Dissertations   Theses and Dissertations

2015

The Influence of Racial and Crime Stereotypes on Jurors' Representations of Trial Evidence and Verdict Decisions

Katharina Kluwe
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

THE INFLUENCE OF RACIAL AND CRIME STEREOTYPES ON JURORS'

REPRESENTATIONS OF TRIAL EVIDENCE AND VERDICT DECISIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

PROGRAM IN APPLIED SOCIAL PSYCHOLOGY

BY

KATHARINA KLUWE

CHICAGO, ILLINOIS

MAY 2015

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

viii

LIST OF FIGURES

ABSTRACT

Past research has shown that jurors tend to make more severe culpability judgments when defendants are charged with crimes that are associated with their racial group (e.g., Jones & Kaplan, 2013). Although the Story Model (e.g., Pennington & Hastie, 1988) has received much empirical support and currently is the most prominent model in the juror decision making literature, it has not been applied to cases where racially stereotypical crimes might bias verdict decisions. The present research investigated whether the narrative believability of the stories that are suggested at trial, especially story coherence, would explain the effect of race-crime congruence on verdict decisions. Specifically, I attempted to answer the question of whether race-crime congruence affects bias verdict judgments because it leads to an increase in the coherence of the stories jurors form from the evidence. In addition, my goal was to assess whether motivation to control prejudice and stereotype suppression instructions serve as two potential moderating variables. In Study 1 I developed a scale that fully captures the certainty principles affecting acceptance of and confidence in a story (i.e., interpretation of the evidence leading to trial) in the specific context of juror decision making. Specifically, I examined the factor structure underlying a modified version of Yale's (2013) Narrative Believability Scale-12. Study 2 investigated whether the coherence of the prosecution story (and/or the defense story) mediates the relationship between race-crime congruence and verdict judgments in a diverse online sample. In addition, I planned to assess whether high levels

of motivation to control prejudice and the presence of stereotype suppression instructions independently reduce this effect. Moreover, with study 3, I attempted to replicate the results in study 2 with a student sample in the laboratory, using the stories that were actually created by the mock jurors as the mediating mechanisms. Based on the findings of Study 2 and Study 3, an additional study was conducted to investigate whether those null effects that emerged in the other studies were due to order effects. No race-crime congruence effects emerged. Findings and implications were discussed.

CHAPTER 1

RACIAL DISCRIMINATION IN JUROR DECISION MAKING

Purpose of the Current Research

According to the sixth and seventh Amendments to the constitution of the United

States of America as part of the Bill of Rights, all defendants in criminal trials and almost

all defendants in civil trials have the right to a jury trial. Jurors listen to and evaluate the

evidence presented at trial, engage in deliberations with one another, make verdict and

liability decisions, determine compensatory and punitive damage awards, and sometimes

even decide on a sentence. This means that a panel of ordinary citizens has an enormous

amount of responsibility when determining the outcome of a case. Although jurors are

supposed to be impartial, a variety of different extralegal and trial-related factors

influence their decisions. In this current line of research, I investigated how jurors'

stereotypes of both the defendants and crimes affect their verdict decisions and

culpability ratings. The present studies combine the research on stereotypes and the Story

Model of juror decision making in an attempt to understand how stereotypes may affect

decisions in the courtroom, and to identify potential interventions that might be used to

reduce the effects of juror bias on culpability ratings.

The Effects of Race on Juror Decision Making

Race stands out as an important individual difference that affects jurors' verdict

judgments. Past research has shown that, in general, Black defendants are convicted more

1

frequently and receive harsher punishments, when found guilty, than White defendants. For example, in one of the first meta-analyses investigating the effects of race on juror decision making, Sweeney and Haney (1992) found a small significant effect of racial bias on sentencing decisions ($d = .17$). After investigating 14 studies including 2,836 White participants, the results indicated that, on average, White participants gave longer sentences to Black compared to White defendants. This effect was larger when the researchers specified the race of both participants and victims. In addition, Kleider, Knuycky, and Cavrak (2012) investigated the conditions under which this type of discriminatory behavior is more likely to occur. They found that when the case is ambiguous (i.e., the likelihood of conviction or acquittal is equally probable), mock jurors with prejudicial racial attitudes and those under high cognitive load were more likely to give higher verdict-confidence scores to Black, as compared to White, defendants. In contrast to the researchers' expectations, being low on working memory capacity as well as being highly prejudicial also led to higher verdict-confidence scores for Black defendants when the case was strong. Thus, jurors' cognitive resources as well as their racial attitudes might be responsible for the racial bias that is exhibited in the courtroom.

In contrast to the aforementioned results, in their meta-analysis, Mazzella and Feingold (1994) did not find any evidence that jurors judge Black defendants guilty more often and give them longer sentences than they give White defendants. In the light of these apparently inconsistent findings, Devine (2012) emphasized the importance of considering not only defendant race, but the race of the juror as well. He argued that it is

not so much a defendant's race affecting a juror's decision making in isolation; rather, it matters whether or not a juror's race is *different* from the defendant's race. Therefore, White jurors tend to judge Black defendants more harshly than they judge White defendants, while Black jurors are more likely to be unsympathetic towards White jurors in judgments about guilt and sentencing than they are towards Black defendants. Accordingly, a similarity-leniency effect occurs when jurors treat defendants who they perceive to be similar to them more mildly than they treat defendants with who they believe to share fewer attributes.

A recent meta-analysis supports Devine's supposition (Mitchell, Haw, Pfeifer, & Meissner, 2005). The researchers investigated whether jurors treat defendants more harshly when they are part of a racial out-group compared to when they are part of the in-group. They included studies that manipulated defendant race (i.e., Black vs. White) and reported participant/juror race in the same study. The researchers found a small, but significant, effect on racial bias on juror verdicts ($d = .09$) as well as on sentencing decisions ($d = .19$), and these effects appeared to be robust. Participants were more likely to convict the defendant when the defendant was of a different race, especially when ecological validity was low (e.g., continuous measure of guilt, no judicial instructions). Surprisingly, this effect was larger for Black jurors. However, only two studies included Black jurors, a dichotomous guilt measure, as well as standard jury instructions, so caution is warranted when interpreting these results. In addition, the researchers found a similar effect for sentencing decisions. Jurors recommended longer sentences when the defendant was part of a racial out-group. However, the researchers pointed out that only a

small number of null studies would have been needed for this significant effect concerning sentencing judgments to go away. Overall, they concluded that jurors treat defendants more negatively when they are part of the racial out-group (Mitchell et al., 2005).

Racially Stereotypical Crimes

These somewhat contradictory findings of research investigating racial discrimination in the legal decision making domain suggest that other factors might covary with the defendant's race, which then may interact to affect biased verdict decisions. In addition to jurors' race, past research has shown that the type of crime that a defendant was charged with might be one of those factors. Biased decision making might only occur when the jurors associate the type of crime with members of the defendant's race.

Sunnafrank and Fontes (1983) were two of the first researchers to test whether individuals tend to assimilate certain crimes with either a Black or a White criminal. In their first study, they showed participants five pictures of Black and five pictures of White criminals, and then asked them to match each photograph with a crime they believed that the criminal had committed. Although no significance testing was employed, which the researchers justified by stating that the differences were so large that it was not necessary, they discovered that certain crimes were associated with a certain race. Specifically, participants attributed assault-mugging, grand-theft auto, assault of a police officer, and soliciting more frequently to a Black criminal, while they attributed rape, counterfeiting, child molestation, embezzlement, and fraud more often to

a White criminal. Because 45% of participants believed that a Black criminal committed vehicular manslaughter and 55% attributed this crime to a White defendant, Sunnafrank and Fontes concluded that this type of crime is not stereotypical for a certain race (i.e., Black or White). Accordingly, they used this crime in their second study, which was intended to establish whether or not a person's race would affect culpability ratings in the absence of a crime stereotype.

The researchers hypothesized that jurors would be more likely to find a Black defendant who was charged with vehicular manslaughter guilty more often than a White defendant who was charged with the same crime. In contrast to their predictions, the results showed that there was no difference between White and Black defendants in terms of the frequencies of guilty and not guilty verdict decisions (Sunnafrank & Fontes, 1983). It should be noted that their sample was quite small (i.e., only seven participants were included in one condition), more significance testing could have been employed, and the effects of a stereotypical (and not only a non-stereotypical) crime on verdict decisions were not included in the second study. Despite these limitations, their findings were the first ones to suggest that whether or not a crime is stereotypical of the defendant's race might affect jurors' decisions more strongly than the defendant's race does by itself.

Gordon, Brindrim, McNicholas, and Walden (1988) examined the effect of defendant race as well as type of crime on jurors' sentencing decisions in their experiments. They manipulated whether the defendant was Black or White and whether he committed a blue-collar or white-collar crime. As expected, the researchers found that Black defendants who were convicted of burglary received longer jail sentences than

White defendants who were charged with the same crime. However, White embezzlers received longer jail sentences than their Black counterparts. In addition to sentence length, research has shown that when the defendant's race fits the crime stereotype, jurors also are more likely to find the defendant guilty, to have more confidence in their verdict decision, and to suggest harsher punishments (e.g., Gordon & Anderson, 1995; Jones & Kaplan, 2003; Schuller, Kazoleas, & Kawakami, 2008). For example, Jones and Kaplan (2003) found a race-crime congruency effect for White jurors. In their study, White defendants received more guilty verdicts, the jurors were more confident in their decisions, and punished the White defendant harsher than the Black defendant in an embezzlement case. However, when the crime included grand-auto theft, the opposite pattern of results arose; Black defendants were found guilty more often and received harsher punishments than White defendants. In addition, jurors' judgments about culpability did not differ between White and Black defendants, when the crime was not typical for any specific racial group (i.e., vehicular manslaughter).

Although Mazzella and Feingold's (1994) meta-analysis found no overall effect of race on culpability judgments, it is interesting to note that crime type moderated the effect of racial bias on sentencing decisions. On the one hand, Black and White jurors recommended longer sentences for Black compared to White defendants in the case of negligent homicide. On the other hand, jurors gave White defendants harsher punishments than they gave to Blacks, when the charge represented embezzlement or fraud. There were no differences in jurors' judgments for rape, robbery, and assault.

Accordingly, considering the type of crime that the defendant was charged with,

specifically whether or not it fits a juror's racial stereotype, seems to be important when investigating racial bias in juror decision making. Overall, these results show that biased decision making might even be more apparent when the juror associates the crime with members of the defendant's race.

Some researchers have investigated the conditions under which this effect is more apparent. For example, the strength of evidence and the allowed response time both represent moderators in the relationship between the interaction of defendant race and type of crime, and verdict judgments. Specifically, defendants who were charged with stereotypical crimes are punished more harshly when the evidence presented at trial is moderate in strength (Gordon, 1993). In addition, the race-crime congruence effect is stronger when mock jurors are instructed to make their judgments as quickly as possible compared to when they can take a much time as they please (Gordon & Anderson, 1995).

Past research also has attempted to investigate which factors might explain the race-crime congruence effect. Gordon & Anderson (1995) investigated potential mediators for the relationship between the perceived typicality of the defendant (i.e., to what extent does the defendant resemble a typical offender of that particular crime) and punishment recommendation. They found that perceived severity of crime as well as jurors' attributions about the offense (whether they were more situational or internal) were significant mediators. Although they did not directly assess the effect of the interaction between defendant race and whether or not the crime is stereotypical of the defendant's racial group on verdict judgment, their findings might explain this relationship as well. Jones and Kaplan (2003) tested this more directly in their study.

They found that race-crime congruence led to higher culpability ratings because mock jurors made more negative attributions, they engaged in a more limited search for additional pieces of evidence, and looked for more confirmatory evidence. These findings show that jurors might have a specific (negative) picture of the defendant in mind, which confirms the racial stereotype, just because the defendant was charged with a racially stereotypical crime. However, more research is needed examining what factors explain this race-crime congruence effect exists, and what conditions might attenuate this effect.

Conclusion

In sum, several meta-analyses have provided substantial evidence in support of the notion that a defendant's race might lead to biased verdict decisions (Sweeney & Haney, 1992), especially when it is different from the juror's race (Mitchell et al., 2005). Thus, White jurors are more likely to convict a Black defendant than a White defendant. In addition, whether the crime is typical for the defendant's racial group also seems to affect juror verdicts, generally exacerbating these effects. Jurors tend to give higher culpability ratings to defendants who were charged with crimes that are stereotypical of their race (e.g., Jones & Kaplan, 2003). One of the goals of this present research was to replicate this race-crime congruence effect and to examine its underlying mechanism from the perspective of the Story Model of juror decision making.

CHAPTER 2

THE STORY MODEL OF JUROR DECISION MAKING

Overview of the Story Model

The dominant model in the juror decision making literature is the Story Model,

which was developed by Pennington and Hastie in the 1980s. It attempts to explain in

what ways jurors deal with the evidence and come to their individual decisions

(Lieberman & Krauss, 2009). In general, this explanation-based model posits that in

order to make sense of the evidence presented at trial, jurors organize the information

chronologically by creating one or more narratives of the events leading to the trial.

These stories then help them to make a verdict decision by classifying the preferred story

into the best fitting verdict category (Devine, 2012). Devine (2012) defined a story as "a

causal chain of connected actions with an episodic structure" (p. 27). This means that

intentions produce actions with consequences, which then might lead to additional

actions. Furthermore, episodes consist of mental states or psychological reactions that

directly cause action (i.e., initiating events), objectives, behaviors, and effects. For

example, imagine the following scenario: Kevin is suing Anna to recover the cost of his

medical bills. The initial event is that Kevin stayed out late, and came home drunk again.

Anna's psychological reaction was to get angry. She wanted Kevin to apologize and to

promise her that he would quit drinking (goal). Their verbal argument escalated and the

couple began to fight physically (action). The consequence was that Anna pushed Kevin,

which led to him falling down the stairs and breaking his arm. This example portrays a story that explains the event (Kevin's broken arm) and the story makes sense because one action logically leads to the next.

It is expected that all jurors have an understanding about this general structure of stories (Hastie, 1993). Accordingly, multiple episodes are embedded in each story, which are hierarchical in nature. At the highest level, the general episodic schema characterizes the most significant events of the scenario that led to the trial, and has the most influence on the verdict decision (Pennington & Hastie, 1986).

Three major assumptions underlie the Story Model. First, trials represent complex environments in which various sources present jurors with an abundance of information. Moreover, attorneys present their evidence in a fragmented manner because different witnesses generally testify about the case facts in a sequence that is different from the order of the events in the original crime. Thus, one can assume that jurors have difficulty following the presentation of evidence, and making sense of the bulk of information. The second assumption is that the jurors process information in an active manner. Because they have a desire to understand the circumstances leading up to the charge fully, they engage in effortful cognitive processing by looking for cause-and-effect relationships. Finally, when creating their stories jurors rely on their mental scripts that include stereotypes and regular arrangements of events. Accordingly, they use their existing knowledge and beliefs to fill in missing information, to sort out contradictory evidence, and to determine the believability of a story (Devine, 2012). Thus, jurors do not only use case-specific information presented at trial to create their own descriptive

accounts of the events, but they also draw inferences by accessing their existing cognitive structures (Hastie, 1993).

Phases of Cognitive Processing

According to the Story model, jurors go through three phases of cognitive processing when evaluating the evidence and reaching a decision: story construction, verdict representation, and story classification.

Story Construction Phase

The central cognitive process of juror decision making is the *story construction phase*. When constructing their stories, which occurs during the proceedings of the trial and continues until after the presentation of evidence is completed, jurors actively reason about the evidence, evaluate each piece of information, and then create one or more stories that they think might describe what happened during the events that led to trial (Devine, 2012; Hastie, 1993, Pennington & Hastie, 1986). The prosecuting and defense attorneys typically do not present their evidence in temporal or causal order. Instead, they ask their witnesses very specific questions to which the individuals on the stand, one at a time, then respond. In addition, the witnesses need to rely on facts when answering; they are not allowed to make guesses about what they think might have happened. This format of evidence representation is very uncharacteristic of a storyline. Therefore, jurors create their own stories by elaborating on and interpreting all the different pieces of evidence during the course of the trial. Moreover, they organize the information in a way that it makes sense to them in the form of a narrative. Thus, comprehension is seen as a constructive process. During this story construction process, jurors rely on information

that is specific to the trial. For example, they take the opening statements and witness testimony into consideration. If one or more credible sources provide consistent pieces of evidence, these are accepted as potential story elements (i.e., facts) without much elaboration on the part of the juror. Furthermore, jurors rely on their personal knowledge of events that are similar to the ones presented at trial. For instance, they might be aware of a similar offense that recently happened in their neighborhood. In addition, jurors have are knowledgeable about story structures and of what makes a story complete (Devine, 2012).

Essentially, jurors not only make use of inductive reasoning procedures, but they also make inferences from case-specific evidence. In addition, they use their own personal experiences, hypothetical scenarios, and reasoning by contradiction to draw conclusions, especially to fill in gaps of information (Hastie, 1993). Oftentimes, jurors engage in counterfactual thinking—imagining alternative scenarios to "factual" reality (Hastie & Dawes, 200). Because of the jurors' dissimilar world knowledge, individual jurors might create different stories. Therefore, during the course of the trial some jurors reject exploring other alternatives and construct a single "best" story, while other jurors create multiple stories (i.e., at most two or three alternate stories), and then select the "best" story. It important to note that the story the juror chooses ultimately will decide his or her individual verdict (Hastie, 1993).

Verdict Representation Phase

During the *verdict representation phase* jurors learn the verdict categories (e.g., guilty of first-degree murder, guilty of second-degree murder, and guilty of

manslaughter) and the legal definitions of each decision alternative (e.g., first-degree is the premeditated and intentional killing of another person). They must conceptualize and comprehend the different verdict options. This might be a difficult task because the judge's procedural instructions that describe the verdict categories (and judicial instructions in general; e.g., Allison & Brimacombe, 2010) contain unfamiliar language and abstract concepts to describe the requirements. This, in addition to prior knowledge and beliefs regarding crime schemas, might influence the comprehension of the verdict options (Devine, 2012; Hastie, 1993; Lieberman & Krauss, 2009). A juror's mental representation of these verdict categories along with the legal requirements includes category labels and feature lists (Devine, 2012; Hastie, 1993).

Story Classification Phase

The final phase of an individual juror's cognitive processing of information is the *story classification phase*, which represents a matching process. Here, jurors determine which verdict option best matches the story they created and accept it at the end of the trial. During this process, they compare the features of the verdict categories with the attributes of the episode schema (e.g., actions, goals, psychological states) of their best-fitting story (Devine, 2012; Hastie, 1993). The goodness-of-fit governs which verdict option the juror chooses. If the juror decides that no match exists, his or her individual verdict decisions will be "not guilty" (Lieberman & Krauss, 2009).

Hastie (1993) pointed out that during comprehension and encoding of events presented at trial, jurors begin to process the information through attempting to connect the different pieces of evidence causally, and this initial comprehension process

influences subsequent inferences. Moreover, jurors create their story or stories before they reach a decision. Thus, the decision does not influence the story construction posteriori. However, because jurors might go through the processing phases more than once, the decision a jurors is considering at the time can affect the elaboration of certain pieces of evidence.

Certainty Principles

Pennington and Hastie developed four certainty principles that determine the individual jurors' choices in story selection, if competing story alternatives exist, and the verdict decision as well as their confidence with which they make their decisions. How certain a juror feels about the decision depends on the assessment of these principles (Hastie, 1993). *Coverage* refers to the extent to which the story can integrate the various pieces of evidence presented at trial. The more items can be incorporated, the more coverage increases. When coverage is high, a juror will be more likely to accept this story as an explanation of the evidence, and the juror will be more confident in choosing this story (Hastie, 1993; Hastie & Dawes, 2002; Lieberman & Krauss, 2009).

Another certainty principle that controls story acceptability is story *coherence*, which reflects the extent to which the story is consistent with itself and with the juror's world knowledge. In other words, story coherence is high when the explanation is resistant to other logical explanations and one can infer many parts. This certainty principle has three components, which all need to be present for a story to be accepted: consistency, plausibility, and completeness. Consistency is high when parts of the story do not contradict with one another or with other evidence that the juror believes is

accurate. Furthermore, plausibility reflects the extent to which the story does not refute one's personal worldviews about how the social world works. In addition, completeness refers to whether or not a story has all of its parts, which means all pieces of evidence need to be integrated completely, and all the events are causally linked. In sum, story coherence is high when the story lacks internal inconsistencies, it is in accord with one's world knowledge, and the story does not leave any informational gaps (Hastie, 1993; Hastie & Dawes, 2002; Lieberman & Krauss, 2009).

The certainty principle of *uniqueness* influences one's confidence in a story and the verdict decision. It is high when a story provides a unique explanation of evidence, which means that it is the only story that is coherent (Lieberman & Krauss, 2009). Then, a juror will accept this story as the true explanation of evidence. If multiple coherent and complete stories exist, the stories will lack uniqueness, which will make it much more difficult to reach a decision. In addition, confidence will be low (Devine, 2012; Hastie, 1993).

The final certainty principle is *goodness-of-fit*, which refers to the strength of the relationship between the story and the verdict category that matches this explanation the best. If the story does not represent a sufficient number of elements of the verdict category (e.g., the story does not contain a component that proves premeditation), then this verdict category (e.g., first-degree murder) would be rejected, and another decision would have to be made (e.g., manslaughter, "not guilty;" Hastie, 1993).

Empirical Evidence of the Story Model

Over the last three decades, much research has validated the Story Model,

making it the most prominent model of juror decision to date. Pennington and Hastie's

(1986) initial research on the Story Model intended to establish that jurors organize trial

information in the form of a narrative, in which causal event chains are central. They

proposed that this would then facilitate jurors' comprehension of evidence and assist

them in making pre-deliberation verdict judgments. In addition, the researchers

hypothesized that for each verdict alternative jurors would create a corresponding list of

features including the identity, psychological state, circumstance, and actions. Pennington

and Hastie's final prediction was that each verdict would be associated with distinct story

structures. To test their hypotheses they had potential jurors watch a three-hour

reenactment of a capital trial. Afterwards, the researchers conducted semi-structured

interviews in which the jurors talked about the case and how they arrived at their verdict

decisions. They were asked to choose between four decision alternatives ranging from

not guilty to guilty of first-degree murder. Extensive content coding was employed in

order to analyze the interview protocols—the researchers identified verdict decisions and

coded the verdict representations as well as the protocol content. In addition, they

developed a verdict story for each verdict group by using a subset of participants whose

verbal protocols were either the least or the most persuasive.

      The results supported Pennington and Hastie's hypotheses (1986). They found

that juror's mental representations of the evidence exhibited narrative structures.

Specifically, 69% of the protocols included story statements consisting of episode

structures and cause-and-effect relationships, and 85% of those were causally linked.

Jurors generally also inferred missing pieces of information and omitted evidence that

was unrelated to their stories. In addition, jurors who made different verdict decisions also constructed different stories, which varied systematically in content and structure. Although only 10% of the interview protocols included specific verdict statements, as expected, jurors developed feature lists that corresponded to each verdict category, in which the general information was included. The researchers concluded that jurors explain the evidence presented at trial in the form of stories, verdict alternatives are represented as categories, and pre-deliberation verdict decisions vary by type of story. Although these results gave initial support for the Story Model, Pennington and Hastie pointed out that the methodology of conducting interviews might have created a demand for creating a narrative. Moreover, because the study was correlational in nature, it was impossible to conclude that there was a direct causal relationship between verdict decisions and types of stories (Pennington & Hastie, 1986).

In order to address these limitations, Pennington and Hastie (1988) conducted additional experiments. In their first experiment, they investigated whether jurors actually construct stories spontaneously, without prompting them to discuss the reasons for why they chose one specific verdict. The researchers employed the methodology of a recognition memory task, in which they presented potential jurors with a criminal trial that included 119 evidence items. Pilot research (Pennington & Hastie, 1986) had determined which events 80% of jurors who chose a particular verdict believed to be true. Thus, the trial summary consisted of evidence items that could be classified into a guilty verdict story, a not guilty verdict story, both, or neither. During the experiment, participants read a trial, for which each evidence item was presented one at a time. Then,

they chose a guilty or not guilty verdict. Afterwards, they again read a list of evidence items of which about half were not included in the trial. They indicated whether they had seen the sentence in the trial summary before, and how important they thought each specific item was for their verdict decision. Results showed that jurors were more likely to recognize sentences that supported their verdict decision. In addition, they were more likely to falsely recognize evidence items that were not included in the trial summary, but supported their verdict decision. Moreover, participants gave the highest ratings to those items that were part of the story that supported their preferred verdict, while giving the lowest ratings to the items in the other story. This showed that how important a juror thinks an item is depends on whether or not the item is part of the story that the juror constructs. Thus, Pennington and Hastie (1988) concluded that jurors engage in a spontaneous story construction process, that story structures differ between participants who choose divergent verdicts, and that evidence items are believed to be more important when they are part of the participant's story, especially when it causally connects significant events.

Because the participants were assigned to different conditions based on their verdict decisions, the experiment described above could not establish that the type of story an individual construct actually determines their verdict decision. To investigate this claim as well as the research question of to what extent story coherence and uniqueness influence jurors' levels of confidence in their verdict decisions, Pennington and Hastie (1988) conducted as second experiment in which they manipulated the ease of story construction. They presented the information either in "story order" or in "witness-

by-witness" order. On the one hand, when the items were organized in "story" order," the trial presentation of evidence items followed in a temporal and causal sequence, which was expected to make story construction easy. On the other hand, evidence items were presented as they are in a real-world trial, in a "witness-by-witness" order. The researchers expected that order of presentation would make story construction more difficult. The experiment included four conditions, in which the order of presentation did or did not vary for the prosecution and the defense. The researchers hypothesized that jurors more frequently would choose a guilty verdict when the prosecution story was easy to construct, and a not guilty verdict would prevail when the defense organized their evidence items in story order. They also predicted that jurors would rate the evidence items as stronger and would be more confident about their verdict decision when the evidence was presented in story order. Finally, it was hypothesized that when the coherence of the opposing story is low (i.e., the presentation order varies between the prosecution and defense), jurors would be more confident in their judgments.

Results indicated that when only the prosecution evidence was presented in story order, 78% of jurors chose a guilty verdict. However, when only the defense evidence was presented in story order, 31% chose guilty. Moreover, evidence strengths ratings were higher in the story order conditions and lower in the witness order conditions. While these findings supported the researchers' hypothesis, ratings of verdict confidence did not follow the expected pattern. It was found that confidence was highest when both the prosecution and defense evidence was in story order. In addition, confidence ratings only significantly differed between this condition and when neither set

of evidence was in story order. These findings revealed that story coherence, as determined by the order of evidence presentation, influences verdict judgments, and evidence strength as well as confidence ratings. Specifically, story coherence mediates the relationships of perceived evidence strength and confidence ratings. Moreover, the results indicated that story construction occurs spontaneously. When stories are easily constructed, they dominate the decisions in the expected direction (Pennington & Hastie, 1988). Pennington and Hastie (1986, as cited in Hastie, 1993) found similar results in another study. After listening to evidence, which was presented in legal issue or story order, participants were asked to recall the information freely. High story clustering of items occurred when they heard the evidence organized in story order, while participants more frequently clustered the evidence items by legal issue when the evidence was presented in legal issue order. Furthermore, participants were more likely to choose the verdict that corresponded with a story order presentation of evidence.

Pennington and Hastie (1992) replicated their own results and ruled out the possibility that those stories which are easier to construct would result in more verdict decisions consistent with these stories because it would be easier for jurors to memorize the evidence items presented through a storyline. The researchers found that story organization had the expected effect on verdict decisions, even under conditions of minimal witness credibility. Differences in memorability of the evidence did not have an effect on the results. Accordingly, evidence that is summarized in a story-like fashion mediates the relationship between evidence and verdict decisions (Pennington & Hastie, 1992).

The studies cited above found support for the Story Model of juror decision making. In sum, jurors attempt to make sense of complex information by instinctively imposing a story structure on the evidence presented at trial. This type of organization facilitates comprehension of the material, and leads to the construction of one or more stories. The type of story that is created then leads to the juror choosing the verdict category that best fits in with the story. This classification process occurs after jurors attempt to understand the verdict categories along with their legal requirements that the judge provides to them via instructions. Overall, the ease of story constructions mediates perceived evidence strength, confidence ratings, and the impact of witness credibility. In sum, the stories jurors construct ultimately cause their decisions (Pennington & Hastie, 1992).

Mathematical Models of Juror Decision Making

Before the development of the Story Model, mathematical models of (juror) decision making were prominent. When employing *Bayesian models*, jurors assess the probability of each new piece of evidence independently. The value then is multiplied by the juror's prior opinion about his or her belief about the probability of the defendant's guilt. Therefore, jurors constantly update their beliefs about the probability of guilt throughout the trial. At the end of the trial, they compare their final probability of guilt value to the probability of culpability that is required for a guilty verdict, and they base their verdict decision on this comparison (Lieberman & Krauss, 2009).

Concerning *algebraic models* (e.g., basic weighted average/information integration, sequential averaging model), the meaning of evidence presented at trial also

is represented in quantitative terms. Here, a juror assigns weights to each piece of information, and these weights depend on how important and relevant this information is to the final verdict decision. By summing all weighted numbers, a global average evidence value is obtained. This value represents how strongly the juror believes that the defendant is guilty (i.e., subjective probability of guilt; Hastie, 1993; Lieberman & Krauss, 2009). Algebraic models differ from Bayesian models in the way that the former assume that pieces of evidence are added, not multiplied. Moreover, jurors determine their verdicts through probabilities that are not bounded by "0" and "1" (Lieberman & Krauss, 2009). However, the two models are similar in that each piece of evidence is evaluated one by one, and that the order in which they are present during trial is irrelevant.

The temporal ordering of evidence during presentation plays a role in *stochastic processing models*. As jurors sequentially process the information, they weigh the pieces of evidence and sum up these values to obtain probabilities regarding the defendant's culpability. In contrast to Bayesian and algebraic models, they do not update these probabilities until the end of the trial. Rather, the decision making process is halted at some point during the trial because of cognitive load, fatigue, or a particularly diagnostic item of evidence. At this point, the juror already has reached a judgment, and new pieces of evidence will not influence this verdict decision. The likelihood of this critical event occurring varies from individual to individual, and thus is modeled as a random variable (Devine, 2012).

As outlined above, it becomes clear that mathematical models of juror decision

making are very logical and rational in a sense that they estimate the likelihood of each

verdict decision through mathematical formulas. The underlying assumption of these

models is that each juror has an initial belief about the defendant's culpability.

Throughout the trial, this probability value is updated through the assessment of each new

piece of evidence. Finally, it is compared to a certainty threshold (Devine, 2012).

Comparisons between the Story Model and Its Major Predecessors

Pennington and Hastie's Story Model is very different from these major

predecessors because it utilizes an explanation-based approach toward juror decision

making. While Bayesian probabilities models, algebraic models, and stochastic process

models use mathematical algorithms to *prescribe* how jurors ought to make judgments,

the Story Model *describes* how jurors typically process information in complex

environments (i.e., during a trial) and arrive at their individual verdict decisions (Devine,

2012). The latter is associated with research on mental representation and interpretation

during text comprehension, and thus provides a cognitive account of decision making

(Pennington & Hastie, 1990).

The Story Model relies on the assumption that jurors do not evaluate each piece

of evidence in isolation, and the story or stories they create capture(s) these

interdependencies. In contrast, mathematical models presume that separate evaluations of

each item are made. Moreover, in the Story Model the narrative explanation (i.e. story)

mediates the relationship between the evidence and the verdict decision, which means

that the story actually causes the judgments, while the effect of the evidence on the

judgment is much more direct in mathematical models (Pennington & Hastie, 1990).

Pennington and Hastie (1992) experimentally assessed whether jurors engage in an on-line updating approach or construct stories to reach their judgments when processing the information. The researchers varied the points in time at which the participants were asked to assess the evidence: item-by-item assessments (i.e., on-line processing strategy), local assessments after the presentation of blocks of evidence (i.e., after each witness testimony), and one global assessment at the end of the trial. Results showed that the Bayesian model did not describe any of the judgment processes, while the anchor-and-adjust model explained the behavior of participants who made a judgment after each item. Evidently, the Story Model was the only model that described the global assessments; in addition, it prescribed the jurors' assessments after each block of evidence (Pennington & Hastie, 1992). Because jurors are always asked to provide a global assessment of their beliefs about a defendant's culpability at the end of the trial (and never during the proceedings), it is not surprising that the Story Model has gained prominence over mathematical models of juror decision making.

To sum up, in situations where many complex pieces of evidence are present and these components are symbiotic, the decision maker engages in explanation-based cognitive processing. In short, individuals create a causal model that explains the relationships between the available facts. They then base their decisions on these causal explanations of the evidence (Lieberman & Krauss, 2009; Pennington & Hastie, 1992). Because reaching a verdict decision represents a complex cognitive task, an explanation-based model undoubtedly does a better job in describing a juror's cognitive processing strategies compared to mathematical models. In contrast to its predecessors, the Story

Model takes into account that people's working memories are limited in capacity, which might restrict their cognitive resources concerning attention and computation (Hastie, 1993). Therefore, the interpretation of evidence is preferred over computation (Pennington & Hastie, 1986).

Race-Crime Congruence and Stories

As described above, stories play an important role in juror decision making. The Story Model posits that during and after the presentation of evidence at trial, jurors engage in story construction. During this phase, they individually create one or more stories that explain the evidence. Story selection, then, determines a juror's verdict decision. In two of the only studies investigating the causal mechanism in the form of stories that have been conducted until this point in time, Huntley and Costanzo (2003) showed that stories mediate the relationship between juror gender and verdict decisions in a sexual harassment case. Thus, it becomes apparent that the story a juror creates and selects can explain why a juror makes certain verdict judgments. However, research investigating this theoretical preposition directly is surprisingly rare.

According to Pennington and Hastie (1988), the coherence and coverage of the story determine which story a juror chooses. In addition, the uniqueness of the selected story influences how confidently the juror is in the verdict decision. In order to select a verdict with much confidence, all of these certainty principles need to be high. In addition, their fourth certainty principle, goodness-of-fit, which relates to the extent of which the selected story matches a verdict category, needs to be high as well. However, goodness-of-fit was neglected in the present research because I only distinguished

between a "not guilty" and "guilty" verdict decision, therefore not including multiple verdict categories.

I suggested that to what extent jurors think that the prosecution, defense, and their own stories describing the events leading to the trial are believable can explain the effect of race-crime congruence. The coherence of the story should be the certainty principle that is influenced the most because when the crime "matches" the defendant's race in the eyes of the jurors, their worldview about how society works is "in order" because it confirms their expectations. Thus, their stories should be high in consistency, plausibility, and completeness; in other words, the stories would make sense to the jurors. On the other hand, when the crime does not match the defendant's race, a juror's story coherence should be impaired because the story then might refute a juror's personal worldview. Therefore, this juror might be more hesitant to choose a guilty verdict. In addition, his or her confidence should be lower.

Conclusion

Although the Story Model of juror decision making has received much empirical support, it is surprising that it has not been applied to racially stereotypical crimes biasing verdict decisions. I tested whether the representation of trial evidence (i.e., prosecution, defense, and the juror's stories) can explain this effect. Specifically, I attempted to answer the question of whether race-crime congruence leads to biased verdict judgment because it leads to an increase in the coherence of the prosecution and juror stories.

CHAPTER 3

POTENTIAL MODERATORS

The existence of the race-crime congruence effect on verdict judgments undoubtedly is a concern because it epitomizes the justice system as unfair. Truly innocent people might be found guilty just because their race is stereotypical of an offender who typically commits this type of crime. At the same time, truly guilty defendants might be acquitted because their race does not fit the crime stereotype. Therefore, it is important to investigate potential factors that might curtail these effects.

Motivation to Control Prejudiced Reactions

One of these potential moderators might be a jurors' motivation to control prejudiced reactions (MCPR), an individual difference variable. Past research has showed that MCPR might influence how people answer explicit questions about racially related hypothetical behaviors. For instance, Towles-Schwen and Fazio (2003) demonstrated that people with higher levels of MCPR are more likely to agree to spend time with a Black person (versus a person of unspecified race) in a hypothetical situation, compared to people with lower levels of MCPR. This could mean that people who are high on MCPR are aware of the racial stereotype and might be prejudicial as well. However, they try to avoid engaging in actions (or reactions) that others might interpret as prejudicial. On the other hand, individuals low on MCPR might not have a reason for refraining from engaging prejudicial response, and thus express their negative attitudes toward Blacks

27

openly. High levels of MCPR not only affect individuals' explicit racial prejudice, but also influence more implicit, automatic, responses (Plan, Amodio, Harmon-Jones, & Vance, 2002). In addition, Maddux, Barden, Brewer, and Petty (2005) showed that whether people favor their in-group or out-group does not only depend on their level of MCPR, but on the context as well. They primed participants with a context in which either racially prejudicial cues were absent (e.g., church) or present (e.g., foggy street), measured their MCPR, and then implicitly measured their in-group vs. out-group biases. The results demonstrated that participants with low MCPR and participants with high MCPR favored the in-group independent of whether the context was threatening or not. However, when the target in the context was seen as threatening (e.g., jail), participants with high MCPR automatically inhibited their negative responses towards Blacks to the extent that they actually favored Blacks over Whites.

In light of these findings, one might argue that a courtroom represents a contextual threat; jurors might see it as a dangerous setting that is consistent with race and crime stereotypes. Therefore, jurors with high levels of MCPR might try unconsciously (or consciously) to prevent stereotypes from affecting their decisions. Accordingly, MCPR might moderate the effects of race-crime congruence on their verdict judgments. In addition, story coherence might only mediate these race-crime congruence effects when MCPR is low.

Stereotype Suppression Instructions

Another potential moderator might be the presence (vs. absence) of stereotype suppression instructions during the trial. Past research has shown that sometimes

stereotype suppression instructions lead to a decrease in stereotyping, while under other conditions they foster the application of stereotypes. In one of the empirical investigation of stereotype suppression, Macrae, Bodenhausen, Milne, and Jetten (1994) showed participants a photograph of skinhead and asked them to write about a typical day in the life of the individual. They also instructed half of the participants to try to avoid thinking about the person in stereotypical ways. Results showed that stereotype suppression instructions were effective; participants who did not receive stereotype suppression instructions were more likely to use stereotypes in their passage than participants who received stereotype suppression instructions. However, the effect disappeared, and actually led to more stereotyping, on a subsequent task for which the participants did not hear stereotype suppression instructions. For example, when participants were told that they would interact with the skinhead about whom they had just written, participants who had heard the stereotype suppression instructions chose to take a seat further away from where the skinhead would sit, compared to participants who did not hear any stereotype suppression instructions. Thus, the researcher concluded that, while stereotype suppression instructions might have positive effects immediately after they are stated, they might exacerbate the use of stereotypes in later judgments and behaviors.

Although Macrae, Bodenhausen, Milne, and Wheeler (1996) found similar results, they argued that this "rebound effect" might disappear when more highly sensitive target groups (Blacks, not skinheads or construction workers) are evaluated. Accordingly, Monteith, Spicer, and Tooman (1998) used a hypothetical gay couple in their stimulus materials in order to test these assumptions. As expected, participants who

heard stereotype suppression instructions used fewer stereotypes in the first, as well as in subsequent tasks, than participants who did not receive this type of instruction. This was found for both low-prejudicial and high-prejudicial participants.

Because racial discrimination is a sensitive topic and people are generally more subtle in expressing their stereotypes about Blacks (if they have any), it is likely that instructing jurors to ignore their stereotype-based thoughts and beliefs as a source of information when forming impressions of the defendant might have its intended effects. Thus, stereotype suppression instructions might attenuate the effect of race-crime congruence on verdict judgments. When stereotype suppression instructions are present, the interaction between race and type of crime should not affect story congruence.

Conclusion

Undoubtedly, it is critical to identify factors that have the potential to decrease the application of stereotypical thoughts about race in the courtroom, so that jurors make verdict that are free of any biases. In the current research, I investigated whether MCPR, an individual difference variable, as well as stereotype suppression instructions, as a trial-related factor, affect the strength of the relation between race-crime congruence (vs. incongruence), story coherence, and jurors' verdict decisions.

CHAPTER 4

PRESENT RESEARCH

Overview

In Study 1 I examined the factor structure that underlies edited versions of the

Narrative Believability Scale (NBS-12; Yale, 2013), so that I could use the final version

to measure the extent to which a story (e.g., prosecution statement) includes the certainty

principles that, according to Pennington and Hastie (Hastie, 1993), determine individual

jurors' choices in story select (e.g., coherence). During the initial phase of developing the

NBS-12, Yale (2013) found that a four-factor model including coverage, plausibility,

consistency, and completeness provides a better fit than a two-factor model with

coverage and coherence as the factors, and it fits better than the unidimensional model.

Yale (2013) originally developed the NBS-12 to measure narrative believability

of a presentation in a variety of different domains, including, but not limited to, the legal

domain. Because alternative narratives need to be present in order to be able to

investigate uniqueness, which is an additional certainty principle, of a story or

communication, this dimension was not included in the original scale. Because the

proposed research exclusively focuses on legal decision making, story uniqueness

appears to be an important principle. Thus, items attempting to measure uniqueness were

included in the edited versions of the NBS-12.

Although goodness-of-fit is the fourth certainty principle in the Story Model

31

(Lieberman & Krauss, 2009), items measuring this construct were not included in the new measure because goodness-of-fit is concerned with the choices of verdict options and whether or not one matches one's story, rather with whether or not a story is accepted as the explanation of the evidence.

After editing the items in a way that they related to the prosecution story (as opposed to a general presentation) and adding items that might measure an additional factor (i.e. uniqueness) to the original scale, my goal was to investigate whether one, three correlated (i.e., coherence, coverage, and uniqueness), or five correlated factors (i.e., plausibility, consistency, completeness, coverage, and uniqueness) make up the modified version of the NBS-12, which I called MNBS-P, and that includes 18 items.

In addition, I conducted three studies with the objective to test whether jurors' representation of the trial in the form of stories, specifically story coherence, can explain the race-crime congruence effects on culpability ratings. I also investigated under which conditions story coherence is more or less effective in making participants' judgments unbiased. As an individual difference variable, MCPR has been linked to a decrease in the application of stereotypes (e.g., Towles-Schwen & Fazio, 2003). In addition, past research has shown that stereotype suppression instructions represent an effective means to reduce one's use of stereotypes in general (e.g., Macrae et al., 1994), suggesting that stereotype suppression instructions might be a successful tool for judges and attorneys during a trial.

Because an individual's tendency to engage in and enjoy effortful thinking has been linked to a more successful completion of various cognitive tasks (e.g., information

seeking, recall, decision making; Cacioppo, Petty, Feinstain, Jarvis, & Blair, 1996), need for cognition might affect how thoroughly the trial evidence was processed. Therefore, need for cognition was included as a potential control variable in the analyses.

Although the Story Model is the most prominent model in the juror decision making literature, to date, no research has directly examined the effects of racial and crime stereotypes on culpability judgments through story coherence. Although many different racial and crime stereotypes exist and are worthy of study, I focused on investigating jurors' verdict decisions and culpability ratings about Black and White defendants who either were charged with a stereotypical or non-stereotypical crime concerning their race.

Research Questions

The present research investigated the following research questions in four studies.

1) What is the factor structure of the modified version of the Narrative Believability Scale (MNBS-P)? Specifically, do both the three-factor (coherence, coverage, and uniqueness) and the five-factor (plausibility, consistency, completeness, coverage, and uniqueness) model provide an adequate fit for the data, or does one provide a better fit than the other does?

2) Does a crime that is stereotypical of a defendant's race (as opposed to a crime that is non-stereotypical of a defendant's race) increase jurors' beliefs about how coherent the presentation of trial evidence is (i.e., story coherence), and produce biased culpability judgments?

3) Does race-crime congruence have less of an effect on story coherence and

decreases biased verdict judgments, when motivation to control prejudiced reactions is high? Conversely, does race-crime congruence have more of an effect on story coherence and biased culpability judgments, when jurors' level of motivation to control prejudiced reactions is low?

4) Does race-crime congruence have less of an effect on story coherence and decreases biased verdict judgments, when jurors receive stereotype suppression instructions? In contrast, does race-crime congruence have a stronger effect on story coherence and biased culpability judgments, when jurors' receive stereotype suppression instructions?

CHAPTER 5

STUDY 1

Overview

The overall purpose of Study 1 was to develop a scale that fully captures the

certainty principles affecting story acceptance and confidence in a story in the specific

context of juror decision making. Accordingly, the first goal of Study 1 was to test the

underlying factor structure of the Modified Narrative Believability Scale-P (MNBS-P)

that included 18 items by using confirmatory factor analysis with one random subsample

of the data. The second goal was to reduce the set of items for the uniqueness subscale,

while maximizing factor loadings as maintaining adequate statistical fit. The third goal

was to confirm the latent factor structure that had emerged with one random subsample of

the data with the other set of study participants using confirmatory factor analysis.

The final goal was to control for acquiescence, which is also called agreement

bias, agreeing-response bias, or response style effect. There is considerable evidence

suggesting that respondents who fill out a measurement instrument that includes identical

response scales for its items tend to select those responses that indicate a preference for

either agreement or disagreement. Specifically, respondents who score low on

acquiescence tend to disagree with all survey items, whereas respondents who score high

on acquiescence usually agree with all survey items. Thus, although not all researchers

agree with this notion, many interpret the occurrence this acquiescence bias as a

personality variable or cognitive style that statistically needs to be controlled for when using measurement instruments that include questions and statements that are written in the typical "Likert" format" (Billiet & McClendon, 2000). In scales that include a different number of items that are worded in the positive direction and items that are worded in the negative direction, this response bias can inflate the variance and reliability estimates. In addition, it leads to either an over- or underestimation of the relationship between the construct that is measured with this unbalanced scale and other constructs (Mirowsky & Ross, 1991).

To conduct Study 1, the participants were randomly assigned to read one of six opening statements of the prosecution. While one version of the opening statements was high on all certainty principles, each one of the other five versions was low on one certainty principle: plausibility, consistency, completeness, coverage, or uniqueness. Thus, in addition to investigating the factor structure underlying the MNBS-P, I also assessed the discriminant validity of the factors underlying the given measurement models. After reading the statement, each participant rated the narrative believability of the prosecution story by using the MNBS-P.

<div align="center">

Hypotheses

</div>

Hypothesis 1a

I predicted that both a three-factor model with coherence, coverage, and uniqueness as the three factors and a five-factor model with plausibility, completeness, consistency, coverage, and uniqueness would fit the data better than a global one-factor model because, according to the Story Model, these constructs represent the certainty

principles that affect story acceptability and confidence. This three-factor model would fit the data better than a one-factor model because Yale (2013) found that the latter did not provide a good fit. In line with Yale's (2013) findings, the five-factor model including plausibility, consistency, completeness, coverage, and uniqueness as the factors also would provide an adequate fit. However, the three-factor model might be preferred because of its greater parsimony.

In studies 2, 3, and 4 I used different versions of the final MNBS-P in order to measure the narrative believability of the defense story (MNBS-D) and each mock juror's story (MNBS-J and MNBS-D). These three versions only differed from the MNBS-P in terms of some wording of the questions. For example, every time "prosecution" was mentioned in the MNBS-P, I replaced it with "defense" for the MNBS-D. Because these were only minor changes, I expected the factor structure that underlies the MNBS-P also to underlie the MNBS-D, MNBS-J, and MNBS-R. Therefore, I only investigated the factor structure of the MNBS-P in Study 1.

Hypothesis 1b

If a three-factor emerged as the best fit for the data, I predicted that coherence and coverage will be more strongly correlated than each of these factors with uniqueness because Pennington and Hastie (1984, 1986, 1988) argued that story coherence and coverage both predict story acceptance, while uniqueness predicts one's confidence in the selected story (Devine, 2012).

I also predicted that if a five-factor model measurement model emerged, plausibility, completeness, and consistency should be strongly correlated because

coherence consists of these three factors. For the same reason as mentioned above, plausibility, completeness, and consistency each should be more strongly correlated with coverage than with uniqueness. Accordingly, I expected the correlation between uniqueness and each of the other factors to be the weakest correlation.

Method

Participants

For Study 1, participants were recruited via Amazon's Mechanical Turk (MTurk), which is a web-based crowdsourcing platform, where researchers can get access to anonymous "workers" (i.e., participants) who complete research studies for relatively small sums of money (Crump, McDonnell, & Gureckis, 2013). MTurk workers are generally more demographically diverse than members of psychology (or university) subject pools (Buhrmester, Kwang, & Gosling, 2011). In addition, the average age is greater (i.e., 30s), and women and men are fairly equally represented (Mason & Suri, 2011; Paolacci, Chanderl, & Iperotis, 2010). Research has shown that reasonable compensation rates do not affect the quality of the data (Buhrmester et al., 2010). In addition, data obtained from MTurk samples usually have the same pattern as data collected via other online methods or in the laboratory (Buhrmester et al., 2010; Crump et al., 2013; Paolacci et al., 2010).

Bentler (1985) suggested to have at the very least five observations for each parameter one is estimating in a statistical model. My largest model included five correlated factors with 18 items. Following Bentler's informal guidelines, I needed 230 participants at the minimum. Because I planned to randomly split my sample in half in

order to explore the factor structure of the Modified NBS-P with one half, and then to

replicate my findings with only the items that will have loaded the best with the other half

of participants, I had to double the amount of data I would collect. I also planned to

collect an additional 24 (15%) participants to account for unusable data due to hypothesis

guessing or missing data. Accordingly, I proposed to include 530 participants in Study 1.

In total, data were collected from 614 MTurk workers who received $0.50 in

exchange for their participation in the current research study. I omitted those participants

from the analyses who either failed ($n = 23$) or skipped ($n = 1$) the attention check item.

In addition, 11 pairs of cases as well as one triad of cases were identified for which the

I.P. addresses were the same. Because doubt arose about the quality of these data, all

cases of the replicas were dropped if the start and end times of the pairs were the same (8

pairs; $n = 16$), or if those were only a few minutes apart (one pair and one triad; $n = 5$).

For two pairs, the start and end times of survey completion were much different from one

another. I deleted the case with the later start date ($n = 2$) to ensure that every participant

who would be included in further statistical analyses partook only once in the study and

had not participated in the study before that specific point in time. An additional two

participants were removed from the data set because of multiple missing data points.

Accordingly, the present sample included 565 participants, of which 50.4% were men.

Their ages ranged from 18 to 73 years with a mean age of 33.28 ($SD = 11.69$); two

individuals did not report their age. Furthermore, 80.00% classified themselves as

White/European American, 6.70% as African American/Black, 5.50% as Asian, 3.90% as

belonging to multiple ethnic groups, 2.50% as Latino/a, .40% as Native American, and

0.20% as Middle Eastern; 0.90% of the participants did not disclose their race/ethnicity.

The majority of participants (34.50%) had received some college education at the time of

the study. For a graphical depiction of the participants' level of education, see Figure 1.

Figure 1. Study 1 – Level of Education
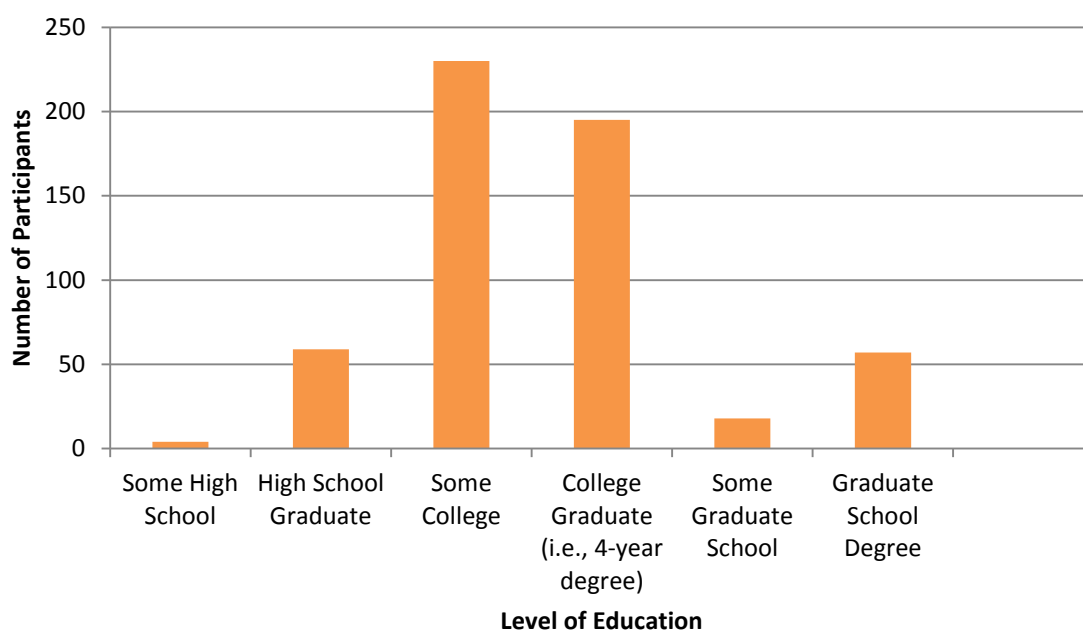
Materials

    Opening statements. The case State v. Lawrence involves a purse-snatching

incident**,** and the prosecution argues that the jurors should find the defendant guilty of 2nd

Degree Theft and Robbery. The opening statements by the prosecution were modified

versions of the narratives Yale (2013) developed; specifically, I changed the dates in

order to make the crime more recent. In total, I included six versions of the prosecution

narrative. The master prosecution case was high in five certainty principles, which relate

to narrative believability: coverage, consistency, plausibility, completeness, and uniqueness. The other five narratives are each low on one of the factors. Narrative coverage was reduced by adding information suggesting that somebody else, and not the defendant, committed the crime. Another manipulation reduced consistency by including contradictory information regarding the victim's identification of the perpetrator and the perpetrator's motive. The defendant not having a reasonable incentive for having committed the crime reduced plausibility. Completeness was manipulated by presenting the prosecution statement in the form of bullet points and direct quotations from witness and defendant interviews with the police. Because Yale did not include uniqueness as a factor that might affect narrative believability, I developed an additional version of the opening statement(s), which was intended to reduce uniqueness. The final manipulation included a defense narrative in addition to the original master prosecution narrative (see Appendices A-F).

Modified Narrative Believability Scale – Prosecution (MNBS-P). In order to measure narrative believability of the prosecution statement, Yale's (2013) Narrative Believability Scale (NBS-12) was edited. The original NBS-12 includes twelve items. Ten items are measured on a seven-point scale ranging from *strongly disagree* to *strongly agree*. For example, participants indicate their responses to the statement, "All of the facts in this presentation agreed with each other." In addition, two items include a seven-point Likert scale ranging from 1 *very low* to 7 *very high*. For example, participants read, "The 'consistency' of a presentation refers to the extent to which a presentation does not contradict itself or contradict other things you know to be true or false. How would you

rate this presentation in terms of 'consistency'?" Participants respond to items measuring

consistency, plausibility, completeness, and coverage of a presentation. Higher scores

correspond to higher levels of the construct. The NBS-12 has high internal reliability, as

well as good criterion-related, predictive, and construct validity (Yale, 2013).

Six additional items attempting to measure uniqueness were included at the end of

the original NBS-12. Participants responded to the newly created items on a seven-point

scale ranging from 1 *strongly disagree* to 7 *strongly agree*. For example, participants

rated the following uniqueness items; "The prosecution story provides a unique

explanation of the evidence," and "Multiple unique stories can explain the evidence." In

addition, minor adjustments to the wording of the items were made (i.e., "prosecution

story" was replaced with "presentation". Overall, the MNBS-P includes 18 items

attempting to measure narrative believability (see Appendix G).

Demographic questionnaire. Participants were asked to indicate their gender,

age, ethnicity, and level of education (see Appendix H).

Procedure

On MTurk, the study was advertised as "Evaluating Opening Statements," and

participants were ensured that it would not take longer than 30 minutes to complete, with

a pay of US $0.50. Using MTurk's enrollment rules, only participants with United States-

based IP addresses were included in the study. In addition, every participant's HIT

(human intelligence task) approval rate, which reflects a person's ability to complete a

HIT successfully, needed to be at least 90% in order to be able to take part in the study.

After selecting the HIT, participants were redirected to the web-based survey program

Survey Monkey (SurveyMonkey Inc.), through which they then completed the task. On

the first page, participants saw the informed consent form. They were told that by

continuing with the survey they are indicating their consent to participate in the study,

and they were ensured that all their responses would be anonymous. Next, each

participant was randomly assigned to read one of six opening statements: master

prosecution statement, prosecution statement/low coverage, prosecution statement/low

consistency, prosecution statement/low plausibility, prosecution statement/low

completeness, or prosecution and defense statements/low uniqueness. Afterwards, they

read the following instructions:

> "Now think about the case you just read from the perspective of the prosecution.
>
> Think about what the basic story of the prosecution was and rate the prosecution
>
> story by answering the following questions. Please circle the response that best
>
> reflects your opinion about the prosecution's description of what happened (i.e.,
>
> prosecution story). There are no correct answers."

Participants then responded to the MNBS-P. Finally, they filled out demographic

information.

<div align="center">

Results

</div>

MNBS-P

I conducted confirmatory factor analyses (CFA) employing LISREL 8.80 for

Windows (Jöreskog & Sörbom 2006) to examine the factor structure of the MNBS-P.

Yale (2013) showed that the plausibility, consistency, and completeness items of the

original Narrative Believability Scale form three separate factors instead of one (i.e.,

coherence). In contrast, after testing several measurements models, my results revealed that the plausibility, consistency, and completeness items load on one factor; thus, the MNBS-P is best explained by three latent factors including coherence, coverage, and uniqueness.

All of the measurement models were fit to the MNBS-P. Specifically, I tested three different measurement models for narrative believability in order to rule out the possibility that narrative believability is a unitary construct and to determine its multidimensionality. In addition, I established whether it should be defined by three or five separate factors.

Following Tabachnik and Fidell (2007) I used five different measures to assess the goodness-of-fit of the CFA models mentioned above—(1) maximum-likelihood goodness-of-fit chi-square, (2) root mean square error of approximation (RMSEA), (3) standardized root mean square residual (SRMR), (4) comparative fit index (CFI), and (5) non-normed fit index (NNFI).The first three statistics are measures of absolute fit, meaning they show how close to perfect the tested model is. The latter ones are indices of relative fit; these measures show how close the tested model fits compared to the null model (which has no factors). The null model (e.g., the worst model possible) assumes that sampling error alone explains the covariation among observed measures. Accordingly, the null model assumes that there is no common variance among observed measures. The following represent descriptions of the goodness-of-fit statistics that were used. These indices of relative and absolute fit are commonly used to determine model fit in confirmatory factor analyses.

Maximum-Likelihood Goodness-of-Fit Chi-Square (χ^2). While a smaller chi-square value and a larger p-value reflect better goodness-of-fit, a larger chi-square value and smaller p-value reflect worse goodness-of-fit. Because a larger sample size inflates the chi-square value, this value should be interpreted with caution. Thus, this measure of absolute fit always should be supplemented with other measures of fit (Bollen & Long, 1993).

Root Mean Square Error of Approximation (RMSEA). The RMSEA represents a measure of fit based on the population discrepancy (i.e., measure of absolute fit). RMSEA ≤ .05 represents "close fit;" .05 < RMSEA ≤ .08 represents "reasonably close fit;" .08 < RMSEA < .10 represents "acceptable fit;" RMSEA > .10 represents "unacceptable fit" (Steiger, 1989).

Standardized Root Mean Square Residual (SRMR). SRMR is a measure of absolute fit, which is expressed in standard deviation units. Because it is expressed in Z-score units (i.e., it is unit-free), it is interpretable across different data sets. SRMR ≤ .08 represents "good model fit," and SRMR ≤ .05 represents "great model fit" (Hu & Bentler, 1998; Joreskorg & Sorbom, 1981)

Comparative Fit Index (CFI). CFI, a measure of relative fit, adjusts for the number of parameters estimated in the model. When the chi-square value for the model is less than its degrees of freedom, it will exceed 1. A CFI value > .90 is "acceptable" (Bentler, 1990).

Non-Normed Fit Index (NNFI). NNFI, also called Tucker-Lewis Coefficient (TLC), is not distorted by sample size. It is also not constrained to a value between 0 and 1. When the chi-square value for the model is less than its degrees of freedom, it will exceed 1. A NNFI value > .90 is "acceptable" (Bentler & Bonnett, 1980).

To address the problem of agreement-response bias, I utilized confirmatory factor analysis controlling for acquiescence (Billiet & McClendon, 1998, 2000). For each model under investigation, this procedure involves specifying a style factor in addition to the other content factor(s). Thus, I fixed all the loadings on the style factor +1, so that respondents who tend to agree with most of the scale items score high and respondents that tend to disagree with most of the scale items score low on the acquiescence factor. Also, I set the covariance(s) between the style factor and the content factor(s) to 0 because I did not expect acquiescence to correlate with specific attitudes. By including the acquiescence factor in the model, any possible agreement-response bias portrayed by the respondents did not bias the final solution of each measurement model because I statistically controlled for it (Billiet & McClendon, 1998, 2000).

In addition, when two models were nested, I contrasted the goodness-of-fit chi-squares in order to determine whether one model fit the data better than the other model. Accordingly, I calculated $\Delta\chi^2 = \chi^2_1 - \chi^2_2$ with $\Delta df = df_1 - df_2$, where χ^2_1 and df_1 represent the goodness-of-fit statistics for the model with the higher chi-square value, and χ^2_2 and df_2 represent the goodness-of-fit statistics for the lower chi-square value.

In addition, I analyzed item factor loadings in order to ensure that uniqueness

items with poor loadings were not included in the final version of the MNBS-P.

Before conducting the analyses, I performed a random split on the data, which allowed me to explore the factor structure of the Modified Narrative Believability Scale – Prosecution (MNBS-P) with one half of the data first, and then to validate my findings by confirming the factor structure of the MNBS-P after having removed the uniqueness item that had the smallest factor loading with the other half of the data. The first random subsample included 283 participants, and the second random subsample included 282 participants. I first tested the reliability of the scale and its subscales for each sample, and then explored the factor structure of the MNBS-P by using confirmatory factor analysis (CFA) for each sample.

Reliability Analyses of MNBS-P for First Random Subsample (*n* = 283)

I calculated Cronbach's alpha to determine the internal consistency of the MNBS-P and its presumed subscales. Overall, the 18 items of the MNBS-P demonstrated a Cronbach's α of .93, which shows excellent internal reliability. The analyses also indicated that deleting any of the items would not significantly increase Cronbach's α. I expected that either three or five subscales would make up the MNBS-P. All of these possible subscales showed acceptable reliability. Specifically, Cronbach's α = .87 for the three items that presumably measure plausibility; Cronbach's α = .77 for the three items that presumably measure completeness; Cronbach's α = .87 for the three items that presumably measure consistency; Cronbach's α = .77 for the three items that presumably measure coverage; Cronbach's α = .84 for the six items that presumably measure uniqueness; and Cronbach's α = .90 for the nine items (plausibility, completeness, and

consistency items combined) that presumably measure coherence.

Confirmatory Factor Analyses of MNBS-P for First Random Subsample (*n* = 283)

To rule out the possibility that the MNBS-P is best represented by one overall factor, I first tested a global one-factor model of narrative believability with 18 self-report items. The results indicated that the one-factor model provided a poor fit to the data, $\chi^2(134, N = 283) = 1353.42$, $p < .001$, RMSEA = .18, SRMR = .10, CFI = .90, NNFI = .89. Only one of the indices of goodness-of-fit was at an acceptable level (CFI $\geq$ .90).

In line with Pennington and Hastie's (1988) assertion that coherence is made up of other constructs, but represents a certainty principle by itself, I conducted a correlated (i.e., oblique) three-factor model of narrative believability with 18 self-report items that measure coverage (COV1, COV2, and COV3), coherence (PL1, PL2, PL3, CN1, CN2, CN3, CM1, CM2, and CM3), and uniqueness (UNI1, UNI2, UNI3, UNI4, UNI5, and UNI6). The results demonstrated a better fit compared to the global model, $\chi^2(131, N = 283) = 584.03$, $p < .001$, RMSEA = .12, SRMR = .08, CFI = .94, NNFI = .93 (Table 1). Although the RMSEA and SRMR are slightly above the acceptable cut-off scores, the goodness-of-fit measures overall show that this model provides a good fit to the data.

Yale (2013) concluded that narrative believability is best reflected by five factors. Therefore, I also tested a correlated five-factor model of narrative believability with 18 self-report items that measure coverage (CV1, CV2, and CV3), plausibility (PL1, PL2, and PL3), consistency (CN1, CN2, and CN3), completeness (CM1, CM2, and CM3), and uniqueness (UN1, UN2, UN3, UN4, UN5, and UN6). The results showed that the model provided a tolerable fit to the data, $\chi^2(126, N = 283) = 586.35$, $p < .001$, RMSEA = .11,

SRMR = .23, CFI = .94, NNFI = .93 (Table 1). For most of the goodness-of-fit measures the results were similar to the three-factor model; however, SRMR greatly deviated from the acceptable cut-off score.

Table 1. Goodness-of-fit Statistics for Measurement Models of MNBS-P with 18 items (N = 283)

Model	χ^2	df	RMSEA	SRMR	CFI	NNFI
One global factor	934.09	134	.18	.10	.90	.89
3 oblique factors controlling for acquiescence	584.50	131	.12	.08	.94	.93
5 oblique factors controlling for acquiescence	586.35	126	.11	.23	.94	.93

In addition to examining the fit indices, I contrasted the goodness-of-fit chi-squares of the nested models. I found that the correlated three-factor model fit the data significantly better than the one-factor model, $\Delta\chi^2(3, n = 283) = 349.58$, $p < .00001$. In addition, chi-square difference testing revealed that the correlated three-factor model and the correlated five-factor model fit the data equally well, $\Delta\chi^2(5, n = 283) = 1.85$, $p = .85$ (Table 2). Thus, the more parsimonious correlated three-factor model was preferred over the latter.

Table 2. Contrasted Nested Measurement Models of MNBS-P with 18 items ($N = 283$)

	$\Delta\chi^2$	Δdf	p
One global factor vs. 3 oblique factors controlling for acquiescence	349.58	3	$< .00001$
3 oblique factors vs. 5 oblique factors (controlling for acquiescence)	1.85	5	$= .87$

In terms of factor correlations for the correlated three-factor model of narrative believability, coverage and coherence were strongly correlated, $r = .72$ (sharing 51.84% of their variance), coverage and uniqueness were moderately correlated, $r = .64$ (sharing 40.96% of their variance), and coherence and uniqueness were also moderately correlated, $r = .52$ (sharing 27.04% of their variance).

Reliability Analyses for Second Random Subsample ($n = 282$)

After examining the standardized factor loadings of the oblique three-factor solution, I decided to remove one of the uniqueness items (i.e., The prosecution story provides a unique explanation of the evidence) from the scale because its loading was much lower (0.52) than the loadings of all the other uniqueness items (> 1.04). In order to verify that the oblique three-factor model would still fit the data the best after omitting one of the items that was originally hypothesized to measure uniqueness, I dropped this item for the analyses of the other random half of the data. Thus, the remaining analyses were conducted with 17, not 18, items.

To determine the internal consistency of the final version of the MNBS-P and its three subscales I calculated Cronbach's alpha. The 17 items of the MNBS-P showed a Cronbach's α of .93, which shows excellent internal reliability. In addition, the reliability analyses indicated that deleting any of the items would not significantly increase Cronbach's α. The three subscales, coherence (nine items), coverage (three items), and uniqueness (five items), all demonstrated acceptable reliability—Cronbach's α = .91, Cronbach's α = .77, Cronbach's α = .87, respectively.

Confirmatory Factor Analyses for Second Random Subsample (*n* = 282)

In order to validate my findings that emerged from the data with the first random subsample and to create a final version of the MNBS-P, I attempted to replicate the results from the first set of confirmatory analyses with now only 17 items. Accordingly, I tested a global one-factor mode measuring narrative believability, a correlated three-factor model, and a correlated five-factor model, while controlling for acquiescence in all tested models.

First, I conducted a confirmatory analysis testing a global one-factor model of Narrative Believability with 17 self-report items. As before, the results demonstrated that the one-factor model provided a poor fit to the data, $\chi^2(118, N = 282) = 870.09$, $p < .001$, RMSEA = .18, SRMR = .11, CFI = .91, NNFI = .89. All goodness-of-fit indices except from CFI (> .90) were at unacceptable levels (Table 3).

Next, I tested a correlated three-factor model of Narrative Believability with 17 self-report items that measure coverage (CV1, CV2, and CV3), coherence (PL1, PL2, PL3, CN1, CN2, CN3, CM1, CM2, and CM3), and uniqueness (UN1, UN2, UN3, UN4,

and UN5). The results showed a good fit with the data, χ^2(115, $N = 282$) = 501.64, $p <$ .001, RMSEA = .11, SRMR = .07, CFI = .95, NNFI = .94. All goodness-of-fit measures showed acceptable fit, except from RMSEA because it was > .10 (Table 3).

Finally, I tested a correlated five-factor model of Narrative Believability with 17 self-report items that measure coverage (CV1, CV2, and CV3), plausibility (PL1, PL2, and PL3), consistency (CN1, CN2, and CN3), completeness (CM1, CM2, and CM3), and uniqueness (UN1, UN2, UN3, UN4, and UN5). As expected the results showed that the model provided only a tolerable fit to the data, χ^2(110, $N = 282$) = 534.44, $p < .001$, RMSEA = .11, SRMR = .24, CFI = .95, NNFI = .93 (Table 3). For a majority of the goodness-of-fit measures the results were similar to the three-factor model; however, SRMR was much higher than the acceptable cut-off score (Table 3).

Table 3. Goodness-of-fit Statistics for Measurement Models of MNBS-P with 17 items ($N = 282$)

Model	χ^2	df	RMSEA	SRMR	CFI	NNFI
One global factor	870.09	118	.18	.11	.91	.89
3 oblique factors controlling for acquiescence	501.64	115	.11	.07	.95	.94
5 oblique factors controlling for acquiescence	534.44	110	.11	.24	.95	.93

As before, I also tested the differences between the chi-squares of the nested

models. The results revealed that that correlated three-factor model fit the data significantly

better than the global model, $\Delta\chi^2(3, n = 282) = 368.45, p < .0001$. The chi-square of the

correlated five-factor model was larger than the chi-square value for the correlated three-

factor model. Thus, I refrained from chi-square difference testing because the results

would have been not meaningful due to a negative chi-square value that would have been

produced. Because the chi-square value was larger in the five-factor model, the correlated

three-factor model was preferred. The large deviation of SRMR from the cut-off score ($<$

.07), corroborated this interpretation.

Table 4. Contrasted Nested Measurement Models of MNBS-P with 17 items ($N = 282$)

	$\Delta\chi^2$	Δdf	p
One global factor vs. 3 oblique factors controlling for acquiescence	368.45	3	$< .00001$
3 oblique factors vs. 5 oblique factors (controlling for acquiescence)	[-32.80]	---	---

For the correlated three-factor model measuring narrative believability,

correlational analyses revealed that coverage and coherence were strongly correlated, $r =$

.78 (sharing 60.84% of their variance), coverage and uniqueness were moderately

correlated, $r = .58$ (sharing 33.64% of their variance), and coherence and uniqueness

were moderately correlated, $r = .50$ (sharing 25.00% of their variance). For a visual

depiction of the conceptual correlated three-factor model, see Appendix W, Figure 2.

Discussion

The goal of Study 1 was to determine how many items the modified versions of

Yale's (2010) original Narrative Believability Scale (NBS-12) should include. In

addition, I investigated how many factors underlie the MNBS-P. Confirmatory factor

analyses revealed that 17, and not 18 items, should be included in the MNBS-P. Thus, all

other versions of this scale (i.e., MNBS-D, MNBS-J, and MNBS-P) also should include

17 items. In addition, I found that a correlated three-factor structure underlies the 17

items of the MNBS-P; the three factors include coverage (3 items), coherence (9), and

uniqueness (5 items). Although a correlated five-factor model also emerged as good fit,

the three-factor model was more parsimonious, and thus, the latter model was preferred.

As expected, all three subscales were positively correlated, while I found the strongest

relationship between coverage and coherence.

In the following three studies, the 17-item scale was employed, and a three-factor

structure was assumed for the analyses that included the MNBS-P, or any other version

of the scale.

CHAPTER 6

STUDY 2

Overview

The purpose of Study 2 was to replicate the race-crime congruence effects on

jurors' judgments about the defendant's culpability that were found in past research

(Kaplan & Jones, 2003), so I could then explain these findings in terms of the coherence

of the prosecution as well as defense story. In addition, the goal was to investigate

whether a juror's low motivation to control prejudiced reactions (MCPR) or asking the

jurors to refrain using stereotypes in their decision making would eliminate those race-

crime congruence effects.

In Study 2, participants were randomly assigned to read about a White or Black

defendant (race conditions) who was charged with either grand-theft auto or

embezzlement (type of case conditions). Half of the participants received stereotype

suppression instructions (SSI), while the other half did not receive this type of

instructions. Participants rated the prosecution story and the defense story in terms of

narrative believability. They then provided culpability judgments, and filled out the

Motivation to Control Prejudiced Reactions Scale (MCPRS) and the Need for Cognition

Scale (NCS).

The predictions outlined below were made for White jurors only. I did not make

any specific predictions concerning the main effects. However, if the main effects had

been significant, in addition to the interaction effects, on verdict judgments, I would have expected story coherence to mediate each main effect. I also would have assumed that the same moderators would play a role.

Hypotheses

Hypothesis 2a

I predicted a significant interaction between defendant race (Black, White) and type of crime (grand-theft auto, embezzlement) on the verdict decision (not guilty vs. guilty), when controlling for the main effects of defendant race and type of crime. Defendants charged with crimes stereotypical to their race would be judged guilty more frequently than defendants who do not fit that crime stereotype. Specifically, White jurors would render guilty verdicts more often for a Black defendant who was charged with grand-theft auto than for a White defendant who was charged with grand-theft auto. In addition, White jurors would render guilty verdicts more often for a White defendant who was charged with embezzlement than for a Black defendant who was charged with embezzlement (see Figure 2).

Hypothesis 2b

I predicted that the significant interaction effect between defendant race (Black, White) and type of crime (grand-theft auto, embezzlement) on the verdict decision would be mediated by the coherence of the prosecution story, when controlling for the main effects of defendant race and type of crime. Specifically, I predicted that race-crime congruence, as opposed to race-crime incongruence, would be associated with increased coherence of the prosecution story. Increased coherence of the prosecution story, in turn,

should be associated with an increase in guilty verdict decisions.

I predicted that the key mediator is the coherence of the prosecution story, and not of the defense story. Because the burden of proof lies with the prosecution, good jurors always should render a not guilty verdict whenever the prosecution does not tell a good (e.g., coherent) story. However, the coherence of the defense story also could mediate the interaction effect between defendant race (Black, White) and type of crime (grand-theft auto, embezzlement) on jurors' verdict decisions, so I had planned to test this as well. I expected that the coherence of the defense story would be associated with an increase in not guilty verdicts (see Figure 2).

Hypotheses 3a and 3b

I expected the same effects as described in hypotheses 2a and 2b for jurors' verdict-confidence scores as the outcome variable (see Figure 2).

Hypotheses 4a and 4b

I expected the same effects as described in hypotheses 2a and 2b for jurors' probability of guilt ratings as the outcome variable (see Figure 2).

Figure 2. Statistical Diagram for Hypotheses 2a-4b
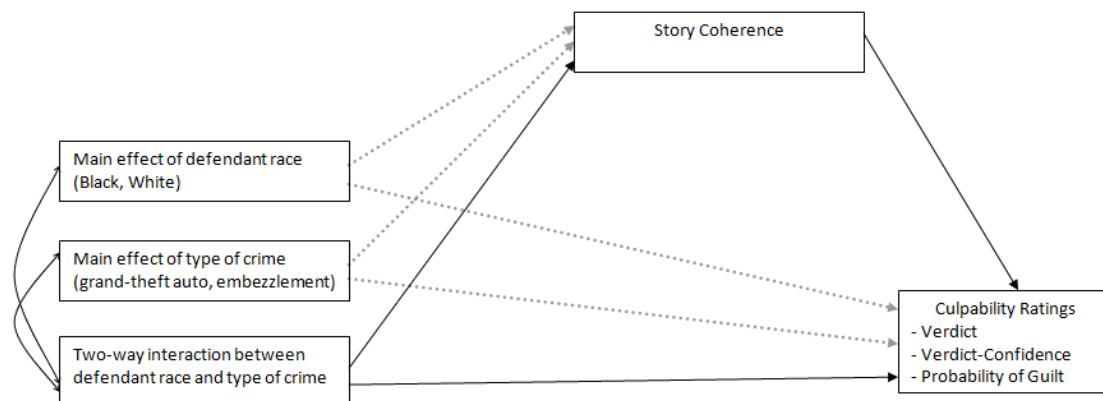
Figure 2. The residual disturbance terms for both the mediator and the outcome variable, reflecting the unexplained variance in each of these endogenous variables, were omitted from the diagram to simplify presentation.

Hypothesis 5a

I expected a significant Defendant Race x Type of Crime x Motivation to Control Prejudiced Reactions (MCPR) interaction predicting jurors' verdict decisions, when controlling for the main effects of defendant race, type of crime, and MCPR on jurors' verdict decision. I hypothesized that MCPR would moderate the interaction effect between defendant race and type of crime on verdict decisions. When White jurors have low levels of MCPR, defendants charged with crimes stereotypical to their race would be judged guilty more frequently than defendants who do not fit this crime stereotype. Specifically I predicted that these jurors would render guilty verdicts more often for a Black defendant who was charged with grand-theft auto than for a White defendant who was charged with grand-theft auto. In addition, White jurors with low levels of MCPR would render a guilty verdict more frequently for a White defendant who was charged with embezzlement than for a Black defendant who was charged with embezzlement. In

contrast, I hypothesized that when jurors have high levels of MCPR, their verdict decisions will not differ between Black defendants who were charged with grand-theft auto and White defendants who were charged with grand-theft auto.

The MCPRS specifically measures an individual's prejudice toward *Black* people. This means that this scale is not concerned with one's reactions toward people of different races. Thus, I did not make specific predictions about the way that jurors' high level of MCPR (toward Black people) might affect differences in verdict decisions between a Black defendant who was charged with a crime stereotypical of Whites (i.e., embezzlement) and a White defendant who was charged with a crime stereotypical of Whites (i.e., embezzlement; see Figure 3).

Hypothesis 5b

I predicted that the interaction effect between defendant race (Black, White), type of crime (grand-theft auto, embezzlement), and MCPR on jurors' verdict decisions will be mediated by the coherence of the prosecution story. Specifically, I predicted that race-crime congruence, as opposed to race-crime incongruence, would be associated with increased coherence of the prosecution story, although this effect should only be present when jurors have low levels of MCPR, as opposed to when they have high levels of MCPR. In turn, increased coherence of the prosecution story should be associated with more frequent guilty verdicts. Here, a significant indirect effect would demonstrate that the effect of race-crime congruence on verdicts is moderated by jurors' levels MCPR, and this interaction's effect on verdicts of guilt would be due to increased coherence of the prosecution story (see Figure 3).

Hypotheses 6a and 6b

I expected the same effects as described in hypotheses 2a and 2b for jurors'

verdict-confidence scores as the outcome variable (see Figure 3).

Hypotheses 7a and 7b

I expected the same effects as described in hypotheses 2a and 2b for jurors'

probability of guilt ratings as the outcome variable (see Figure 3).

Figure 3. Statistical Diagram for Hypotheses 5a-7b
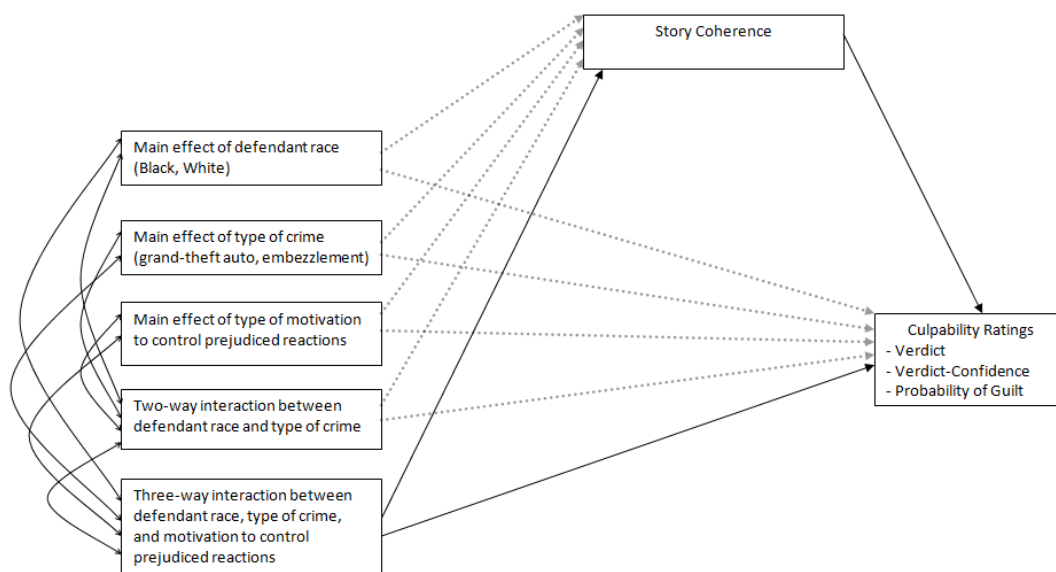
Figure 3. The residual disturbance terms for both the mediator and the outcome variable, reflecting the unexplained variance in each of these endogenous variables, were omitted from the diagram to simplify presentation.

Hypothesis 8a

I predicted a significant three-way interaction between defendant race (Black,

White), type of crime (grand-theft auto, embezzlement), and stereotype suppression

instructions (SSI; SSI, control) on the verdict decision, when controlling for the main

effects of defendant race, type of crime, and SSI. I expected that SSI would moderate the interaction effect between defendant race and type of crime on jurors' verdict decisions. When White jurors receive no SSI, defendants charged with crimes stereotypical to their race would be judged guilty more frequently than defendants who do not fit that crime stereotype. Specifically, White jurors would render guilty verdicts more often for a Black defendant who was charged with grand-theft auto than for a White defendant who was charged with grand-theft auto. In addition, White jurors will render guilty verdict more frequently for a White defendant who was charged with embezzlement than for a Black defendant who was charged with embezzlement. In contrast, I hypothesized that when jurors receive SSI, their verdict decisions would not differ between defendants who were charged with crimes that are stereotypical to their race (i.e., a White defendant who was charged with embezzlement; a Black defendant who was charged with grand-theft auto) and defendants who were charged with crimes that are not stereotypical to their race (i.e., a Black defendant who was charged with embezzlement; a White defendant who was charged with grand-theft auto; see Figure 4).

Hypothesis 8b

I predicted that the interaction effect between defendant race (Black, White), crime (grand-theft auto, embezzlement), and stereotype suppression instructions (stereotype suppression instructions, control) would be mediated by story coherence. I predicted that race-crime congruence, as opposed to race-crime incongruence, would be associated with increased coherence of the prosecution story, although this effect should only be present when jurors do not hear any SSI, as opposed to when they hear SSI.

Increased coherence of the prosecution story, in turn, should be associated with an increase in guilty verdicts. A significant indirect effect would demonstrate that the effect of race-crime congruence on verdicts is moderated by whether or not jurors hear SSI, and this interaction's effect on the number of guilty will be due to increased coherence of the prosecution story (see Figure 4).

Hypothesis 9a and 9b

I expected the same effects as described in hypotheses 2a and 2b for jurors' verdict-confidence scores as the outcome variable (see Figure 4).

Hypotheses 10a and 10b

I expected the same effects as described in hypotheses 2a and 2b for jurors' probability of guilt ratings as the outcome variable (see Figure 4).

Figure 4. Statistical Diagram for Hypotheses 8a-10b
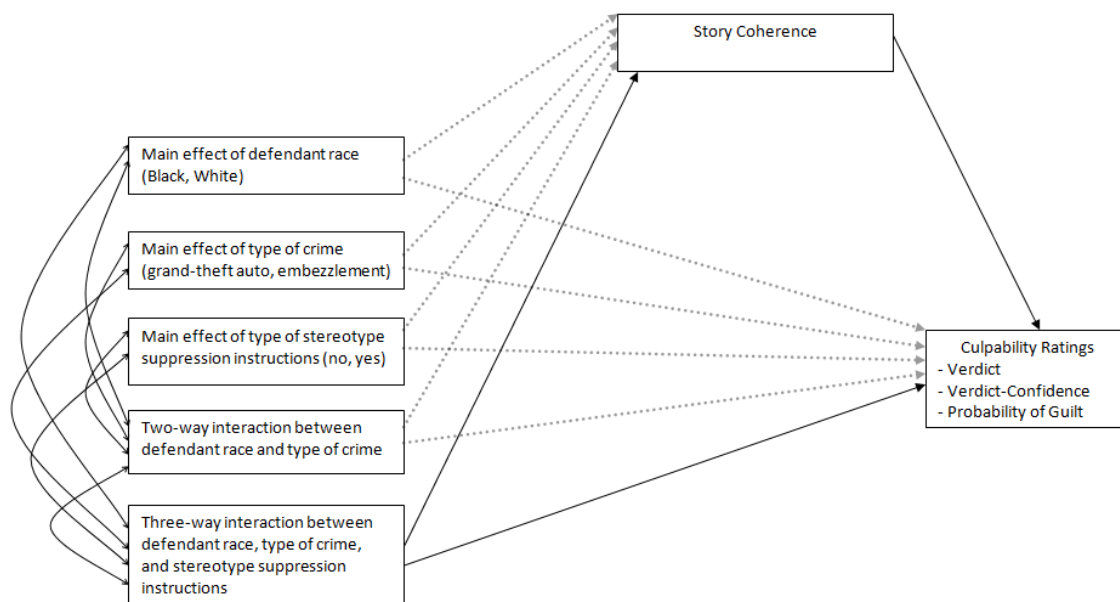
Figure 4. The residual disturbance terms for both the mediator and the outcome variable, reflecting the unexplained variance in each of these endogenous variables, were omitted from the diagram to simplify presentation.

Method

Participants

For Study 2, I planned to recruit 505 participants via MTurk. This sample size was based on a prospective power analysis to achieve 90 % power using Power Analysis and Sample Size (PASS 11) statistical software (Hintze, 2011) as well as Fritz and MacKinnon's (2007) empirical estimates of sample size needed for 80% power for testing mediation. Past research using meta-analysis demonstrated a small effect ($d = .09$, $d = .17$) of race on culpability ratings (Mitchell et al., 2005; Sweeney & Haney, 1992), and Jones and Kaplan (2003) found a medium effect of the interaction between defendant

race and type of crime on verdict-confidence scores ($\eta^2 = .17$). Given these findings and to provide a conservative estimate of sample size, I had used a small effect size to determine sample size. I had determined that a sample size of 200 for a significant two-way interaction and of 184 for a significant three-way would achieve 90% power (Hintze, 2011). Regarding statistical mediation, when using the bias-corrected method and assuming that the standardized regression coefficient linking the independent variable to the mediator is at least ß = .14 and the standardized regression coefficient linking the independent variable to the mediator is at least ß = .26, a sample size of 377 is needed to achieve 80% power to detect a meditational effect of the given independent variable at two-tailed $p < .05$ (Fritz & MacKinnon, 2007, Table 3, p. 237). Berinsky, Huber, and Lenz (2012) showed that generally approximately 83.5% of MTurk workers are White. My goal was to have a large enough sample to have sufficient power detect significant effects, if they exist, for White participants only. However, I did not want to restrict data collection to only Whites because I planned to conduct some exploratory analyses with non-White participants, if a sufficient number of non-White participants were to take part in the study. Therefore, I planned to collect an additional 62 participants (377 + 16.5 % of 377 = 439). In addition, in order to account for any data that would have needed to be dropped from the analysis, I planned to collect an additional 15% ($n = 66$). Thus, my goal was to have a total of 505 participations. The inclusion criteria were the same as in Study 1. In addition, only participants who identified themselves as citizens of the United States were included in the study because, by law, only this population is allowed to serve as a juror.

In total, data were collected from 559 MTurk workers who received $0.75 in exchange for their participation in the research study. I omitted the data of those participants who had stopped responding to the questionnaire items before making a verdict decision ($n = 14$), and those who did not respond to any of the manipulation check items ($n = 13$). When participants had the same I. P. address, I omitted the response with the latest start date from the analyses ($n = 9$). European Americans represented 78.68% of the sample; thus, I excluded the 110 participants whose ethnicity was different from "European American/White" (see Appendix X, Figure 20). I also omitted the data of those participants who indicated that they were not United States citizens ($n = 17$), failed the attention check item ($n = 17$), failed the manipulation check for defendant race ($n = 22$), and for those who failed the manipulation check for crime type ($n = 16$). If any of the participant's responses to what they believe the study was all about and whether they were suspicious included the following words, their data also was excluded from the analysis: "prejudice," "Black," "profiling," "stereotypes," "person's color," "African American," "color of skin" and "White man" ($n = 79$).

Overall, after data cleaning, the final data set included 299 White participants, of which 54.5% were female. The participants were between 18 and 80 years of age ($M = 35.21$, $SD = 12.04$). The majority of participants (43.48%) had received some higher education (Figure 5).

Figure 5. Study 2 – Level of Education
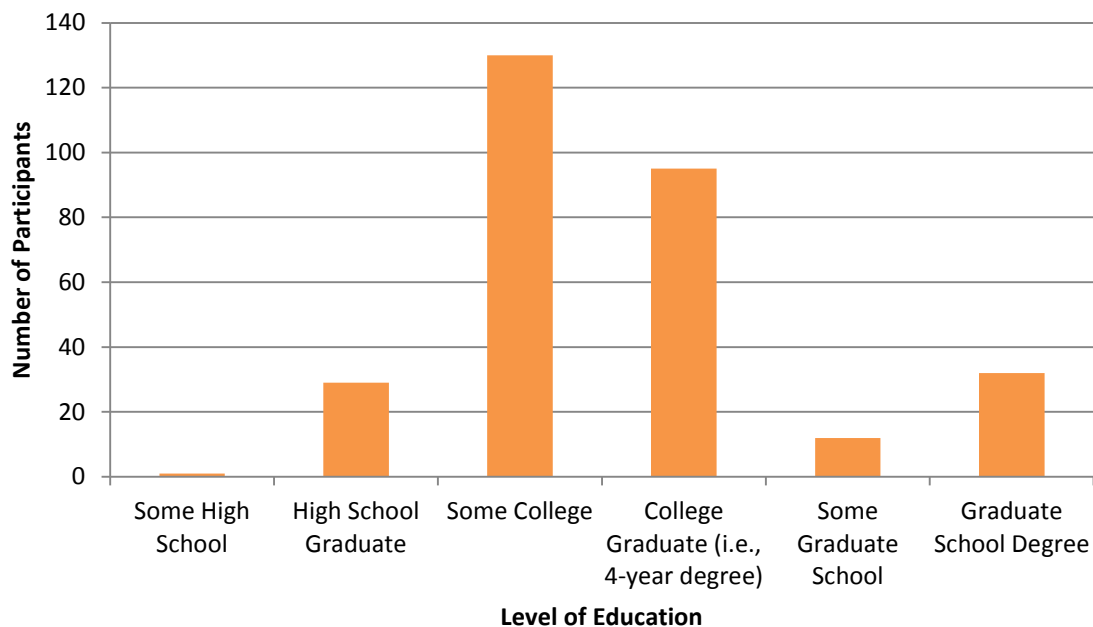
Materials

  Defendant profile. Two pictures, one of a White male and one of a Black male, were selected from the CAL/PAL face database (Minear & Park, 2004). Because I attempted to use photographs that could be used as mug shots, I only inspected photographs of men who had a neutral facial expression. In addition, I ensured that the two males were similar in age at the time to photograph was taken. (i.e., the Black person was 39 years old and the White person was 41 years old in the picture). A brief profile of the defendant, including his name, race, sex, age, charge, and arrest date, appeared next to each photograph. In order to make race even more salient, the defendant's first name was varied for the Black (Jamal) and White (Jack) defendant, while the last name was held constant (Williams). I created four versions of the defendant profile, so that the

defendant's race as well as the type of crime would vary across the conditions (see Appendix I).

Judicial instructions. The judicial instructions on the summary of charge, the function of the court and the jury, and consideration of evidence that were used in the present study were adapted from sample court materials. The word-for-word instructions were included in the Procedure section.

Case summaries. Each participant read one of two criminal case summaries. Half of the participants read about a defendant, who was charged with grand-theft auto, while the other half read about a defendant who is on trial for embezzlement. The case summaries were edited versions of Jones and Kaplan's (2003) original materials; in their study, embezzlement represented a crime that was stereotypical of White defendants, and grand-theft auto generally was associated with Black defendants. In the embezzlement case (563 words in length), the defendant was charged with approving loans for fictitious people and taking bank funds for his personal use in order to pay off debt (among other things). The defendant claims that the president of the bank authorized him to do these purchases. In the grand-theft auto case (544 words in length), the defendant was charged with stealing a Cadillac Escalade from a dealership. He was driving the stolen car until he was stopped for speeding. The defendant states that a friend had lent him the car, and that he has never been to the car dealership. For complete versions of the case summaries, see Appendix J).

Modified Narrative Believability Scale – Prosecution (MNBS-P). In order to measure narrative believability of the prosecution story, the MNBS-P was used. After

examining the factor loadings of each item in Atudy 1, one uniqueness item was dropped

from the scale because it did not load very highly onto the uniqueness factor (i.e., "The

defense story provides a unique explanation of the evidence."). Thus, I used an edited

(final) version of the MNBS-P that includes 17 items (see Appendix K).

Modified Narrative Believability Scale – Defense (MNBS-D). In order to

measure narrative believability of the defense, the MNBS-D was employed. The MNBS-

D included the same items as the final version of the MNBS-P; however, the word

"defense" replaced "prosecution" in the wording of the each item and the instructions

(see Appendix L).

Explanations of the case law. I provided the participants with the legal

explanation of the specific offense the read about in their case summary. The legal

definitions of embezzlement (California Penal Code § 487 d, as cited in Pickens v.

Miller, 2002) and grand-theft auto (18 U.S. Code § 656, www.law.cornell.edu) were

abbreviated and simplified, so that the participants would have no doubt about what the

crime they would be reading about legally involves, and what must be proven in order to

find a defendant guilty of the crime (see Appendix M).

Verdict judgments. The participants' verdict judgments represented the outcome

measures. Participant made individual verdict decisions (not guilty or guilty), and

indicated how confident they are in their verdict decision on a scale from 1 *not confident*

at all to 7 *completely confident*. In addition, the participants rated the probability of this

particular defendant having committed this specific crime. They provided a number

between 0 *the defendant could not possibly have committed the crime* and 100 *the*

defendant committed the crime with absolute certainty (see Appendix N).

Questions probing for suspicion. Two questions probed the participants for suspicion. For example, they were asked to guess what the study was about. In addition, they were asked whether they had ever participated in a study that included the same cases summary (see Appendix O).

Motivation to Control Prejudiced Reactions Scale (MCPRS). In order to measure each participant's motivation to control prejudiced reactions (MCPR) toward Blacks, the MCPRS (Dunton & Fazio, 1997) was used. In general, it conceptualizes an individual's amount of MCPR. The MCPRS has two subscales measuring concern with acting prejudice and restraint to avoid dispute. Participants indicate the extent to which they agree or disagree with the 17 items using a bipolar rating scale ranging from -3 *strongly disagree* to +3 *strongly agree*. For example, participants state their level of agreement to the following statement, "I think it is important to speak one's mind rather than to worry about offending someone." After reverse scoring the negatively worded items, the total score reflects a person's MCPR, with higher scores representing higher levels of motivation. The MCPRS has high internal reliability, with Cronbach's α ranging from .76 to .81 for student samples and with a Cronbach's α of .74 for paid participant samples (Dunton & Fazio, 1997). Because the original scale primarily was developed for its use with student samples, and my sample included not only students, but also adults that are not students, item two was modified, so it includes discussions at school *and* at work. Specifically, the description of the item was changed from, "If I were participating in a class discussion and a Black student expressed an opinion with which I disagreed, I

would be hesitant to express my own viewpoint" to "If I were participating in a general discussion, either at work or at school, and a Black person expressed an opinion with which I disagreed, I would be hesitant to express my own viewpoint." Please refer to Appendix P for the MCPRS as it was used in the proposed research.

Need for Cognition Scale (NC). The Need for Cognition Scale is a self-report questionnaire, which measures "the tendency for an individual to engage in and enjoy thinking" (Cacioppo & Petty, 1982, p. 116). The shortened version includes 18 items, and respondents are asked to indicate whether or not each statement is characteristic of them or what they believe using a Likert scale ranging from 1 *extremely uncharacteristic of me* to 5 *extremely characteristic of me* (Cacioppo, Petty, & Kao, 1984). Sample items include "I find satisfaction in deliberating hard and for long hours," "Thinking is not my idea of fun," and "I prefer my life to be filled with puzzles I must solve." After reverse scoring nine items, a participant's composite score refers to his or her level of need of cognition; the higher one's total score, the higher the level of need for cognition. Past research has shown that is the Need for Cognition Scale is highly reliable and valid in measuring a person's need for cognition (e.g., Cacioppo & Petty, 1982; Sadowski, 1993; Sadowski & Gulgoz, 1995b; see Appendix Q).

Demographic questionnaire. Participants reported their age, gender, ethnicity, and level of education. They also indicated whether they were a United States citizen (see Appendix R).

Final questionnaire. The final questionnaire included some manipulation check items. Specifically, the participants were asked to name the defendant's race as well as

his charge. In addition, participants indicated whether they remembered the SSI

instructions that were embedded in the judicial instructions. They were also asked to

indicate to what degree they would associate this crime with a Black or White defendant.

The response options ranged from 1 *associated with Whites* to 7 *associated with Blacks*.

The final questionnaire also included a question about the perceived seriousness of the

crime; the Likert scale ranged from 1 *not serious at all* to 7 *very serious* (see Appendix

S).

Procedure

Participants signed up for the study via MTurk. The study was advertised as

"Juror Decision Making," and potential workers were informed that it would take them

no longer than 30 minutes to complete the study. Furthermore, they would earn US $0.50

for their participation. Workers who selected this HIT were redirected to a web-based

survey, which was designed with Survey Monkey (SurveyMonkey Inc.). After providing

electronic consent, participants read the following general instructions:

> For the purpose of the study, please imagine that you have been selected to serve
> on a jury for a criminal trial. So while completing this study, please think of
> yourself as an individual juror who will have to come a verdict decision at the end
> of a trial.

Participants were randomly assigned to one of four conditions. They saw a mug shot and

a profile of a Black defendant who was charged with grand-theft auto, a White defendant

who was charged with embezzlement, a Black defendant who was charged with

embezzlement, or a White defendant who was charged with grand-theft auto. The former

two conditions represented race-crime congruence, while the latter two conditions

represented race-crime incongruence. Afterwards, all participants read judicial

instructions that summarized the charge:

> The prosecution brings one charge in this case. The prosecution claims that the defendant, Jamal R. Williams [Jack R. Williams] committed grand-theft auto [embezzlement]. The defendant pleads not guilty. To succeed on this charge, the prosecution must prove beyond a reasonable doubt that the defendant Jamal R. Williams [Jack R. Williams] committed grand-theft auto [embezzlement]. If you find that the prosecution has proved this claim beyond a reasonable doubt, then you should find for the prosecution, and go on to find the defendant guilty of this charge. If, on the other hand, you find that the prosecution has failed to prove this claim beyond a reasonable doubt, then you should find for defendant Jamal R. Williams [Jack R. Williams] and you will find him not guilty.

Participants then were provided with judicial instructions on the function of the court and

the jury, which explained their duties as jurors in this case:

> Members of the jury, you will hear all the evidence and arguments of the prosecution and defense. At this point, I will instruct you on the law. You have two duties as a juror. Your first duty is to decide the facts for the evidence in the case. This is your job and yours alone. Your second duty is to apply the law that I give you to the facts. You must follow these instructions, even if you disagree with them. Each of the instructions is important, and you must follow all of them. Perform these duties fairly and impartially. Do not allow sympathy to influence you.

The following judicial instructions provided participants with information on how they

should consider the evidence:

> In determining whether any fact has been proved, you should consider all of the evidence bearing on the question regardless of who introduced it. You should use common sense in weighing the evidence and consider the evidence in light of your own observations in life. In our lives, we often look at one fact and conclude from it that another fact exists. In law, we call this "inference." Jurors are allowed to make reasonable inferences. Any inference you make must be reasonable and must be based on the evidence in the case.

Participants were randomly assigned to either receive or not receive stereotype

suppression instructions (SSI). Therefore, half of the participants did not receive any

additional instructions, while the other half of participants read the following SSI:

> However, previous psychological research has shown that our impressions and
> evaluations of others are commonly influenced by stereotype-based thoughts and
> beliefs. As a good juror, please try to ignore this source of information when
> forming an impression of the defendant.

The participants then read a summary of the criminal case. Depending on the condition

they were randomly assigned to at the beginning of the experiment, participants read

about either an embezzlement case or a grand-theft auto case. In order to make race

salient again, the defendant's photograph as well as his profile appeared at the beginning

of the case summary. Afterwards, they were asked to respond to questions on the MNBS-

P and MNBS-D scales, thus rating the narrative believability of both the prosecution as

well as defense story. I counterbalanced the order in which the participants responded to

these two scales. The participants were instructed to think about the case they just read

from the perspective of the prosecution, and the defense (or vice versa). They were

instructed to think about what the basic story of the prosecution (or defense) was and to

rate each story. Afterwards, the participants saw an attention check item. Specifically,

they were asked to "select the bubble located all the way to the left, which is 'never'."

The response options ranged from 1 *never* to 7 *always*.

Following a legal explanation of the charge (embezzlement or grand-theft auto),

the participants were asked to make their verdict decision, state how confidence they

were in this decision, and made a probably of guilt rating. In order to ensure that the participants would not change their responses to the verdict measures after they might guess that the study is about prejudice (after they see the MCPRS), they were then told not to use the "prev" (i.e., previous page) button on the screen. Next, the participants were probed for suspicion and they were asked whether they have ever seen the case summary before.

Afterwards, the participants indicated the extent to which they agreed or disagreed with each item on the MCPRS. They then filled out the Need for Cognition Scale. Finally, participants answered demographic questions and filled out the final questionnaire, which included, for example, manipulation check items.

Results

Pilot Test

According to Jones and Kaplan (2003), embezzlement represents a crime that is stereotypical of White defendants, and grand-theft auto generally is associated with Black defendants. Pilot tests were conducted in order to replicate their findings in a student sample (in the lab and online), and to confirm those with MTurk workers (online). In addition, my goal was to ensure that the crimes did not differ in perceived seriousness, and that guilt ratings are the same for each crime.

Perceived associations. After reading either the embezzlement or grand-theft auto case for which defendant race was not specified, participants were asked the following question, "To what degree would you typically associate this crime with a White or Black defendant?" Respondents answered on a 9-point scale with "White

defendant" and "Black defendant" as the end points. Results showed that MTurk workers associated the crime of embezzlement more with a White defendant ($M = 2.66$, $SD = 1.68$) and the case of grand-theft auto more with a Black defendant ($M = 5.98$, $SD = 2.10$), $t(85) = -11.03$, $p < .001$. Similarly, college students who saw hard-copies of the cases indicated that they associate embezzlement with a White defendant ($M = 3.58$, $SD = 1.62$) and grand-theft auto with a Black defendant ($M = 5.73$, $SD = 1.54$), $t(105) = -9.56$, $p < .001$. In addition, college students who responded to the brief survey online indicated that they would typically associate embezzlement with a White defendant ($M = 2.55$, $SD = 1.60$) and grand-theft auto more with a Black defendant ($M = 6.12$, $SD = 1.90$), $t(41) = -8.73$, $p < .001$ (Figure 6).

Figure 6. Perceived Associations between Defendant Race and Type of Crime
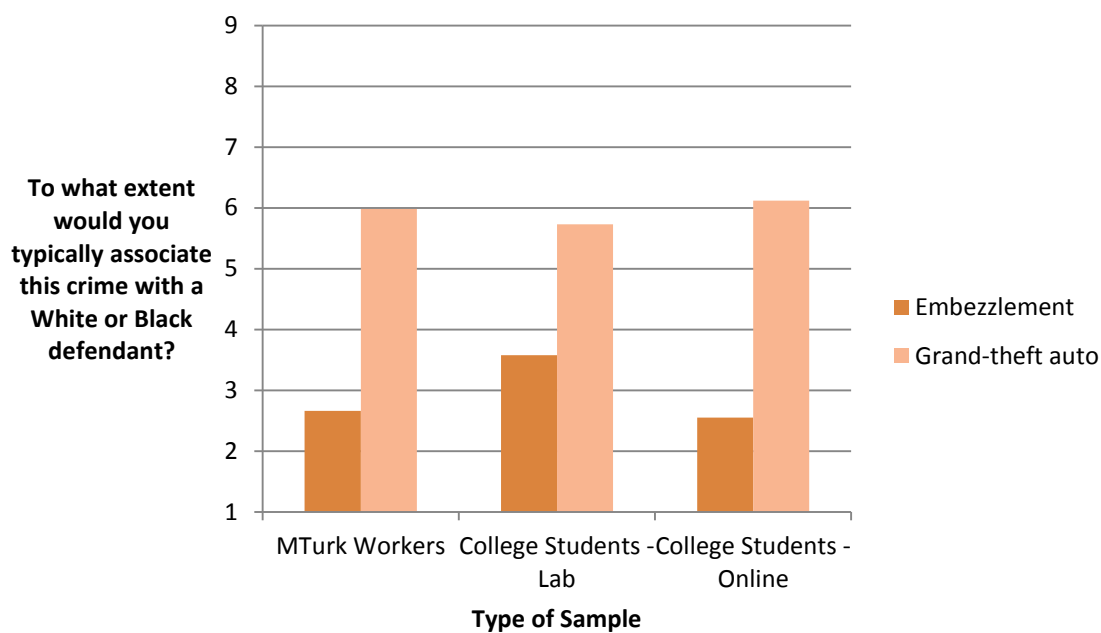
Perceived crime seriousness. In addition, participants responded to "How serious do you think the crime was?" on a 7-point scale ranging from 1 *not serious* at all to 7 *very serious*. I found that MTurk workers believed that the embezzlement case ($M = 5.62$, $SD = 1.13$) depicted a significantly more serious crime than the grand-theft auto case ($M = 5.23$, $SD = 1.33$), $t(85) = 2.98$, $p < .01$. College students in the lab also believed that embezzlement ($M = 5.35$, $SD = 0.92$) was a significantly more serious crime than grand-theft auto ($M = 5.08$, $SD = 0.93$), $t(105) = 2.33$, $p = .02$. However, college students who responded to this question via an online survey indicated that the embezzlement case ($M = 5.17$, $SD = 1.10$) and the grand-theft auto case ($M = 4.88$, $SD = 1.23$) had similar levels of seriousness, $t(41) = 1.27$, $p = .21$ (Figure 7).

Figure 7. Perceived Crime Seriousness
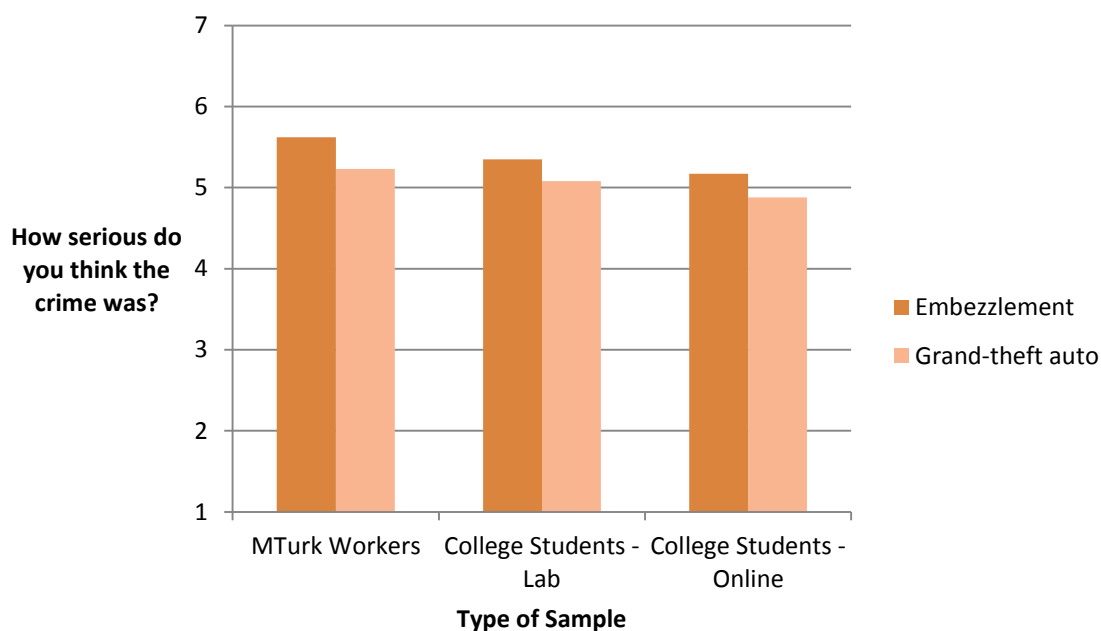
Guilt ratings. Participants also were asked "How guilty does the defendant appear?" They gave their responses on a 10-point scale ranging from 1 *not guilty* to 10 *guilty*. Results showed that MTurk workers believed that the defendant charged with grand-theft auto ($M = 8.22$, $SD = 1.89$) appeared significantly guiltier than the defendant charged with embezzlement ($M = 7.71$, $SD = 2.01$), $t(85) = -2.37$, $p = 02$. On average, college students in the lab also indicated that they thought that the defendant who was charged with grand theft auto ($M = 7.63$, $SD = 1.95$) appeared significantly guiltier than the defendant who was charged with embezzlement ($M = 7.07$, $SD = 1.68$), $t(105) = -2.36$, $p = .02$. However, college students who responded to the question online, did not indicate that the car thief ($M = 7.93$, $SD = 2.11$) appeared guiltier than the embezzler ($M = 7.48$, $SD = 1.80$), $t(41) = -1.25$, $p = .21$ (Figure 8).

Figure 8. Perceived Guilt Ratings
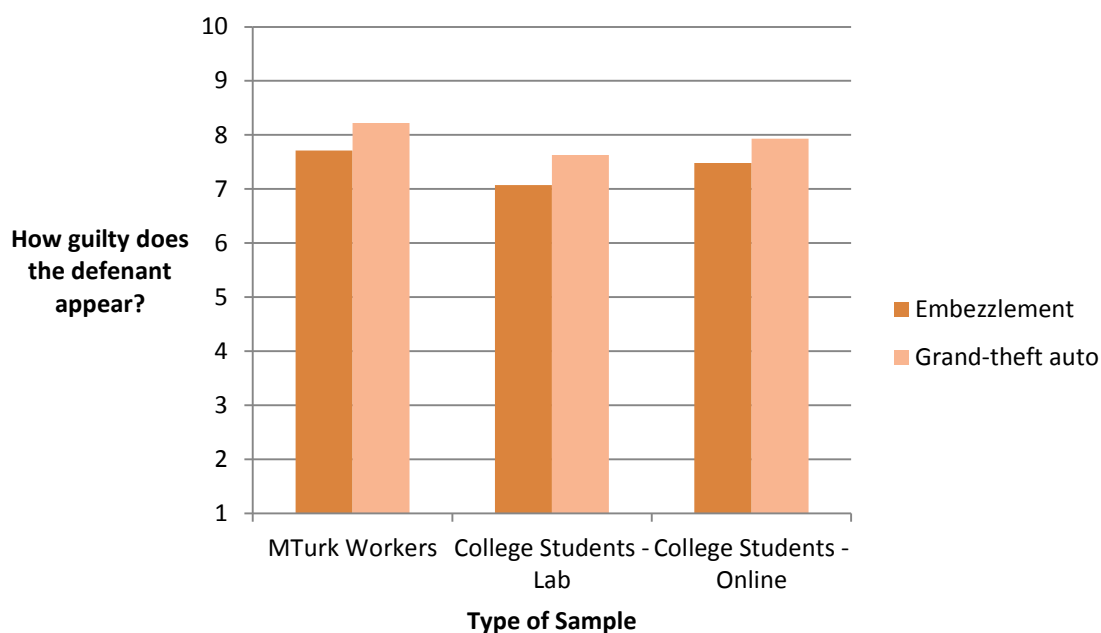
Overall, these results from the pilot tests show that people believe that, generally,

Black individuals are more likely to get charged with grand-theft auto, while White

individuals are more likely to be charged with embezzlement. With regards to the specific

crimes used in these studies, most participants viewed embezzlement as a more serious

crime than grand-theft auto, while the defendant in the grand-theft auto case appeared

guiltier than the defendant in the embezzlement case (independent of race). In an attempt

to hold perceived seriousness as well as perceived guilt constant, and to make the

associations more extreme, slight adjustments were made to both cases.

Reliability Analyses

I calculated Cronbach's alphas to determine the internal consistency of all the

measures (i.e., MNBS-P, MNBS-D, MCPRS, and NCS), and of the subscales of the

MNBS-P and MNBS-D.

The 17 items of the MNBS-P demonstrated a Cronbach's alpha of .86. The analyses also indicated that deleting any of the items would not significantly (i.e., by at least .05) increase the alpha. The three subscales of the MNBS-P, coherence, coverage, and uniqueness, all demonstrated acceptable scale reliabilities ($\geq$ .70)—α = .78, α = .75, α = .86—respectively. Deleting one of items of the coherence subscale (i.e., "If I were given the prosecution story, I would have organized it differently.") would have increased the Cronbach's α to .90. However, because confirmatory analyses had determined that the factor coherence loaded on it, I kept this item as part of the scale for subsequent analyses.

The 17 items of the MNBS-D also demonstrated acceptable reliability with a Cronbach's alpha of .81. In addition, the analyses showed that deleting any of the items would not significantly increased the reliability of the scale. The subscale coherence demonstrated an acceptable level of reliability (α = .80); however, as for the coherence subscale of the MNBS-P, dropping one item would significantly increase the alpha. The subscales coverage and uniqueness did not reach acceptable reliabilities—α = .65 and α = .64—respectively.

In addition, the MCPRS (Motivation to Control Prejudiced Reactions Scale; 17 items) and the NCS (Need for Cognition Scale; 18 items) both exhibited acceptable scale reliabilities —α = .82 and α = .94—respectively.

Manipulation Checks

I conducted an independent samples t-test in order to ensure that neither one of the

of the crimes was perceived as more serious than the other. Results showed that participants perceived embezzlement ($M = 5.36$, $SD = 1.23$) as a significantly more serious crime than grand-theft auto ($M = 5.07$, $SD = 1.11$), $t(297) = -2.06$, $p = .04$.

As expected, results from another independent samples t-test showed that participants typically associated grand-theft auto ($M = 4.19$, $SD = 1.22$) significantly more with a Black defendant than they associated embezzlement with a Black defendant ($M = 2.82$, $SD = 1.42$), $t(297) = 8.94$, $p < .001$.

Descriptive Statistics

Across all conditions, 67.20% of mock jurors found the defendant guilty. Means and standard deviations were computed for all main variables (Table 5).

Table 5. Descriptive Statistics – Study 2

	Mean	Standard Deviation
Verdict-Confidence	2.41	5.00
Probability of Guilt	74.85	24.78
Coherence – Prosecution	43.00	7.75
Coverage – Prosecution	14.12	3.64
Uniqueness – Prosecution	18.38	5.97
Coherence – Defense	37.71	8.49
Coverage – Defense	10.85	3.41
Uniqueness – Defense	14.93	4.23
MCPRS	67.55	14.21
NCS	63.10	14.09

Correlational Analyses

Bivariate correlational analyses were conducted between all variables measuring narrative believability, and guilt. As seen in Table 6, many of the correlations were significant. In addition, all of the significant correlations were in the expected direction. Specifically, the subscales of the MNBS-P were positively correlated, and the subscales of the MNBS-D were positively correlated. In addition, the subscales of the MNBS-P and the subscales of the MNBS-D were negatively correlated. Moreover, the correlation between verdict-confidence ratings and probability of guilt was positive. Finally, the

subscales of the MNBS-P were positively correlated with verdict-confidence as well as with probability of guilty, while the subscales of the MNBS-D were negatively correlated with both guilt measures.

Table 6. Correlations between Variables Measuring Narrative Believability and Verdict Judgments (Study 2)

Variable	1	2	3	4	5	6	7	8
1. Coherence/Prosecution	1.00							
2. Coverage/Prosecution	.66**	1.00						
3. Uniqueness/Prosecution	.34**	.51**	1.00					
4. Coherence/Defense	-.23**	-.37**	-.55**	1.00				
5. Coverage/Defense	-.30**	-.09	-.28**	.60**	1.00			
6. Uniqueness/Defense	-.37**	-.16**	.18**	.09	.28**	1.00		
7. Verdict x Confidence	.33**	.40**	.50**	-.55**	-.34**	-.05	1.00	
8. Probability of Guilt	.38**	.44**	.45**	-.53**	-.39**	-.14*	.83**	1.00

Note: *p < .05 **p < .01

Outcome Measures

In total, I used three different outcome measures in my main analyses – verdict decision (guilty, not guilty), verdict-confidence, and probability of guilt. When a participant decided on a "not guilty" verdict, I coded this verdict decision as -1. In addition, I coded "guilty" verdicts as +1. Then, I multiplied this value by the participant's confidence rating in order to obtain a continuous measure of verdict-confidence. Accordingly, this final measure ranged from -7, meaning that *the participant was very confident that the defendant is not guilty*, to +7, meaning that *the participant was very confident that the defendant is guilty*.

Verdict Decision as the Outcome Variable

I utilized separate logistic regression analyses with three main predictors and one categorical outcome variable (verdict decision: not guilty, guilty) to test the hypotheses that predicted significant two-way interactions and three-way on the verdict decision (hypotheses 2a, 5a, and 8a). Specifically, I tested the hypotheses that whether or not a defendant is found guilty or not depended on whether or not the type of crime is stereotypically congruent with defendant race. In addition, I tested whether this interaction depends on an individual's level of MCPR and on whether or not SSI were received.

Verdict decision as the outcome, with MCPR in the model. In order to determine whether or not a defendant was found guilty, I conducted a simultaneous logistic regression analysis with the verdict variable (dichotomous) as the outcome and with defendant race (dichotomous), type of crime (dichotomous), and motivation to

control prejudiced reactions (MCPR, continuous) as the predictor variables. The interaction terms were included into the model as additional predictors.

The omnibus test of model coefficients indicated that the model was marginally statistically significant, $\chi^2(7) = 13.19$, $p = .07$. This means that the predictor variables as a group predicted the verdict decision variable because it shows that probability of obtaining the chi-square, if there is no effect of all the predictor variables together on the criterion variable.

Defendant race was not a significant predictor for the verdict decision variable, $B = .34$, $\chi^2(1) = .89$, $p = .35$. Crime type also was not a significant predictor for the verdict decision, $B = -.13$, $\chi^2(1) = .12$, $p = .73$. In addition, MCPR did not significantly predict verdict decision, $B = -.01$, $\chi^2(1) = .49$, $p = .48$. There was no significant two-way interactions between type of crime and defendant race, or between defendant race and MCPR on the verdict decision—$B = -.27$, $\chi^2(1) = .27$, $p = .60$; $B = 0.1$, $\chi^2(1) = .31$, $p = .58$, respectively. However, there was a significant interaction between type of crime and MCPR on the verdict decision, $B = .07$, $\chi^2(1) = 5.34$, $p = .02$. To probe this significant interaction, I assessed the effect of type of crime on the verdict decision when MCPR was at one standard deviation above and when MCPR was at one standard deviation below the mean in two additional analyses. Results showed that when mock jurors had low levels of MCRP, type of crime was not as significant predictor of the verdict decision, $B = .36$, $\chi^2(1) = .94$, $p = .33$. However, when mock jurors had high levels of MCPR, type of crime significantly predicted their verdict decision, $B = -.91$,

$\chi^2(1) = 6.63$, $p = .01$. The odds ratio for type of crime was -.60 (.40 − 1 = -.60), which means that when moving from grand-theft auto to embezzlement, there is a 60.00% decrease in the odds that the defendant will be found guilty of embezzlement compared to grand-theft auto. Finally, there was not a significant three-way interaction between defendant race, type of crime, and MCPR, $B = -.04$, $\chi^2(1) = .94$, $p = .33$ (see Appendix X, Table 11).

Verdict decision as the outcome, with SSI in the model. In order to determine whether or not a defendant was found guilty, I conducted a simultaneous logistic regression analysis with the verdict variable (dichotomous) as the outcome and with defendant race (dichotomous), type of crime (dichotomous), and stereotype suppression instructions (SSI, categorical) as the predictor variables. The interaction terms were included into the model as additional predictors.

The omnibus test of model coefficients indicated that the model was not statistically significant, $\chi^2(7) = 3.95$, $p = .79$. This means that the predictor variables as a group did not predict the verdict decision variable because it shows that probability of obtaining the chi-square, if there is no effect of all the predictor variables together on the criterion variable.

Results showed that none of the predictors significantly predicted the verdict decision variable. For a summary of the results, see Appendix X, Table 12.

Verdict decision as the outcome, with NC as a control variable. Although I did not make any specific hypotheses with regards to need for cognition, I included it in the

analyses in order to rule out the possibility that need for condition (NC) accounts for any

of the possible effects. Accordingly, I conducted a simultaneous logistic regression

analysis with the verdict variable (dichotomous) as the outcome, and with defendant race

(dichotomous), type of crime (dichotomous), and NC (continuous) as the predictor

variables. I also included all the interactions terms as additional predictors in the model.

The omnibus test of model coefficients indicated that the model was not

statistically significant, $\chi^2(7) = 4.71$, $p = .68$. This means that the predictor variables as a

group did not predict the verdict decision variable because it shows that probability of

obtaining the chi-square, if there is no effect of all the predictor variables together on the

criterion variable.

Results showed that none of the predictors significantly predicted the verdict

decision variable. For a summary of the results, see Appendix X, Table 13.

Verdict-Confidence Scores as the Outcome

I conducted multiple Analyses of Covariance (ANCOVA) to test the hypotheses

that predicted significant two-way interactions and three-way on verdict-confidence (i.e.,

hypotheses 3a, 6a, and 9a). Specifically, I tested whether how confident a mock juror was

in his or her verdict decision depended on whether race-crime congruence exists, and

whether this interaction was affected by one's level of MCPR and/or the presence of SSI.

Verdict-confidence as the outcome, with MCPR in the model. I conducted a 2

(Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement)

ANCOVA on verdict-confidence scores, controlling for MCPR. Results revealed that

there was no significant main effect of defendant race on verdict-confidence, $F(1, 291) =$ 0.24, $p = .62$. This means that whether the defendant was Black ($M = 2.25$, $SD = 5.10$) or White ($M = 2.53$, $SD = 4.93$) did not affect how confident mock jurors felt about their verdict decision. There was also no significant main effect of type of crime on verdict-confidence, $F(1, 291) = 1.99$, $p = .16$. Whether the charge was embezzlement ($M = 1.94$, $SD = 5.14$) or grand-theft auto ($M = 2.84$, $SD = 4.83$) did not make a difference to mock jurors' verdict confidence scores. However, the covariate, MCPR, was significantly related to the mock jurors' verdict scores, $F(1, 291) = 4.27$, $p = .04$. A scatterplot of the covariate and the outcome showed that as a mock juror's level of MCPR increases, so does his or her confidence in a guilty verdict decision. In contrast to my predictions, the interaction between defendant race and type of crime on verdict-confidence was not significant, $F(1, 291) = -0.03$, $p = .87$ (Figure 9).

Figure 9. Non-significant Interaction between Defendant Race and Type of Crime on Verdict-Confidence (Study 2)
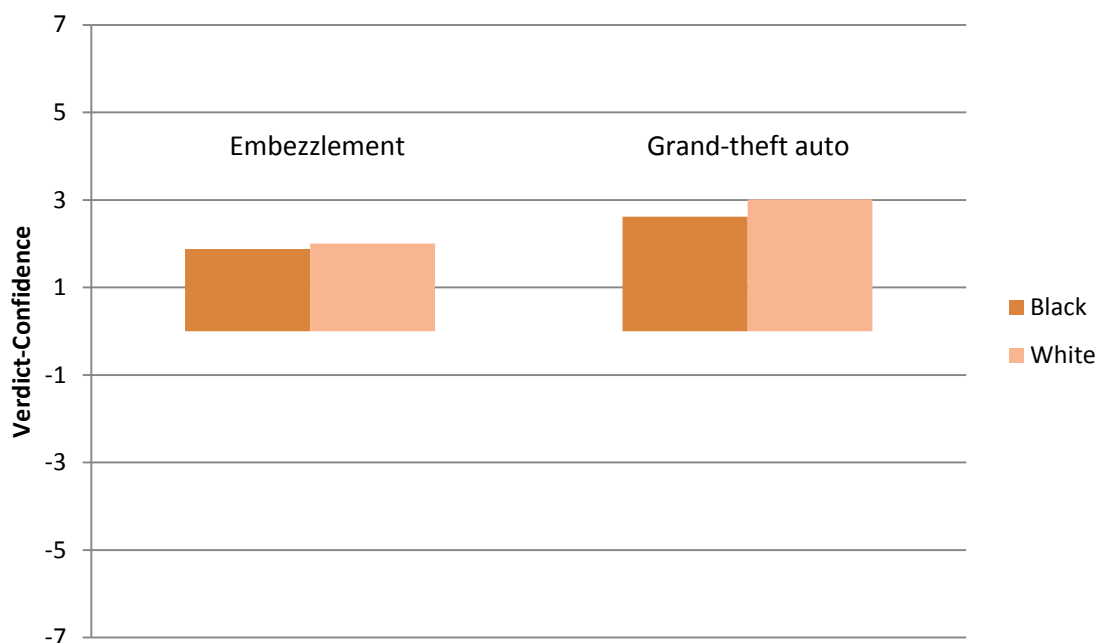
Moreover, there was no significant effect between defendant race and MCPR on verdict-confidence, $F(1, 291) = 0.00$, $p = .99$. However, there was a significant interaction between type of crime and MCPR on verdict-confidence, $F(1, 291) = 6.72$, $p = .01$. To probe this significant interaction, I assessed the effect of type of crime on verdict-confidence when MCPR was at one standard deviation above and when MCPR was at one standard deviation below the mean in two additional analyses. Results from the regression analyses revealed that when mock jurors had low levels of MCRP, type of crime was not a significant predictor of mock juror's verdict-confidence, $B = 0.58$, $\beta = 0.06$, $t(3) = .72$, $p = .47$. However, when mock jurors had high levels of MCPR, type of crime significantly predicted their verdict-confidence, $B = -2.23$, $\beta = -0.23$, $t(3) = -2.79$, $p = .01$. This suggests that for mock jurors with high levels of MCPR, those who read about

an embezzlement case showed were less confident in a guilty verdict that those who read about grand-theft auto. Finally, the three-way interaction between defendant race, type of crime, and MCPR was not significant, $F(1, 291) = 0.92$, $p = .34$.

Verdict-confidence as the outcome, with SSI in the model. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on verdict-confidence scores. Results showed that there were no significant main effects, interactions, or covariates (see Appendix X, Table 14).

Verdict-confidence as the outcome, with NC as a control variable. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on verdict-confidence scores, controlling for NC. Results showed that there were no significant main effects, interactions, or covariates (see Appendix X, Table 15).

Probability of Guilt Ratings

I conducted multiple Analyses of Covariance (ANCOVA) to test the hypotheses that predicted significant two-way and three-way interactions on probability of guilt (i.e., hypotheses 4a, 7a, and 10a). Specifically, I tested whether how probably it is that the defendant is guilty of the crime he was charged with depends on whether the type of crime and the defendant race are stereotypically consistent with one another, and whether this interaction is affected by one's level of MCPR and/or the presence of SSI.

Probability of guilt as the outcome, with MCPR in the model. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement)

ANCOVA on probability of guilt ratings, controlling for MCPR. Results revealed that there was no significant main effect of defendant race on probability of guilt, $F(1, 291) = 0.00$, $p = .97$. There was also no significant main effect of MCPR on probability of guilt, $F(1, 291) = 2.15$, $p = 0.14$. However, there was a significant main effect of type of crime on probability of guilt, $F(1, 291) = 10.09$, $p < .01$. Mock jurors indicated that it is more probably that the defendant is guilty when the charge was grand-theft auto ($M = 79.30$, $SD = 21.59$) compared to when it was embezzlement ($M = 70.12$, $SD = 21.59$). In contrast to my predictions, there was no significant interaction between defendant race and type of crime on probability of guilt ratings, $F(1, 291) = 0.08$, $p = .77$ (Figure 10).

Figure 10. Non-significant Interaction between Defendant Race and Type of Crime on Probability of Guilt (Study 2)
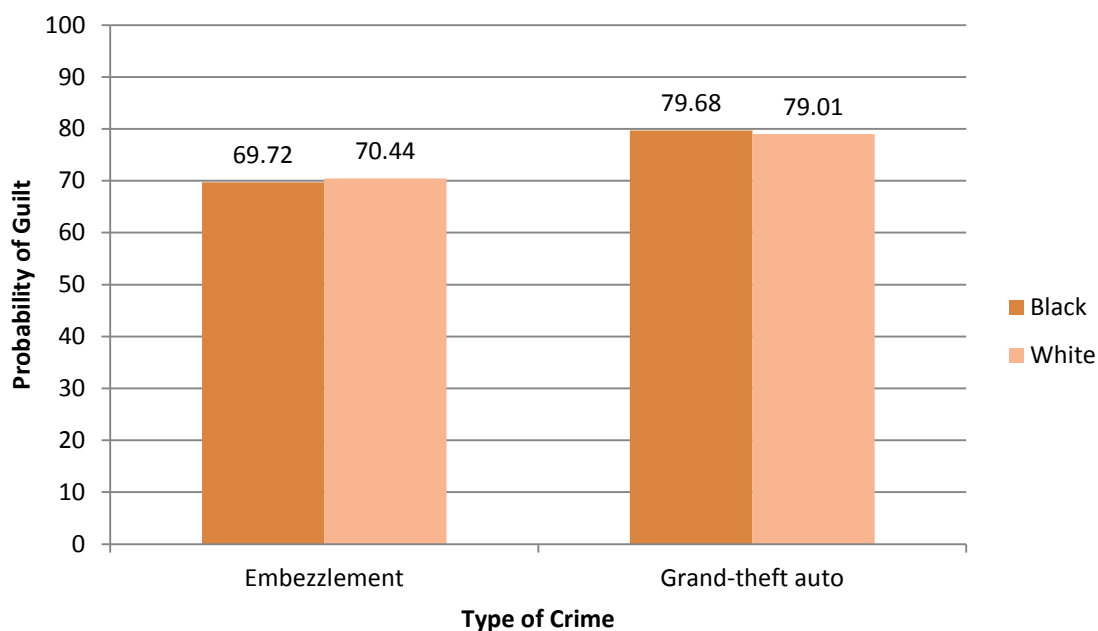
There was a significant interaction effect between type of crime and MCPR on probability of guilt ratings, $F(1, 291) = 5.41$, $p = .02$. As before, to probe this significant interaction, I assessed the effect of type of crime on verdict-confidence when MCPR was at one standard deviation above and when it was at one standard deviation below the mean in two additional analyses. Results from the regression analyses revealed that when mock jurors had low levels of MCRP, type of crime was not a significant predictor of mock juror's probability of guilt ratings, $B = -2.72$, $\beta = -.06$, $t(3) = -0.68$, $p = .50$. However, when mock jurors had high levels of MCPR, type of crime significantly predicted their probability of guilt ratings, $B = -15.19$, $\beta = -.307$, $t(3) = -2.79$, $p < .001$. This suggests that for mock jurors with high levels of MCPR, those who read about an embezzlement case believed it was less probably that the defendant committed the crime than those who read about a grand-theft auto case. In addition, the effects of two-way interaction between defendant race and MCPR and the three-way interaction between defendant race, type of crime, and MCPR on probability of guilt were both not significant—$F(1, 291) = 0.00$, $p = .99$; $F(1, 291) = 1.04$, $p = .31$, respectively.

Probability of guilt as the outcome, with SSI in the model. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on probability of guilt ratings. Results showed that there was a significant main effect of type of crime on probability of guilt, $F(1, 291) = 9.87$, $p = < .01$. On average, mock jurors who read about the grand-theft auto case ($M = 79.30$, $SD = 21.59$) believed that it was more probable that the

defendant is guilty than those who read about the embezzlement case ($M = 70.12$, $SD = 27.06$). All other main effects and interactions were not significant (see Appendix X, Table 16).

Probability of guilt as the outcome, with NC as a control variable. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on probability of guilt ratings, controlling for NC. Results revealed that there was a significant main effect of type of crime on probability of guilt, $F(1, 291) = 10.25$, $p < .01$. When controlling for NC, mock jurors who read about a grand-theft auto case ($M = 79.30$, $SD = 21.59$) believed that it was more probable that the defendant is guilty than mock jurors who read about an embezzlement case ($M = 70.12$, $SD = 27.06$). However, there were no other significant main effect, covariates, or interactions (see Appendix X, Table 17).

Narrative Believability as the Outcome

I predicted that there was a significant interaction between defendant race and type of crime, which then would be moderated by MCPR and SSI. I hypothesized that two-and three-way interactions would be mediated by narrative believability, especially coherence of the prosecution story (hypotheses 2b-10b). Unexpectedly, none of these two- and three-way interactions had significant effects on the verdict judgments in the prior analyses. However, mediation could still occur, if the independent variables were related to the potential mediator. If this was not the case, then there would not be a need to run any mediational analyses. In order to test whether mediational analyses would be warranted, I conducted multiple ANCOVAs with the three factors of narrative

believability, coherence, coverage, and uniqueness, as separate outcome variables. In order to be all-inclusive, I conducted these analyses for the narrative believability of both the prosecution and the defense story.

Coherence of the prosecution story. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on story coherence (MNBS-P), controlling for MCPR. Results showed that there was a significant main effect of type of crime on story coherence, $F(1, 291) = 29.96$, $p < .001$. This means that mock jurors rated the prosecution story for the grand-theft auto case ($M = 45.32$, $SD = 7.44$) as more coherent than the prosecution story for the embezzlement case ($M = 40.53$, $SD = 7.31$). There was not a significant interaction between defendant race and type of crime on story coherence, $F(1, 291) = 0.11$, $p = .89$. All other main effects and interactions were also not significant (see Appendix X, Table 18).

I also conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on story coherence (MNBS-P). Results showed that there was a significant main effect of type of crime on story coherence, $F(1, 291) = 31.41$, $p < .001$. Mock jurors believed that the prosecution story for the grand-theft auto case ($M = 45.32$, $SD = 7.44$) is more coherent than the prosecution story for the embezzlement case ($M = 40.53$, $SD = 7.31$). There was also a significant defendant race x SSI interaction on story coherence, $F(1, 291) = 10.85$, $p < .01$. There were significant simple main effects of defendant race predicting story coherence in both the SSI and the control condition—$F(1, 130) = 4.71$, $p = .03$, $F(1, 165) = 4.85$, $p = .03$, respectively (Figure 11).

Figure 11. Significant Interaction between Defendant Race and SSI on Coherence of the Prosecution Story (Study 2)
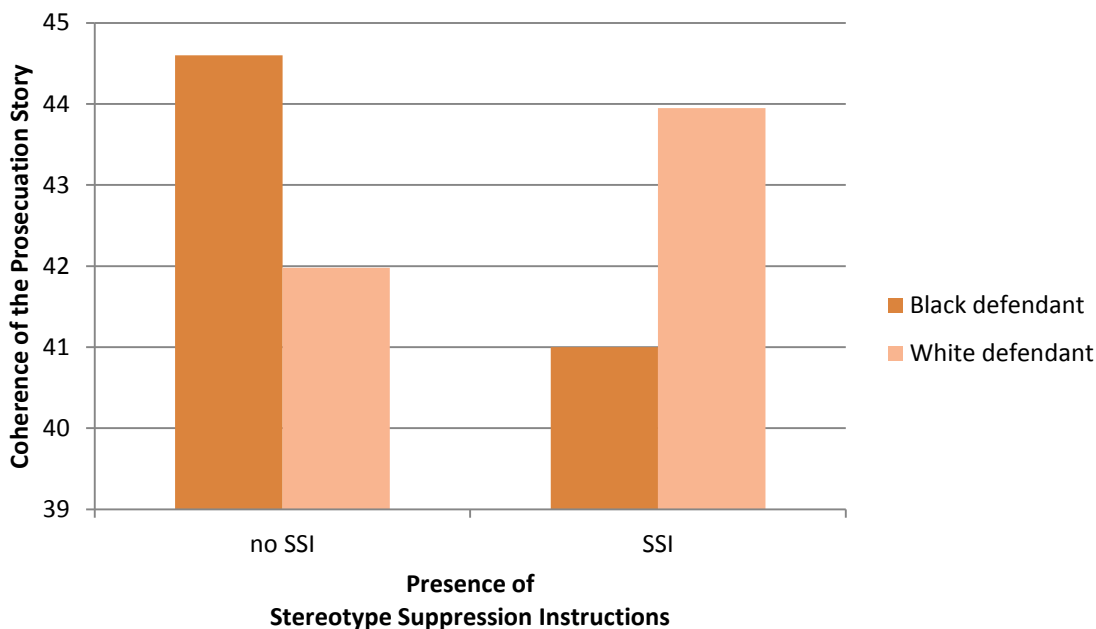
There was not a significant interaction between defendant race and type of crime on story coherence, $F(1, 291) = 0.00$, $p = .998$. All other main effects and interactions were also not significant (see Appendix X, Table 19).

In addition, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime: Grand-theft Auto, Embezzlement) ANCOVA on story coherence (MNBS-P), controlling for mock juror's need for cognition (NC). There was a significant main effect of type of crime on story coherence, $F(1, 291) = 30.99$, $p < .001$. Mock jurors rated the prosecution story for the grand-theft auto case ($M = 45.33$, $SD = 7.44$) as more coherent than the prosecution story for the embezzlement case ($M = 40.53$, $SD = 7.31$). In addition, the covariate, NC was significantly related to story coherence, $F(1, 291) = 4.58$, $p = .03$. A scatterplot of the covariate and the outcome showed that as a mock juror's level of NC

increases, so does his or her ratings of the story coherence of the prosecution story. The interaction effect between defendant race and type of crime on story coherence was not significant, $F(1, 291) = 0.00$, $p = .95$, as were none of the other main effects and interactions (see Appendix X, Table 20).

Because the results showed that defendant race and type of crime did not interact to predict story coherence of the prosecution story, I did not test for mediation.

Coverage of the prosecution story. Next, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on story coverage (MNBS-P), controlling for MCPR. Results showed that there was a significant main effect of type of crime on story coverage, $F(1, 291) = 17.41$, $p < .001$. This means that mock jurors believed that the prosecution story for the grand-theft auto case ($M = 14.95$, $SD = 3.98$) covered more evidence items than the prosecution story for the embezzlement case ($M = 13.25$, $SD = 3.02$). There was not a significant interaction between defendant race and type of crime on story coverage, $F(1, 291) = 0.94$, $p = .33$. All other main effects and interactions were also not significant (see Appendix X, Table 21).

In addition, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on story coverage (MNBS-P). Results showed that there was a significant main effect of type of crime on story coverage, $F(1, 291) = 19.05$, $p < .001$. Mock jurors believed that the prosecution story for the grand-theft auto case ($M = 14.95$, $SD = 3.98$) covered more evidence items than the prosecution story for the embezzlement case ($M =$

13.25, $SD = 3.02$). Moreover, there was not a significant interaction between defendant race and type of crime on story coverage, $F(1, 291) = 1.09$, $p = .30$. All other main effects and interactions were also not significant (see Appendix X, Table 22).

I also I conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime: Grand-theft Auto, Embezzlement) ANCOVA on story coverage (MNBS-P), controlling for mock juror's need for cognition (NC). There was a significant main effect of type of crime on story coverage, $F(1, 291) = 18.13$, $p < .001$. Mock jurors believed that the prosecution story for the grand-theft auto ($M = 14.95$, $SD = 3.98$) has higher coverage than the prosecution story for the embezzlement case ($M = 13.25$, $SD = 3.02$). The interaction effect between defendant race and type of crime on story coverage was not significant, $F(1, 291) = 0.64$, $p = .42$, as were none of the other main effects and interactions (see Appendix X, Table 23).

Because the results showed that defendant race and type of crime did not interact to predict story coverage of the prosecution story, there was no need to test for mediation.

Uniqueness of the prosecution story. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on story uniqueness (MNBS-P), controlling for MCPR. Results revealed that there was a significant main effect of type of crime on story uniqueness, $F(1, 291) = 18.53$, $p < .001$. On average, mock jurors believed that the prosecution story for the grand-theft auto case ($M = 19.79$, $SD = 6.43$) was more unique than the prosecution story for the embezzlement case ($M = 16.89$, $SD = 5.04$). There was not a significant interaction between defendant race and type of crime on story uniqueness, $F(1, 291) = 0.40$, $p = .53$. All other main

effects and interactions were also not significant (see Appendix X, Table 24).

I also conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on story uniqueness (MNBS-P). Results showed that there was a significant main effect of type of crime on story uniqueness, $F(1, 291) = 17.55$, $p < .001$. Mock jurors believed that the prosecution story for the grand-theft auto case ($M = 19.79$, $SD = 6.43$) was more unique than the prosecution story for the embezzlement case ($M = 16.89$, $SD = 5.04$). Moreover, there was not a significant interaction between defendant race and type of crime on story uniqueness, $F(1, 291) = 0.18$, $p = .67$. All other main effects and interactions were also not significant (see Appendix X, Table 25).

Furthermore, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime: Grand-theft Auto, Embezzlement) ANCOVA on story uniqueness (MNBS-P), controlling for mock juror's need for cognition (NC). There was a significant main effect of type of crime on story uniqueness, $F(1, 291) = 19.26$, $p < .001$. Mock jurors believed that the prosecution story for the grand-theft auto case ($M = 19.79$, $SD = 6.43$) was more unique than the prosecution story for the embezzlement case ($M = 16.89$, $SD = 5.04$). The interaction effect between defendant race and type of crime on story uniqueness was not significant, $F(1, 291) = 0.54$, $p = .46$. There were no other main effects or interactions (see Appendix X, Table 26).

Because the results showed that defendant race and type of crime did not interact to predict story coverage of the prosecution story, there was no need to test for mediation.

Coherence of the defense story. I conducted a 2 (Defendant Race: Black, White)

x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on story coherence

(MNBS-D), controlling for MCPR. Results revealed that there was a significant main

effect of type of crime on story coherence, $F(1, 291) = 38.48$, $p < .001$. This means that

mock jurors rated the defense story for the embezzlement case ($M = 40.73$, $SD = 7.38$) as

more coherent than the defense story for the grand-theft auto case ($M = 34.86$, $SD =$

8.50). There was not a significant interaction between defendant race and type of crime

on story coherence, $F(1, 291) = 0.08$, $p = .78$. All other main effects and interactions

were not significant (see Appendix X, Table 27).

I also conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-

theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA

on story coherence (MNBS-D). Results showed that there was a significant main effect of

type of crime on story coherence, $F(1, 291) = 38.70$, $p < .001$. Mock jurors believed that

the defense story for the embezzlement case ($M = 40.73$, $SD = 7.38$) is more coherent

than the defense story for the grand-theft auto case ($M = 34.86$, $SD = 8.50$). No other

main effects or interactions had significant effects on story coherence (see Appendix,

Table 28).

Moreover, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime:

Grand-theft Auto, Embezzlement) ANCOVA on story coherence (MNBS-D), controlling

for mock juror's need for cognition (NC). There was a significant main effect of type of

crime on story coherence, $F(1, 291) = 37.27$, $p < .001$. Mock jurors rated the defense

story for the embezzlement case ($M = 40.73$, $SD = 7.38$) as more coherent than the

defense story for the grand-theft auto case ($M = 34.86$, $SD = 8.50$). The interaction effect

between defendant race and type of crime on story coherence was not significant, $F(1,$
$291) = 0.04$, $p = .85$, as were none of the other main effects and interactions (see
Appendix X, Table 29).

Because the results showed that defendant race and type of crime did not interact
to predict story coherence of the defense story, I did not test for mediation.

Coverage of the defense story. Next, I conducted a 2 (Defendant Race: Black,
White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on story
coverage (MNBS-D), controlling for MCPR. Results revealed that there was a significant
main effect of type of crime on story coherence, $F(1, 291) = 33.63$, $p < .001$. This means
that mock jurors believed that the defense story for the embezzlement case ($M = 11.98$
$SD = 3.12$) covered more evidence items than the prosecution story for the grand-theft
auto case ($M = 9.78$, $SD = 3.33$). There was not a significant interaction between
defendant race and type of crime on story coverage, $F(1, 291) = 0.46$, $p = .85$. All other
main effects and interactions were also not significant (see Appendix X, Table 30).

In addition, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime:
Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes)
ANCOVA on story coverage (MNBS-D). Results showed that there was a significant
main effect of type of crime on story coverage, $F(1, 291) = 30.92$, $p < .001$. Mock jurors
believed that the defense story for the embezzlement case ($M = 11.98$, $SD = 3.13$)
covered more evidence items than the prosecution story for the grand-theft auto case (M
$= 9.78$, $SD = 3.24$). Moreover, there was not a significant interaction between defendant
race and type of crime on story coverage, $F(1, 291) = .08$, $p = .77$. All other main effects

and interactions were also not significant (see Appendix X, Table 31).

I also conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime: Grand-theft Auto, Embezzlement) ANCOVA on story coverage (MNBS-D), controlling for mock juror's need for cognition (NC). There was a significant main effect of type of crime on story coverage, $F(1, 291) = 31.16$, $p < .001$. Mock jurors believed that the defense story for the embezzlement case ($M = 11.98$, $SD = 3.13$) covered more evidence items than the defense story for the grand-theft auto case ($M = 9.78$, $SD = 3.33$). The interaction effect between defendant race and type of crime on story coverage was not significant, $F(1, 291) = 0.10$, $p = .75$. Moreover, there was a significant three-way interaction between defendant race, type of crime, and NC on story coverage, $F(1, 291) = 5.26$, $p = .02$. There were no other significant main effects or interactions on story coverage (see Appendix X, Table 32).

Because the results showed that defendant race and type of crime did not interact to predict story coverage of the defense story, mediational analyses would have been redundant.

Uniqueness of the defense story. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on story uniqueness (MNBS-D), controlling for MCPR. Results revealed that the main effect of MCPR was significant, $F(1, 291) = 4.21$, $p = .04$. A scatterplot of MCPR and uniqueness of the defense story showed that as a mock juror's level of MCPR increases, so does his or her ratings of story uniqueness. In addition, there was not a significant interaction between defendant race and type of crime on story uniqueness, $F(1, 291) = 0.09$, $p = .77$.

All other main effects and interactions were also not significant (see Appendix X, Table 33).

I also conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on story uniqueness (MNBS-D). Results showed that there was a significant main effect of defendant race on uniqueness ratings of the defense story, $F(1, 291) = 4.25$, $p = .04$. Mock jurors rated the defense story as more unique when the defendant was Black ($M = 15.45$, $SD = 4.53$) compared to when he was White ($M = 14.52$, $SD = 3.95$). Moreover, there was not a significant interaction between defendant race and type of crime on story uniqueness, $F(1, 291) = 0.10$, $p = .75$. None of the other main effects or interactions was significant (see Appendix X, Table 34).

Finally, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime: Grand-theft Auto, Embezzlement) ANCOVA on story uniqueness (MNBS-D), controlling for mock juror's need for cognition (NC). The covariate, NC was significantly related to the mock juror's uniqueness ratings of the defense story, $F(1, 291) = 7.05$, $p = .01$. NC was negatively related to story uniqueness; as mock jurors' levels of NC increased, their uniqueness ratings of the defense story decreased. The interaction effect between defendant race and type of crime on story coverage was not significant, $F(1, 291) = 0.26$, $p = .61$. There were no other main effects or interactions (see Appendix X, Table 35).

Because the results showed that defendant race and type of crime did not interact to predict story uniqueness of the defense story, there was no need to test for mediation.

Discussion

Findings from the current study showed that the type of crime determined how probable mock jurors thought it was that the defendant was guilty of the crime he was charged with. Specifically, mock jurors who read the grand-theft auto case believed that it was more probable that the defendant was guilty compared to mock jurors who read the embezzlement case.

Mock jurors who read about defendant charged with grand-theft auto also thought that the story which the prosecution presented was more coherent, more unique, and covered more of the evidence items. In addition, they rated the defense story for the grand-theft auto case as less coherent, and indicated that it covered fewer of the evidence items. The type of crime especially predicted mock jurors' verdict decisions, their confidence in that decision, and their probability of guilty ratings when they had high levels of MCPR. Specifically, when they were motivated to control their prejudiced reactions, grand-theft auto led to more guilty verdict decisions, higher confidence, and higher probability of guilt ratings, compared to embezzlement.

Although these results did not support any of my hypotheses, they were in line with my findings from the pilot study. Because the embezzlement case was seen as more serious than the grand-theft auto case, participants may have been more cautious in finding the defendant who was charged with embezzlement guilty because it would lead to more severe consequences (e.g., longer sentence). In addition, results from the pilot study and manipulation checks showed that the defendant who was charged with grand-theft auto appeared guiltier than the defendant that was charged with embezzlement.

Thus, assuming that participants in the pilot study were representative of the participants of the Study 2, then it is no surprise that grand-theft auto evoked harsher verdict judgments, and the prosecution story was more believable.

Contrary to my predictions, racially stereotypically consistent crimes did not produce stronger verdict judgments. Specifically, White jurors did not find Black defendants charged with grand-theft auto (a stereotypically consistent crime) more often guilty, were not more confident in their guilty verdict decisions, and also did not believe it was more probable that the defendant was guilty than when mock jurors made their verdict judgments about a White person who was charged with grand-theft auto (a stereotypically inconsistent crime). Similarly, White jurors did not make higher guilt ratings about a White defendant who was charged with embezzlement than a Black defendant charged with the same crime. These results were surprising because they did not replicate findings from past studies (e.g., Jones & Kaplan, 2003). In addition, I had predicted that those crimes which are more consistent with a racial stereotype would lead jurors to believe that the story presented at trial by the prosecution is more coherent than a story which tries to incorporate a defendant whose racial stereotype does not fit the crime he was a charged with. However, unexpectedly, racially stereotypically consistent crimes did not produce higher coherence ratings of the prosecution story than racially stereotypically inconsistent crimes.

One reason for why the consistency between defendant race and type of crime may not have affected verdict judgments and coherence ratings is that over the past decades jurors' attitudes may have been changing in a positive direction. It could be that

jurors no longer cognitively map the crime onto the defendant's race, and evaluate whether or not the story "fits" with regards to racial and crime stereotypes. Although this would be an extremely encouraging development for the civil rights movement in the United States, this interpretations is doubtful. Past studies have shown over and over again that White jurors find Black defendants more frequently guilty and sentence them longer than White defendants (Hetey & Eberhardt, 2014). Granted that the disproportionality between African American and White prison inmates can be attributed to a variety of different factors, jurors' perceptions of and attitudes toward the defendants on trial undoubtedly still play a role.

It is more likely that the present findings are due to the research methodology that was employed, and the sample of study participants completing the experiment. MTurk workers may be more susceptible to picking up on certain cues about what the studies they are participating in are trying to investigate because they have much practice in completing experiments online. Because of a lack of restrictions in the online environment of Amazon's MTurk, workers are highly likely to complete many similar studies, and even to participate in slight modifications of the same experiment (Chandler et al., 2013). Thus, participant reactance, instead of changes in race and crime stereotypes, may be responsible for the current findings. Just seeing a picture of a Black defendant might trigger the MTurk worker to believe that the researchers are interested in the effects of racial prejudice, just because they have been exposed to similar manipulations in other participants they had participated in prior to the current study. This assumption is supported by the fact that more than twice as many participants

guessed the general purpose of the study when they saw a picture of a Black defendant than compared to those that saw a picture of a White defendant in the current experiment. Although I guaranteed anonymity, the study participants may not have wanted to appear prejudicial, and thus, they responded to certain questions in ways that may be uncharacteristic of their actual feelings and attitudes, as soon as they realize or guess that their racial stereotypes are being examined. Unfortunately, this then can lead to those null effects that were found in the present study.

Because defendant race did not interact in predicting culpability judgments, my hypotheses regarding coherence explaining those effects could not be tested. However, it is worth noting that the all three subscales of the MNBS-P, coherence, coverage, and uniqueness turned out to be reliable measurement scale of the narrative believability of prosecution story. In addition, while the subscale coherence of the MNBS-D reliable measured the construct, the coverage and uniqueness subscales fell somewhat short of being reliable measures.

CHAPTER 7

STUDY 3

Overview

The purpose of Study 3 was to replicate the race-crime congruence effects on

culpability judgments from past research (e.g., Kaplan & Jones, 2003), and to test the

same hypotheses as in Study 2 with a student sample. Furthermore, the goal to measure

story coherence in a few different ways that are more in line with what Pennington and

Hastie (1984, 1986, 1988) actually posited. They suggested that it is not so much about

how coherent jurors believe the prosecution and defense stories are. Instead, jurors

construct their own stories during and after the trial, and then evaluate their individual

stories in terms of coherence (as well as coverage and uniqueness). And the degree to

which their stories are coherent then determines whether a juror believes that the

defendant is guilty or not. Thus, instead of asking participants to just rate the coherence

of the prosecution and defense stories, I asked them to develop and rate their own stories.

In addition, in order to get more objective ratings of the participants' stories, each story

was coded for coherence, coverage, and uniqueness.

As in Study 2, participants were randomly assigned to read about a White or

Black defendant (race conditions) who was charged with either grand-theft auto or

embezzlement (type of case conditions). Half of the participants received SSI, while the

other half did not receive this type of instruction. Participants wrote their own stories

107

about what they think would be an accurate description of what happened in the events leading to the trial. Then, participants rated their own stories in terms of narrative believability. Afterwards, they provided culpability judgments, and filled out the MCPRS and NCS. Coders who were blind to the conditions rated the participants' stories in terms of narrative believability after data collection was completed.

As for Study 2, my predictions for Study 3 were for White jurors only, and I did not make specific predictions concerning the main effect of race.

I expected the same findings for Study 3 as I do for Study 2. The only difference was that participants' ratings of the case summaries did not determine story coherence. Instead, participants' ratings of the narrative believability of their own stories as well as unbiased coders' ratings of the narrative believability of their stories were used independently as measures of story coherence.

Method

Participants

For study 3, my goal was to recruit at least 230 undergraduate students via the university participant pool. According to the prospective power analysis that was described above (see Study 2 Participants), a sample size of 200 was needed to detect significant two-way and three-way interactions with 80% power. Instead of using a very conservative estimate of the sample size for the meditational effect, I made a more liberal estimate for this study due to the difficulty of obtaining large samples when conducting studies in the laboratory in a reasonable amount of time. Accordingly, approximately 150 participants were needed to achieve 80% power, if the path coefficients were

approximately 0.26 (Fritz & MacKinnon, 2007). In order to account for data that may have to be omitted due to failing to pass manipulation checks, or too much missing data, I planned to collect an additional 15% of participants, for a total sample size of 173. The inclusion criteria were the same as in study 2.

In total, data were collected from 291 undergraduate students who receive class credit for partaking in the study. I excluded 36.43% of the participants from the final data file because they indicated that their ethnicity was different from "European American/White" (see Appendix Y, Figure 21). Of the remaining 185 White participants, all of them were citizens of the United States. I omitted the data for those that did not pass both attention checks ($n = 9$), failed the manipulation check for crime type ($n = 11$), and those who failed the manipulation check for defendant race ($n = 9$). The final data set included 175 White participants, of which 81.71% were female. The majority of participants were freshmen in college (see Figure 12). Their ages ranged between 17 and 28 ($M = 18.70$, $SD = 1.19$).

Figure 12. Study 3 – Year in School
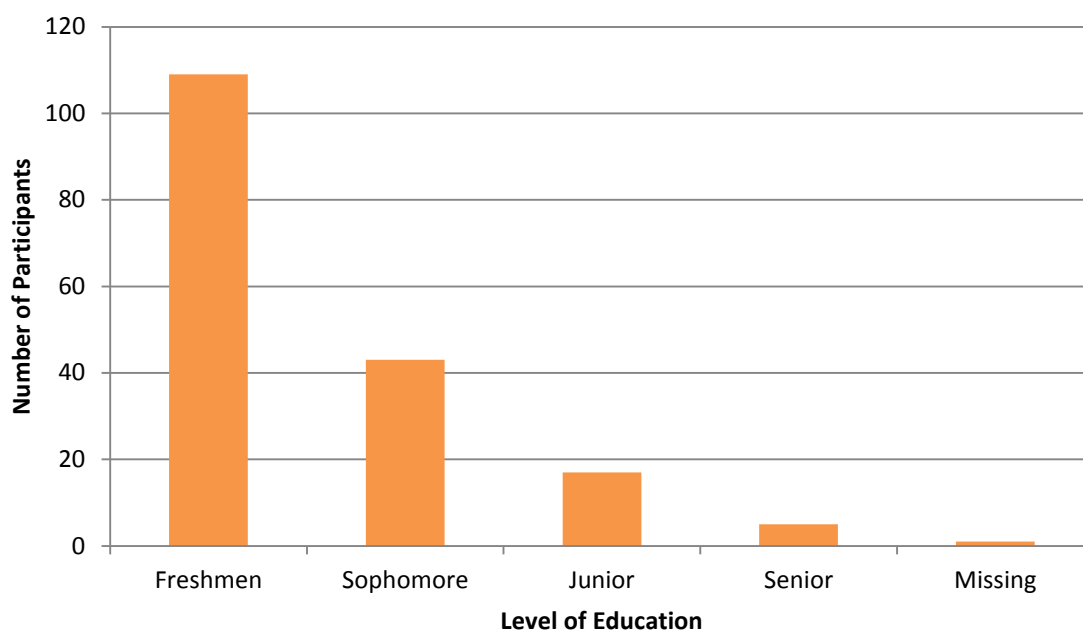
Materials

Study 3 included the same materials as study 2 with the exception of the MNBS-P and MNBS-D. Instead, participants were asked to rate their own stories by using the Modified Narrative Believability Scale-Jurors (MNBS-J). This scale included the same items as the other two scales. However, the wording was changed from "prosecution story" (or "defense story") to "your story." In addition, three items were modified, so these would make more sense in terms of mock jurors' ratings their own stories. For example, "If I were to write my own story again, I would organize it differently" replaced "If I were given the prosecution [defense] story, I would have organized it differently" (see Appendix T). As I explained in more in detail below, independent raters assessed the narrative believability of the participants' stories with the Modified Narrative

Believability Scale-Raters (MNBS-R). Therefore, the items used "the story" instead of

the "prosecution story"/"defense story"/"your story" (see Appendix U).

The demographic questionnaire included age, gender, ethnicity, and year in

school. Participants also indicated whether they are a United States citizen (see Appendix

V).

Procedure

I employed the software program MediaLab to conduct Study 3. This method of

data collection had several advantages. First, using computer software contribute to a

speedy data collection because six to seven students were able to participate in the

experiment at the same time, whereas a maximum number of four participants would

have fitted into a regular-sized lab room, if I had used a standard paper-and-pencil

format. Second, because participants were sitting in individual cubicles, it is very likely

that their feelings of anonymity increased. This may have reduced evaluation

apprehension, especially when participants were filling out the verdict measures and the

MCPRS. Finally, when creating their own stories (as was explained in more detail

below), participants were able to type up their stories, which was much easier and less

time-consuming than handwriting the responses; in addition, handwriting could have led

to participants writing shorter and lower quality stories because it is more effortful. The

participants followed the instructions, completed the tasks, and responded to the question

on individual computer screens.

The procedure for Study 3 was the same as for Study 2. However, in addition to

participants rating the prosecution and defense story with the MNBS-P and MNBS-D,

they were also asked to create their own stories based on the case summary they read.

The instructions were as follows:

> Jurors usually have their own story in mind of what happened in the events leading to trial. We are interested in your story. Please think about the case summary you just read. Based on the information presented in this summary, try to put all the pieces of evidence together. Please come up with a description of what do you think may truly have happened in the events leading to the trial. In addition, with your story, try to answer the question of why the defendant should be found guilty or not guilty. You description should explain the evidence and ultimately lead to your verdict decision. Please be as specific as possible in your story.

Participants were then asked to rate the narrative believability of their own stories with the MNBS-J. As in Study 2, participants made their verdict decisions, responded to questions that probed for their suspicion, filled out the MCPRS and Need for Cognition Scale, answered demographic questions, and responded to manipulation check items. In order to obtain relative objective ratings of the narrative believability of the stories, independent raters who were not aware of each participant's condition rated the narrative believability of each story with the MNBS-R.

Results

Reliability Analyses

I calculated Cronbach's alphas for all the standardized measures (i.e., MNBS-P, MNBS-D, MNBS-J, MNBS-R, MCPRS, and NCS), and of the subscales of the all four scales that measure narrative believability, to determine consistency of each scale.

The 17 items of the MNBS-P demonstrated a Cronbach's alpha of .85. The analyses also indicated that deleting any of the items would not significantly (i.e., by at least .05)

increase the alpha. The three subscales of the MNBS-P, coherence, coverage, and uniqueness, all demonstrated acceptable scale reliabilities ($\geq$ .70)—α = .76, α = .70, α = .84—respectively. Deleting one of items of the coherence subscale (i.e., "If I were given the prosecution story, I would have organized it differently.") would increase the Cronbach's α to .88. However, because confirmatory analyses had determined that the factor coherence loaded on that item, I kept this item as part of the scale for subsequent analyses.

The 17 items of the MNBS-D also demonstrated acceptable reliability with a Cronbach's alpha of .77. Moreover, the analyses showed that deleting any of the items would not significantly have increased the reliability of the scale. The subscale coherence demonstrated an acceptable level of reliability (α = .73); however, as for the coherence subscale of the MNBS-P, dropping one item would significantly increase the alpha. The subscales coverage and uniqueness did not reach acceptable reliabilities—α = .63 and α = .68—respectively.

Furthermore, the 17 items of the MNBS-J showed acceptable reliability, Cronbach's alpha = .78. As before, dropping that one particular item would significantly increase the reliability of the scale (to .85). The subscale coherence displayed an acceptable level of reliability (α = .77); again, dropping an item would increase the alpha. The subscales coverage and uniqueness reached acceptable reliabilities—α = .77 and α = .76, respectively.

Two independent raters coded a subsample of the stories that participants wrote

using the MNBS-R, and I established interrater reliability (see Appendix Y, Table 36). Then, we conducted reliability analyses, which showed that the 17 items of the MNBS-R were reliably measuring narrative believability, Cronbach's alpha = .79. Deleting one of the items (i.e., "If I were given the prosecution story, I would have organized it differently.") would increase the scale to α = .88. The coherence subscale did not read an acceptable level of reliability, α = .53. However, the coverage subscale did, α = .95.

While the MCPRS (Motivation to Control Prejudiced Reactions Scale; 17 items) did not exhibited an acceptable alpha, (α = .68), the NCS (Need for Cognition Scale; 18 items) represented a reliable measurement scale, α = .88.

Manipulation Checks

As in Study 2, I conducted an independent samples t-test in order to ensure that neither one of the crimes was perceived as more serious than the other. Results showed that participants perceived embezzlement (M = 4.59, SD = 1.09) as a significantly more serious crime than grand-theft auto (M = 4.19, SD = 1.21), $t(173)$ = -2.28, p = .02.

As expected, results from another independent samples t-test showed that participants typically associated grand-theft auto (M = 5.60, SD = 1.17) significantly more with a Black defendant than they associated embezzlement with a Black defendant (M = 4.78, SD = 1.53), $t(173)$ = 4.13, $p < .001$.

Descriptive Statistics

Across all conditions, 69.70% of mock jurors found the defendant guilty. The means and standard deviations of all main variables are displayed in Table 7.

Table 7. Descriptive Statistics – Study 3

	Mean	Standard Deviation
Verdict-Confidence	2.37	4.80
Probability of Guilt	70.70	19.35
Coherence – Prosecution	39.18	7.69
Coverage – Prosecution	12.64	3.20
Uniqueness – Prosecution	16.19	5.15
Coherence – Defense	36.90	7.04
Coverage – Defense	10.41	3.24
Uniqueness – Defense	14.55	4.12
Coherence – Juror	39.38	8.29
Coverage – Juror	10.76	3.91
Uniqueness – Juror	14.55	4.93
Coherence – Rater	39.78	6.37
Coverage – Rater	11.56	5.06
MCPRS	70.31	10.71
NCS	61.49	11.53

Correlational Analyses

Bivariate correlational analyses were conducted between all variables measuring narrative believability (4 scales), and guilt on a continuum (2 scales). As seen in Table 8,

many of the correlations were significant. In addition, all of the significant correlations were in the expected direction. When subscales of a scale had significant relationships, they were correlated in a positive direction. In addition, all subscales of the MNBS-P and the uniqueness subscale of the MNBS-J were positively and coherence of the MNBS-D was negatively c correlated with probability of guilt. Only the coverage and uniqueness subscales of the MNBS-P were correlated with verdict-confidence. None of the two subscales of the MNBS-R showed significant relationship with the verdict judgments.

Table 8. Correlations between Variables Measuring Narrative Believability and Verdict Judgments (Study 3)

	Verdict Confidence	Guilt Prob.	MNBS-P COH	MNBS-P COV	MNBS-P UNQ	MNBS-D COH	MNBS-D COV	MNBS-D UNQ	MNBS-J COH	MNBS-J COV	MNBS-J UNQ	MNBS-R COH	MNBS-R COV
Verdict Confidence	1.0	---	---	---	---	---	---	---	---	---	---	---	---
Guilt Prob.	.74**	1.0	---	---	---	---	---	---	---	---	---	---	---
MNBS-P COH	.13	.26**	1.0	---	---	---	---	---	---	---	---	---	---
MNBS-P COV	.22*	.28**	.64**	1.0	---	---	---	---	---	---	---	---	---
MNBS-P UNQ	.26*	.31*	.36**	.48**	1.0	---	---	---	---	---	---	---	---
MNBS-D COH	-.12	-.24**	-.20**	-.36**	-.37**	1.0	---	---	---	---	---	---	---
MNBS-D COV	-.01	-.13	-.21*	-.02	-.10	.50**	1.0	---	---	---	---	---	---
MNBS-D UNQ	.14	.04	-.14	.05	.44**	.09	.34**	1.0	---	---	---	---	---
MNBS-J COH	.09	.14	.24**	.07	-.10	.04	-.09	-.29**	1.0	---	---	---	---
MNBS-J COV	-.04	.07	.07	.23*	.01	-.04	.07	-.22*	.49*	1.0	---	---	---
MNBS-J UNQ	.12	.15*	.06	.27*	.55**	-.27**	.07	.35**	.06	.20**	1.0	---	---
MNBS-R COH	-.04	.09	.34**	.27**	.14	-.09	-.08	-.10	.36**	.35**	.16**	1.0	---
MNBS-R COV	-.05	.08	.19*	.18*	.20*	-.09	-.07	-.02	.24**	.33**	.12	.69**	1.0

Note: *p < .05 **p < .01

Outcome Measures

As for Study 2, I used three different outcome measures in my main analyses −
verdict decision (guilty, not guilty), verdict-confidence, and probability of guilt. I created
the verdict-confidence measure by multiplying either -1 (i.e., "not guilty verdict
decision") or +1 (i.e., "guilty verdict decision") by the mock juror's confidence rating.
Thus, this final measure ranged from -7, meaning that *the participant was very confident
that the defendant was not guilty*, to +7, meaning that *the participant was very confident
that the defendant was guilty*.

Verdict Decision as the Outcome Variable

I conducted separate logistic regression analyses with three main predictors and
one categorical outcome variable (verdict decision: not guilty, guilty) to test the
hypotheses that predicted significant two-way interactions and three-way on the verdict
decision (hypotheses 2a, 5a, and 8a). Specifically, I tested the hypotheses that whether or
not a defendant is found guilty or not depended on whether or not the type of crime is
stereotypically congruent with defendant race. In addition, I tested whether this
interaction depends on an individual's level of MCPR and whether or not SSI were
received.

Verdict decision as the outcome, with MCPR in the model. In order to
determine whether or not a defendant was found guilty, I conducted a simultaneous
logistic regression analysis with the verdict variable (dichotomous) as the outcome and
with defendant race (dichotomous), type of crime (dichotomous), and motivation to
control prejudiced reactions (MCPR, continuous) as the predictor variables. All

interaction terms were included into the model as additional predictors.

The omnibus test of model coefficients indicated that the model was statistically significant, $\chi^2(7) = 19.07$, $p = .01$. This means that the predictor variables as a group predicted the verdict decision variable.

Crime type was a significant predictor of the verdict decision, $B = 1.15$, $\chi^2(1) = 4.53$, $p = .03$. In addition, MCPR did not significantly predict verdict decision, $B = -.01$, $\chi^2(1) = .49$, $p = .48$.. The odds ratio for type of crime was 3.15, which means that when moving from grand-theft auto to embezzlement, mock jurors are 3.15 times more likely to find the defendant guilty. No other predictors were significant (see Appendix Y, Table 37).

Verdict decision as the outcome, with SSI in the model. In order to determine whether or not a defendant was found guilty, I conducted a simultaneous logistic regression analysis with the verdict variable (dichotomous) as the outcome and with defendant race (dichotomous), type of crime (dichotomous), and stereotype suppression instructions (SSI, categorical) as the predictor variables. The interaction terms were included into the model as additional predictors.

Results showed that none of the predictors significantly predicted the verdict decision variable. For a summary of the results, see Appendix Y, Table 38).

Verdict decision as the outcome, with NC as a control variable. I conducted one additional analysis for predicting the verdict decision variable in order to rule out the possibility that need for condition (NC) accounts for any of the possible effects.

Specifically, I conducted a simultaneous logistic regression analysis with the verdict

variable (dichotomous) as the outcome, and with defendant race (dichotomous), type of

crime (dichotomous), and NC (continuous) as the predictor variables. I also included all

the interactions terms as additional predictors in the model.

Results showed that none of the predictors significantly predicted the verdict

decision variable. For a summary of the results, see Appendix Y, Table 39.

Verdict-Confidence as the Outcome Variable

I conducted multiple Analyses of Covariance (ANCOVA) to test the hypotheses

that predicted significant two-way interactions and three-way on verdict-confidence (i.e.,

hypotheses 3a, 6a, and 9a). Specifically, I tested whether mock jurors' level of

confidence with regards to their verdict decision depended on whether or not race-crime

congruence existed. In addition, I planned to test whether this interaction was affected by

the juror's level of MCPR and/and or the presence of SSI.

Verdict-confidence as the outcome, with MCPR in the model. I conducted a 2

(Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement)

ANCOVA on verdict-confidence scores, controlling for MCPR. Results revealed that

there was a significant main effect of type of crime on verdict-confidence, $F(1, 167) =$

14.68, $p = < .001$. Participants who read about embezzlement ($M = 3.81$, $SD = 3.75$) had

higher verdict-confidence than participants who read about grand-theft auto ($M = 1.13$,

$SD = 5.25$). However, there was no significant main effect of defendant race on verdict-

confidence, $F(1, 167) = 0.00$, $p = .97$. This means that whether the defendant was Black

($M = 2.39$, $SD = 4.75$) or White ($M = 2.35$, $SD = 4.87$) did not affect how confident mock

jurors felt about their verdict decisions. In addition, MCPR was not significantly related to the mock jurors' verdict-confidence scores, $F(1, 167) = 0.47$, $p = .50$. Moreover, in contrast to my predictions, the interaction between defendant race and type of crime on verdict-confidence was not significant, $F(1, 167) = 0.58$, $p = .45$ (Figure 13).

Figure 13. Non-significant Interaction between Defendant Race and Type of Crime on Verdict-Confidence (Study 3)
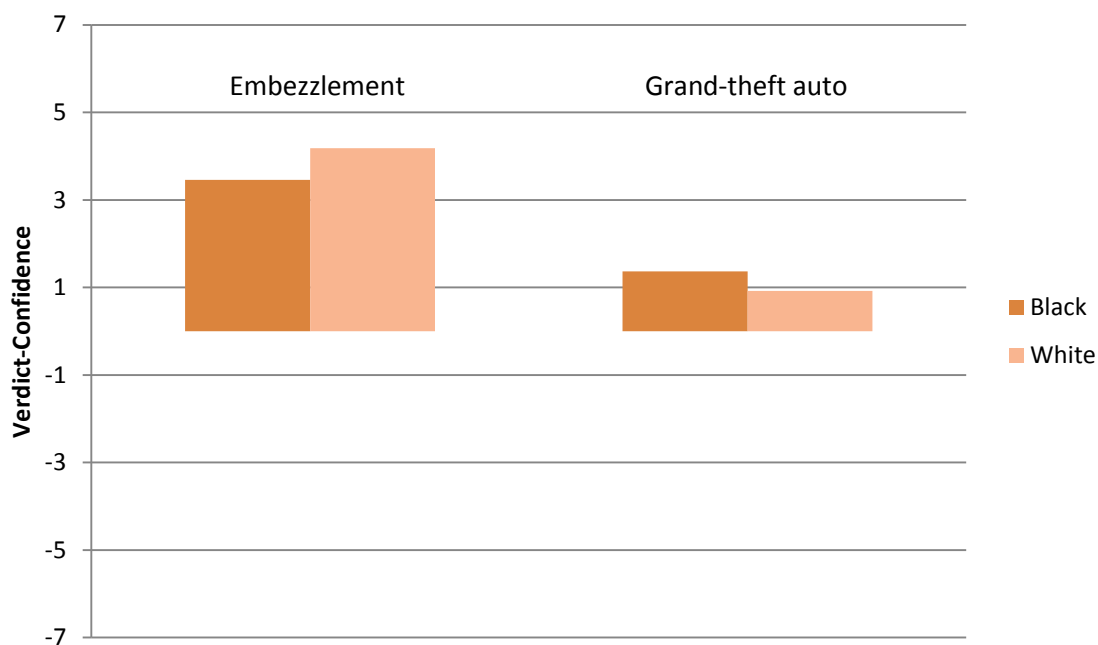
Moreover, there was no significant effect between defendant race and MCPR on verdict-confidence, $F(1, 167) = 0.92$, $p = .34$. There was also no significant interaction between type of crime and MCPR on verdict-confidence, $F(1, 167) = 0.22$, $p = .64$. Finally, the three-way interaction between defendant race, type of crime, and MCPR was not significant, $F(1, 167) = 0.43$, $p = .51$.

Verdict-confidence as the outcome, with SSI in the model. I conducted a 2

(Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on verdict-confidence scores. Results showed that type of crime was a significant predictor for verdict-confidence scores, $F(1, 167) = 14.58$, $p < .001$. Participants who saw the embezzlement case ($M = 3.81$, $SD = 3.75$) were more confident in their guilty verdict decision than participants who saw the grand-theft auto case ($M = 1.13$, $SD = 5.25$) However, there were no other significant main effects, interactions, or covariates (see Appendix Y, Table 40).

Verdict-confidence as the outcome, with NC as a control variable. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on verdict-confidence scores, controlling for NC. Again, Results showed type of crime significantly predicted mock jurors' verdict-confidence scores, $F(1, 167) = 11.85$, $p < .01$. As before, participants who saw the embezzlement case ($M = 3.81$, $SD = 3.75$) had higher verdict-confidence scores than participants who saw the grand-theft auto case ($M = 1.13$, $SD = 5.25$) However, there were no other significant main effects, interactions, or covariates (see Appendix Y, Table 41).

Probability of Guilt Ratings

I conducted multiple Analyses of Covariance (ANCOVA) to test the hypotheses that predicted significant two-way and three-way interactions on probability of guilt (i.e., hypotheses 4a, 7a, 10a). Specifically, I tested whether how probably it is that the defendant is guilty of the crime he was charged with depends on whether the type of crime and the defendant race are stereotypically consistent with one another, and whether

this interaction was affected by one's level of MCPR and/or the presence of SSI.

Probability of guilt as the outcome, with MCPR in the model. I conducted a 2

(Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement)

ANCOVA on probability of guilt ratings, controlling for MCPR. In contrast to my

predictions, results revealed that there were no significant interaction effects between

defendant race and type of crime on probability of guilt ratings, $F(1, 167) = 0.55$ $p = .46$

(Figure 14; see appendix Y).

Figure 14. Non-significant Interaction between Defendant Race and Type of Crime on Probability of Guilt (Study 3)
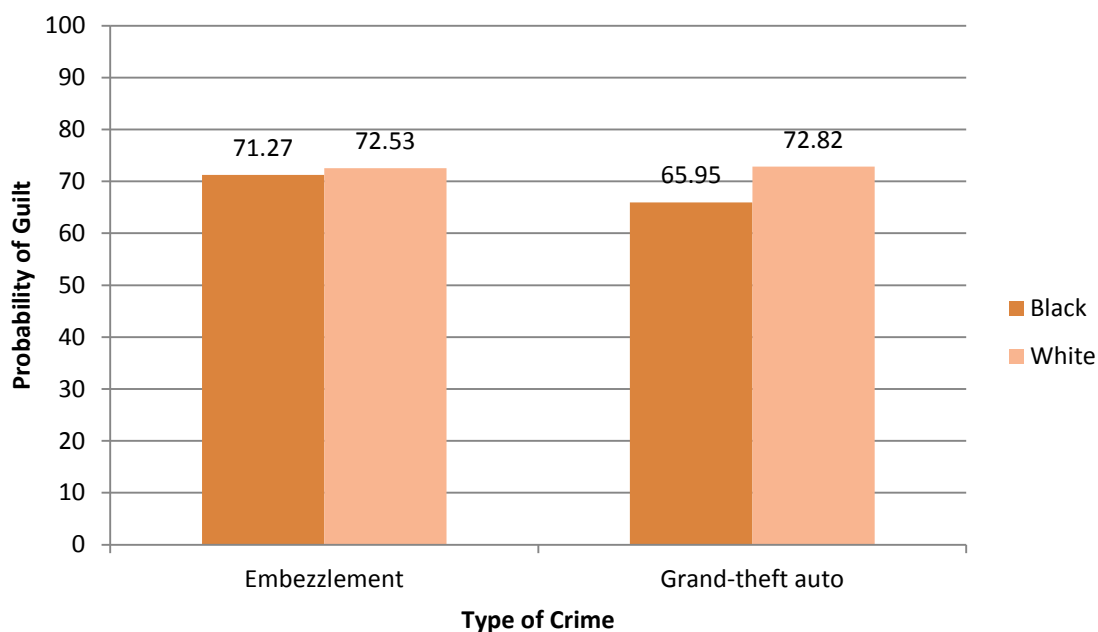
All other main effects and interactions were also not significant (see Appendix Y, Table

42).

Probability of guilt as the outcome, with SSI in the model. I conducted a 2

(Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on probability of guilt ratings. Results revealed that the interaction between defendant race and type of crime was not significant, $F(1, 167) = 0.94$, $p = .33$. All other main effects and interactions were not significant (see Appendix Y, Table 43).

Probability of guilt as the outcome, with NC as a control variable. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on probability of guilt ratings, controlling for NC. Results showed that there was no significant interaction effect between defendant race and crime type on how probable mock jurors believed it was that the defendant is guilty, $F(1, 167) = 0.88$, $p = .35$. All other main effect, covariates, and interactions were not significant predictors either (see Appendix Y, Table 44).

Story Coherence as the Outcome

I predicted that there was a significant interaction between defendant race and type of crime, which then would be moderated by MCPR and SSI. I hypothesized that two-and three-way interactions would be mediated by narrative believability. I believed that especially the coherence of the participants' own stories would be a strong mediator (hypotheses 2b-10b). Unexpectedly, none of these two- and three-way interactions had significant effects on the verdict judgments in the prior analyses. However, mediation could still occur, if the independent variables were related to the potential mediator. If this was not the case, then there would not be a need to run any mediational analyses. In order to test whether mediational analyses would be warranted, I conducted multiple

ANCOVAs with coherence as the outcome variable. In contrast to the analyses in Study 2, I refrained from conducting the analyses with coverage and uniqueness as the outcome variables in addition to coherence because I was mainly interested in coherence as the mediator. In order to be all-inclusive, I conducted these analyses for the coherence of the prosecution story (MNBS-P), defense story (MNBS-D), and juror stories (MNBS-J and MNBS-R).

Coherence of the prosecution story. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on story coherence (MNBS-P), controlling for MCPR. Results showed that there was a significant main effect of type of crime on story coherence, $F(1, 167) = 14.06$, $p < .001$. This means that mock jurors rated the prosecution story for the grand-theft auto case ($M = 41.30$, $SD = 7.79$) as more coherent than the prosecution story for the embezzlement case ($M = 36.72$, $SD = 6.83$). In addition, there was a significant defendant race x MCPR interaction, $F(1, 167) = 4.21$, $p = .04$. To probe this significant interaction, I investigated the effect of defendant race on coherence of the prosecution story when MCPR was at one standard deviation above and when it was on one standard deviation below the mean in two additional analyses. Results from the regression analyses revealed that when mock jurors had low levels of MCRP, defendant race was not a significant predictor of story coherence, $B = -1.44$, $\beta = -0.09$, $t(3) = -0.87$, $p = .38$. However, when mock jurors had high levels of MCPR, type of crime significantly predicted how much they felt that the prosecution story was coherent, $B = 4.44$, $\beta = 1.60$, $t(3) = 2.78$, $p = .01$. This suggests that

for White mock jurors with high levels of MCPR who saw a White defendant the prosecution story was more coherent than those who saw a Black defendant. However, there was not a significant interaction between defendant race and type of crime on story coherence, $F(1, 291) = 0.11$, $p = .89$. All other main effects and interactions were also not significant (see Appendix Y, Table 45).

I also conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on story coherence (MNBS-P). Results showed that there was a significant main effect of type of crime on story coherence, $F(1, 167) = 16.42$, $p < .001$. Mock jurors believed that the prosecution story for the grand-theft auto case ($M = 41.30$, $SD = 7.79$) was more coherent than the prosecution story for the embezzlement case ($M = 36.72$, $SD = 6.84$). However, there was no significant interaction effect between defendant race and type of crime on story coherence, $F(1, 167) = 1.12$, $p = .29$. No other significant main or interaction effects became apparent (see Appendix Y, Table 46).

In addition, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime: Grand-theft Auto, Embezzlement) ANCOVA on story coherence (MNBS-P), controlling for mock juror's need for cognition (NC). There was a significant main effect of type of crime on story coherence, $F(1, 167) = 15.48$, $p < .01$. Mock jurors rated the prosecution story for the grand-theft auto case ($M = 41.30$, $SD = 7.79$) as more coherent than the prosecution story for the embezzlement case ($M = 36.72$, $SD = 6.84$). In addition, the covariate, NC was significantly related to story coherence, $F(1, 291) = 4.58$, $p = .03$. Contrary to my predictions, the interaction effect between defendant race and type of

crime on story coherence was not significant, $F(1, 167) = 1.06$, $p = .30$, as were none of the other main effects and interactions (see Appendix Y, Table 47).

Because the results showed that defendant race and type of crime did not interact to predict story coherence of the prosecution story, I did not test for mediation.

Coherence of the defense story. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on story coherence (MNBS-D), controlling for MCPR. Results revealed that there was a significant main effect of type of crime on story coherence, $F(1, 167) = 14.30$, $p < .001$. This means that mock jurors rated the defense story describing the embezzlement case ($M = 39.02$, $SD = 6.03$) as more coherent than the defense story outlining the grand-theft auto case ($M = 35.09$, $SD = 7.37$). There also was as significant effect of MCPR on the story coherence of the defense story, $F(1, 167) = 4.92$, $p = .03$. A scatterplot of MCPR and the outcome showed that as an individual mock juror's level of MCPR increases, his beliefs about the coherence of the defense story become less strong. Contrary to my predictions, there was not a significant defendant race by type of crime interaction on the coherence of the defense story, $F(1, 167) = 0.01$, $p = .78$. All other main effects and interactions were not significant (see Appendix Y, Table 48).

In addition, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on story coherence (MNBS-D). Results showed that there was a significant main effect of type of crime on story coherence, $F(1, 167) = 13.93$, $p < .001$. Specifically, mock jurors who saw the embezzlement case ($M = 39.02$, $SD = 6.03$) rated the coherence

of the defense story higher than mock jurors who saw the grand-theft auto case ($M =$ 35.09, $SD = 7.37$). Although there was no significant defendant race by type of crime interaction, $F(1, 167) = 0.10$, $p = .75$, there was a significant interaction between defendant race and SSI on coherence ratings of the defense story, $F(1, 167) = 4.68$, $p = .03$. This suggests that the relation between defendant race and mock jurors' coherence ratings of the defense story depended on the presence (or absence) of SSI. When mock jurors did not receive SSI, there was not a significant simple main effect of defendant race on how much mock jurors believed that the defense story was coherent, $F(1, 85, = 1.97, = .16$. Mock jurors who saw a White defendant (M = 38.33, SD = 6.70) did not significantly differ in their coherence ratings from mock jurors who saw a Black defendant ($M = 36.21$, $SD = 7.38$). , $F(1, 85) = 1.97$. In addition, when mock jurors received SSI, there was not a significant simple main effect of defendant race predicting their coherence ratings of the defense story, $F(1, 85) = 2.20$. Mock jurors who saw a White defendant ($M = 35.46$, $SD = 7.12$) gave similar coherence ratings to the defense story than mock jurors who saw a Black defendant ($M = 37.57$, $SD = 6.82$). No other main effects or interactions had significant effects on story coherence (see Appendix Y, Table 49).

I also conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime: Grand-theft Auto, Embezzlement) ANCOVA on story coherence (MNBS-D), controlling for mock juror's need for cognition (NC). There was a significant main effect of type of crime on coherence ratings of the defense story, $F(1, 167) = 15.72$, $p < .001$. Mock jurors rated the defense story for the embezzlement case ($M = 39.09$, $SD = 6.03$) as more

coherent than the defense story for the grand-theft auto case ($M = 35.09$, $SD = 7.37$). As in prior analyses, the interaction effect between defendant race and type of crime on story coherence was not significant, $F(1, 167) = 0.00$, $p = .83$, as were none of the other main effects and interactions (see Appendix Y, Table 50).

Because none of the defendant race x type of crime interactions had significant effects on the coherence ratings of the defense story, no mediational analyses were warranted.

"Subjective" coherence of mock jurors' stories. I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on mock jurors' coherences ratings of their own stories (MNBS-J), controlling for MCPR. There were no significant main or interaction effects (see Appendix Y, Table 51).

I also conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on mock jurors' coherence ratings of their own stories (MNBS-J). Again, there were no significant main or interaction effects on coherence (see Appendix Y, Table 52).

Furthermore, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime: Grand-theft Auto, Embezzlement) ANCOVA on mock jurors' coherence ratings of their own stories (MNBS-J), controlling for mock juror's need for cognition (NC). As before, none of the main or interaction effects were significant (Appendix Y, Table 53).

Because the main hypothesis was not supported, any mediational analyses were not warranted.

"Objective coherence" of mock jurors' stories. I conducted a 2 (Defendant

Race: Black, White) x 2 (Type of Crime: Grand-theft auto, Embezzlement) ANCOVA on

independent raters' coherence ratings they made about the mock jurors' stories (MNBS-

R), controlling for MCPR. Results showed that there was a significant main effect of type

of crime on how coherent mock jurors' stories were rated, $F(1, 167) = 14.06$, $p < .001$.

Mock jurors who read about the grand-theft auto case (M 42.30, $SD = 7.78$) rated the

stories as more coherent than mock jurors who read about the embezzlement case ($M =$

36.72, $SD = 6.84$). In addition, there was a significant defendant race x MCPR interaction

on story coherence, $F(1, 167) = 4.21$, $p = .04$. To probe this significant interaction, I

investigated the effect of defendant race on coherence of the mock jurors' stories when

MCPR was at one standard deviation above and when it was on one standard deviation

below the mean in two additional analyses. Results from the regression analyses revealed

that when mock jurors had low levels of MCRP, defendant race was not a significant

predictor of raters' coherence ratings of the mock jurors' stories, $B = -1.44$, $\beta = -0.09$, $t(3)$

$= -0.87$, $p = .39$. However, when mock jurors had high levels of MCPR, defendant race

significantly predicted how much they felt that mock jurors' stories were coherent, $B =$

4.44, $\beta = 1.60$, $t(3) = 2.78$, $p = .01$. This suggests that mock jurors' stories where more

coherent when the jurors who had high levels of MCPR saw a White instead of a Black

defendant. However, there was not a significant interaction between defendant race and

type of crime on story coherence, $F(1, 167) = 0.81$, $p = .37$. All other main effects and

interactions were also not significant (see Appendix Y, Table 54).

In addition, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of Crime:

Grand-theft auto, Embezzlement) x 2 (Stereotype Suppression Instructions: no, yes) ANCOVA on story coherence (MNBS-R). Results revealed that there was a significant main effect of type of crime on raters' coherence ratings of the mock jurors' stories, $F(1, 167) = 16.42$, $p < .001$. Specifically, mock jurors who saw the grand-theft auto case ($M = 41.30$, SD = 7.78) rated the jurors' stories as more coherent than mock jurors who saw the embezzlement case ($M = 36.72$, $SD = 6.84$). The interaction between defendant race and type of crime on story coherence was not significant, $F(1, 167) = 1.12$, $p = .29$ (see Appendix Y, Table 55).

Finally, I conducted a 2 (Defendant Race: Black, White) x 2 (Type of crime: Grand-theft Auto, Embezzlement) ANCOVA on story coherence (MNBS-R), controlling for mock juror's need for cognition (NC). There was a significant main effect of type of crime on coherence ratings of mock jurors' stories, $F(1, 167) = 15.48$, $p < .001$. The stories of mock jurors' who saw the grand-theft auto case ($M = 41.30$, $SD = 7.79$) were rated as more coherent than the stories of mock jurors who read about embezzlement ($M = 36.72$, $SD = 6.84$). As before, the interaction effect between defendant race and type of crime on how coherence the mock jurors' stories were rated was not significant, $F(1, 167) = 1.06$, $p = .30$, as were none of the other main effects and interactions (see Appendix Y, Table 56).

Because none of the defendant race x type of crime interactions had significant effects on the coherence ratings of mock jurors' stories, no mediational analyses were warranted.

Discussion

As in Study 2, the type of crime that the mock jurors read about predicted their culpability judgments. However, in contrast to the findings in Study 2, jurors found the defendant more frequently guilty, were more confident in that decision, and believed that it was more probable that the defendant committed the crime when the charge was embezzlement, not grand-theft auto. These findings are surprising because the results from the pilot tests showed that participants generally perceived the defendant who was charged with grand-theft auto as guiltier than the than the defendant who was charged with embezzlement. However, they saw embezzlement as the more serious crime. Thus, at first glance, it would have made more sense for results from Study 2 would replicate— because mock jurors saw the defendant that was charged with grand-theft auto guiltier, they found him more often guilty and made harsher culpability judgments.

The only major difference between Study 2 and 3 up until when participants were asked to make their verdict decisions was the type of sample that was used. Thus, those differences in results could only be due to the differences that exist between those two groups. Although the results from the pilot test are fairly consistent, it is possible that the college students in Study 3 did not view grand-theft auto a crime that deserves punishment, even if he is seen as guilty of stealing a car. This could be due to the fact that the student population was much younger in age than the MTurk workers in Study 2, who were, on average in their mid-30s. College students tend to play more video and computer games as older people, and video games that are currently popular include playing grand-theft auto—stealing cars and committing street crimes. Because of this

frequent exposure, grand-theft auto may not be a big deal to them. In addition, college students tend to have more financial constraints than older adults, thus viewing embezzling money as a crime that deserves serious punishment.

In addition, the type of crime predicted how coherent participants viewed the story that the prosecution came up with when interpreting the evidence leading to the trial. Specifically, they believed that the prosecution story was more coherent when it explained the grand-theft auto case compared to the embezzlement case. The same results emerged for the coherence of the jurors' stories, which was rated by unbiased coders unaware of the study hypotheses. In addition, when evaluating the defense story was more coherent when the embezzlement represented the crime. These results mirror the findings from the pilot studies and manipulation checks.

As in Study 2, contrary to my predictions, no race-crime congruence effects emerged, although the participants associated grand-theft auto more with a Black defendant and embezzlement more with a White defendant. Thus, whether a defendant is found guilty does not seem to depend on whether the crime he or she committed fits the racial stereotype. Furthermore, mock jurors did not think that a story about a Black person being charged with grand-theft auto is more coherent than a story about a White person being charged with the same crime. Similarly, in the minds of the jurors, a story about a White defendant supposedly committing embezzlement is no more coherent than a story about Black defendant who is charged with embezzlement. This could be due to the fact that the study participants were not only college students who may be more educated than the average juror, but they also were attending a university with the

mission of promoting social justice. Thus, the participants may not have held stereotypes about race and crime all, or refrained from applying them in the currents study.

This explanation also makes sense in light of the finding that participants who had high levels of MCPR, believed that the prosecution story was more coherent when the defendant was White than when he was White. In addition, their own stories reflected similar levels coherence, as rated by an unbiased coder, when they had a high motivation to control their prejudiced reactions. This shows that participants in attempting to not appear prejudicial, they overcompensated by indicating that a story that includes a White defendant that is charged with a crime is more logical and consistent.

Surprisingly, type of crime and defendant race did not predict how participants rated their own stories in terms of coherence, coverage, and uniqueness, although Pennington and Hastie (1984, 1986, 1988) theorized that the those attributes of their own stories ultimately determine their culpability judgments.

A disadvantage of the methodology of Study 3 (and Study 2) was that the order in which the participants completed the study may not mirror juror decision making in real life. In a courtroom, jurors listen to the evidence that the prosecution and defense presents at trial, and they construct their own stories during or after the trial, while simultaneously evaluating the prosecution and defense storied, but before they start to deliberate with their fellow jury members. Thus, it is assumed that the stories are complete, and they evaluate the believability of their stories before they make a verdict decision. While the participants most likely constructed their stories while reading the case materials, I may not have given them enough time to process and evaluate the information sufficiently

because they were asked to make their verdict judgments right away. In addition,

responding to the MNBS-P and MNBS-D before they construct their own stories may

have biased their own stories that they wrote afterwards.

CHAPTER 8

STUDY 4

Overview

Study 4 was a replication of Study 2. The only difference was that instead of

rating the prosecution and defense stories and after they read the trial and before

participants made their verdict judgments, they were asked to fill out the MNBS-P and

MNBS-D right after they made their culpability judgments. Although Study 4 was not a

part of my original proposal, I believed it was important to mimic the cognitive processes

that jurors in real life go through as realistically as possible. According to the Story

Model (Pennington & Hastie, 1984, 1986), jurors make sense of the evidence by listening

to the stories that the prosecution and defense present. In addition, I ensured that the

order of presentation of study materials was similar to Jones & Kaplan's (2003) study.

Therefore, I attempted to rule out the possibility that the order in which participants

completed the research study in studies 2 and 3 was responsible for the absence of race-

crime congruence effects. For these reasons, Study 4 was included in the current line of

research. My hypotheses were the same as in Study 2.

Method

Participants

Data was collected from 132 undergraduate students who received course credit

in exchange for their participation in the research study. Because 7 participants were not

U.S. citizens, and 52 participants reported an ethnicity that was different from European American/White (see Appendix Z, Figure 18), those participants were dropped from the analyses. Accordingly, the present sample included 74 participants, of whom 77% were female. Their ages ranged from 18 to 67 *(M* = 19.41, *SD* = 5.77, median = 18.00). The majority of the sample were sophomores in college (*n* = 62); there were 11 freshmen, and 1 junior.

Materials

The materials were the exact same as in Study 2.

Procedure

The procedure was the same as in Study 2, except for the time in point when participants were asked to rate the prosecution and defense stories. Instead of asking them to fill out the MNBS-P and MNBS-R after they made their verdict judgments (as in Study 2), participants rated the stories right after they read the case, and before they made decisions about the defendant's guilt. This ensured that the cognitive processes mock jurors engage in during the study mimic those that real-life jurors experience during an actual trial as closely as an artificial setting would allow.

<center>**Results**</center>

Reliability Analyses

As in Study 2 and Study 3 I calculated Cronbach's alphas to determine the internal consistency of the MNBS-P, MNBS-D, MCPRS, and NCS. In addition, I conducted reliability analyses for all three subscales of the MNBS-P and MNBS-D.

Results showed that the 17 items of the MNBS-P were reliable, Cronbach's α = .71. In addition, both coherence and coverage demonstrated acceptable scale reliabilities (≥ .70)—Cronbach's α = .79, and Cronbach's α = .78—respectively. However, uniqueness, was not a reliable scale in this study, Cronbach's α = .66. Deleting any of the items would not have significantly increased the reliability of the uniqueness scale.

The 17 items of the MNBS-D also demonstrated acceptable reliability with a Cronbach's alpha of .80. In addition, the analyses showed that deleting any of the items would not significantly increased the reliability of the scale. The subscale coherence demonstrated an acceptable level of reliability (α = .75). The subscales coverage and uniqueness did not reach acceptable reliabilities—α = .65 and α = .69—respectively. Deleting any of the items would not have increase the reliability of the subscales.

In addition, the MCPRS (Motivation to Control Prejudiced Reactions Scale; 17 items) and the NCS (Need for Cognition Scale; 18 items) both exhibited acceptable scale reliabilities —α = .75 and α = .87—respectively.

Manipulation Checks

I conducted an independent samples t-test in order to ensure that neither one of the crimes was perceived as more serious than the other. Results showed that participants perceived embezzlement ($M = 4.90$, $SD = 1.35$) as equally serious of a crime as grand-theft auto ($M = 4.31$, $SD = 1.34$), $t(72) = -1.83$, $p = .07$.

As expected, results from an independent samples t-test showed that participants typically associated grand-theft auto ($M = 4.27$, $SD = 0.92$) significantly more with a Black

defendant than they associated embezzlement with a Black defendant ($M = 2.83$, $SD = 1.44$), $t(72) = 5.26$, $p < .001$.

Descriptive Statistics

Overall, 62.20% of jurors found the defendant guilty. Means and standard deviations were calculated for all main variables (Table 9).

Table 9. Descriptive Statistics – Study 4

	Mean	Standard Deviation
Verdict-Confidence	1.58	4.66
Probability of Guilt	66.03	22.70
Coherence – Prosecution	42.84	8.14
Coverage – Prosecution	12.86	3.58
Uniqueness – Prosecution	16.73	4.17
Coherence – Defense	38.62	8.08
Coverage – Defense	11.64	3.23
Uniqueness – Defense	15.34	4.30
MCPRS	68.41	12.46
NCS	77.78	11.30

Correlational Analyses

Bivariate correlational analyses were conducted between all variables measuring

narrative believability, and guilt. As seen in Table 10, many of the correlations were

significant. In addition, all of the significant correlations were in the expected direction.

Surprisingly, not all the subscales of the MNBS-P and not of all the subscales of the

MNBS-D were correlated with one another; some of the correlations were not significant.

In addition, against expectations, the subscales of the two measurement scales were not

all negatively correlated with one another. However, verdict-confidence scores and

probability of guilt ratings were positively correlated, and these verdict measures also had

significant correlations with most of the subscales of the MNBS-P and MNBS-D,

especially with coherence of the prosecution story.

Table 10. Correlations between Variables Measuring Narrative Believability and Verdict Judgments (Study 4)

Variable	1	2	3	4	5	6	7	8
1. Coherence/Prosecution	1.00							
2. Coverage/Prosecution	.64**	1.00						
3. Uniqueness/Prosecution	.06	.40**	1.00					
4. Coherence/Defense	-.01	-.22	-.49**	1.00				
5. Coverage/Defense	-.05**	.17	-.12**	.43**	1.00			
6. Uniqueness/Defense	-.56**	-.26**	.22	.08	.42**	1.00		
7. Verdict x Confidence	.36**	.37**	.26*	-.25*	-.15	-.20	1.00	
8. Probability of Guilt	.44**	.25**	.08	-.16**	-.25**	-.39*	.73**	1.00

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

Outcome Measures

As in Study 2 and Study 3, verdict decision, verdict-confidence scores, and probability of guilt ratings were the three outcome measures. Because the main goal of the study was to determine whether changing at what time in the study mock jurors will rate the prosecution and defense story will lead to the expected race-crime congruence effect, I only reported the results of the interaction between defendant race and type of crime on the different verdict judgments here.

Verdict Decision as the Outcome Variable

In order to determine whether a defendant was found guilty or not, I conducted a simultaneous logistic regression analysis with verdict (not guilty vs. guilty) as the outcome measure, and with defendant race (Black vs. White) and type of crime (grand-theft auto vs. embezzlement) as the predictor variables. I included the interaction term as an addition predictor.

The omnibus test of model coefficients indicated that the model was not statistically significant, $\chi^2(4) = 1.20$, $p = .74$. This means that the predictor variables as a group did not predict the verdict decision variable.

Results showed that defendant race was not a significant predictor of the verdict decision, B $= 0.07$, $\chi^2(1) = 9.91$, p $= 0.91$. The type of crime also was not a significant predictor of the verdict decision, $B = 0.27$, $\chi^2(1) = 0.14$, $p = .71$. In addition, the defendant race x type of crime interaction was not significant, $B = 0.33$, $\chi^2(1) = 0.10$, $p = .75$ (see Figure 19).

Verdict-Confidence as the Outcome Variable

In addition, I conducted a 2 (Defendant race: Black vs. White) x 2 (Type of Crime: Grand-theft auto vs. embezzlement) Analysis of Variance (ANOVA) on mock jurors' verdict-confidence scores. Results revealed that the main effect of defendant race on verdict-confidence scores was not significant, $F(1, 70) = 0.14$, $p = .71$. This means that there was no difference in verdict-scores between mock jurors who read about a Black defendant ($M = 1.28$, $SD =- 4.88$) and those who read about a White defendant ($M = 1.91$, $SD = 4.45$). In addition, there was no significant main effect of type of crime on verdict-confidence score, $F(1, 70) = 0.72$, $p = .40$. There was no difference in verdict-scores between mock jurors who saw the grand-theft auto case ($M = 1.16$, $SD = 4.45$) and those who saw the embezzlement case ($M = 2.24$, $SD = 4.53$). Finally, the interaction between defendant race and type of crime on verdict-confidence scores was not significant, $F(1, 70) = 0.20$, $p = .93$ (Figure 15).

Figure 15. Non-significant Interaction between Defendant Race and Type of Crime on Verdict-Confidence Scores (Study 4)
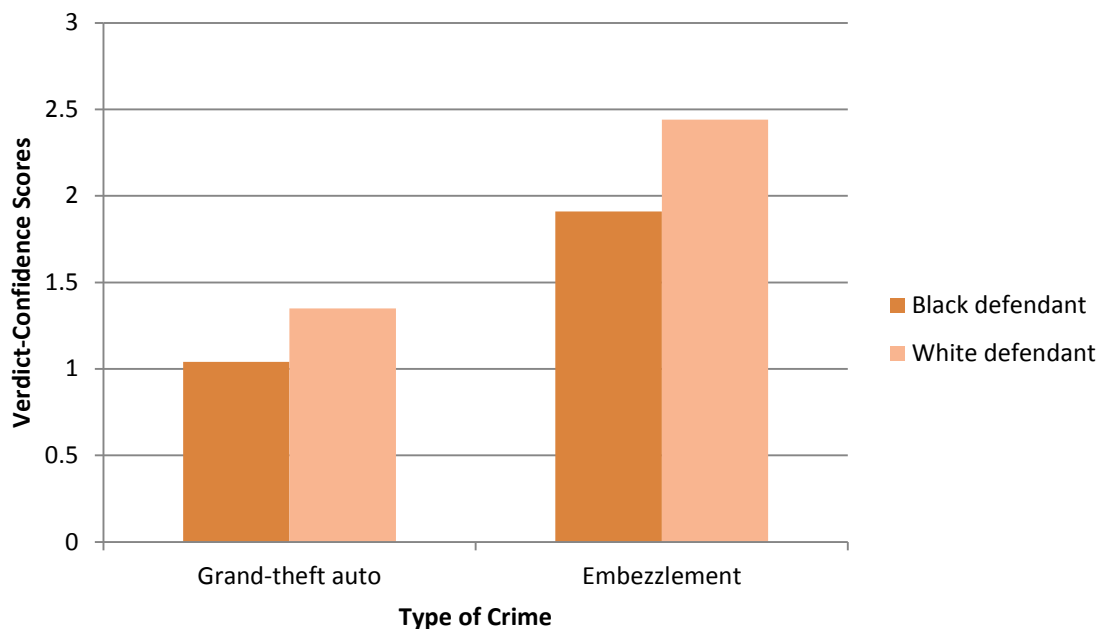
Probability of Guilt Ratings as the Outcome Variable

In addition, I conducted a 2 (Defendant race: Black vs. White) x 2 (Type of Crime: Grand-theft auto vs. embezzlement) Analysis of Variance (ANOVA) on mock jurors' probability of guilt ratings. Results revealed that the main effect of defendant race on probability of guilt ratings was not significant, $F(1, 70) = 0.04$, $p = .84$. This means that there was no difference in probability of guilt ratings between mock jurors who read about a Black defendant ($M = 65.26$, $SD = 23.69$) and those who read about a White defendant ($M = 66.89$, $SD = 21.86$). In addition, there was no significant main effect of type of crime on probability of guilt ratings, $F(1, 70) = 2.72$, $p = .10$. There was no difference in how probable they believed it was that the defendant was guilty between mock jurors who saw the grand-theft auto case ($M = 62.53$, $SD = 24.36$) and those who

saw the embezzlement case ($M = 71.45$, $SD = 19.00$). Finally, the interaction between defendant race and type of crime on probability of guilt ratings was not significant, $F(1, 70) = 0.31$, $p = .58$ (Figure 16).

Figure 16. Non-significant Interaction between Defendant Race and Type of Crime on Probability of Guilt Ratings (Study 4)
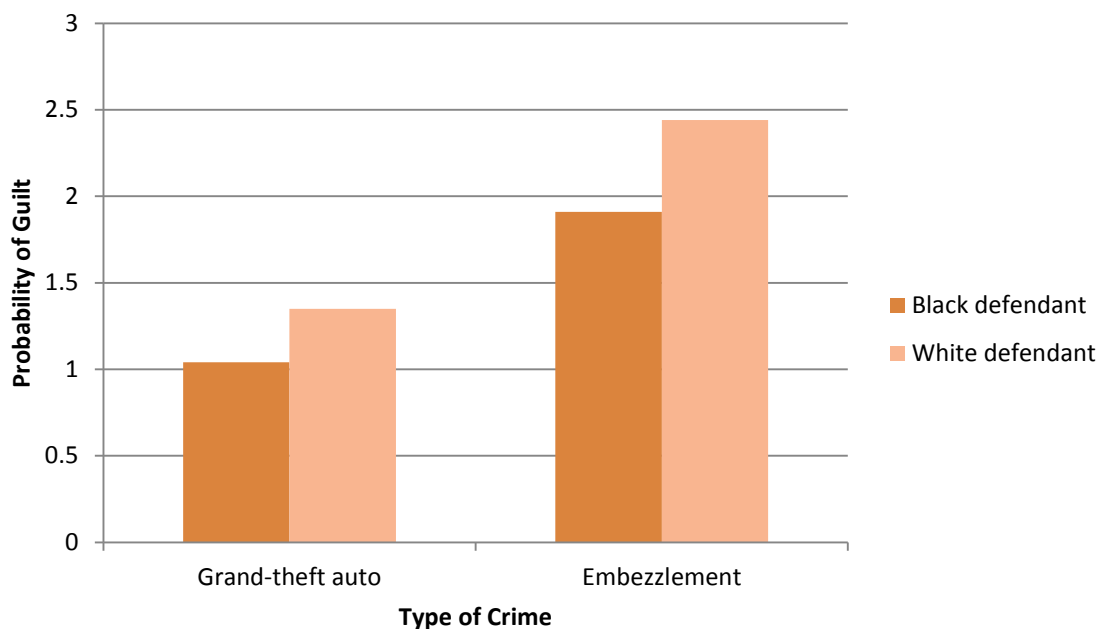
Coherence of the Prosecution Story as the Outcome

I also conducted a 2 (Defendant race: Black vs. White) x 2 (Type of Crime: Grand-theft auto vs. embezzlement) Analysis of Variance (ANOVA) on coherence ratings of the prosecution story. Results revealed that the main effect of defendant race on coherence ratings of the prosecution story was not significant, $F(1, 70) = 0.21$, $p = .64$. This means that there was no difference in coherence ratings of the prosecution story between mock jurors who read about a Black defendant ($M = 42.82$, $SD = 7.53$) and those

who read about a White defendant ($M = 42.86$, $SD = 8.88$). In addition, there was no significant main effect of type of crime on mock jurors' coherence ratings of the prosecution story, $F(1, 70) = 0.17$, $p = .68$. There was no difference in how coherent they believed the prosecution story was between mock jurors who saw the grand-theft auto case ($M = 42.60$, $SD = 7.79$) and those who saw the embezzlement case ($M = 43.21$, $SD = 8.81$). Finally, the interaction between defendant race and type of crime on coherence ratings of the prosecution story was approaching significance, $F(1, 70) = 3.52$, $p = .07$. The results were exactly the opposite of what was predicted—when there was a lack of congruence between stereotypes about defendant race and crime, the prosecution story was rated as more coherent compared to when there was congruence (Figure 17).

Figure 17. Interaction between Defendant Race and Type of Crime on Coherence Ratings of the Prosecution Story (Study 4)
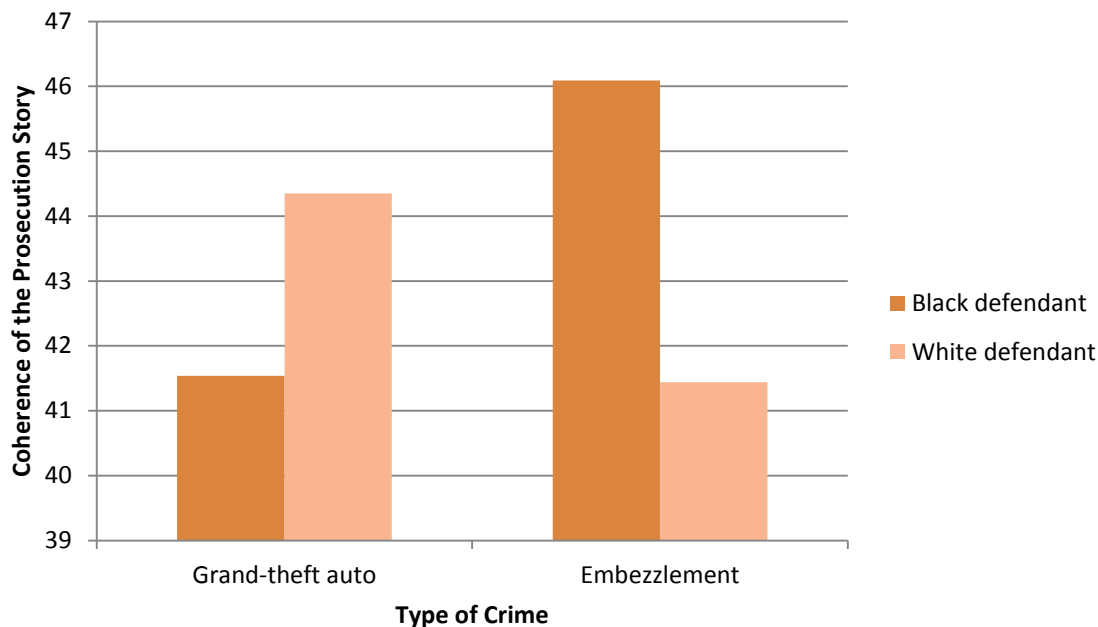
Coherence of the Defense Story Outcome

I also conducted a 2 (Defendant race: Black vs. White) x 2 (Type of Crime: Grand-theft auto vs. embezzlement) Analysis of Variance (ANOVA) on coherence ratings of the defense story. Results revealed that the main effect of defendant race on coherence ratings of the defense story was not significant, $F(1, 70) = 1.70$, $p = .20$. This means that there was no difference in coherence ratings of the defense story between mock jurors who read about a Black defendant ($M = 39.46$, $SD = 7.57$) and those who read about a White defendant ($M = 37.69$, $SD = 8.63$). In addition, the main effect of type of crime on mock jurors' coherence ratings of the defense story was approaching significance, $F(1, 70) = 2.98$, $p = .09$. Mock jurors who read about the embezzlement case ($M = 40.31$, $SD = 7.89$) believed that the defense story was slightly more coherent than mock jurors who read about the grand-theft auto case ($M = 37.53$, $SD = 8.63$). Finally, the interaction between defendant race and type of crime on coherence ratings of the defense story was not significant, $F(1, 70) = 0.00$, $p = .997$ (Figure 18). Accordingly, no additional mediational analyses were warranted.

Figure 18. Non-significant Interaction between Defendant Race and Type of Crime on Coherence Ratings of the Defense Story (Study 4)
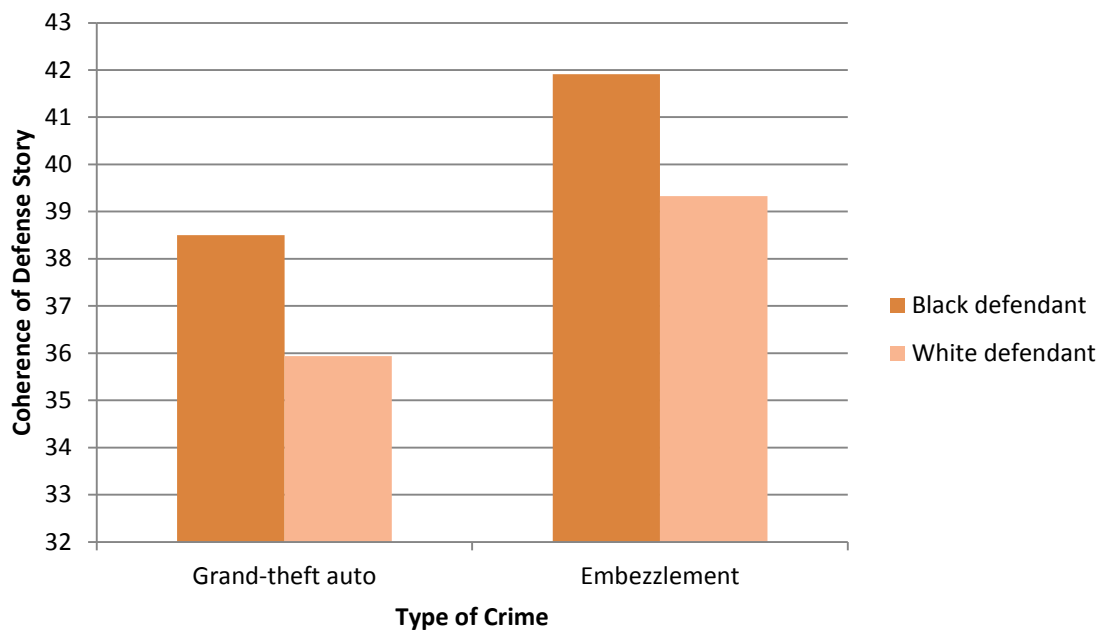
Discussion

Despite of changing the order in which participants were asked to make their culpability judgments and to rate the prosecution and defense stories, no race-crime congruence effects on verdict judgments emerged. Even though participants now had a chance to thoroughly think about the two stories presented at trial by asking them to rate the prosecution and defense stories before deciding on a verdict and making probability of guilt ratings, the results were similar to those in the two prior studies. Whether or not the type of crime matched the racial stereotype about defendants did not make a difference to participants' verdict judgments.

However, type of crime seemed to moderate the effects of defendant race on coherence ratings of the prosecution story. Contrary to what was predicted, participants

who read about a Black defendant who was charged with embezzlement rated the prosecution story higher in coherence than those participants who read about a White defendant who was charged with the same crime. Similarly, participants who read about a grand-theft auto crime with a White defendant believed the prosecution story was more coherent than when a Black defendant was charged with grand-theft auto. It is possible that participants attempted to suppress any stereotypes they may have had throughout the study, but this uses up many cognitive resources, so that they overcorrected their initial attitudes, which led to results that were the direct opposite of what I expected in terms of coherence ratings.

CHAPTER 9

GENERAL DISCUSSION

Every day people from all different walks of life come together in courtrooms to make decisions about guilt and innocence, thus determining the future of numerous defendants who are standing trial. Therefore, it is crucial to understand this very important decision making context, so that specific procedures can be put in place, which ensure that those verdicts are as accurate as possible to avoid the guilty being acquitted and the innocent being convicted. With the current studies, I attempted to explain why jurors may use race and crime stereotypes when making verdict decisions from the perspective of the Story Model, which is the most prominent theoretical model in the juror decision making literature. Moreover, the purpose was to identify person-specific and trial-related factors that may remedy these race-crime congruence effects.

In Study 1, by modifying an existing measurement scale (NBS-12; Yale, 2013), I developed a new self-report instrument that allowed me to measure the perceived coherence, coverage, and uniqueness of the story that the prosecution tells during trials (i.e., MNBS-P). I modified this scale in order to create three other scales that can be used to measure the story for the defense (i.e., MNBS-D) and a juror's individual story about what he or she believes happened in the events leading to trial (i.e., MNBS-J and MNBS-R).

Study 2 was an experimental study, in which MTurk workers were asked to take

on the role of a juror during a criminal trial. I randomly assigned participants to read a

case summary about a Black or a White defendant who was charged with either grand-

theft auto or embezzlement. In addition, half of the participants heard stereotype

suppression instructions, while the other half did not. Participants then made verdict

judgments, and rated the prosecution and defense stories.

Study 3 was a direct replication of Study 2 except from the fact that the sample

included college students, and that the participants were asked to put their own stories,

which they, according to the Story Model, had naturally developed while reading the case

summaries. Afterwards, they rated the coherence, coverage, and uniqueness of their

stories, before making their verdict judgments. In addition, after all the data was

collected, an unbiased coder rated each story in terms of its coherence, coverage, and

uniqueness.

A disadvantage of Study 2 and Study 3 was that participants rated the prosecution

and defense stories (and wrote down and rated their own stories in Study 3), after, and not

before, they made their final verdict judgments. Therefore, they may not have been able

to thoroughly process all the information and to evaluate each story appropriately before

they were asked to decide about guilt or innocence. Thus, in Study 4 I switched the order

around, so that the participants made their verdict judgments after they had the chance to

think about the stories that evolved during the trial.

In general, the results from Study 2, Study 3, and Study 4 did not support any of

my hypotheses. Although race-crime congruence effects on verdict judgments have been

found in past studies, these findings did not replicate in the current research. Across all

studies, White jurors were not more likely to find the defendant guilty more frequently

and to make higher culpability judgments when a Black defendant, compared to a White

defendant, committed grand-theft auto, a crime that is stereotypical of Blacks. Similarly,

White jurors did not view a White defendant who was charged with embezzlement

guiltier than a Black defendant who was charged with the same crime, one that is

typically associated with Whites. In addition, contrary to my predictions, when the crime

was consistent with the racial stereotype, neither prosecution story nor the jurors' own

stories were seen as more coherent than when no race-crime congruence occurred.

Overall, neither defendant race nor type of crime predicted verdict judgments.

At first glance, these findings seem positive because they suggest that race and

crime stereotypes seem to fade in juror decision making. It could be argued that

individuals simply may not hold those stereotypes any longer. And even if these are still

present, they are not applied in this type of setting. However, this interpretation is

questionable because past studies in the social sciences have shown that people do

engage in stereotyping in the courtroom.

College students as well as MTurk workers tend to be more educated than the

average juror. Not only may they be aware of their stereotypes much more and actively

try to counter them because "it is the right thing to do," but also are much more exposed

to similar studies that examine the effects of stereotypes. For example, MTurk workers

get paid to participate in research studies, and therefore, many of them try to receive

make as much money as possible by participating in a variety of different studies over a

short period of time. In addition, psychology students learn about prejudice and

stereotypes in their classes, and thus, they may try to figure out the purpose of each study they participate in. Just seeing a picture of a Black (or White) defendant might alert them about the purpose of the study. This could then lead to participant reactance; participants may not want to be seen as prejudicial, and therefore answer certain questions in ways that might not mirror how they would react to the same situation in real life.

Because race-crime congruence effects did not emerge in the present research, it did not allow me to explore any possible explanations for these effects (e.g., story coherence), and whether a juror's motivation to control prejudiced reactions and/or stereotype suppression instructions could alleviate these effects.

Although this was not the main focus of the research, one of my objectives was to modify the NBS-12 and validate it, so it measures the certainty principles for the Story Model. Contrary to Yale's (2013) results who concluded that five factors should be included in the model, I found that the three-factor solution of the MNBS-P provided a better fit. Accordingly, I concluded that the factors coherence, coverage, and uniqueness measure the narrative believability of a story that is present at trial, thus validating the full measure of the certainty principles outlined in the Story Model.

In addition, I found that the three principles predicted verdict judgments quite well, especially when these pertained to the prosecution story. For example, in Study 2, coherence, coverage, and uniqueness of the prosecution story were positively correlated with verdict-confidence scores as well as with probability of guilt. When jurors believe that the story presented by the prosecution is coherent, covers many of the evidence items, and appears to be the only story that makes sense, they will believe the

prosecution's narrative, and thus, support the prosecutor's point of view by finding the defendant guilty. Moreover, the coherence, coverage, and uniqueness of the defense story had negative relationships with verdict judgments. This showed that jurors who believe the narrative of the defense because it is coherent, covers many of the evidence items, and is unique are less likely to find the defendant guilty.

While the certainty principles underlying the prosecution as well as the defense story were related to a juror's verdict judgments in Study 2, results in Study 3 revealed that it appears that is mainly the believability of the prosecution story, and not so much the defense story, that drives the verdict judgments. In other words, when jurors rate the coherence, coverage, and uniqueness of the prosecution story high, they are more likely to make guilty verdict judgments. However, whether they think the defense story is coherent, unique, and covers many evidence items does not seem to influence their verdict judgments very much. This makes sense when considering the fact that the justice system of the United States is built upon the principle that "every man is innocent until proven guilty" (i.e., presumption of innocence). Therefore, the legal burden of proof lies with the prosecution—the prosecution must show that the defendant is guilty beyond a reasonable doubt. If the prosecution does not provide a "good enough" story, the defendant should be acquitted. As shown in the present research, jurors seem to fully comprehend these notions because their interpretation of the prosecution's story, and not so much of the defense's story, impacts their verdict judgments.

Overall, these findings nicely map on to Pennington & Hastie's (1984, 1986, 1988) theoretical model, thus encompassing the full spectrum of the certainty principles

that are relevant for making verdict judgments, and validating the modified versions of the NBS.

Limitations

Contrary to my predictions, the data suggests mock jurors in a laboratory setting do not employ any race and crime stereotypes when making verdict judgments. As mentioned above, although this may be good news for the criminal justice system, these findings should be interpreted with caution because several limitations may be responsible for these null effects.

One limitation of my studies is that I used college students and MTurk workers as participants. There may not have been any race-crime congruence effects on verdict judgments because my participants were able to pick up on cues about what the purpose of the study was, because they get exposed to similar studies on a regular basis. Thus, another limitation of the current research is that I employed an experimental research design, which decreases the external validity of my findings. The laboratory is artificial by nature, and thus might not lead to those "natural" behaviors that are seen in the real world.

In addition, despite my efforts to make the crimes as similar as possible in terms of seriousness and guilt appearance through conducting multiple pilot studies, participants in all four studies viewed embezzlement as the more serious of the cases, and they felt that the defendant in the grand-theft auto case appeared guiltier than the defendant who was charged with embezzlement. However, because none of the hypothesized effects emerged, this limitation is secondary because the differences in

crime seriousness and guilt appearance could not have tainted null effects.

Recommendations for Future Research

Although none of my hypotheses were supported by the results of the present research, it would be a premature decision for researchers to refrain from pursuing this line of research in futures studies. Instead of concluding that these race-crime effects do not exist (anymore), researchers should explore those in different ways, for example, by utilizing crimes that are more racially stereotypical in nature. While embezzlement or fraud are probably the most extreme White collar crimes on the crime stereotype continuum, crimes other than grand-theft auto may be seen as more stereotypical of Black defendants (e.g., selling drugs).

In addition, it would be interesting to see whether the same results would emerge when the sample consists of members of the community, who better represent the jury pool, and who do not regularly participate in research studies. Furthermore, noting the defendant race in a real-life case, coding the crime they were charged with as stereotypically Black or White, and asking jurors about their stories or about how coherent they believe the prosecution or defense stories were after the conclusion of a trial may help to get more insight into race-crime congruence effects.

Finally, it may be beneficial to use opening and closing arguments by the prosecution and defense to include in the study materials, instead of just providing the participants with a summary of the case. This may foster their understanding about what distinguishes the prosecution from the defense stories, and what evidence items help the prosecution and/or defense case.

Conclusion

In conclusion, a modified version of the NBS was created and validated; thus,

juror and jury researchers can benefit from using this measurement instrument in future

studies. In addition, the present research attempted to fill an important gap in the

literature on juror decision making and racial stereotypes. Although many non-significant

results emerged, it opened up questions about the right ways to study these topics. To

date, this was the first application of the most prominent model of juror decision making,

the Story Model, to racial prejudice in the courtroom, and its validity should be tested.

Indisputably, it is important "beyond a reasonable doubt" that social science research

continues to focus on improving the fairness of our justice system, especially for

members of minority groups, by exploring those person- as well as trial-specific factors

that could decrease the application of stereotypes.

APPENDIX A

PROSECUTION STATEMENT – MASTER NARRATIVE

Twenty-two year old James Lawrence is on trial with charges of 2nd Degree Theft and Robbery as the result of a purse-snatching incident on the evening of Friday, April 4, 2014. On that night, after a date with his girlfriend, James, who had been having money problems, saw an opportunity to make some quick cash by assaulting a woman and stealing her purse.

The evidence will show that on the evening of April 4, Ms. Fitzgerald was working late at her office and did not leave until around 9:15pm. She caught the bus outside of her office building and got off the bus at the stop at the corner of 5th St. and Main St. at approximately 9:45pm. From the bus stop, she walked east on Main St. toward her apartment at 406 Main St. At about the middle of the block where her apartment building is located, she heard footsteps behind her and felt a sharp pull on her purse. She struggled briefly with her attacker, who threw her to the ground and ran off with the purse across the street and back toward 5th St.

Immediately after the incident, Ms. Fitzgerald, the victim of the robbery, returned to her apartment and called the police, who came and took her statement about the robbery. The next day, Ms. Fitzgerald's purse was found in a mailbox across the street from Dorothy's Bar, at 816 Main St.

Two days later, on April 6, Ms. Fitzgerald went down to the police station to sign the statement that she had given on the night of the robbery and to identify the purse found in the mailbox. She identified the purse and determined that all of the cash inside (approximately $300) was missing. At that time, she was shown several books of mug shots of known thieves, but she was not able to identify her attacker form the photographs.

On April 18, 2014, James Lawrence was arrested in the 300-block of Davenport Avenue after attempting to steal the purse of an off-duty police officer, Ms. Sonia Henderson. Mr. Lawrence's mug shot was taken after he was booked into jail for the crime. Later that day, Ms. Fitzgerald returned to the police station and looked at the books of mug shots again. On this occasion, Ms. Fitzgerald quickly identified James Lawrence as her attacker.

James Lawrence claims that he had nothing to do with the theft of Ms. Fitzgerald's purse on the night of April 4. However, the evidence tells a different story.

The evidence will show that on the night of April 4, 2014, James Lawrence was out on a date with his high-school sweetheart, Chelsea Williams. Ms. Williams will testify that on

that evening, James met here when she got off work around 5:30 and that they went out to Dairy Queen for dinner before heading to the theater for a movie. When they arrived at Dairy Queen, James got upset because he didn't have enough money to pay for both his meal and Chelsea's. According to Ms. Williams, James never seemed to have very much money from the odd yard work and cleanup jobs he does, especially after paying rent. This evening, he seemed especially upset that he couldn't treat Ms. Williams to dinner. Even James admitted in an interview that he was short on money and couldn't pay for Chelsea's dinner on the night in question.

After dinner, the couple went to Varsity Cinemas to see the 1962 movie "To Kill a Mockingbird," which was playing as part of the Summer Classic Theater Series. According to Ms. Williams' testimony, after the movie, James walked her home, kissed her goodnight, and left, around 9:30pm. This timeline would have given James plenty of time to walk the 10 blocks to the location on Main St. where Ms. Fitzgerald was attacked.

Ms. Fitzgerald's purse was found in a mailbox near Dorothy's Bar. During an interview after his arrest for the attempted theft of Officer Henderson's purse, James admitted that he occasionally went to Dorothy's, but not very often because "it's pretty expensive." The bartender at Dorothy's recognized James form a photograph, but was not able to state with certainty whether or not James was at the bar on the night of April 4.

James Lawrence is clearly guilty of this crime. He was admittedly short on cash, and was upset because he wasn't able to provide for his girlfriend that night. On his way home, he saw an opportunity to make some easy money, and he took it, and then he went to an expensive bar on his way home to celebrate. In fact, it was such an easy way to make money that a few weeks later, Mr. Lawrence tried it again, but this time, his target was an off-duty police officer who was able to restrain and arrest him. Once the new mug shot was placed in the book of suspects, Ms. Fitzgerald was able to immediately identify James Lawrence as her attacker. This is not a complex case. As a juror, it is your duty to vote to convict James Lawrence.

APPENDIX B

PROSECUTION STATEMENT – LOW COVERAGE

Twenty-two year old James Lawrence is on trial with charges of 2nd Degree Theft and Robbery as the result of a purse-snatching incident on the evening of Friday, April 4, 2014. On that night, after a date with his girlfriend, James, who had been having money problems, saw an opportunity to make some quick cash by assaulting a woman and stealing her purse.

The evidence will show that on the evening of April 4, Ms. Fitzgerald was working late at her office and did not leave until around 9:15pm. She caught the bus outside of her office building and got off the bus at the stop at the corner of 5th St. and Main St. at approximately 9:45pm. From the bus stop, she walked east on Main St. toward her apartment at 406 Main St. At about the middle of the block where her apartment building is located, she heard footsteps behind her and felt a sharp pull on her purse. She struggled briefly with her attacker, who threw her to the ground and ran off with the purse across the street and back toward 5th St.

Immediately after the incident, Ms. Fitzgerald, the victim of the robbery, returned to her apartment and called the police, who came and took her statement about the robbery. The next day, Ms. Fitzgerald's purse was found in a mailbox across the street from Hawg Heaven Saloon, a bar frequented by motorcyclists.

Two days later, on April 6, Ms. Fitzgerald went down to the police station to sign the statement that she had given on the night of the robbery and to identify the purse found in the mailbox. She identified the purse and determined that all of the cash inside (approximately $300) was missing. At that time, she was shown several books of mug shots of known thieves, but she was not able to identify her attacker form the photographs.

On April 18, 2014, James Lawrence was arrested in the 300-block of Davenport Avenue after attempting to steal the purse of an off-duty police officer, Ms. Sonia Henderson. Mr. Lawrence's mug shot was taken after he was booked into jail for the crime. Later that day, Ms. Fitzgerald returned to the police station and looked at the books of mug shots again. On this occasion, Ms. Fitzgerald quickly identified James Lawrence as her attacker.

James Lawrence claims that he had nothing to do with the theft of Ms. Fitzgerald's purse on the night of April 4. However, the evidence tells a different story.

The evidence will show that on the night of April 4, 2014, James Lawrence was out on a date with his high-school sweetheart, Chelsea Williams. Ms. Williams will testify that on

that evening, James met here when she got off work around 5:30 and that they went out to Dairy Queen for dinner before heading to the theater for a movie. When they arrived at Dairy Queen, James got upset because he didn't have enough money to pay for both his meal and Chelsea's. According to Ms. Williams, James never seemed to have very much money from the odd yard work and cleanup jobs he does, especially after paying rent. This evening, he seemed especially upset that he couldn't treat Ms. Williams to dinner. Even James admitted in an interview that he was short on money and couldn't pay for Chelsea's dinner on the night in question.

After dinner, the couple went to Varsity Cinemas to see the 1962 movie "To Kill a Mockingbird," which was playing as part of the Summer Classic Theater Series. According to Ms. Williams' testimony, after the movie, James walked her home, kissed her goodnight, and left, around 9:30pm.

According to the statement given by James Lawrence, after leaving Chelsea's apartment, he went to the 7-Eleven on 9th Street and browsed the magazines before heading home. Security camera footage confirms that James was in the store from 9:36pm to 9:40pm.

Ms. Fitzgerald's purse was found in a mailbox near Hawg Heaven Saloon. During an interview after his arrest for the attempted theft of Officer Henderson's purse, James stated that the "never went in that place. If you're not a biker, that's not a good place to go." The bartenders at Hawg Heaven did not recognized James from a photograph.

The purse was dusted for fingerprints, and no prints from James Lawrence were present on the purse or the wallet inside. There were fingerprints of an unidentified person on the purse handles and wallet. We believe that someone else may have found the purse and checked for money inside the wallet before the purse was found by the owner of the bar.

James Lawrence is clearly guilty of this crime. He was admittedly short on cash, and was upset because he wasn't able to provide for his girlfriend that night. On his way home, he saw an opportunity to make some easy money, and he took it, and then he went to a bar to celebrate. In fact, it was such an easy way to make money that a few weeks later, Mr. Lawrence tried it again, but this time, his target was an off-duty police officer who was able to restrain and arrest him. Once the new mug shot was placed in the book of suspects, Ms. Fitzgerald was able to immediately identify James Lawrence as her attacker. This is not a complex case. As a juror, it is your duty to vote to convict James Lawrence.

APPENDIX C

PROSECUTION STATEMENT – LOW CONSISTENCY

Twenty-two year old James Lawrence is on trial with charges of 2nd Degree Theft and Robbery as the result of a purse-snatching incident on the evening of Friday, April 4, 2014. On that night, after a date with his girlfriend, James, who had been having money problems, saw an opportunity to make some quick cash by assaulting a woman and stealing her purse.

The evidence will show that on the evening of April 4, Ms. Fitzgerald was working late at her office and did not leave until around 9:15pm. She caught the bus outside of her office building and got off the bus at the stop at the corner of 5th St. and Main St. at approximately 9:45pm. From the bus stop, she walked east on Main St. toward her apartment at 406 Main St. At about the middle of the block where her apartment building is located, she heard footsteps behind her and felt a sharp pull on her purse. She struggled briefly with her attacker, who threw her to the ground and ran off with the purse across the street and back toward 5th St.

Immediately after the incident, Ms. Fitzgerald, the victim of the robbery, returned to her apartment and called the police, who came and took her statement about the robbery. At that time, she stated that her attacker was between 5'8" and 5'10". The next day, Ms. Fitzgerald's purse was found in a mailbox across the street from Hawg Heaven Saloon, at 310 Newton.

Two days later, on April 6, Ms. Fitzgerald went down to the police station to sign the statement that she had given on the night of the robbery and to identify the purse found in the mailbox. She identified the purse and determined that all of the cash inside (approximately $300) was missing. At that time, she was shown several books of mug shots of known thieves. Even though James Lawrence's picture from a year earlier was in the book of mug shots, Ms. Fitzgerald stated:

> "Whoever it was that attacked me isn't in this book. None of these men look familiar to me."

On April 18, 2014, Ms. Fitzgerald returned to the police station and looked at the books of mug shots again. On this occasion, Ms. Fitzgerald quickly identified James Lawrence as her attacker, stating:

> "Well, I wouldn't put my life on it, it was dark you know, but this one definitely looks like him. The face, there was barely enough light for me to see his face, and it was only for a few seconds. This looks like the guy. No, I'm sure this is the guy. It must be."

James Lawrence claims that he had nothing to do with the theft of Ms. Fitzgerald's purse on the night of April 4. However, the evidence tells a different story.

The evidence will show that on the night of April 4, 2014, James Lawrence was out on a date with his high-school sweetheart, Chelsea Williams. Ms. Williams will testify that on that evening, James met here when she got off work around 5:30 and that they went out to Applebee's for dinner before heading to the theater for a movie. James paid for dinner and the movie. Chelsea specifically remembered that James made a big show of paying for the date. She stated:

> "James never seems to have very much money from the landscaping and yard work he does, but he'd just gotten back-pay from a customer who was behind, so He'd been able to pay his rent early and had lots of money in the bank. He doesn't like it, but a lot of times when we go out I pay for myself because he doesn't have enough money."

James made a sizable deposit into his bank account (about $1500) on the morning of April 4. When asked about the deposit in an interview, James said:

> "That money was from Andrew Miller. I do landscaping and yard work for him. He was about four months behind paying me for some work I had done on his property, and he finally paid me.

Mr. Miller confirmed that he had paid James for landscaping work on April 4.

After dinner, the couple went to Varsity Cinemas to see the 1962 movie "To Kill a Mockingbird," which was playing as part of the Summer Classic Theater Series. According to Ms. Williams' testimony, after the movie, James walked her home, kissed her goodnight, and left, around 9:30pm. This timeline would have given James plenty of time to walk the 10 blocks to the location on Main St. where Ms. Fitzgerald was attacked.

Ms. Fitzgerald's purse was found in a mailbox near the Hawg Heaven Saloon. During an interview after his arrest for the attempted theft of Officer Henderson's purse, James stated that he "never went in that place. If you're not a biker, that's not a good place to go." The bartenders at Hawg Heaven did not recognize James from a photograph. However, James Lawrence is 6'2" tall, and would probably not be afraid of bikers.

James Lawrence is clearly guilty of this crime. He had a history of money problems and was often upset when he didn't have money to pay for dates with his girlfriend. On his way home from his date that night, he saw an opportunity to make some easy money, and

he took it, and then he went to an expensive bar on his way home to celebrate. Ms. Fitzgerald was able to identify James Lawrence as her attacker. This is not a complex case. As a juror, it is your duty to vote to convict James Lawrence.

APPENDIX D

PROSECUTION STATEMENT – LOW PLAUSIBILITY

Twenty-two year old James Lawrence is on trial with charges of 2nd Degree Theft and Robbery as the result of a purse-snatching incident on the evening of Friday, April 4, 2014. On that night, after a date with his girlfriend, James, who had been having money problems, saw an opportunity to make some quick cash by assaulting a woman and stealing her purse.

The evidence will show that on the evening of April 4, Ms. Fitzgerald was working late at her office and did not leave until around 9:15pm. She caught the bus outside of her office building and got off the bus at the stop at the corner of 5th St. and Main St. at approximately 9:45pm. From the bus stop, she walked east on Main St. toward her apartment at 406 Main St. At about the middle of the block where her apartment building is located, she heard footsteps behind her and felt a sharp pull on her purse. She struggled briefly with her attacker, who threw her to the ground and ran off with the purse across the street and back toward 5th St.

Immediately after the incident, Ms. Fitzgerald, the victim of the robbery, returned to her apartment and called the police, who came and took her statement about the robbery. The next day, Ms. Fitzgerald's purse was found in a mailbox across the street from Dorothy's Bar, at 816 Main St.

Two days later, on April 6, Ms. Fitzgerald went down to the police station to sign the statement that she had given on the night of the robbery. At that time, she was shown several books of mug shots, one of which contained a photo of James Lawrence. During this interview, Ms. Fitzgerald did not identify James as her attacker.

On April 18, 2014, two weeks after the incident, Ms. Fitzgerald returned to the police station to look at the mug shots again. This time, she positively identified him as the thief. Here's what she said during that interview:

> "Well, I wouldn't put my life on it, it was dark you know, but I definitely recognize the tattoo on his right arm. His arm was right next to me. The face, there was barely enough light for me to see his face, and it was only for a few seconds. This looks like the guy. No, I'm sure this is the guy. It must be."

James Lawrence claims that he had nothing to do with the theft of Ms. Fitzgerald's purse on the night of April 4. However, the evidence tells a different story.

The evidence will show that on the night of April 4, 2014, James Lawrence was out on a date with his high-school sweetheart, Chelsea Williams. Ms. Williams will testify that on that evening, James met here when she got off work around 5:30 and that they went out

to Applebee's for dinner before heading to the theater for a movie. James paid for dinner.

After dinner, the couple went to Varsity Cinemas to see the 1962 movie "To Kill a Mockingbird," which was playing as part of the Summer Classic Theater Series. James paid for the movie and popcorn and candy. According to Ms. Williams' testimony, after the movie, James walked her home, kissed her goodnight, and left, around 9:30pm. This timeline would have given James plenty of time to walk the 10 blocks to the location on Main St. where Ms. Fitzgerald was attacked.

When interviewed James insisted that he had nothing to do with the purse snatching. He stated that he has worked very hard over the past several years to turn his life around. For the past year and a half, he's been volunteering on weekends at the Boys and Girls Club, helping with their program for kids. He started his own lawn-care and landscaping business a little more than a year ago, and although he's no millionaire, he makes a decent living. Just last year James finally saved enough money for a down payment on his own place here in town, and was able to move out of his parents' house. He says that he had no reason to go back jeopardize everything he's accomplished by stealing Ms. Fitzgerald's purse.

Ms. Fitzgerald's purse was found in a mailbox near Dorothy's Bar. During an interview after his arrest for the attempted theft of Officer Henderson's purse, James admitted that he occasionally went to Dorothy's, but not very often because "it's pretty expensive." The bartender at Dorothy's recognized James form a photograph, but was not able to state with certainty whether or not James was at the bar on the night of April 4.

James Lawrence is clearly guilty of this crime. He has a history of petty crime, which is why his picture was in the mug shot book in the first place. On his way home, he saw an opportunity to make some easy money, and he took it, and then he went to an expensive bar on his way home to celebrate. Ms. Fitzgerald was able to positively identify James Lawrence as her attacker. This is not a complex case. As a juror, it is your duty to vote to convict James Lawrence.

APPENDIX E

PROSECUTION STATEMENT – LOW COMPLETENESS

- James Lawrence is on trial with charges of 2nd Degree Theft and Robbery.
- Ms. Gale Fitzgerald worked late and left her office around 9:15 PM on April 4, 2014.
- Ms. Fitzgerald rode the bus home, getting off the bus at 5th and Main St. at 9:45 PM.
- From the bus stop, Ms. Fitzgerald walked toward her apartment at 406 Main ST.
- Halfway down the block, she heard footsteps and felt a sharp pull on her purse.
- Ms. Fitzgerald struggled briefly with her attacker who threw her to the ground.
- The attacker ran off with her purse back toward 5th St.
- Ms. Fitzgerald returned to her apartment after the attack and took her statement.
- Ms. Fitzgerald's purse was found, without the money in it, in a mailbox near Dorothy's Bar.
- Ms. Fitzgerald went to the police station on April 6, 2014, to identify the purse found in the mailbox.
- Ms. Fitzgerald looked at several books of mug shots of known thieves, but she was not able to identify her attacker from the photographs.
- James Lawrence's mug shot was in the book that Ms. Fitzgerald examined on April 6.
- James Lawrence was arrested on April 18, 2014 after attempting to steal the purse of an off-duty police officer, Ms. Sonia Henderson.
- After the attempted purse-snatching on April 18, 2014, James Lawrence had new mug shots taken when he arrived at the jail.
- Ms. Fitzgerald returned to the police station on April 18, 2014, to look at mug shots again. James Lawrence had brand new mug shots in the book on April 18, 2014, from his arrest from the attempted purse snatching of Officer Henderson.
- On April 18, 2014, Ms. Fitzgerald identified James Lawrence as the man who attacked her and stole her purse.

Statements of Gale Fitzgerald

- "I am twenty-eight years old."
- "I am single and live in an apartment at 406 Main St."
- "On April 4, 2014, I worked at the office until a little after 9 PM."
- "I left the office at 9:15 PM and caught a bus right away outside the building."
- "I got off the bus at the stop at 5th and Main St. around 9:45pm."
- "I would say that the visibility was fairly good given that it was nighttime."

- "As I got to a fire hydrant about the middle of the block, I heard fast footsteps behind me."
- "A sharp tug on my purse turned me around and I was facing my attacker."
- "I know it was stupid, but I struggled with him."
- "There was more than $300 in my purse."
- "We struggled for few seconds, and he threw me to the ground."
- "My attacker was a white man, approximately 5'8" to 5'10" in height, and 160-175 points, with dark hair (dark brown or black), wearing dark pants, a white tank-top shirt, and running shoes."
- "I have been shown several books of mug shot photographs and have not been able to identify my attacker among them."
- "I wouldn't put my life on it, it was dark you know, but I definitely recognize the tattoo on his right arm."
- "The face, there was barely enough light for me to see his face, and it was only for a few seconds."
- "This looks like the guy. No, I'm sure this is the guy. It must be."

Statements of Chelsea Williams

- I am nineteen years old and live with two girlfriends in an apartment at 1030 Elm St."
- "I currently work during the day as a clerk at the Sunglass Hut at the mall."
- "I go to school at night at the community college for cosmetology classes."
- "I have known James Lawrence since high school."
- "James dropped out of high school."
- "A few months ago, James came by the Sunglass Hut and struck up a conversation with me."
- "He asked me to meet him after work to talk some more. We hit it off pretty good."
- "James told me he has his own business doing yard work and clean up."
- "I know I am here at the police station because James is in some kind of trouble."
- "James called me yesterday and asked me questions about the date we had a few weeks ago when we went to see "To Kill a Mockingbird" at the Varsity Theater."
- "He asked me what night we saw the movie, but I couldn't remember at first."
- "I figured out that we saw the movie on a Friday, which would have been the beginning of the month."

- "I have cosmetology classes on Monday, Tuesday, and Wednesday nights, so I couldn't have been out with James on those nights."
- "That night, James met me after work and we walked to Dairy Queen for dinner."
- "James doesn't make much money, and even though it upsets him, I insist on paying my way."
- "That night, James had just paid his rent."
- "The movie started around 7:00 PM."
- "We got back to my apartment after the movie around 9:15 or 9:30. I was really tired, so James only stayed for 10 or 15 minutes."
- "I don't remember what James was wearing that night."
- "James normally wears jeans, air sneakers, a tee shirt or tank top, and a jacket, kind of dark in color."

Statements of James Lawrence

- "I am twenty-one years old and I live at 523 Maple St."
- "I am just over six feet talk and I weigh 175 pounds."
- "I have never stolen a purse before tonight when I got caught trying to steal a purse from Officer Henderson."
- "On April 4th, I was with my girlfriend, Chelsea Williams. We saw "To Kill a Mockingbird" at the Varsity Theater near her apartment."
- "After the movie, I walked her home. I stayed at her apartment until 9:30 or 10 PM. Then I walked home."
- "I always walk the same way home from Chelsea's: I walk over to 8th Avenue, down 8th to Maple St., and then over to my apartment at 523 Maple."
- "I like to walk down 8th Avenue because there are some interesting stores and I like to window shop."
- "I've been to Dorothy's, but I'm not a regular or anything. It's pretty expensive. I don't drink much, maybe a beer or two, but that's it. Especially at those prices."
- "I don't think I was at Dorothy's on April 4th. I was really short of money a couple of weeks ago."
- "I do lawns and yard clean-up for people, and some of my customers had been slow paying."
- "I had to pay my rent for the month on April 4th."
- "I couldn't afford to pay for Chelsea's meal at some fast food place on that night, so I doubt if I stopped in to Dorothy's."
- "I have never heard of Gale Fitzgerald."

- "I might have been to her [Ms. Fitzgerald's] neighborhood. It's only a couple of blocks from my apartment, but I have no reason to go to that block."
- "If she [Ms. Fitzgerald] said that I attacked her, she must be mistaken, and it's dark at 9:45 at night and there are lots of trees in that neighborhood. Just houses, no stores. There aren't many street lights either."

APPENDIX F

PROSECUTION STATEMENT – LOW UNIQUENESS

James Lawrence had nothing to do with the theft of Ms. Fitzgerald's purse on the night of April 4, 2014. The evidence will show that on that evening James Lawrence did not assault Ms. Fitzgerald because he was about to be paid, and because he had an injury that would have prevented him from attacking someone, even if he had wanted to. In addition, James was not in near the scene of the crime that night.

There is no doubt that James Lawrence went to dinner and a movie with his girlfriend, Chelsea Williams, that night. There is also no doubt about the fact that James did not have enough money, and got verbally upset about not being able to pay. The prosecution argues that James' motive for attacking Ms. Fitzgerald was that he was short on money. However, James expected to get a paycheck in the amount of $535 from a client the following day. This client, Mr. Henderson, confirmed that he and James Lawrence had a meeting scheduled for the following morning, and he planned to give him the money.

James Lawrence collided with a car while riding home from work on his bike on March, 30, five days before Ms. Fitzgerald was attacked by an unknown person. The accident caused him a sprained ankle and multiple lacerations on the left side of his body. James left his girlfriend's house at around 9:30, according to James himself and Ms. Williams. Ms. Fitzgerald stated that she was attacked briefly after she got off the bus at approximately 9:45pm. It is impossible that James could have walked with a sprained ankle the 10 blocks from Ms. William's house to 5th and Main St. in 15 minutes. Ms. Williams testified that James was in pain all evening, and was walking with a limp. He was also walking with crutches. It actually took James to walk 45 minutes from Ms. William's apartment to his apartment, which is a distance of 12 blocks. During an interview, his next-door neighbor, Teresa Johnson, said that she ran into James at around 10:15pm when he was on her way out and he had just gotten home. In her testimony, Ms. Fitzgerald never mentioned that a man with crutches attacked her.

James Lawrence admits that he tried to steal the police officer's purse on April 18, 2014. He will plead guilty to that charge. However, this incident is irrelevant for this case. James Lawrence is clearly innocent of stealing Ms. Fitzgerald's purse. He was in no physical condition to attack someone and steal money on April 4 because he was walking with crutches, and expected to be paid the following day. As a juror, it is your duty to vote to acquit James Lawrence.

APPENDIX G

MODIFIED NARRATIVE BELIEVABILITY SCALE – PROSECUTION (MNBS-P)

Instructions: Now think about the case you just read from the perspective of the prosecution. Think about what the basic story of the prosecution was and rate the prosecution story by answering the following questions. Please circle the response that best reflects your opinion about the prosecution's description of what happened (i.e., prosecution story). There are no correct answers.

	strongly disagree					strongly agree	
I believe the prosecution story could be true.	1	2	3	4	5	6	7
It was easy to follow the prosecution story from beginning to end.	1	2	3	4	5	6	7
The information presented in the prosecution story was consistent.	1	2	3	4	5	6	7
There was important information missing from the prosecution story.	1	2	3	4	5	6	7
The prosecution story was plausible.	1	2	3	4	5	6	7
It was hard to follow the prosecution story.	1	2	3	4	5	6	7
All of the facts in the prosecution story agreed with each other.	1	2	3	4	5	6	7
There were a lot of "holes" in the prosecution story.	1	2	3	4	5	6	7
The prosecution story seems to be true.	1	2	3	4	5	6	7
If I were given the prosecution story, I would have organized it differently.	1	2	3	4	5	6	7
The prosecution story stands out as the only possible description of what happened.	1	2	3	4	5	6	7
I could generate a number of additional stories from the evidence.	1	2	3	4	5	6	7
The prosecution story is irreplaceable.	1	2	3	4	5	6	7
The prosecution story provides a unique explanation of the evidence.	1	2	3	4	5	6	7

	strongly disagree						strongly agree

Multiple unique stories can explain the evidence.	1	2	3	4	5	6	7
At least one alternative explanation exists.	1	2	3	4	5	6	7

The "coverage" of a story refers to the extent to which the story accounts for all of the information presented in the story. How would you rate the prosecution's story in terms of "coverage"?

1	2	3	4	5	6	7
very low						very high

The "consistency" of a story refers to the extent to which a story does not contradict itself or contradict other things you know to be true or false. How would you rate the prosecution's story in terms of "consistency"?

1	2	3	4	5	6	7
very low						very high

APPENDIX H

STUDY 1 DEMOGRAPHIC QUESTIONNAIRE

Instructions: Please answer the following questions about yourself.

What is your age? _____

What is your gender?   Male   Female

What is your ethnicity?

       European American / White
       African American / Black
       Asian
       Latino/a
       Middle Eastern
       Multi-racial (please explain): _____
       Other (please indicate): _____

What is your level of education?

       Some high school
       High school graduate
       Some college
       college graduate (i.e., 4-year degree)
       Some graduate school
       Graduate school degree

APPENDIX I

DEFENDANT PROFILES

The defendant Jamal R. Williams is charged with grand-theft auto.

Name: Jamal R. Williams

Sex: Male

Race: Black

Age: 40

Charge: Grand-theft auto

Date of arrest: 05/06/2014

The defendant Jamal R. Williams is charged with embezzlement.

Name: Jamal R. Williams

Sex: Male

Race: Black

Age: 40

Charge: Embezzlement

Date of arrest: 05/06/2014

The defendant Jack R. Williams is charged with grand-theft auto.

Name: Jack R. Williams

Sex: Male

Race: White

Age: 40

Charge: Grand-theft auto

Date of arrest: 05/06/2014

The defendant Jack Williams is charged with grand-theft auto.

Name: Jack R. Williams

Sex: Male

Race: White

Age: 40

Charge: Grand-theft auto

Date of arrest: 05/06/2014

APPENDIX J

CASE SUMMARIES

Embezzlement

R. Campbell was charged with embezzlement. The charge alleges that Campbell, as vice president and trust officer of Citizens Trust Company of Atlanta, Georgia, had embezzled and willfully misapplied bank funds and made false entries in bank statements and records for the purpose of deceiving bank officers and bank examiners. According to testimony provided by bank officials, Campbell allegedly made and approved loans to fictitious persons in several transactions, and used the proceeds to purchase real estate. On other occasions, the officials testified that Campbell took bank funds, converted them into cashier's checks, and exchanged the checks for bank stock from L. D. Milton, the president and chief executive officer of Citizens Trust. They said Campbell also debited the bank's operating accounts and deposited the proceeds to his personal checking account to meet debts incurred as a result of his political campaign for city alderman.

For most of his life, Campbell had been a protégé of the bank's president, L. D. Milton. Mr. Milton served as the bank's president for fifty years. At trial, Campbell contended that the transactions had been performed at Milton's direction. Campbell testified that the real estate was acquired on behalf of the bank and the fictitious names were used to obtain the loans and purchase the real estate because the bank, according to regulatory agencies, was already over-invested in real estate and Milton wanted to conceal these further acquisitions. Campbell contended that the other transactions were loans of bank funds to Milton and that Campbell was holding the stock as collateral for the bank. He also explained that the transfers of bank funds to reduce the overdrafts in his personal checking accounts were made at Milton's suggestion. Campbell also claimed

that he was not in debt from his political campaign, and investigation disclosed that his campaign was in debt for less than $300.00.

Milton testified that he did not recall telling or suggesting to Campbell that he use bank funds for any of the transactions mentioned. He further stated that Campbell had come to him in desperation over his mounting campaign costs, saying he needed to pay his creditors back as soon as possible. He further testified that Campbell had approached him early in the year and had asked him for a personal loan. When he refused, Campbell persisted and asked him if he would be willing to make him a loan through the bank. Again Milton refused, saying he could not breach the rules of the bank. However, on cross examination it was discovered that Milton had on a few occasions approved bank operations that violated minor banking regulations or were otherwise irregular. Some of these irregularities included allowing bank employees to overdraw on their personal accounts, provided that they cover the check within a specified time.

T. S. Davis, a real estate agent in Atlanta, testified that Campbell had approached him and inquired about purchasing land that could be developed. After showing him several sites, Campbell purchased four lots of land. When Davis asked him what he intended to do with the land, Campbell replied that he knew of several out of state developers that were interested in building office buildings in the Atlanta area. Davis stated that all his correspondence with Campbell was on the bank's stationery. However, he did not know if these transactions were for Campbell' personal gain or for the bank.

Grand-Theft Auto

R. Campbell was charged with grand-theft auto. According to court records, he

went to Parker Cadillac Company in Murray, Kentucky, on the morning in question.

After looking at the used cars on the lot, he decided that he liked a 2004 Cadillac

Escalade the best. He drove this car around the lot, and then came back to talk to the sales

man for Parker Cadillac. P. T. Rogers, the salesman for Parker Cadillac, testified that the

defendant identified himself as Campbell, and asked him if he could take the car off the

lot in order for a mechanic to look at it. Rogers checked with the owner, F. Parker, who

agreed to let Campbell take the car off the lot, providing he left the phone number of the

mechanic. Later that day, when Campbell had not returned with the car, Rogers called the

mechanic to see what the problem was. The mechanic said that he did not know a Mr.

Campbell, nor did Campbell bring in an Escalade for him to look at. Parker immediately

reported the car stolen. Additional evidence further showed that Campbell, driving a 2004

Cadillac Escalade, bearing the same vehicle identification number as the one reported

stolen in Murray, Kentucky, was arrested in Ellsworth, Kansas, for speeding. After a

routine check, it was learned that the car had been reported stolen from Murray,

Kentucky. Campbell was then arrested.

Campbell testified that he had never been on the used car lot on the day in

question nor had he talked to Rogers. He also claimed that he first saw the car when a

friend of his, L. L. Worthy, came to his residence in Murray, Kentucky, and claimed that

the car was his girlfriend's car. Campbell stated next that Worthy then asked him to

accompany him to Colorado so that he (Worthy) could look for a job, which Campbell

did. They went to Colorado and at this time he dropped Worthy off. Worthy then asked Campbell if he would mind taking the car back to his girlfriend in Kentucky. Campbell agreed and proceeded to go back to Kentucky. This is when he was stopped for speeding in Kansas. The defense concluded that Campbell had, at this time, no reason to believe that the car was stolen since he assumed that the car belonged to Worthy's girlfriend and that Worthy had permission to drive it.

Rogers, the salesman at Parker Cadillac, identified Campbell as the person who had been on the car lot on the morning in question and that he was the one who had driven the car off of the lot. However, under cross-examination, Rogers, who had admitted to dealing with many people that day, could not make a positive identification.

Officer Tabor, who had arrested Campbell in Ellsworth, Kansas, identified him as the driver of the stolen Escalade and testified that Campbell did not mention anything about the car belonging to the girlfriend of L. L. Worthy.

No L. L. Worthy could be located in Colorado at the address at which Campbell said he dropped him off, though it was found that an L. L. Worthy had lived in Murray, Kentucky.

Worthy's girlfriend testified that she and Worthy had recently split up and she had not seen him since.

APPENDIX K

FINAL VERSION OF THE MODIFIED NARRATIVE BELIEVABILITY SCALE

(MNBS-P)

Instructions: Now think about the case you just read from the perspective of the prosecution. Think about what the basic story of the prosecution was and rate the prosecution story by answering the following questions. Please circle the response that best reflects your opinion about the prosecution's description of what happened (i.e., prosecution story). There are no correct answers.

	strongly disagree					strongly agree	
I believe the prosecution story could be true.	1	2	3	4	5	6	7
It was easy to follow the prosecution story from beginning to end.	1	2	3	4	5	6	7
The information presented in the prosecution story was consistent.	1	2	3	4	5	6	7
There was important information missing from the prosecution story.	1	2	3	4	5	6	7
The prosecution story was plausible.	1	2	3	4	5	6	7
It was hard to follow the prosecution story.	1	2	3	4	5	6	7
All of the facts in the prosecution story agreed with each other.	1	2	3	4	5	6	7
There were a lot of "holes" in the prosecution story.	1	2	3	4	5	6	7
The prosecution story seems to be true.	1	2	3	4	5	6	7
If I were given the prosecution story, I would have organized it differently.	1	2	3	4	5	6	7
The prosecution story stands out as the only possible description of what happened.	1	2	3	4	5	6	7
I could generate a number of additional stories from the evidence.	1	2	3	4	5	6	7
The prosecution story is irreplaceable.	1	2	3	4	5	6	7
Multiple unique stories can explain the evidence.	1	2	3	4	5	6	7
At least one alternative explanation exists.	1	2	3	4	5	6	7

The "coverage" of a story refers to the extent to which the story accounts for all of the information presented in the story. How would you rate the prosecution's story in terms of "coverage"?

1	2	3	4	5	6	7
very low						very high

The "consistency" of a story refers to the extent to which a story does not contradict itself or contradict other things you know to be true or false. How would you rate the prosecution's story in terms of "consistency"?

1	2	3	4	5	6	7
very low						very high

APPENDIX L

MODIFIED NARRATIVE BELIEVABILITY SCALE – DEFENSE (MNBS-D)

Instructions: Now think about the case you just read from the perspective of the prosecution. Think about what the basic story of the defense was and rate the defense story by answering the following questions. Please circle the response that best reflects your opinion about the defense's description of what happened (i.e., defense story). There are no correct answers.

	strongly disagree					strongly agree	
I believe the defense story could be true.	1	2	3	4	5	6	7
It was easy to follow the defense story from beginning to end.	1	2	3	4	5	6	7
The information presented in the defense story was consistent.	1	2	3	4	5	6	7
There was important information missing from the defense story.	1	2	3	4	5	6	7
The defense story was plausible.	1	2	3	4	5	6	7
It was hard to follow the defense story.	1	2	3	4	5	6	7
All of the facts in the defense story agreed with each other.	1	2	3	4	5	6	7
There were a lot of "holes" in the defense story.	1	2	3	4	5	6	7
The defense story seems to be true	1	2	3	4	5	6	7
If I were given the defense story, I would have organized it differently.	1	2	3	4	5	6	7
The defense story stands out as the only possible description of what happened.	1	2	3	4	5	6	7
I could generate a number of additional stories from the evidence.	1	2	3	4	5	6	7
The defense story is irreplaceable.	1	2	3	4	5	6	7
Multiple unique stories can explain the evidence.	1	2	3	4	5	6	7
At least one alternative explanation exists.	1	2	3	4	5	6	7

The "coverage" of a story refers to the extent to which the story accounts for all of the information presented in the story. How would you rate the defense's story in terms of "coverage"?

1	2	3	4	5	6	7
very low						very high

The "consistency" of a story refers to the extent to which a story does not contradict itself or contradict other things you know to be true or false. How would you rate the defense story in terms of "consistency"?

1	2	3	4	5	6	7
very low						very high

APPENDIX M

EXPLANATIONS OF THE CASE LAW

Grand-Theft Auto:

Before rendering a verdict decision, it is important that you understand the relevant case law. Please briefly familiarize yourself with the following legal definition of grand-theft auto.

Every person who steals, takes, or drives away the automobile of another with the specific intent to deprive him or her permanently of his or her property is guilty of the crime of grand theft of an automobile. In order to prove this crime, each of the following elements must be proved:

(1) A person took or drove away an automobile belonging to another person; and

(2) When the person took or drove away the automobile, he or she had the specific intent to deprive the owner permanently of his or her property.

Embezzlement:

Before rendering a verdict decision, it is important that you understand the relevant case law. Please briefly familiarize yourself with the following legal definition of or embezzlement.

An individual associated as an insider with a covered institution who embezzles, abstracts, purloins, or willfully misapplies any of the moneys, funds or credits of such institution or any moneys, funds, assets, or securities entrusted to the custody of care of such institution, or to the custody or care of any such insider is guilty of the crime of embezzlement. In order to prove this crime, each of the following elements must be proved:

(1) That the defendant was an officer, director, agent or employee of, or connected in any capacity with a bank; and

(2) That the defendant embezzled, abstracted, purloined, or willfully misapplied the bank's funds or funds entrusted to the bank or to the defendant; and

(3) That the defendant acted knowingly, with a specific intent to injure or defraud the bank.

APPENDIX N

VERDICT JUDGMENTS

Instructions: After reviewing all the information presented in this case, please answer the following questions.

What is your final verdict decision?   not guilty   guilty

Please rate how confident you are in your verdict decision.

1	2	3	4	5	6	7
not confident at all						completely confident

Please rate how probable you think it is that this particular defendant committed this specific crime. Please write a number between 0 and 100 in the space below.

- 0 = The defendant could not possibly have committed the crime.

- 100 = The defendant committed the crime with absolute certainty.

Please enter a number between 0 and 100: _____

APPENDIX O

QUESTIONS PROBING FOR SUSPICION

Instructions: From this point on, please <u>do not</u> use the "prev" button.

If you had to guess, what do you think we were trying to find with this study?

Were you suspicious at any point during this study? If so, please describe about what you were suspicious.

Before today, have you ever participated in a study that used the same case summary? No   Yes

APPENDIX P

MOTIVATION TO CONTROL PREJUDICED REACTIONS SCALE (MCPRS)

Instructions: Please indicate the extent to which you agree or disagree with each of the following statements.

	strongly disagree						strongly agree
1. In today's society it is important that one not be perceived as prejudiced in any manner.	-3	-2	-1	0	+1	+2	+3
2. I always express my thoughts and feelings, regardless of how controversial they might be.	-3	-2	-1	0	+1	+2	+3
3. I get angry with myself when I have a thought or feeling that might be considered prejudiced.	-3	-2	-1	0	+1	+2	+3
4. If I were participating in a general discussion, either at work or at school, and a Black person expressed an opinion with which I disagreed, I would be hesitant to express my own viewpoint.	-3	-2	-1	0	+1	+2	+3
5. Going through life worrying about whether you might offend someone is just more trouble than it's worth.	-3	-2	-1	0	+1	+2	+3
6. It's important to me that other people not think I'm prejudiced.	-3	-2	-1	0	+1	+2	+3
7. It's important to behave according to society's standards.	-3	-2	-1	0	+1	+2	+3
8. I'm careful not to offend my friends, but I don't worry about offending people I don't know or don't like.	-3	-2	-1	0	+1	+2	+3
9. I think that it is important to speak one's mind rather than to worry about offending someone.	-3	-2	-1	0	+1	+2	+3

	strongly disagree						strongly agree
10. It's never acceptable to express one's prejudices.	-3	-2	-1	0	+1	+2	+3
11. I feel guilty when I have a negative thought or feeling about a Black person.	-3	-2	-1	0	+1	+2	+3
12. When speaking to a Black person, it's important to me that he/she not think I'm prejudiced.	-3	-2	-1	0	+1	+2	+3
13. It bothers me a great deal when I think I've offended someone, so I'm always careful to consider other people's feelings.	-3	-2	-1	0	+1	+2	+3
14. If I have a prejudiced thought or feeling, I keep it to myself.	-3	-2	-1	0	+1	+2	+3
15. I would never tell jokes that might offend others.	-3	-2	-1	0	+1	+2	+3
16. I'm not afraid to tell others what I think, even when I know they disagree with me.	-3	-2	-1	0	+1	+2	+3
17. If someone who made me uncomfortable sat next to me on a bus, I would not hesitate to move to another seat.	-3	-2	-1	0	+1	+2	+3

APPENDIX Q

NEED FOR COGNITION SCALE

Instructions: For each of the statements below, please indicate whether or not the statement is characteristic of you or of what you believe. For example, if the statement is extremely uncharacteristic of you or of what you believe about yourself (not at all like you), please place a "1" on the line to the left of the statement. If the statement is extremely characteristic of you or of what you believe about yourself (very much like you), please place a "5" on the line to the left of the statement. You should use the following scale as you rate each of the statements below.

1	2	3	4	5
extremely uncharacteristic of me	somewhat uncharacteristic of me	uncertain	somewhat characteristic of me	extremely characteristic of me

1. _____ I would prefer complex to simple problems.

2. _____ I like to have the responsibility of handling a situation that requires a lot of thinking.

3. _____ Thinking is not my idea of fun.

4. _____ I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.

5. _____ I try to anticipate and avoid situations where there is a likely chance I will have to think in depth about something.

6. _____ I find satisfaction in deliberating hard and for long hours.

7. _____ I only think as hard as I have to.

8. _____ I prefer to think about small, daily projects to long-term ones.

9. _____ I like tasks that require little thought once I've learned them.

10. _____ The idea of relying on thought to make my way to the top appeals to me.

11._____ I really enjoy a task that involves coming up with new solutions to problems.

12._____ Learning new ways to think doesn't excite me very much.

13._____ I prefer my life to be filled with puzzles I must solve.

14. _____ The notion of thinking abstractly is appealing to me.

15._____ I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

16._____ I feel relief rather than satisfaction after completing a task that required a lot of mental effort.

17._____ It's enough for me that something gets the job done; I don't care how or why it works.

18. I usually end up deliberating about issues even when they do not affect me personally.

APPENDIX R

STUDY 2 DEMOGRAPHIC QUESTIONNAIRE

Instructions: Please answer the following questions about yourself.

Are you a United States citizen?   Yes   No

What is your age? _____

What is your gender?   Male   Female

What is your ethnicity?

       European American / White
       African American / Black
       Asian
       Latino/a
       Middle Eastern
       Multi-racial (please explain): _____
       Other (please indicate): _____

What is your level of education?

       Some high school
       High school graduate
       Some college
       college graduate (i.e., 4-year degree)
       Some graduate school
       Graduate school degree

APPENDIX S

FINAL QUESTIONNAIRE

Instructions: Please you <u>do not</u> use the "prev" button.

To what degree would you typically associate this crime with a White or Black defendant? (Please circle one.)

1	2	3	4	5	6	7	8	9

associated
with Whites

associated
with Blacks

How serious do you think the crime was? (Please circle one.)

1	2	3	4	5	6	7

not serious
at all

very serious

What was the defendant's race? _____

What crime was the defendant charged with? _____

Was the following statement included in the judicial instructions that you read at the

beginning of the study?

"Previous psychological research has shown that our impressions and evaluations of

others are commonly influenced by stereotype-based thoughts and beliefs. As a good

juror, please try to ignore this source of information when forming an impression of the

defendant."

                    No   Yes

APPENDIX T

MODIFIED NARRATIVE BELIEVABILITY SCALE – JUROR (MNBS-J)

Instructions: Now think about the story you just created. Think about what your basic story was and rate your own story by answering the following questions. Please circle the response that best reflects your opinion about your own description of what happened (i.e., your story). There are no correct answers.

	strongly disagree						strongly agree
I believe my story could be true.	1	2	3	4	5	6	7
It is easy to follow my story from beginning to end.	1	2	3	4	5	6	7
The information presented in my story was consistent.	1	2	3	4	5	6	7
There was important information missing from my story.	1	2	3	4	5	6	7
My story was plausible.	1	2	3	4	5	6	7
It was hard to follow my own story.	1	2	3	4	5	6	7
All of the facts in my story agreed with each other.	1	2	3	4	5	6	7
There were a lot of "holes" in my story.	1	2	3	4	5	6	7
My story seems to be true.	1	2	3	4	5	6	7
If I were to write my own story again, I would organize it differently.	1	2	3	4	5	6	7
My story stands out as the only possible description of what happened.	1	2	3	4	5	6	7
In addition to my story, I could generate a number of additional stories from the evidence.	1	2	3	4	5	6	7
My story is irreplaceable.	1	2	3	4	5	6	7
Multiple unique stories can explain the evidence.	1	2	3	4	5	6	7
At least one alternative explanation exists.	1	2	3	4	5	6	7

The "coverage" of a story refers to the extent to which the story accounts for all of the information presented in the story. How would you rate your own story in terms of "coverage"?

1	2	3	4	5	6	7
very low						very high

The "consistency" of a story refers to the extent to which a story does not contradict itself or contradict other things you know to be true or false. How would you rate your own story in terms of "consistency"?

1	2	3	4	5	6	7
very low						very high

APPENDIX U

MODIFIED NARRATIVE BELIEVABILITY SCALE – RATERS (MNBS-R)

Instructions for Raters: Think about what the basic story of the individual was and rate that story by answering the following questions. Please circle the response that best reflects your opinion about the individual's description of what happened (i.e., the story). There are no correct answers.

	strongly disagree					strongly agree	
I believe the story could be true.	1	2	3	4	5	6	7
It was easy to follow the story from beginning to end.	1	2	3	4	5	6	7
The information presented in the story was consistent.	1	2	3	4	5	6	7
There was important information missing from the story.	1	2	3	4	5	6	7
The story was plausible.	1	2	3	4	5	6	7
It was hard to follow the story.	1	2	3	4	5	6	7
All of the facts in the story agreed with each other.	1	2	3	4	5	6	7
There were a lot of "holes" in the prosecution story.	1	2	3	4	5	6	7
The story seems to be true.	1	2	3	4	5	6	7
If I were given the story, I would have organized it differently.	1	2	3	4	5	6	7
The story stands out as the only possible description of what happened.	1	2	3	4	5	6	7
I could generate a number of additional stories from the evidence.	1	2	3	4	5	6	7
The story is irreplaceable.	1	2	3	4	5	6	7
Multiple unique stories can explain the evidence.	1	2	3	4	5	6	7
At least one alternative explanation exists.	1	2	3	4	5	6	7

The "coverage" of a story refers to the extent to which the story accounts for all of the information presented in the case. How would you rate the story in terms of "coverage"?

1	2	3	4	5	6	7
very low						very high

The "consistency" of a story refers to the extent to which a story does not contradict itself or contradict other things you know to be true or false. How would you rate the story in terms of "consistency"?

1	2	3	4	5	6	7
very low						very high

APPENDIX V

STUDY 3 DEMOGRAPHIC QUESTIONNAIRE

Instructions: Please answer the following questions about yourself.

Are you a United States citizen?　　Yes　　　　No

What is your age? _____

What is your gender?　　　Male　　　　　Female

What is your ethnicity?

　　　European American / White
　　　African American / Black
　　　Asian
　　　Latino/a
　　　Middle Eastern
　　　Multi-racial (please indicate): _____
　　　Other (please indicate): _____

What year in college are you?

　　　Freshman
　　　Sophomore
　　　Junior
　　　Senior
　　　Other (please explain): _____

APPENDIX W

FIGURE 19: CONCEPTUAL PREFERRED OBLIQUE THREE-FACTOR MODEL

Note: The acquiescence factor was omitted from the conceptual model.
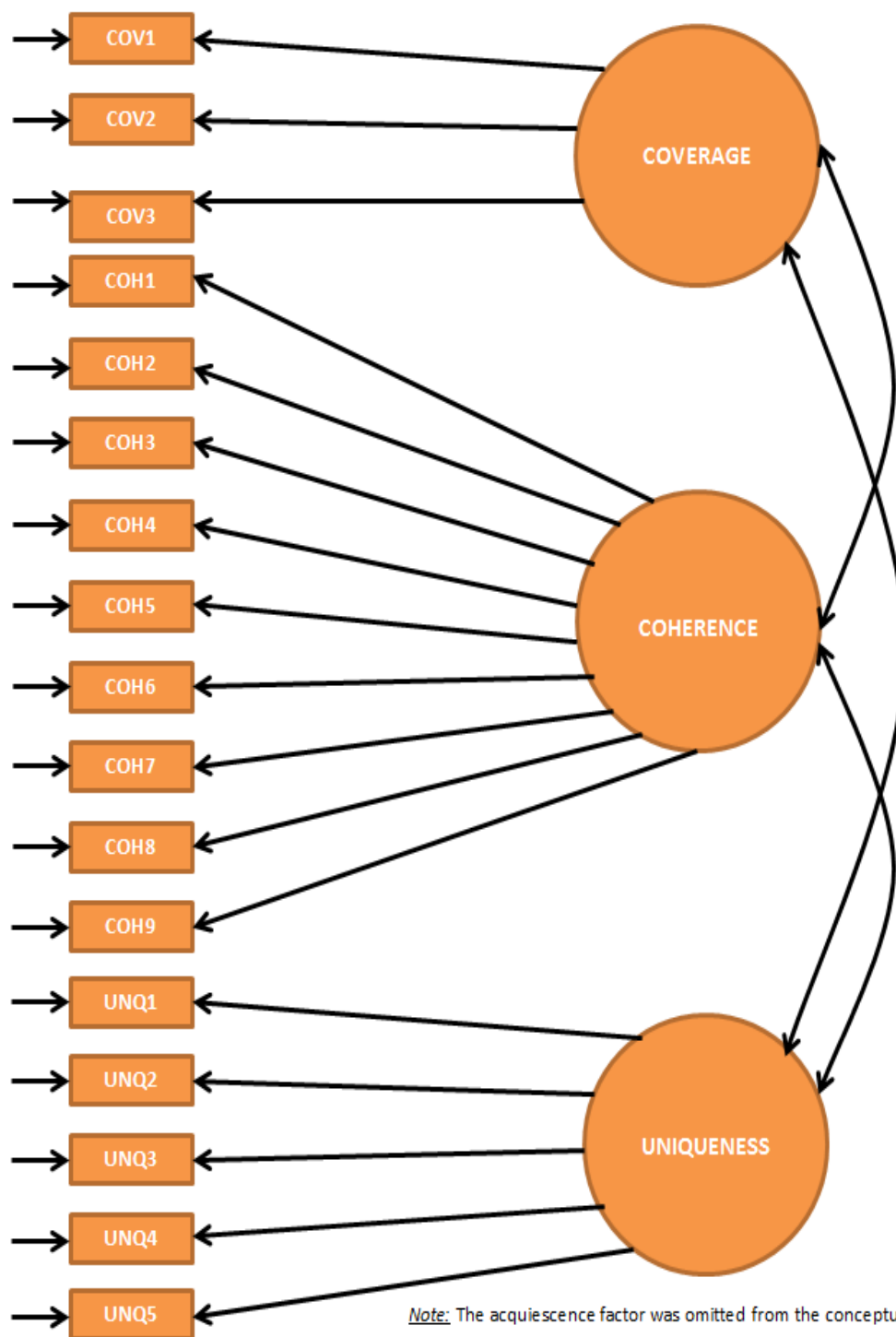
APPENDIX X

STUDY 2: FIGURES AND TABLES

Figure 20. Mock Jurors' Ethnicity (MTurkers)
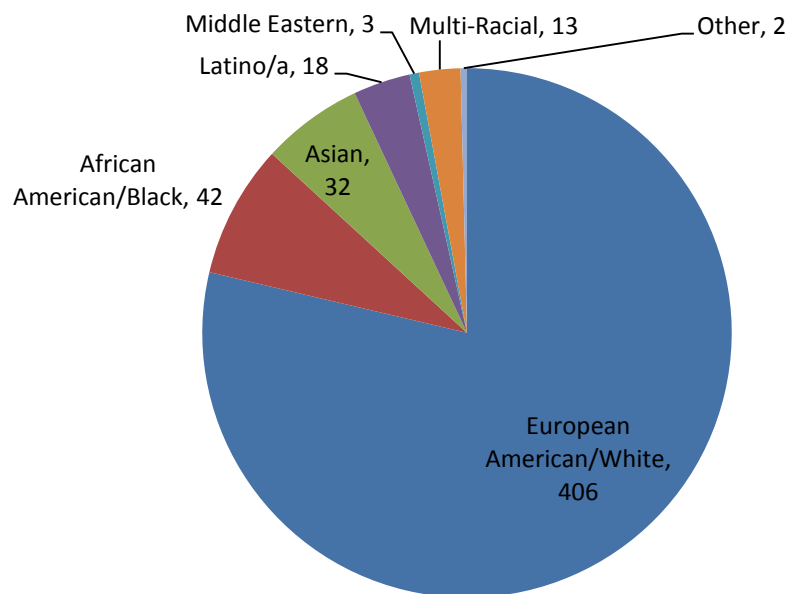
Table 11. Logistic Regression Analysis with Verdict Decision as the Outcome and MCPR in the Model (Study 2)

	B (SE)	Exp(B)
Defendant Race	0.34 (0.36)	1.40
Type of Crime	-0.13 (0.38)	0.89
Motivation to Control Prejudiced Reactions (MCPR)	-0.01 (0.02)	0.99
Defendant Race x Type of Crime	-0.27 (0.51)	0.77
Defendant Race x MCPR	0.01 (0.03)	1.01
Type of Crime x MCPR	0.07 (0.03)*	1.07
Defendant Race x Type of Crime x MCPR	-0.04 (0.04)	0.96

Note: $*p < .05$

Table 12. Logistic Regression Analysis with Verdict Decision as the Outcome and SSI in the Model (Study 2)

	B (SE)	Exp(B)
Defendant Race	0.61 (0.47)	1.84
Type of Crime	-0.09 (0.48)	0.92
Stereotype Suppression Instructions (SSI)	0.10 (0.54)	1.10
Defendant Race x Type of Crime	-0.56 (0.68)	0.57
Defendant Race x SSI	-.59 (0.73)	0.56
Type of Crime x SSI	-0.23 (0.75)	0.79
Defendant Race x Type of Crime x SSI	0.66 (1.01)	1.93

Note: None of the effects were significant at $p < .05$.

Table 13. Logistic Regression Analysis with Verdict Decision as the Outcome and NC in the Model (as a Control) (Study 2)

	B (SE)	Exp(B)
Defendant Race	0.34 (0.36)	1.40
Type of Crime	-0.21 (0.37)	0.81
Need for Cognition (NC)	-0.00 (0.02)	1.00
Defendant Race x Type of Crime	-0.24 (0.50)	0.79
Defendant Race x NC	-0.00 (0.03)	1.00
Type of Crime x NC	-0.02 (0.20)	0.98
Defendant Race x Type of Crime x NC	0.04 (0.04)	1.04

Note: None of the effects were significant at $p < .05$.

Table 14. ANCOVA with Verdict-Confidence as the Outcome and SSI in the Model (Study 2)

	F-value	*p*
Defendant Race	0.24	.62
Type of Crime	2.11	.15
Stereotype Suppression Instructions (SSI)	0.76	.38
Defendant Race x Type of Crime	0.03	.86
Defendant Race x SSI	0.10	.76
Type of Crime x SSI	0.00	1.00
Defendant Race x Type of Crime x SSI	0.23	.63

Note: None of the effects were significant at $p < .05$.

Table 15. ANCOVA with Verdict-Confidence as the Outcome and NC in the Model (as a Control) (Study 2)

	F-value	*p*
Defendant Race	0.31	.58
Type of Crime	2.42	.12
Need for Cognition (NC)	0.12	.73
Defendant Race x Type of Crime	0.03	.86
Defendant Race x NC	0.77	.33
Type of Crime x NC	0.59	.46
Defendant Race x Type of Crime x NC	1.82	.18

<u>*Note:*</u> None of the effects were significant at $p < .05$.

Table 16. ANCOVA with Probability of Guilt as the Outcome and SSI in the Model (Study 2)

	F-value	*p*
Defendant Race	0.03	.86
Type of Crime	9.87*	$< .01$
Stereotype Suppression Instructions (SSI)	0.48	.49
Defendant Race x Type of Crime	0.03	.86
Defendant Race x SSI	1.23	.26
Type of Crime x SSI	0.03	.86
Defendant Race x Type of Crime x SSI	0.04	.85

*significant

Table 17. ANCOVA with Probability of Guilt as the Outcome and NC in the Model (as a Control) (Study 2)

	F-value	*p*
Defendant Race	0.02	.89
Type of Crime	10.25*	< .01
Need for Cognition (NC)	0.00	.99
Defendant Race x Type of Crime	0.05	.82
Defendant Race x NC	0.01	.92
Type of Crime x NC	1.18	.28
Defendant Race x Type of Crime x NC	.85	.36

*significant

Table 18. ANCOVA with Story Coherence (MNBS-P) as the Outcome and MCPR in the Model (Study 2)

	F-value	*p*
Defendant Race	0.11	.75
Type of Crime	29.97*	< .001
Motivation to Control Prejudiced Reactions (MCPR)	0.13	.72
Defendant Race x Type of Crime	0.02	.89
Defendant Race x MCPR	1.03	.31
Type of Crime x MCPR	0.27	.61
Defendant Race x Type of Crime x MCPR	0.05	.83

*significant

Table 19. ANCOVA with Story coherence (MNBS-P) as the Outcome and SSI in the Model (Study 2)

	F-value	p
Defendant Race	0.00	.99
Type of Crime	31.41*	< .001
Stereotype Suppression Instructions (SSI)	0.29	.59
Defendant Race x Type of Crime	0.00	.998
Defendant Race x SSI	10.85*	< .01
Type of Crime x SSI	0.05	.82
Defendant Race x Type of Crime x SSI	1.46	.23

*significant

Table 20. ANCOVA with Story Coherence (MNBS-P) as the Outcome and NC in the Model (as a Control) (Study 2)

	F-value	p
Defendant Race	0.09	.76
Type of Crime	30.99*	< .001
Need for Cognition (NC)	4.58*	.03
Defendant Race x Type of Crime	0.00	.95
Defendant Race x NC	0.91	.33
Type of Crime x NC	0.07	.79
Defendant Race x Type of Crime x NC	0.50	.48

*significant

Table 21. ANCOVA with Story Coverage (MNBS-P) as the Outcome and MCPR in the Model (Study 2)

	F-value	p
Defendant Race	1.40	.24
Type of Crime	17.21*	< .001
Motivation to Control Prejudiced Reactions (MCPR)	0.11	.74
Defendant Race x Type of Crime	0.94	.33
Defendant Race x MCPR	3.79	.05
Type of Crime x MCPR	0.23	.63
Defendant Race x Type of Crime x MCPR	1.31	.25

*significant

Table 22. ANCOVA with Story Coverage (MNBS-P) as the Outcome and SSI in the Model (Study 2)

	F-value	p
Defendant Race	1.25	.27
Type of Crime	19.05*	< .001
Stereotype Suppression Instructions (SSI)	0.07	.79
Defendant Race x Type of Crime	1.09	.30
Defendant Race x SSI	0.43	.52
Type of Crime x SSI	0.94	.33
Defendant Race x Type of Crime x SSI	1.95	.16

*significant

Table 23. ANCOVA with Story Coverage (MNBS-P) as the Outcome and NC in the Model (as a Control) (Study 2)

	F-value	*p*
Defendant Race	1.21	.27
Type of Crime	18.13*	< .001
Need for Cognition (NC)	1.89	.17
Defendant Race x Type of Crime	0.64	.42
Defendant Race x NC	1.62	.20
Type of Crime x NC	0.00	.96
Defendant Race x Type of Crime x NC	0.06	.81

*significant

Table 24. ANCOVA with Story Uniqueness (MNBS-P) as the Outcome and MCPR in the Model (Study 2)

	F-value	*p*
Defendant Race	0.03	.86
Type of Crime	18.53*	< .001
Motivation to Control Prejudiced Reactions (MCPR)	0.14	.71
Defendant Race x Type of Crime	0.40	.53
Defendant Race x MCPR	0.06	.06
Type of Crime x MCPR	0.01	.91
Defendant Race x Type of Crime x MCPR	0.38	.54

*significant

Table 25. ANCOVA with Story Uniqueness (MNBS-P) as the Outcome and SSI in the Model (Study 2)

	F-value	*p*
Defendant Race	0.00	.97
Type of Crime	17.55*	< .001
Stereotype Suppression Instructions (SSI)	0.04	.84
Defendant Race x Type of Crime	0.18	.67
Defendant Race x SSI	0.56	.46
Type of Crime x SSI	0.07	.77
Defendant Race x Type of Crime x SSI	0.77	.38

*significant

Table 26. ANCOVA with Story uniqueness (MNBS-P) as the Outcome and NC in the Model (as a Control) (Study 2)

	F-value	*p*
Defendant Race	0.01	.92
Type of Crime	19.26*	< .001
Need for Cognition (NC)	4.00	.046
Defendant Race x Type of Crime	0.54	.46
Defendant Race x NC	0.50	.48
Type of Crime x NC	1.41	.24
Defendant Race x Type of Crime x NC	0.50	.48

*significant

Table 27. ANCOVA with Story Coherence (MNBS-D) as the Outcome and MCPR in the Model (Study 2)

	F-value	p
Defendant Race	1.36	.25
Type of Crime	38.48*	< .001
Motivation to Control Prejudiced Reactions (MCPR)	0.08	.78
Defendant Race x Type of Crime	0.00	.97
Defendant Race x MCPR	0.37	.55
Type of Crime x MCPR	0.05	.82
Defendant Race x Type of Crime x MCPR	0.37	.55

*significant

Table 28. ANCOVA with Story Coherence (MNBS-D) as the Outcome and SSI in the Model (Study 2)

	F-value	p
Defendant Race	1.17	.28
Type of Crime	38.70*	< .001
Stereotype Suppression Instructions (SSI)	2.04	.15
Defendant Race x Type of Crime	0.03	.88
Defendant Race x SSI	0.01	.93
Type of Crime x SSI	0.02	.88
Defendant Race x Type of Crime x SSI	0.32	.57

*significant

Table 29. ANCOVA with Story Coherence (MNBS-D) as the Outcome and NC in the Model (as a Control) (Study 2)

	F-value	*p*
Defendant Race	1.90	.17
Type of Crime	37.27*	< .001
Need for Cognition (NC)	0.04	.85
Defendant Race x Type of Crime	0.04	.84
Defendant Race x NC	2.30	.13
Type of Crime x NC	3.67	.06
Defendant Race x Type of Crime x NC	.99	.32

*significant

Table 30. ANCOVA with Story Coverage (MNBS-D) as the Outcome and MCPR in the Model (Study 2)

	F-value	*p*
Defendant Race	2.03	.16
Type of Crime	33.63*	< .001
Motivation to Control Prejudiced Reactions (MCPR)	0.46	.50
Defendant Race x Type of Crime	0.04	.85
Defendant Race x MCPR	0.06	.81
Type of Crime x MCPR	0.02	.90
Defendant Race x Type of Crime x MCPR	0.06	.81

*significant

Table 31. ANCOVA with Story Coverage (MNBS-D) as the Outcome and SSI in the Model (Study 2)

	F-value	*p*
Defendant Race	2.31	.13
Type of Crime	30.92*	< .001
Stereotype Suppression Instructions (SSI)	0.00	.95
Defendant Race x Type of Crime	0.08	.77
Defendant Race x SSI	0.98	.32
Type of Crime x SSI	1.65	.20
Defendant Race x Type of Crime x SSI	0.00	.98

*significant

Table 32. ANCOVA with Story Coverage (MNBS-D) as the Outcome and NC in the Model (as a Control) (Study 2)

	F-value	*p*
Defendant Race	2.59	.11
Type of Crime	31.16*	< .001
Need for Cognition (NC)	0.82	.37
Defendant Race x Type of Crime	0.10	.75
Defendant Race x NC	2.39	.124
Type of Crime x NC	0.31	.58
Defendant Race x Type of Crime x NC	5.26*	.02

*significant

Table 33. ANCOVA with Story Uniqueness (MNBS-D) as the Outcome and MCPR in the Model (Study 2)

	F-value	p
Defendant Race	3.62	.06
Type of Crime	1.83	.18
Motivation to Control Prejudiced Reactions (MCPR)	4.21*	.04
Defendant Race x Type of Crime	0.09	.77
Defendant Race x MCPR	1.31	.253
Type of Crime x MCPR	0.90	.344
Defendant Race x Type of Crime x MCPR	1.31	.25

*significant

Table 34. ANCOVA with Story Uniqueness (MNBS-D) as the Outcome and SSI in the Model (Study 2)

	F-value	p
Defendant Race	4.25*	.04
Type of Crime	2.00	.16
Stereotype Suppression Instructions (SSI)	0.00	.98
Defendant Race x Type of Crime	0.10	.75
Defendant Race x SSI	3.89	.05
Type of Crime x SSI	0.76	.39
Defendant Race x Type of Crime x SSI	0.03	.87

*significant

Table 35. ANCOVA with Story Uniqueness (MNBS-D) as the Outcome and NC in the Model (as a Control) (Study 2)

	F-value	*p*
Defendant Race	3.73	.06
Type of Crime	1.30	.26
Need for Cognition (NC)	7.05*	.01
Defendant Race x Type of Crime	.26	.61
Defendant Race x NC	0.07	.79
Type of Crime x NC	0.00	.93
Defendant Race x Type of Crime x NC	0.71	.40

*significant

APPENDIX Y

STUDY 3: FIGURES AND TABLES

Figure 21. Mock Jurors' Ethnicity (Students; Study 3)
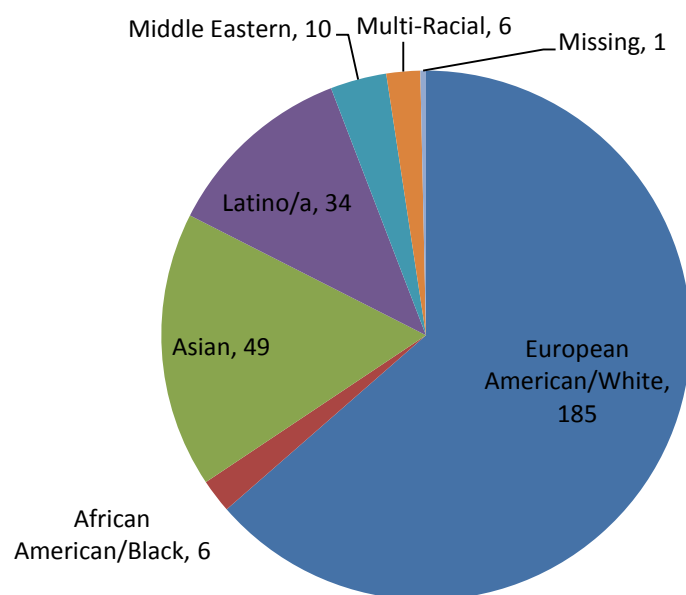
Table 36. Interrater Reliability: Correlations between Coders' Ratings of Each Item of the MNBS-R (Study 3)

Coder 1 / Coder 2	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12
Item 1	.76*	---	---	---	---	---	---	---	---	---	---	---
Item 2	---	.72*	---	---	---	---	---	---	---	---	---	---
Item 3	---	---	.94*	---	---	---	---	---	---	---	---	---
Item 4	---	---	---	.71*	---	---	---	---	---	---	---	---
Item 5	---	---	---	---	.66*	---	---	---	---	---	---	---
Item 6	---	---	---	---	---	.77*	---	---	---	---	---	---
Item 7	---	---	---	---	---	---	.84*	---	---	---	---	---
Item 8	---	---	---	---	---	---	---	.75*	---	---	---	---
Item 9	---	---	---	---	---	---	---	---	.67*	---	---	---
Item 10	---	---	---	---	---	---	---	---	---	.76*	---	---
Item 11	---	---	---	---	---	---	---	---	---	---	.72*	---
Item 12	---	---	---	---	---	---	---	---	---	---	---	.80*

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

Table 37. Logistic Regression Analysis with Verdict Decision as the Outcome and MCPR in the model (Study 3)

	B (SE)	Exp(B)
Defendant Race	-0.25 (0.43)	0.78
Type of Crime	1.15 (0.54)*	3.15
Motivation to Control Prejudiced Reactions (MCPR	0.03 (0.03)	1.03
Defendant Race x Type of Crime	0.60 (0.78)	1.83
Defendant Race x MCPR	-0.04 (0.04)	.96
Type of Crime x MCPR	0.03 (0.05)	1.03
Defendant Race x Type of Crime x MCPR	0.12 (0.72)	1.01

Note: *$p < .05$

Table 38. Logistic Regression Analysis with Verdict Decision as the Outcome and SSI in the Model (Study 3)

	B (SE)	Exp(B)
Defendant Race	-0.83 (0.64)	0.44
Type of Crime	0.53 (0.74)	1.70
Stereotype Suppression Instructions (SSI)	-0.92 (0.64)	0.40
Defendant Race x Type of Crime	1.23 (1.05)	3.42
Defendant Race x SSI	1.12 (0.86)	3.06
Type of Crime x SSI	0.86 (1.02)	2.35
Defendant Race x Type of Crime x SSI	0.92 (0.48)	2.50

Note: None of the effects were significant at $p < .05$.

Table 39. Logistic Regression Analysis with Verdict Decision as the Outcome and NC in the Model (as a Control) (Study 3)

	B (SE)	Exp(B)
Defendant Race	-0.16 (0.48)	0.85
Type of Crime	0.91 (0.51)	2.49
Need for Cognition (NC)	-0.04 (0.03)	0.96
Defendant Race x Type of Crime	0.67 (0.77)	1.95
Defendant Race x NC	0.00 (0.04)	1.00
Type of Crime x NC	0.06 (0.05)	1.06
Defendant Race x Type of Crime x NC	0.49 (0.33)	1.63

Note: None of the effects were significant at $p < .05$.

Table 40. ANCOVA with Verdict-Confidence as the Outcome and SSI in the Model (Study 3)

	F-value	*p*
Defendant Race	0.03	.86
Type of Crime	14.58*	< .001
Stereotype Suppression Instructions (SSI)	0.00	.98
Defendant Race x Type of Crime	0.81	.37
Defendant Race x SSI	1.70	.20
Type of Crime x SSI	0.43	.51
Defendant Race x Type of Crime x SSI	0.34	.56

*significant

Table 41. ANCOVA with Verdict-Confidence as the Outcome and NC in the Model (as a Control) (Study 3)

	F-value	p
Defendant Race	0.11	.74
Type of Crime	11.85*	< .01
Need for Cognition (NC)	1.91	.17
Defendant Race x Type of Crime	0.38	.54
Defendant Race x NC	0.29	.59
Type of Crime x NC	1.63	.20
Defendant Race x Type of Crime x NC	0.01	.92

*significant

Table 42. ANCOVA with Probability of Guilt as the Outcome and MCPR (Study 3)

	F-value	p
Defendant Race	2.04	.16
Type of Crime	1.34	.25
Motivation to control prejudiced reactions (MCPR)	0.35	.56
Defendant Race x Type of Crime	0.55	.46
Defendant Race x MCPR	3.67	.06
Type of Crime x MCPR	0.32	.57
Defendant Race x Type of Crime x MCPR	0.02	.89

Note: None of the effects were significant at $p < .05$.

Table 43. ANCOVA with Probability of Guilt as the Outcome and SSI in the Model (Study 3)

	F-value	p
Defendant Race	1.97	.16
Type of Crime	0.67	.41
Stereotype Suppression Instructions (SSI)	0.19	.66
Defendant Race x Type of Crime	0.94	.33
Defendant Race x SSI	0.00	.95
Type of Crime x SSI	1.50	.22
Defendant Race x Type of Crime x SSI	0.05	.82

Note: None of the effects were significant at $p < .05$.

Table 44. ANCOVA with Probability of Guilt as the Outcome and NC in the Model (as a Control) (Study 3)

	F-value	p
Defendant Race	2.00	.16
Type of Crime	0.53	.47
Need for Cognition (NC)	0.88	.35
Defendant Race x Type of Crime	0.88	.35
Defendant Race x NC	0.28	.60
Type of Crime x NC	1.54	.22
Defendant Race x Type of Crime x NC	0.28	.60

Note: None of the effects were significant at $p < .05$.

Table 45. ANCOVA with Story Coherence (MNBS-P) as the Outcome and MCPR in the Model (Study 3)

	F-value	p
Defendant Race	1.81	.18
Type of Crime	14.06*	< .001
Motivation to Control Prejudiced Reactions (MCPR)	0.06	.81
Defendant Race x Type of Crime	0.81	.37
Defendant Race x MCPR	4.21*	.04
Type of Crime x MCPR	2.95	.09
Defendant Race x Type of Crime x MCPR	0.23	.64

*significant

Table 46. ANCOVA with Story Coherence (MNBS-P) as the Outcome and SSI in the Model (Study 3)

	F-value	p
Defendant Race	1.59	.21
Type of Crime	16.42*	< .001
Stereotype Suppression Instructions (SSI)	0.50	.48
Defendant Race x Type of Crime	1.12	.29
Defendant Race x SSI	0.09	.77
Type of Crime x SSI	2.27	.13
Defendant Race x Type of Crime x SSI	0.51	.48

*significant

Table 47. ANCOVA with Story Coherence (MNBS-P) as the Outcome and NC in the Model (as a Control) (Study 3)

	F-value	*p*
Defendant Race	0.81	.37
Type of Crime	15.48*	< .001
Need for Cognition (NC)	0.65	.42
Defendant Race x Type of Crime	1.06	.30
Defendant Race x NC	0.25	.62
Type of Crime x NC	0.37	.54
Defendant Race x Type of Crime x NC	0.77	.38

*significant

Table 48. ANCOVA with Story Coherence (MNBS-D) as the Outcome and MCPR in the Model (Study 3)

	F-value	*p*
Defendant Race	0.01	.91
Type of Crime	14.30*	< .001
Motivation to Control Prejudiced Reactions (MCPR)	4.92*	.03
Defendant Race x Type of Crime	0.01	.92
Defendant Race x MCPR	1.01	.32
Type of Crime x MCPR	0.30	.56
Defendant Race x Type of Crime x MCPR	0.14	.71

*significant

Table 49. ANCOVA with Story Coherence (MNBS-D) as the Outcome and SSI in the Model (Study 3)

	F-value	p
Defendant Race	0.00	.98
Type of Crime	13.93*	< .001
Stereotype Suppression Instructions (SSI)	0.40	.53
Defendant Race x Type of Crime	0.10	.75
Defendant Race x SSI	4.68*	.03
Type of Crime x SSI	1.06	.31
Defendant Race x Type of Crime x SSI	2.86	.09

*significant

Table 50. ANCOVA with Story Coherence (MNBS-D) as the Outcome and NC in the Model (as a Control) (Study 3)

	F-value	p
Defendant Race	0.09	.77
Type of Crime	15.72*	< .001
Need for Cognition (NC)	1.85	.18
Defendant Race x Type of Crime	0.00	.95
Defendant Race x NC	0.00	.97
Type of Crime x NC	0.05	.83
Defendant Race x Type of Crime x NC	2.28	.13

*significant

Table 51. ANCOVA with Story Coherence (MNBS-J) as the Outcome and MCPR in the Model (Study 3)

	F-value	*p*
Defendant Race	0.42	.52
Type of Crime	2.67	.10
Motivation to Control Prejudiced Reactions (MCPR)	2.39	.12
Defendant Race x Type of Crime	0.57	.45
Defendant Race x MCPR	0.02	.89
Type of Crime x MCPR	0.04	.85
Defendant Race x Type of Crime x MCPR	0.39	.53

Note: None of the effects were significant at $p < .05$.

Table 52. ANCOVA with Story Coherence (MNBS-J) as the Outcome and SSI in the Model (Study 3)

	F-value	*p*
Defendant Race	0.42	.52
Type of Crime	2.64	.11
Stereotype Suppression Instructions (SSI)	0.97	.33
Defendant Race x Type of Crime	0.31	.58
Defendant Race x SSI	0.01	.93
Type of Crime x SSI	0.10	.76
Defendant Race x Type of Crime x SSI	0.15	.70

Note: None of the effects were significant at $p < .05$.

Table 53. ANCOVA with Story Coherence (MNBS-J) as the Outcome and NC in the Model (as a control) (Study 3)

	F-value	p
Defendant Race	0.61	.44
Type of Crime	2.21	.14
Need for Cognition (NC)	0.55	.46
Defendant Race x Type of Crime	0.36	.55
Defendant Race x NC	0.10	.76
Type of Crime x NC	0.92	.34
Defendant Race x Type of Crime x NC	0.02	.89

Note: None of the effects were significant at $p < .05$.

Table 54. ANCOVA with Story Coherence (MNBS-R) as the Outcome and MCPR in the Model (Study 3)

	F-value	p
Defendant Race	1.81	.18
Type of Crime	14.06*	$< .001$
Motivation to Control Prejudiced Reactions (MCPR)	0.06	.81
Defendant Race x Type of Crime	0.81	.37
Defendant Race x MCPR	4.21*	.04
Type of Crime x MCPR	2.95	.09
Defendant Race x Type of Crime x MCPR	0.23	.64

Note: None of the effects were significant at $p < .05$.

Table 55. ANCOVA with Story Coherence (MNBS-R) as the Outcome and SSI in the Model (Study 3)

	F-value	*p*
Defendant Race	1.59	.21
Type of Crime	16.42*	< .001
Stereotype Suppression Instructions (SSI)	0.50	.48
Defendant Race x Type of Crime	1.12	.29
Defendant Race x SSI	0.09	.77
Type of Crime x SSI	2.27	.13
Defendant Race x Type of Crime x SSI	0.51	.48

<u>Note:</u> None of the effects were significant at $p < .05$.

Table 56. ANCOVA with Story Coherence (MNBS-R) as the Outcome and NC in the Model (as a control) (Study 3)

	F-value	*p*
Defendant Race	0.81	.37
Type of Crime	15.48	< .001
Need for Cognition (NC)	0.65	.42
Defendant Race x Type of Crime	1.06	.30
Defendant Race x NC	0.25	.62
Type of Crime x NC	0.39	.54
Defendant Race x Type of Crime x NC	0.77	.38

<u>Note:</u> None of the effects were significant at $p < .05$.

APPENDIX Z

STUDY 4: FIGURE

Figure 22. Mock Jurors' Ethnicity (Students; Study 4)
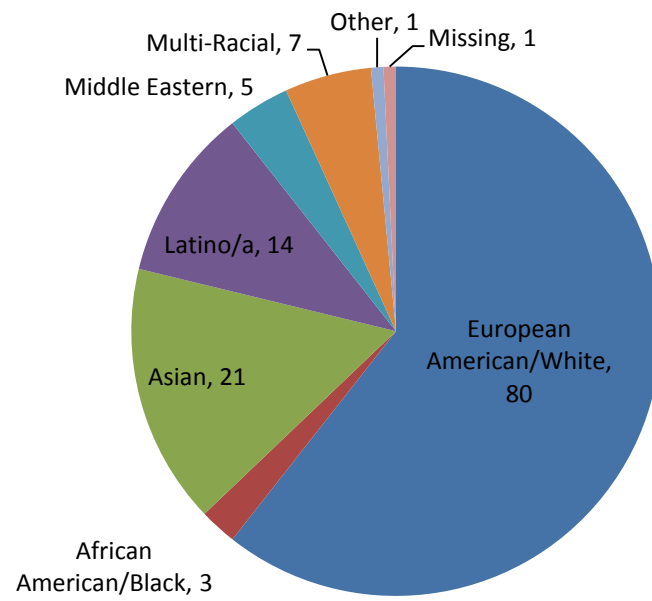
REFERENCE LIST

Allison, M., & Brimacombe, C. A. (2010). Alibi believability: The effect of prior convictions and judicial instructions. *Journal of Applied Social Psychology*, *40*, 1054-1084. doi:10.1111/j.1559-1816.2010.00610.x

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research – conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182. doi: 10.1037/0022-3514.51.6.1173

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M. (1995). EQS structural equations program manual. Encino, CA: Multivariate Software.

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-606. doi:10.1037/0033-2909.88.3.588

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20*, 351-368. doi: 10.1093/pan/mpr057

Billiet, J. B., & McClendon, M. J. (1998). On the identification of acquiescence in balanced sets of items using structural models. *Advances in Methodology, Data Analysis, and Statistics, 14*, 129-150.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. Structural Equal Modeling, 7, 608-628. doi:10.1207/S15328007SEM0704_5

Bollen, K., & Long, J. S. (1993). *Testing structural equation models.* Newbury Park, CA: Sage.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3-5. doi:10.1177/1745691610393980

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116–131. doi:10.1037/0022-3514.42.1.116

Cacioppo, J. T., & Petty, R. E. (1984). The need for cognition: Relationships to attitudinal processes. In R. P. McGlynn, J. E. Maddux, C. Stoltenberg, & J. H. Harvey (Eds.), *Social perception in clinical and counseling psychology*. Lubbock, Texas Tech University Press.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., Jarvis, W., Blair, G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 2, 197-253. doi: 10.1037/0033-2909.119.2.197

Cornell University Law School. 18 U.S. Code § 656. Retrieved on March 17, 2014 from the Legal Information Institute: https://owl.english.purdue.edu/owl/resource/560/10/

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research, *PLOS ONE, 8*, e57410. doi:10.1371/journal.pone.0057410

Devine, D. J. (2012). Jury decision making – The state of the science. NY: New York University Press.

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82*, 835-848. doi:10.1037/0022-3514.82.5.835

Dinizulu, S. M., Grant, K. E., Bryant, F. B., Boustani, M. M., Tyler, D., & McIntosh, J. M. (2014). Nondisclosure as a mediator of the effects of exposure to community violence on psychological distress. *Child and Youth Forum, 43*, 41-61. doi:10.1007/s10566-013-9224-z

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 23*, 316-326. doi:10.1177/014616797233009

Gordon, R. A. (1993). The effect of strong versus weak evidence on the assessment of race stereotypic and race nonstereotypic crimes. *Journal of Applied Social Psychology, 23*, 734-749. doi:10.1111/j.1559-1816.1993.tb01112.x

Gordon, R. A., Bindrim, T. A., McNicholas, M. L., & Walden, T. L. (1988). Perceptions of blue-collar and white-collar crime: The effect of defendant race on simulated juror decisions. *The Journal of Social Psychology, 128*, 191-197. doi:10.1080/00224545.1988.9711362

Gordon, R. A., & Anderson, K. S. (1995). Perceptions of race-stereotypic and race-nonstereotypic crimes: The impact of response-time instructions on attributions and judgments. *Basic and Applied Social Psychology, 16*, 455-470. doi:10.1207/s15324834basp1604_4

Hastie, R. (1994). Inside the juror. Cambridge, England: Cambridge University Press.

Hastie, R., & Dawes (2009). Rational choice in an uncertain world: The psychology of judgment and decision making. Thousand Oaks, CA: SAGE Publications.

Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process: A regression-based approach. New York: The Guilford Press.

Hetey, R. C., & Eberhardt, J. L. (2014). Racial disparities in incarceration increase acceptance of punitive policies. *Psychological Science, 25*, 1949-1954. doi: 10.1177/0956797614040307

Hintze, J. (2011). Power analysis and sample size (PASS) 11. NCSS, LLC. Kaysville, Utah.

Hu, L., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424-453. doi: doi: 10.1037/1082-989X.3.4.424

Huntley, J. E., & Costanzo, M. (2003). Sexual harassment stories: Testing a story mediated model of juror decision-making in civil litigation. *Law and Human Behavior*, *27*, 29-51. doi: 10.1023/A:1021674811225

Jones, C. S., & Kaplan, M. F. (2003). The effects of stereotypical crimes on juror decision making and information-processing strategies. *Basic and Applied Social Psychology,* 25, 1-13. doi:10.1207/S15324834BASP2501_1

Joreskog, K.G., & Sorbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources.

Kleider, H. M., Knuycky, L. R., & Cavrak, S. E. (2012). Deciding the fate of others: The cognitive underpinnings of racially biased juror decision making. *The Journal of*

General Psychology, 139, 175-193. doi: 10.1080/00221309.2012.686462

Lieberman, J. D., & Krauss, D. A. (2009). Jury psychology: Social aspects of the trial process. Burlington, VT: Ashgate.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67*, 808-817. doi: 10.1037/0022-3514.67.5.808

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Wheeler, V. (1996). On resisting the temptation for simplification: Counterintentional effects of stereotype suppression on social memory. *Social Cognition, 14*, 1-20. doi: 10.1521/soco.1996.14.1.1

Mirowsky, J., & Ross, C. E., (1991). Eliminating defense and agreement bias from measures of the sense of control: A 2 x 2 index. Social Psychology Quarterly, 54, 127-145. doi: 10.1037/0893-3200.22.2.212

Monteith, M. J., Spicer, C. V., & Tootman, G. D. (1998). Consequences of stereotype suppression: Stereotypes on AND not the rebound. *Journal of Experimental Social Psychology*, 34, 355-377. doi: 10.1177/0146167203259935

Maddux, W. W., Barden, J. Brewer, M. B., & Petty, R. E. (2005). Saying no to negativity: The effect of context and motivation to control prejudice on automatic evaluative responses. *Journal of Experimental Social Psychology, 41*, 19-35. doi:10.1016/j.jesp.2004.05.002

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk, *Behavioral Research Methods, 44*, 1-23. doi: 10.3758/s13428-011-0124-6.

Mazzella, R., & Feingold A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta- analysis. *Journal of Applied Social Psychology, 24*, 1315-1338. doi: 10.1111/j.1559-1816.1994.tb01552.x

Minear, M., & Park, D.C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*, 630–633. doi: 10.3758/BF03206543

Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior, 29*, 621-637. doi:10.1007/s10979-005-8122-9

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and

mediation is moderated. *Journal of Personality and Social Psychology, 89*, 852–863. doi:10.1037/0022-3514.89.6.852

Paolacci, G., Chandler, J., & Ipeirotis (2010). Running experiments on Amazon Mechanical Turk, *Judgment and Decision Making, 5*, 411-419.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*, 242-258. doi:1037/0022-3514.51.2.242

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 521-533. doi:10.1037/0278-7393.14.3.521

Pennington, N., & Hastie, R. (1990). Practical implications of psychological research on juror and jury decision making. *Personality and Social Psychology Bulletin*, *16*, 90-105. doi:10.1177/0146167290161007

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62,* 189-206. doi:10.1037/0022-3514.62.2.189

Pickens v. Miller. California, 216 F.Supp.2d 1011 (2002)

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36*, 717-731. doi:10.3758/BF03206553

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891. doi:10.3758/BRM.40.3.879

Sadowski, C. J. (1993). An examination of the short need for cognition scale. *The Journal of Psychology, 127,* 451-454. doi:10.1177/00131649921969983

Sadowski, C. J., & Gulgoz, S. (1992b). Internal consistency and test-retest reliability of the Need for Cognition Scale. *Perception and Motor Skills, 74,* 610. doi:10.2466/pms.1992.74.2.610

Schuller, R. A., Kazoleas, V., & Kawakami, K. (2009). The impact of prejudice screening procedures on racial bias in the courtroom. *Law and Human Behavior, 33*, 320-328. doi:10.1007/s10979-008-9153-9

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290-312. doi:10.2307/270723.

Steiger, J.H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: SYSTAT.

Sunnafrank, M., & Fontes, N. E. (1983). General and crime related racial stereotypes and influence on juridic decisions. *Cornell Journal of Social Relations, 17*, 1-15.

Sweeney, L. T, & Haney, C. (1992). The influence of race on sentencing: A meta-analytic review of experimental studies. *Behavioral Sciences & the Law, 10*, 179-195. doi:23701900204

Tabachnick, B. G., & Fidell, L. S. (2006). Using multivariate statistics (5th ed.). Boston, MA: Pearson.

Towles-Schwen, T., & Fazio, R. H. (2003). Choosing social situations: The relation between automatically activated racial attitudes and anticipated comfort interacting with African Americans. *Personality and Social Psychology Bulletin, 29*, 170-182. doi:10.1177/0146167202239042

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminancy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. Psychological Methods, 6, 371-386. doi:10.1037//1082-989X.6.4.371

Yale, R. N. (2013). Measuring narrative believability: Development and validation of the narrative believability scale (NBS-12). *Journal of Communication, 3*, 578-599. doi:10.1111/jcom.12035

VITA

Dr. Kluwe was born and raised in Berlin, Germany. Before attending Loyola University Chicago for her Ph.D. in Applied Social Psychology in 2010, Dr. Kluwe attended Northeastern Illinois University in Chicago, IL, where she earned a Bachelor of Arts in Psychology with honors distinction and summa cum laude, in 2009. In 2012, Dr. Kluwe received a Master of Arts Applied Social Psychology from Loyola University Chicago.

While at Loyola, Dr. Kluwe was awarded a Graduate Research Assistantship from 2010 to 2014 and the Advanced Doctoral Fellowship for the 2014-2015 academic year.