00	1.00	0.96	0.00
x Strength of evidence	3.85	1.00	0.05*	0.02
Order * Strength of evidence	1.27	1.00	0.26	0.01
Severity of charge * Order * Strength of evidence	1.55	1.00	0.21	0.01
Seriousness * Order * Verdict	0.26	1.00	0.61	0.00
Severity of charge * Strength of evidence * Verdict	0.26	1.00	0.61	0.00
Order * Strength of evidence * Verdict	0.03	1.00	0.87	0.00
Severity of charge *Order * Strength of evidence * Verdict	0.28	1.00	0.59	0.00
Error		156		

Note. *$p < .05$, **$p < .01$

Table 22. Effects of Seriousness, Order, and Strength of Evidence on Guilty Certainty

Source	F	df	Sig.	Partial Eta Squared
Severity of charge	2.057	1.00	0.15	0.01
Order	0.00	1.00	.997	0.00
Strength of evidence	3.13	1.00	0.08	0.02
Rationality	0.83	1.00	0.37	0.01
Experiential	0.62	1.00	0.43	0.00
Severity of charge *Order	3.25	1.00	0.07	0.02
Severity of charge * Strength of evidence	1.09	1.00	0.30	0.01
Order * Strength of evidence	2.91	1.00	0.09	0.02
Severity of charge *Order * Strength of evidence	5.90	1.00	0.02*	0.03
Error		165		

Note. *$p < .05$. Strength of evidence indicates strength of evidence (Medium v. High)

Table 23. The Indirect Effects of severity of charge on verdict and coherence shift through Accuracy motivation measures

Mediation Source	Data Indirect effect	Boot Indirect effect	Standard error	95% confidence interval LL	UL
Verdict as the dependent variable					
If you reached the wrong verdict, how much would you regret your decision?	.15	.17	.13	-.04	.48
How hard did you try to judge the verdict accurately?	-.0003	.03	.09	-.17	.18
How hard did you try to judge fairly?	-.01	-.01	.08	-.28	.09
Were you motivated to make your judgments accurately?	.07	.07	.09	-.03	.34
Was it more important to reach a fair judgment or an accurate judgment?	.02	.02	.15	-.28	.33
Coherence shift as the defendant variable					
If you reached the wrong verdict, how much would you regret your decision?	-.02	-.02	.06	-.14	.09
How hard did you try to judge the verdict accurately?	.01	-.003	.06	-.04	.18
How hard did you try to judge fairly?	-.04	-.04	.06	-.24	.02
Were you motivated to make your judgments accurately?	.03	.02	.05	-.03	.17
Was it more important to reach a fair judgment or an accurate judgment?	.004	.002	.03	-.04	.07

Note. The data coefficient came from the regression weight and the boot coefficient from the bootstrapping.

No indirect effects was significant

Appendixes

Appendix A
The Rational-Experiential Inventory (REI; Pacini & Epstein, 1999)

Instruction: Please indicate the extent to which each statement describes you

1: Completely False

2: False

3: Neither False nor True

4: Moderately True

5: Completely True

1. I have a logical mind.
2. I prefer complex problems to simple problems.
3. I believe in trusting my hunches.
4. I am not a very analytical thinker.
5. I trust my initial feelings about people.
6. I try to avoid situations that require thinking in depth about something.
7. I like to rely on my intuitive impressions.
8. I don't reason well under pressure.
9. I don't like situations in which I have to rely on intuition.
10. Thinking hard and for a long time about something gives me little satisfaction.
11. Intuition can be a very useful way to solve problems.
12. I would not want to depend on anyone who described himself or herself as intuitive.
13. I am much better at figuring things out logically than most people.
14. I usually have clear, explainable reasons for my decisions.
15. I don't think it is a good idea to rely on one's intuition for important decisions.
16. Thinking is not my idea of an enjoyable activity.
17. I have no problem thinking things through carefully.
18. When it comes to trusting people, I can usually rely on my gut feelings.
19. I can usually feel when a person is right or wrong, even if I can't explain how I know.
20. Learning new ways to think would be very appealing to me.

21. I hardly ever go wrong when I listen to my deepest gut feelings to find an answer.

22. I think it is foolish to make important decisions based on feelings.

23. I tend to use my heart as a guide for my actions.

24. I often go by my instincts when deciding on a course of action.

25. I'm not that good at figuring out complicated problems.

26. I enjoy intellectual challenges.

27. Reasoning things out carefully is not one of my strong points.

28. I enjoy thinking in abstract terms.

29. I generally don't depend on my feelings to help me make decisions.

30. Using logic usually works well for me in figuring out problems in my life.

31. I think there are times when one should rely on one's intuition.

32. I don't like to have to do a lot of thinking.

33. Knowing the answer without having to understand the reasoning behind it is good enough for me.

34. Using my gut feelings usually works well for me in figuring out problems in my life.

35. I don't have a very good sense of intuition.

36. If I were to rely on my gut feelings, I would often make mistakes.

37. I suspect my hunches are inaccurate as often as they are accurate.

38. My snap judgments are probably not as good as most people's.

39. I am not very good at solving problems that require careful logical analysis.

40. I enjoy solving problems that require hard thinking.

Appendix B
Informed Consent

Identification of Project: Social Cognition in Recall of Life Events, assessing social issues, and legal decision making

You are invited to participate in research that examines social cognition in recall of life events, assessing social issues, and legal decision making. A group of researchers studying social cognition is interested in a number of different types of information processing, and to facilitate data collection we have combined several of their experiments. You were selected as a participant because you have signed up to participate in web-based research. To complete this research, you must be at least 18 years of age. Researchers at the University of Nebraska-Lincoln are conducting this research under the direction of Yimoon Choi, a graduate student in the Department of Psychology, and Dr. Richard Wiener, professor in the Department of Psychology.

This research will consist of three separate experiments. Each experiment will take approximately 20 minutes to complete. All three experiments will take place on the internet. Experiment 1 will examine how individuals write about life events. You will be asked to write about past or present life events, and then will complete a few other short questionnaires. In Experiment 2, you will answer some questions regarding your opinions about social issues. You will read short stories and answer some questions. Experiment 3 will examine legal decision making. You will read about a case in which a construction company has initiated a disciplinary procedure against its employee. You will play the role of an arbitrator in a disciplinary case and you will determine whether the defendant is guilty and answer several other questions about the case.

The potential benefits of the current research outweigh any cost that may accompany participation. Knowledge generated through these studies will help researchers better understand individual differences in information processing about life events, which will contribute to the study of social cognition. This research will also help researchers better understand legal decision making, which can lead to improvements in the legal system. The risks of this research are minimal. Some participants may experience some minimal discomfort when writing about life events in Experiment 1. Some participants may also experience some minimal discomfort when reading about the crimes described in Experiments 3. In addition, this research, like much other research in Psychology, may contain some questions that you may find sensitive or personal.

If you do not feel comfortable answering a question during the study, you may choose not to respond. You will not be penalized for skipping any questions that you do not want to answer. You may skip any questions or stop completing any survey without penalty. The alternative to this research is non-participation. Your participation is voluntary. You are free to decide not to participate in this research or to withdraw at any time without adversely affecting your relationship with the investigators or the University of Nebraska-Lincoln. Your refusal to participate will involve no penalty to you or loss of any benefits to which you are otherwise entitled.

The results of this research may be published, but your name and identity will not be revealed and all of the data and information collected from you will remain anonymous. All data will be identified with numbers that have no links to you as a research participant and will be kept in a locked, secure lab in Burnett hall at the University of Nebraska for a period of five years after which it will be destroyed. Nonetheless, some of the questions on the demographic

sheet ask about, among other things, your gender, your ethnicity, and your age. Feel free to leave any of those items unanswered if you feel that the answers may reveal your identity.

Yimoon Choi (graduate student) and Dr. Richard L. Wiener are conducting this research. They will be happy to answer any questions or concerns about the research. You may contact Yimoon Choi at (402) 326-8439 or yimoon@gmail.com. You may contact Dr. Wiener at (402) 472-1137 or rwiener2@unl.edu. To obtain more information about your rights as a research participant or to report any concerns about this research, please contact the University of Nebraska-Lincoln Institutional Review Board (IRB) for Human Research at (402) 472-6965.

If you wish to participate in this research, please read the following statement and provide your electronic signature by clicking at the bottom of the webpage. It is suggested that you print a copy of this informed consent form for your personal records.

I have read and understood the information presented above. If I have any questions before I begin, I may contact the researchers. Otherwise, my concerns have been answered to my satisfaction via this consent form. I consent to take part in this research.

____ I agree, and I consent to take part in this research

____ I disagree, and I do not wish to take part in this research

Name and Phone Number of Investigators:

Yimoon Choi (402) 326-8439

Dr. Richard Wiener (402) 472-1137

Appendix C
Emotion Manipulation (Anger/Sad)
Life Events Questionnaire

Instructions: The researchers are interested in how individuals write about life events. Please answer the questions below, taking the time to give complete and detailed responses.

Question 1: What are the three to five things that make you most angry (Sad)? Please write two to three sentences about each thing that makes you angry (Sad).

_____

_____

_____

_____

_____

_____

_____

Question 2: Now we'd like you to describe in more detail the one situation that makes you (or has made you) most angry (Sad). This could be something you are presently experiencing or something from the past. Begin by writing down what you remember of the anger (Sadness) -inducing event(s) and continue by writing as detailed a description of the event(s) as is possible.

If you can, please write your description so that someone reading this might even get angry (sad) just from learning about the situation. What is it like to be in this situation? Why does it make you so angry (sad)?

_____

_____

_____

_____

_____

_____

_____

_____

Appendix D

Appraisal Questionnaire

Instructions: Please consider the situations and experiences you wrote about on the previous pages when answering the following questions.

1. Regarding the events that you described on the previous pages, how well did you understand what was happening in those situations?

Not at All — Somewhat — Extremely

1 2 3 4 5 6 7 8 9

2. Regarding the events that you described on the previous pages, to what extent did you typically feel that someone other than yourself had the ability to influence what was happening?

Not at All — Somewhat — Extremely

1 2 3 4 5 6 7 8 9

3. Regarding the events that you described on the previous pages, how uncertain were you about what would happen in these situations?

Not at All — Somewhat — Extremely

1 2 3 4 5 6 7 8 9

4. Regarding the events that you described on the previous pages, to what extent did you typically feel that someone else was to blame for what was happening in those situations?

Not at All Somewhat Extremely

1 2 3 4 5 6 7 8 9

5. Regarding the events that you described on the previous pages, how well could you typically predict what was going to happen next?

Not at All Somewhat Extremely

1 2 3 4 5 6 7 8 9

6. Regarding the events that you described on the previous pages, to what extent were the events beyond anyone's control?

Not at All Somewhat Extremely

1 2 3 4 5 6 7 8 9

Appendix E

Emotion Manipulation Check

Instructions: Please rate the extent to which you feel each of the following emotions right now.

1. Happy

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

2. Angry

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

3. Disgusted

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

4. Sad

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

5. Fearful

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

6. Surprised

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

7. Blue

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

8. Downhearted

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

9. Hostile

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

10. Nervous

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

11. Alone

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

12. Irritable

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

13. Alert

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

14. Lonely

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

15. Scornful

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

16. Loathing

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

17. Proud

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

18. Excited

Not at all				Moderately				Extremely
1	2	3	4	5	6	7	8	9

Appendix F.

Assessing Social Issues

Please give your opinion about the following questions. The questions concern social situations, business issues, and legally-related topics.

For each question, you will receive a brief summary of relevant information followed by a statement about that information. You will then rate the extent to which you agree or disagree with the statements. You are not expected to have any expert knowledge. Simply use your general knowledge and common sense in making the ratings. The issues are unrelated, so consider each issue independently. You might find that the information provided is less than you would like to have; nonetheless, respond as best you can based on that information. For each of the statements about issues, please rate the extent to which you agree or disagree. Answer all questions using the scales that follow the items.

Michelle is 24 years old, and she works in marketing for a retailing chain. She is sharp and fun and most of her closer friends insist that she's very pleasant. Others find her to be overly ambitious. Since her sophomore year, Michelle has been involved in a number of lengthy relationships, but no matter how good things seemed, she ultimately broke off the relationships. In her senior year she broke up a long relationship with Robert, who was generally considered to be one of the most desirable and popular guys around. She explained then that Robert was a great guy, but that 'we simply weren't suited for each other.' Zoe, a girl friend from college, believes that Michelle has a problem with commitment, and that she is not about to commit herself in a relationship in the near future. Michelle's close friend Jessica holds a different opinion. She says

that Michelle never had a problem with commitment; she simply hadn't found the right guy. Jessica believes that with the right partner, Michelle would be happy to commit herself.

a) Zoe's assessment of Michelle is correct. The fact that Michelle broke up previous relationships suggests that she is not going to commit herself to a stable relationship in the near future.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

b) Generally speaking, people who break off a number of lengthy relationships do so because they have a problem making commitments.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

c) Jessica's assessment of Michelle is correct. Michelle broke up previous relationships because she had not met the right partner.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

d) As a general matter, when people are in a relationship with the right partner, they never break off the relationship.

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

Wendy works as a computer programmer for a large insurance company. One evening, after most of the employees had left, she was walking by the office of the accounting department. She noticed a man rushing into the office and leaving a bouquet of flowers on the desk of Jessica Myers, one of the company's accountants. Jessica is a very shy person, and has experienced great difficulty in forging romantic relationships with men. The next day Wendy noticed that Jessica was distraught. Jessica told her that there was no note on the flowers and that she's eager to learn who had left them for her. When she learned that Wendy had seen the man, she became visibly excited. Jessica said that she suspected that the person might be Dale Brown, who works for a travel agency located on the ground floor of the building. Wendy offered to go down to the travel agency to see if she could recognize Dale. Jessica waited nervously in the office. When Wendy returned, she told Jessica that she recognized Dale as the man she saw. She explained that she was 'at least 90%' confident that it was Dale. She added that she had seen Dale around the building once or twice before.

a) Wendy's identification makes it likely that it was Dale who left the flowers on Jessica's desk.

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

b) In general, when people identify someone whom they've already seen once or twice before the identifications are accurate.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

Beth, a doctoral student at a school of education, was scheduled to present her thesis at an important conference to be held at UNL. When she got to the conference, she was devastated to learn that she had forgotten her slides at home. The time was 6:45 PM, and her presentation was scheduled for 7:30 PM. She called home and spoke to her husband who had just walked in from walking the dog. He promised to rush the slides over to the campus. Beth requested that he try to throw on a pair of slacks and a jacket somewhere along the way. At that time of day, it typically takes about 40-45 minutes to get to campus from the house.

a) It is unlikely that Beth's husband could make it in time for the 7:30 PM presentation.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

b) When driving in evening traffic, no matter how aggressively one drives, it is very difficult to shorten the travel time substantially.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

Eric and Daniella are in their twenties. They have been dating for some months. Eric's parents live on the East Coast, and they have never met Daniella. Eric's favorite aunt Rachel flew

in from Boston to spend a weekend in LA. Daniella and Eric took her for a day trip to Santa Barbara. Aunt Rachel was very curious about Eric's relationship with Daniella, and she closely observed the couple throughout the day. Daniella knew that Aunt Rachel was very close to Eric's mother, and that she was likely to report back to Eric's parents. Overall, Rachel perceived that Eric with Daniella were loving and respectful of one another, even though she noticed them bickering once or twice over silly issues. She found Daniella to be charming, and she was optimistic about the couple's relationship.

a) Aunt Rachel's favorable view of the relationship was correct.

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

b) Generally speaking, if you closely observe a couple for an entire day, you can get a good sense of their relationship.

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

c) Aunt Rachel's favorable view of the relationship was influenced by the fact that on that day Daniella displayed particular affection towards Eric.

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

d) In general, one cannot really read other people's emotions. This is especially true when one doesn't know the other person well.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

Lori works as a salesperson for an electric supplies company. Over the past four months she has accumulated a debt of more than $8,000 to her credit card company. The credit card company had threatened to take legal action. Soon thereafter, Lori pays back her debt in full, and there is some doubt as to the source of that money. She explains that she assumed the debt to help her brother Bill finance the flower store he operates in Santa Barbara. Lori says that she repaid the debt from the money Bill returned her. She explains that she has no bank documents to prove these transfers because in the flower business, financial transactions are typically done in cash.

a) Lori repaid her debt to the credit card company with the money she received from her brother Bill.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

b) In the flower business, financial deals are typically done in cash.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

Steve and Lisa are in their late twenties. They have been dating for some months. Their relationship seems quite serious to some of their friends, though they have not yet made any

common plans for the future. The other night Steve invited Lisa over to watch a video. Lisa showed up two hours late, and seemed to be in a downbeat mood. She said that as she was walking out the door, she received a long-distance phone call from Mike, an ex-boyfriend with whom she had an intensive relationship in her junior year in college. Mike's mother was just diagnosed with breast cancer and he was very depressed. She apologized for coming late, but explained that Mike really needed to talk to someone who could understand him.

a) The phone conversation with Mike affected Lisa because she stills feels intimately connected to him.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

b) Old love affairs often interfere with new relationships.

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

A software company discovered that late one night, one of its employees had accessed the company's accounting system from home, and had stolen about $7,000. Hank Lynch, an employee of the company, was put under investigation. The investigators found that the accounting system was accessed through an internet provider called LinkNet. Hank used LinkNet as his internet provider. At the time, 6% of the local market of internet users in Orange County used LinkNet as their internet provider.

The fact that Hank used LinkNet as his internet provider, makes it likely that it was Hank who had accessed the accounting system.

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

Generally speaking, people who have committed a crime, are likely to commit additional crimes somewhere down the line.

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

Linda Owens works as a marketing instructor for a chain of gas stations. She spends a couple of days a week visiting the company's locations throughout Los Angeles County. Every month Linda submits her expenses for reimbursement. One day Linda's boss angrily chastised her for submitting unallowable expenses. Linda complained that many of her colleagues submit the same expenses, and that she was picked on unfairly. Her boss did not agree. He refused to reimburse her for the expenses and delayed her eligibility for promotion by a year. Linda was deeply offended by this incident. In the following weeks, she was seen working longer hours.

a) Linda was looking for a way to get back at the company for the disciplinary measures taken against her.

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

b) Generally speaking, when people feel that they have been treated unjustly, they often do something spiteful to get back.

Strongly Disagree				Neither Agree Nor Disagree					Strongly Agree	
-5	-4	-3	-2	-1	0	1	2	3	4	5

Appendix G.

Booster manipulation

Life Events Questionnaire (angry / sad)

Instructions: Before starting next study, the researchers are interested in how individuals write about life events.

Please answer the questions below, taking the time to give complete and detailed responses.

Question 1: In previous study, you described in more detail the one situation that makes you (or has made you) most **angry (Sad)**. Please think back to your answer and summarize it below.

_____

_____

_____

_____

_____

_____

_____

Next, we would like you to explain why this situation made you angry (sad) in the space provided below.

_____

_____

_____

_____

_____

_____

_____

Appendix H

The Case of Jason Wells

- General Instructions (Preponderance of Evidence/ Beyond a Reasonable Doubt)

In this experiment, you will assume the role of an arbitrator in a disciplinary case brought by a company against one of its employees.

A construction company called Big Buildings Inc. has initiated a disciplinary procedure against its employee Jason Wells. The company charges that Jason stole $5,200 from the company's safe. The company is seeking to terminate Jason's contract and to order him to return the money.

The employment contract stipulates that in situations like this, the parties can agree to have the issue settled by an arbitrator rather than resorting to the court system. You have been asked to serve as the arbitrator. There are no other arbitrators, so you will be expected to make the decision by yourself.

The arbitration procedure is as follows: you will be presented with the evidence (which is not disputed), and then with the arguments made by the lawyers representing both parties. Your task is straightforward: based on the evidence and arguments, you must determine whether or not the person who took the money from the company's safe was the defendant, Jason Wells.

The company needs to prove its case by the "**PREPONDERANCE OF EVIDENCE (BEYOND A REASONABLE DOUBT).**"

Note that to complete this task, you are not expected to have any prior legal knowledge.

Appendix I

Case Description (Medium probability-50%/ High probability-99%)

Background: Jason Wells

Jason Wells is thirty-four years old. He lives in Omaha with his wife May and two children. Jason works for Big Buildings Inc., a large construction company based in Douglas County. After working as a foreman for over two years, Jason complained to his superiors that the work was causing him back pain. Rather than risk a lawsuit, Jason's boss offered him a job at the company's main office and assigned him to the position of supervisor. As supervisor, his job is to oversee the progress of construction projects and to coordinate the different work teams involved. Jason is generally regarded as a hard worker. His peers testified also that he is a reserved guy who, at times, can be pretty moody.

At the end of every day, the company's bookkeeper places all of the company's cash in a safe. The safe is located in the back of the bookkeeper's office. The safe is used also for safeguarding other kinds of sensitive information, including pending bids and project reports. In addition to the bookkeeper and her assistant, the supervisors, senior sales people, and executives access the safe. In total, **about 18 people (8 people), including Jason,** have access to the safe. The safe has a timing mechanism that records the time and date every time it is opened and closed.

One morning, the company's bookkeeper discovered that cash totaling $5,200 was missing from the safe. The timing mechanism showed that the safe was last opened the previous night at 7:14 PM. Following an investigation conducted by a former FBI agent, the company has initiated this disciplinary procedure against Jason Wells.

You will now be presented with the evidence presented by the parties. All witnesses took an oath to tell the truth, and were warned that false testimony could result in a criminal prosecution for perjury. The evidence is not in dispute.

Read the evidence carefully and be sure that you understand it. There is no need to memorize the evidence; you will be able to look back at it whenever you like.

Summary of the Evidence.

A video camera mounted near the entrance to Big Buildings' office showed a car screaming out of the parking lot at 7:17 that evening. However, the image was smudged and the investigator could not identify the car's license plate. The video revealed that the car was a white sedan. Jason drives a white sedan, and he was seen driving it to work that day. DMV records obtained by the investigator show that **only 6% (0.1%) of cars registered in Omaha are white sedans**.

The investigator discovered that the day after the disappearance of the money, Jason repaid a debt of $4,870 to his credit card company. The debt had been mounting over the previous three months, and the credit card company had threatened to take legal action. Jason testified that he assumed the debt to help his sister-in-law Lynn with the flower store she operates in San Bernardino, and that he repaid the debt from money she returned to him. He explained that he could not show bank documents to prove these transfers because in the flower business, financial transactions are typically done in cash.

Lory, an executive at Big Buildings testified that she saw Jason at 8 PM that night, as they were both picking up their children from a high school swim meet. At that time, Jason was

dressed in sweats. Lory stated that around those hours of the night, it typically takes about 40-45 minutes to get from the office to their neighborhood.

Jason testified that he has kept a clean criminal record for the last sixteen years. At the age of eighteen he was caught trying to break into an apartment, and was convicted for a misdemeanor. Since then, he has had no problems whatsoever with the law.

A couple of months before the incident, Jason was summoned by his boss to discuss expenses that he had submitted for reimbursement. The boss angrily chastised him for submitting unallowable expenses. Jason protested that many other supervisors do the same routinely, and that it was unfair that he was picked on. Jason's boss was unconvinced. She refused to reimburse him for his expenses and delayed his eligibility for promotion by a year. Jason was deeply offended by this incident. In the following weeks, Jason was seen working late more frequently.

A technician who was called in to repair the photocopying machine testified that sometime around 7:15 PM, he saw a person rushing out of the bookkeeper's office carrying a bag. The next day, the investigator took the technician to Jason's office. The technician identified Jason as the person he saw. He explained that he **was 'at least 80% (95%)' confident that it was Jason**. He added that he had briefly seen Jason in the office once or twice before.

The Parties' Arguments.

You will now be presented with the arguments made by the company's lawyer followed by the arguments made by Jason's lawyer. In light of these arguments, you will later be asked to assess the facts of the case.

The Company's Arguments.

- The fact that only **6% (0.1%)** of cars in the area are white sedan makes it very likely that it was Jason who was filmed leaving the parking lot.
- It is no coincidence that Jason repaid his debt to the credit card company just one day after the money was stolen. He paid the debt with the money he stole from the company's safe.
- As a matter of fact, it is not true that financial deals in the flower business are done only in cash.
- Jason could have easily rushed to make it to the swim meet in time for the 8 PM pickup.
- Even in bad traffic, if one drives aggressively enough, one can shorten travel time substantially.
- Generally speaking, people who have already committed a crime, are likely to commit crimes somewhere down the line.
- Jason was angered by his boss' reprimand. Stealing the money from the company safe was a way of getting back at the company.
- Generally speaking, when people feel that they have been treated unjustly, they often do something spiteful to get back.
- The fact that the technician was more than **80% (95%)** confident in his identification of Jason as the person seen hurrying out of the bookkeeper's office proves that it was Jason who stole the money.
- In general, people are very accurate in recognizing other people, especially if they've seen them before.

The Defense Arguments.

- The fact that a high **6% (0.1%)** of cars in the area are white sedan makes it less likely that it was Jason who was filmed leaving the parking lot.
- Jason repaid the debt to the credit card company with the money he received from his sister-in-law Lynn.
- As a matter of fact, in the flower business people deal mostly in cash.
- It was virtually impossible for Jason to have driven all the way from the office and change his clothes in time for the 8 PM pickup.
- When driving in bad traffic, no matter how aggressively one drives, it is very difficult to shorten the travel time substantially.
- It is not true that people who have committed a crime continue to commit additional crimes.

- Jason was not trying to get back at the company for being reprimanded. On the contrary, he worked even harder to prove himself to his boss.
- Generally speaking, when people feel that they have been treated unjustly, they try to prove themselves by doing even better at their job.
- The fact that the technician was not certain about his identification of Jason means that the person he saw could have been somebody else.
- In general, when people think that they can identify other people, particularly when they have seen them only once or twice, they often make mistakes.

Appendix J.

Manipulation of jury instructions (The official model jury instructions of the Ninth Circuit)

1) **By a preponderance of the evidence**

"You should decide by a preponderance of the evidence, it means you must be persuaded by the evidence that the claim is more probably true than not true. You should base your decision on all of the evidence, regardless of which party presented it."

2) **Proof beyond a reasonable doubt**

"Please note that in these cases accused persons are particularly protected. They should only be convicted if the evidence is so convincing that there is no reasonable doubt that the person is guilty. Proof beyond a reasonable doubt is proof that leaves you firmly convinced that the defendant is guilty. It is not required to prove guilt beyond all possible doubt. A reasonable doubt is a doubt based upon reason and common sense and is not based purely on speculation. It may arise from a careful and impartial consideration of all the evidence, or from lack of evidence."

Appendix K

Deciding the verdict

You have heard all the evidence and the parties' arguments. You are now requested to decide your verdict. Base your decision as soundly and fairly as possible on the facts and arguments made in the case.

[Preponderance of evidence version] According to the rules governing this arbitration, the company must prove its case **by a preponderance of the evidence. Preponderance of the evidence means 'more likely than not.' In other words, the company should win the case if it is more likely that Jason took the money than that he did not. If it is unlikely that Jason took the money or if the evidence is evenly balanced, the decision should go in Jason's favor.**

[Beyond a reasonable doubt version] According to the rules governing this arbitration, the company must prove its case **beyond a reasonable doubt. Proof beyond a reasonable doubt is proof that leaves you firmly convinced of the defendant's guilt. There are very few things in this world that we know with absolute certainty, and in criminal cases the law does not require proof that overcomes every possible doubt. In other words, the company should win the case if you are firmly convinced that Jason took the money than that he did not. If your are not firmly convinced that Jason took the money, the decision should go in Jason's favor.**

Take as much time as you need. You may look back at the evidence and the arguments as much as you like. Please think through the entire case as thoroughly as possible.

Please report your judgment: did Jason Wells steal the money from the safe?

- Yes, it was Jason Wells who stole the money from the company's safe
- No, it was not Jason Wells who stole the money from the company's safe

Using the provided scale, please indicate how confident you are that this is the correct conclusions (click one number):

Completely Uncertain					Neither certain nor uncertain					Completely Certain
0	1	2	3	4	5	6	7	8	9	10

Please estimate the probability that Jason Wells had stolen the money from the safe

( ) %

Please estimate the probability that another person took the money

( ) %

Please specify the level of probability necessary for convicting under the standard of proof given to you

( ) %

Appendix L

Post-test

Please answer the following questions:

1. a) If Jason stole the money, it is unlikely that he could have made it in time for the 8 PM swim meet. **(I)**

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree	
-5	-4	-3	-2	-1	0	1	2	3	4	5

1. b) When driving in evening traffic, no matter how aggressively one drives, it is very difficult to shorten the travel time substantially. **(I)**

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree	
-5	-4	-3	-2	-1	0	1	2	3	4	5

2. a) Jason repaid his debt to the credit card company with the money he received from his sister-in-law Lynn. **(I)**

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree	
-5	-4	-3	-2	-1	0	1	2	3	4	5

2. b) In the flower business, financial deals are typically done in cash. **(I)**

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree	
-5	-4	-3	-2	-1	0	1	2	3	4	5

3. a) The technician's identification of Jason makes it likely that the person hurrying out of the bookkeeper's office was in fact Jason. **(G)**

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

3. b) In general, when people identify someone whom they've already seen once or twice before, the identifications are accurate**. (G)**

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

4. The fact that the car seen leaving the scene was the same as Jason's makes it likely that it was Jason who was captured on the video camera driving away. **(G)**

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

5. Generally speaking, people who have committed a crime are likely to commit additional crimes somewhere down the line. **(G)**

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

6. a) Jason was looking for a way to get back at the company for the disciplinary measures taken against him. **(G)**

Strongly Disagree					Neither Agree Nor Disagree					Strongly Agree
-5	-4	-3	-2	-1	0	1	2	3	4	5

6. b) Generally speaking, when people feel that they have been treated unjustly, they often do something spiteful to get back. **(G)**

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree	
-5	-4	-3	-2	-1	0	1	2	3	4	5

6. c) Jason's reaction to the unfair disciplinary action was to prove himself to his boss by working even harder. **(I)**

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree	
-5	-4	-3	-2	-1	0	1	2	3	4	5

6. d) Generally speaking, when people feel that they have been treated unjustly, they typically try to prove themselves by doing even better at their job. **(I)**

Strongly Disagree					Neither Agree Nor Disagree				Strongly Agree	
-5	-4	-3	-2	-1	0	1	2	3	4	5

Note.

(1) G= guilt item; I = innocence item

(2) Calculating Coherence shift index

Convictor (participants who decide against the defendant) =

Posttest score (sum of guilt items – sum of innocence items) – Pretest score (sum of guilt items – sum of innocence items)

Acquitter (participants who decide in favor of the defendant) =

Posttest score (sum of innocence items – sum of guilt items) – Pretest score (sum of innocence items – sum of guilt items)

Appendix M
Demographics Questionnaire

1. What is your age? ____ years old

2. What is your gender? ____ Male ____ Female

3. What is your ethnic origin and/or race?

_____ African American _____ Asian American

_____ Caucasian _____ Hispanic

_____ Latin American _____ Native American

_____ Other

4. What is the highest level of formal education that you have completed?

_____ Less than high school

_____ High school

_____ Associate's degree

_____ Bachelor's degree

_____ Master's degree

_____ Professional degree (i.e., M.D., J.D., Ph.D.)

5. Is English your primary language?

_____ yes _____ no

6. Have you served as a juror on: (check one answer for each):

a state civil case? _____ yes _____ no

a state criminal case? _____ yes _____ no

a federal civil case? _____ yes _____ no

a federal criminal case? _____ yes _____ no

7. Are you registered to vote?

_____ yes _____ no

8. Do you have a valid driver's license?

_____ yes _____ no

9. Are you a citizen of the United States?

_____ yes _____ no

10. Are you a convicted felon without civil rights?

_ ____ yes _____ no

11. What is your religious preference (if any)?

_____

12. What is your current work status? (check one):

_____ employed full time _____ employed part time _____ unemployed

13. What is your political affiliation? (check one):

_____ Democrat _____ Republican _____ other _____ none

14. When you read the trial summary, what did you think the race of the defendant was?

_____

15. What did you think the purpose of Experiment 2 was? Please explain why you thought this.

_____

_____

_____

_____

16. What did you think the purpose of Experiment 3 was? Please explain why you thought this.

_____

_____

_____

_____

Appendix N

Debriefing

Thank you for participating in this study. The purpose of this study was to examine how individuals evaluate evidence and use the standards of proof to decide cases. One of the main objectives of this study is to shed empirical light on how individuals apply the specified standard of proof to their legal finding. Several jury decision-making models suggest that jurors may apply standards of proof in a normatively appropriate way. However, other models suggest that jurors may use evidence to fit the favored verdict, and jurors may use standards of proof for justification. The goal of this study is to clarify this issue by examining the roles of emotion in individual's judgment.

Dual process theories of cognition posit that human judgment is the product of two concurrent cognitive systems: an intuitive system, which operates automatically and effortlessly, and an analytic system which is more effortful, controlled, and normatively rational. Furthermore, the experience of certain emotions may facilitate one of two systems. Angry individuals tend to rely on intuition when making decisions, while sad people tend to process information more systematically. We expect to find that angry individuals will become increasingly coherent with the emerging decisions when they are required to reach higher standards of proof. However, we expect sad individuals to calibrate their judgments more systematically with higher standards of proof.

Once again, we thank you for your participation; we ask that you not discuss this research with any future participants as it may negatively influence the results of our study. You have put a great deal of time and effort acting as a research participant in our study. If you speak about it with any other undergraduate students, that will invalidate the data that we collected from you and the other students. As a result, the time and effort that you put into this study, as well as our own time and effort, will have been wasted.

If you have any questions or concerns about this project, or if you would like to know the general results of the research upon its completion, feel free to contact Yimoon Choi (yimoon@gmail.com) or Richard Wiener at 402-472-9639.

Appendix O
Informed Consent

Identification of Project: Assessing social issues and legal decision making

You are invited to participate in research that examines social cognition in assessing social issues and legal decision making. A group of researchers studying social cognition is interested in a number of different types of information processing. To facilitate data collection, the researchers have combined several of experiments. You were selected as a participant because you signed up to participate in web-based research. To complete this research, you must be at least 18 years of age. Researchers at the University of Nebraska-Lincoln are conducting this research under the direction of Yimoon Choi, a graduate student in the Department of Psychology, and Dr. Richard Wiener, professor in the Department of Psychology.

This research will consist of three separate experiments. Each experiment will take approximately 30 minutes to complete. All three experiments will take place on the internet. In Experiment 1, you will answer some questions regarding your opinions about social issues. You will read short stories and answer some questions. Experiment 2 will examine legal decision making. You will play the role of a juror in a legal case and you will determine whether the defendant is guilty and answer several other questions about the case.

Knowledge generated through these studies will help researchers better understand individual differences in information processing about life events, which will contribute to the study of social cognition. This research will also help researchers better understand legal decision making, which can lead to improvements in the legal system. The risks of this research are minimal. Some participants may also experience some minimal discomfort when reading about the crimes described in Experiments 2. In addition, this research, like much other research in Psychology, may contain some questions that you may find sensitive or personal.

If you do not feel comfortable answering a question during the study, you may choose not to respond. You will not be penalized for skipping any questions that you do not want to answer. You may skip any questions or stop completing any survey without penalty. The alternative to this research is non-participation. Your participation is voluntary. You are free to decide not to participate in this research or to withdraw at any time without adversely affecting your relationship with the investigators or the University of Nebraska-Lincoln. Your refusal to participate will involve no penalty to you or loss of any benefits to which you are otherwise entitled.

The results of this research may be published, but your name and identity will not be revealed and all of the data and information collected from you will remain anonymous. All data will be identified with numbers that have no links to you as a research participant and will be kept in a secure lab at the University of Nebraska for a period of five years after which it will be destroyed. Nonetheless, some of the questions may ask about, among other things, your gender, your ethnicity, and your age. Feel free to leave any of those items unanswered if you feel that the answers may reveal your identity.

Yimoon Choi (a graduate student) and Dr. Richard L. Wiener are conducting this research.

They can answer any questions or concerns about the research. You may contact Yimoon Choi at (402) 326-8439 or yimoon@gmail.com, or Dr. Wiener at (402) 472-1137 or rwiener2@unl.edu. To obtain more information about your rights as a research participant or to report any concerns about this research, please contact the University of Nebraska-Lincoln Institutional Review Board (IRB) for Human Research at (402) 472-6965.

If you wish to participate in this research, please read the following statement and provide your electronic signature by clicking at the bottom of the webpage. You may print a copy of this informed consent form for your personal records.

I have read and understood the information presented above. If I have any questions before I begin, I may contact the researchers. Otherwise, my concerns have been answered to my satisfaction via this consent form. I consent to take part in this research.

____ I agree, and I consent to take part in this research

____ I disagree, and I do not wish to take part in this research

Name and Phone Number of Investigators:

Yimoon Choi (402) 326-8439

Dr. Richard Wiener (402) 472-1137

Appendix P

The Case of Jason Wells

- General Instructions (Theft / First degree murder)

In this experiment you will play the role of a juror in a criminal case, in which the state charges Jason Wells with the crime of **theft (first degree murder and theft)**.

You will review the evidence (which is not disputed), and then you will review the arguments made by the lawyers representing both sides. Your task is straightforward. Based on the evidence and arguments, you should determine whether or not the person who committed the crime was the defendant.

You are not expected to have any prior legal knowledge. There will be no deliberation with other jurors. You will make your decision by yourself. In this experiment, you will not make any decisions regarding sentencing. You will only determine the defendant's guilt, that is, whether he committed the theft or not.

Please pay close attention to the information in the trial summary. At the conclusion of the trial summary, you will make a decision regarding whether the defendant is guilty or not guilty.

Before the trial begins, the Judge instructs you as follows:

By serving as a juror, you are assuming an obligation to truly try the facts of this case and give a true verdict according to the law and evidence presented in Court.

Theft version: In his opening statements, the prosecutor promises to produce evidence from multiple sources that will show decisively that the person guilty of the theft is the defendant. The prosecutor states that the evidence will leave you no alternative but to convict the defendant.

Murder version: In his opening statements, the prosecutor promises to produce evidence from multiple sources that will show decisively that the person guilty of the theft is the defendant. The prosecutor states that the evidence will leave you no alternative but to convict the defendant of this crime. In addition, the prosecutor promises to produce evidence from multiple sources that will show decisively that the defendant is also guilty of murder. The prosecutor exclaims that this was a most gruesome and atrocious murder. The deceased suffered indescribable pain and degradation and his suffering demands that justice be done. The prosecutor adds that a ruthless, heartless and greedy person committed this brutal act and that the evidence will show that the perpetrator is the defendant. The prosecutor states that the evidence will leave you no alternative but to convict the defendant for his horrific deeds.

In response, the defendant's defense attorney insists that the defendant did not commit this crime. The attorney explains that this is a case in which an innocent person appears to be entangled in a crime only because of unfortunate coincidences. He states that it will be your civic,

legal, and moral duty to make sure that the state does not punish an innocent person for a crime that he did not commit.

You will now review a summary of the evidence presented at trial. The evidence is not in dispute. Read the evidence carefully and be sure that you understand it. Take as much time as you like. There is no need to memorize the evidence; you will be able to look back at it whenever you like.

Appendix Q

Case Description (Medium probability-50%/ High probability-99%)

Summary of the Evidence

Jason Wells is thirty-four years old. He lives in Omaha with his wife May and two children. Jason works for Big Buildings Inc., a large construction company based in Douglas County. After working as a foreman for over two years, Jason complained to his superiors that the work was causing him back pain. Rather than risk a lawsuit, Jason's boss offered him a job at the company's main office and assigned him to the position of supervisor. Jason's job as a supervisor was to oversee the progress of construction projects and to coordinate the different work teams involved. Others generally regard Jason as a hard worker. His peers testified also that he is a reserved person who, at times, can be moody.

At the end of every day, the company's bookkeeper places all of the company's cash in a safe. The safe is located in the back of the bookkeeper's office. Big Buildings uses the safe safeguarding other kinds of sensitive information, including pending bids and project reports. In addition to the bookkeeper and her assistant, supervisors, senior sales people, and executives have access to the safe. In total, about **18 people (8 people), including Jason**, have access to the safe. The safe has a timing mechanism that records the time and date every time someone opens and closes it.

(Theft version: One morning, the company's bookkeeper discovered that cash totaling $5,200 was missing from the safe. The timing mechanism showed that the safe was last opened the previous night at 7:14 PM.)

(**Murder version**: One morning, the company's bookkeeper discovered a horrifying scene in her office. The building's security guard was lying dead in what seemed to be the scene of a most gruesome crime. She also found that cash totaling $5,200 was missing from the safe. The timing mechanism showed that the safe was last opened the previous night at 7:10 PM.

As reconstructed by the police detectives and the medical examiner, the events of the previous night were as follows. The building security guard detected a person attempting to remove the contents of the safe. The suspect assaulted the unarmed guard and struck him repeatedly with a heavy flashlight in his face and over the head. As a result, the guard suffered 9 severe facial and skull bone fractures, as well as numerous skin lacerations. The guard also sustained severe organ damage in the abdominal area, which was presumably caused by forceful kicks to the stomach. Before leaving the scene, the suspect removed the cash from the safe totaling $75,200, tore out the phone line and locked the office door. The likely causes of death were a massive brain hemorrhage brought upon by the severe trauma to the head, and asphyxiation by choking on his blood. Investigators estimate that the guard died two to four hours after the attack.)

Following an investigation conducted by the police detectives, a prosecutor filed criminal charges.

You will now read the evidence presented by the parties. All witnesses took an oath to tell the truth and authorities warned them that false testimony could result in a criminal prosecution for perjury. The evidence is not in dispute.

Read the evidence carefully and be sure that you understand it. There is no need to memorize the evidence; you will be able to look back at it whenever you like.

Summary of the Evidence.

A video camera mounted near the entrance to Big Buildings' office showed a car screaming out of the parking lot at 7:17 that evening. However, the image was smudged and the investigator could not identify the car's license plate. The video revealed that the car was a white sedan. Jason drives a white sedan, and he was seen driving it to work that day. DMV records obtained by the investigator show that **only 6% (0.1%) of cars registered in Omaha are white sedans**.

The investigator discovered that the day after the disappearance of the money, Jason repaid a debt of $4,870 to his credit card company. The debt had been mounting over the previous three months, and the credit card company had threatened to take legal action. Jason testified that he assumed the debt to help his sister-in-law Lynn with the flower store she operates in San Bernardino, and that he repaid the debt from money she returned to him. He explained that he could not show bank documents to prove these transfers because in the flower business, financial transactions are typically done in cash.

Lory, an executive at Big Buildings testified that she saw Jason at 8 PM that night, as they were both picking up their children from a high school swim meet. At that time, Jason was dressed in sweats. Lory stated that around those hours of the night, it typically takes about 40-45 minutes to get from the office to their neighborhood.

Jason testified that he has kept a clean criminal record for the last sixteen years. At the age of eighteen he was caught trying to break into an apartment, and was convicted for a misdemeanor. Since then, he has had no problems whatsoever with the law.

A couple of months before the incident, Jason was summoned by his boss to discuss expenses that he had submitted for reimbursement. The boss angrily chastised him for submitting unallowable expenses. Jason protested that many other supervisors do the same routinely, and that it was unfair that he was picked on. Jason's boss was unconvinced. She refused to reimburse him for his expenses and delayed his eligibility for promotion by a year. Jason was deeply offended by this incident. In the following weeks, Jason was seen working late more frequently.

A technician who was called in to repair the photocopying machine testified that sometime around 7:15 PM, he saw a person rushing out of the bookkeeper's office carrying a bag. The next day, the investigator took the technician to Jason's office. The technician identified Jason as the person he saw. He explained that he **was 'at least 80% (95%)' confident that it was Jason**. He added that he had briefly seen Jason in the office once or twice before.

The Parties' Arguments

You will now be presented with the prosecution's closing arguments followed by the closing arguments made by the defendant's lawyer. In light of these arguments, you will later be asked to assess the facts of the case and to determine whether the defendant is guilty of the crime.

The Prosecution's Arguments

The prosecutor states that the evidence of the defendant's guilt is overwhelming.

- The fact that only **6% (0.1%)** of cars in the area are white sedan makes it very likely that it was Jason who was filmed leaving the parking lot.
- It is no coincidence that Jason repaid his debt to the credit card company just one day after the money was stolen. He paid the debt with the money he stole from the company's safe.
- As a matter of fact, it is not true that financial deals in the flower business are done only in cash.
- Jason could have easily rushed to make it to the swim meet in time for the 8 PM pickup.
- Even in bad traffic, if one drives aggressively enough, one can shorten travel time substantially.
- Generally speaking, people who have already committed a crime, are likely to commit crimes somewhere down the line.
- Jason was angered by his boss' reprimand. Stealing the money from the company safe was a way of getting back at the company.
- Generally speaking, when people feel that they have been treated unjustly, they often do something spiteful to get back.
- The fact that the technician was more than **80% (95%)** confident in his identification of Jason as the person seen hurrying out of the bookkeeper's office proves that it was Jason who stole the money.
- In general, people are very accurate in recognizing other people, especially if they've seen them before.

The prosecutor concludes that the defendant planned to steal the cash out of the company's safe. It is your responsibility to correct the wrong that the defendant did. In sum,

he asserts, it is your moral and legal duty to convict the defendant for the theft as charged [and for murder in the first degree with theft].

The Defense Arguments

The defense attorney insists that it was not the defendant who committed the crime. He argues that the prosecution's case is weak and fragmented, and thus fails to prove the defendant's guilt.

- The fact that a high **6% (0.1%)** of cars in the area are white sedan makes it less likely that it was Jason who was filmed leaving the parking lot.
- Jason repaid the debt to the credit card company with the money he received from his sister-in-law Lynn.
- As a matter of fact, in the flower business people deal mostly in cash.
- It was virtually impossible for Jason to have driven all the way from the office and change his clothes in time for the 8 PM pickup.
- When driving in bad traffic, no matter how aggressively one drives, it is very difficult to shorten the travel time substantially.
- It is not true that people who have committed a crime continue to commit additional crimes.
- Jason was not trying to get back at the company for being reprimanded. On the contrary, he worked even harder to prove himself to his boss.
- Generally speaking, when people feel that they have been treated unjustly, they try to prove themselves by doing even better at their job.
- The fact that the technician was not certain about his identification of Jason means that the person he saw could have been somebody else.
- In general, when people think that they can identify other people, particularly when they have seen them only once or twice, they often make mistakes.

In sum, the attorney explains that the defendant is an innocent person who was charged for a crime [for crimes] only because of a series of unfortunate coincidences. He

implores you to make sure that this innocent person does not get punished for a crime [crimees] he did not commit.

Appendix R

Jury instruction (Theft / Murder)

Deciding the verdict

You have heard all the evidence and the parties' arguments.

[Theft only]: The judge explains that according to the law of the state**, stealing company property constitutes theft. In this state theft includes each of the following elements, which the state must prove beyond a reasonable doubt.**

Please read the following statutory definitions of theft and use this definition to answer the questions below.

28-511. Theft by unlawful taking or disposition. (1) A person is guilty of theft if he or she takes, or exercises control over, movable property of another with the intent to deprive him or her thereof.

[Murder]: The judge explains that according to the law of the state, stealing company property constitutes theft. In this state theft includes each of the following elements, which the state must prove beyond a reasonable doubt. The judge explains that the state has also charged the defendant with the crime of first-degree murder.

In this state murder in the first degree includes each of the following elements, which the state must prove beyond a reasonable doubt. Please read the following statutory definitions of first-degree murder and use this definition to answer the questions below.

28-303. Murder in the first degree: A person commits murder in the first degree if he or she kills another person (1) purposely and with deliberate and premeditated malice, or (2) in the perpetration of or attempt to perpetrate any sexual assault in the first degree, arson, robbery, kidnapping, hijacking of any public or private means of transportation, or burglary, or (3) by administering poison or causing the same to be done; or if by willful and corrupt perjury or subornation of the same he or she purposely procures the conviction and execution of any innocent person.

In this state theft includes each of the following elements, which the state must prove beyond a reasonable doubt. Please read the following statutory definitions of first-degree murder and use this definition to answer the questions below.

28-511. Theft by unlawful taking or disposition. (1) A person is guilty of theft if he or she takes, or exercises control over, movable property of another with the intent to deprive him or her thereof.

The question before you is only whether the defendant was the person who committed the crime. Before reaching your verdict, the judge instructs you as follows:

It is your duty to find the facts from all the evidence admitted in this case. You must not be influenced by any personal likes or dislikes, prejudices or sympathy. That means that you must decide the case solely on the evidence before you and according to the law.

The prosecution has the burden of proving the defendant guilty beyond a reasonable doubt. If you have ever served as a juror in a civil case, you were told that it is only necessary to prove that a fact is more likely true than not true. In criminal cases, the government's proof must be more powerful than that. It must be beyond a reasonable doubt.

Proof beyond a reasonable doubt is proof that leaves you firmly convinced of the defendant's guilt. There are very few things in this world that we know with absolute certainty, and in criminal cases, the law does not require proof that overcomes every possible doubt. If, based on your consideration of the evidence, you are firmly convinced that the defendant is guilty of the crime[s] charged you must find him guilty. If on the other hand, you think there is a real possibility that he is not guilty, you must give him the benefit of the doubt and find him not guilty.

Please make your decision now. You may look back at the evidence and the arguments as much as you like. Please think through the entire case as thoroughly as possible. After reporting your decision, you will be asked some questions about the case.

Appendix S.

Deciding the verdict (Theft / Murder-verdict first questionnaires)

You have heard all the evidence and the parties' arguments. Please decide your verdict now. Base your decision as soundly and fairly as possible on the facts and arguments made in the case.

Take as much time as you need. You may look back at the evidence and the arguments as much as you like. Please think through the entire case as thoroughly as possible.

[Theft version]
Please report your verdict: did Jason Wells steal the money from the safe?

- **Yes, it was Jason who stole the money from the company's safe**
- **No, it was not Jason who stole the money from the company's safe**

Using the provided scale, please indicate how confident you are that this is the correct conclusions (click one number):

Completely Uncertain					Neither certain nor uncertain					Completely Certain
0	1	2	3	4	5	6	7	8	9	10

Please estimate the probability that the accused person had **stolen the money from the safe (killed the security guard)**

( ) %

Please estimate the probability that another person **took the money (killed the security guard)**

( ) %

Please specify the level of probability necessary for convicting beyond a reasonable doubt

( ) %

[Murder version]

Please report your verdict: did Jason Wells kill the security guard?

- **Yes, it was Jason who killed the security guard**
- **No, it was Jason who killed the security guard**

Using the provided scale, please indicate how confident you are that this is the correct conclusions (click one number):

Completely Uncertain					Neither certain nor uncertain					Completely Certain
0	1	2	3	4	5	6	7	8	9	10

Please estimate the probability that the accused person had **killed the security guard**

( ) %

Please estimate the probability that another person **killed the security guard**

( ) %

Please specify the level of probability necessary for convicting beyond a reasonable doubt

( ) %

Please report your verdict: did Jason Wells steal the money from the safe?

- **Yes, it was Jason who stole the money from the company's safe**
- **No, it was not Jason who stole the money from the company's safe**

Using the provided scale, please indicate how confident you are that this is the correct conclusions (click one number):

Completely Uncertain					Neither certain nor uncertain					Completely Certain
0	1	2	3	4	5	6	7	8	9	10

Please estimate the probability that the accused person had **stolen the money from the safe**

( ) %

Please estimate the probability that another person **took the money**

( ) %

Please specify the level of probability necessary for convicting beyond a reasonable doubt

( ) %

Deciding the verdict (Theft / Murder -probability first questionnaires)

You have heard all the evidence and the parties' arguments. Please decide your verdict now. Base your decision as soundly and fairly as possible on the facts and arguments made in the case.

You may look back at the evidence and the arguments as much as you like. Please think through the entire case as thoroughly as possible.

[Theft version]

Please estimate the probability that the accused person had stolen the money from the safe

( ) %

Please estimate the probability that another person took the money

( ) %

Please specify the level of probability necessary for convicting beyond a reasonable doubt

( ) %

Please report your verdict: did Jason Wells steal the money from the safe?

- Yes, it was Jason who stole the money from the company's safe
- No, it was not Jason who stole the money from the company's safe

Using the provided scale, please indicate how confident you are that this is the correct conclusions (click one number):

Completely Uncertain				Neither certain nor uncertain						Completely Certain
0	1	2	3	4	5	6	7	8	9	10

[Murder version]

Please estimate the probability that the accused person had killed the security guard
( ) %

Please estimate the probability that another person killed the security guard
( ) %

Please specify the level of probability necessary for convicting beyond a reasonable doubt
( ) %

Please report your verdict: did Jason Wells kill the security guard?

- **Yes, it was Jason who killed the security guard**
- **No, it was Jason who killed the security guard**

Using the provided scale, please indicate how confident you are that this is the correct conclusions (click one number):

Completely Uncertain					Neither certain nor uncertain					Completely Certain
0	1	2	3	4	5	6	7	8	9	10

Please estimate the probability that the accused person had stolen the money from the safe
( ) %

Please estimate the probability that another person took the money
( ) %

Please specify the level of probability necessary for convicting beyond a reasonable doubt
( ) %

Please report your verdict: did Jason Wells steal the money from the safe?

- **Yes, it was Jason who stole the money from the company's safe**
- **No, it was not Jason who stole the money from the company's safe**

Using the provided scale, please indicate how confident you are that this is the correct conclusions (click one number):

Completely Uncertain					Neither certain nor uncertain					Completely Certain
0	1	2	3	4	5	6	7	8	9	10

Appendix T

Manipulation check for accuracy motivation

How hard did you try to judge the verdict accurately?

Not at all Hard					Moderately hard					Extremely Hard
0	1	2	3	4	5	6	7	8	9	10

How hard did you try to judge the verdict fairly?

Not at all Hard					Moderately hard					Extremely Hard
0	1	2	3	4	5	6	7	8	9	10

Were you motivated to make your judgments accurately?

Not at all Motivated					Moderately Motivated					Extremely Motivated
0	1	2	3	4	5	6	7	8	9	10

Were you motivated to make you judgements fairly?

Was it more important to reach a fair judgment or an accurate judgment?

0	1	2	3	4	5	6	7	8	9	10
It was more important to be fair					It was equally important to be fair and accurate					It was more important to be accurate

What would be the probable penalty associated with the case you decided?

1. 0-1 year
2. 1-5 years
3. 5- 10 years
4. 10-25 years
5. 25 years to life or capital punishment

If you reached the wrong verdict, how much would you regret your decision?

Very little regret					Moderate regret					A great deal of regret
0	1	2	3	4	5	6	7	8	9	10

Appendix U

Debriefing for study 2

Thank you for participating in this study. The purpose of this study was to examine how individuals evaluate evidence and use the standards of proof to decide cases. One of the main objectives of this study is to investigate whether severity of charge influences jurors' information processing. Several studies suggest that seriousness of charge can reduce mock jurors' conviction rates. However, other studies show no effect of seriousness of charge on verdict. The goal of this study is to clarify this issue by examining the roles of automatic and intuitive thinking in individuals' judgment.

Dual process theories of cognition posit that human judgment is the product of two concurrent cognitive systems: an intuitive system, which operates automatically and effortlessly, and an analytic system which is more effortful, controlled, and normatively rational. We expect to find that the severity of a charge might increase systematic or analytic processing because crimes that are more serious are usually more blameworthy. We also expect that once individuals supply a verdict, they will use their intuition to rationalize their verdict.

Once again, we thank you for your participation; we ask that you not discuss this research with any future participants as it may negatively influence the results of our study. You have put a great deal of time and effort acting as a research participant in our study. If you speak about it with any other undergraduate students, that will invalidate the data that we collected from you and the other students. As a result, the time and effort that you put into this study, as well as our own time and effort, will have been wasted.

If you have any questions or concerns about this project, or if you would like to know the general results of the research upon its completion, feel free to contact Yimoon Choi (yimoon@gmail.com) or Richard Wiener at 402-472-9639.