City University of New York (CUNY)

CUNY Academic Works

Dissertations, Theses, and Capstone Projects

CUNY Graduate Center

2-2017

Diversity's Impact on the Quality of Deliberations

Amanda N. Bergold

The Graduate Center, City University of New York

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/gc_etds/1859

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).

Contact: AcademicWorks@cuny.edu

DIVERSITY'S IMPACT ON THE QUALITY OF DELIBERATIONS

by

AMANDA NICHOLSON BERGOLD

John Jay College and the Graduate Center, CUNY

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the Degree of Doctor of Philosophy, The City University of New York

2017

© 2017

AMANDA NICHOLSON BERGOLD

All Rights Reserved

Diversity's Impact on the Quality of Deliberations

by

Amanda Nicholson Bergold

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Margaret Kovera_____

Date

Chair of Examining Committee

Richard Bodnar_____

Date

Executive Officer

Steven Penrod_____

Daryl Wout_____

Jennifer Hunt_____

Samuel Sommers_____
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Diversity's Impact on the Quality of Deliberations

By

Amanda Nicholson Bergold

Advisor: Margaret Kovera

Racial minorities endure unfair treatment in our legal system on a variety of different outcomes, jury decisions in particular. Courts and researchers propose increasing diversity in juries as a method for improving jury deliberations and reducing racially biased outcomes for minority defendants (*Peters v. Kiff*, 1972; Sommers 2006). In the present research, I investigated the impact of diversity on the quality of deliberations, as defined by both sensitivity to case strength, and by more high quality contributions to deliberations. In the first study, both minority group members and majority group members provided more, higher quality, contributions when they deliberated in diverse juries than when they deliberated in non-diverse juries. However, there was no evidence that diversity increased sensitivity to case strength. In the second study, I manipulated power and wealth using a minimal groups paradigm (MGP), and then created groups that were either diverse or homogenous on this dimension. Diverse juries were more likely to acquit the defendant compared to the non-diverse low wealth/power juries. In addition, jurors deliberating in diverse juries provided more high quality contributions than those deliberating in non-diverse juries. Thus, diversity's benefits extend to minority jurors and wealth and power may play a role in the quality of jury deliberation.

Table of Contents

List of Tables	vi
Chapter 1: Introduction	1
Chapter 2: Racial Bias in the Legal System.....	3
Chapter 3: Jury Diversity	10
Chapter 4: Deliberations	18
Chapter 5: Interracial Interactions	21
Chapter 6: Overview of the Proposed Research	25
Chapter 7: Overview of Study One.....	28
Chapter 8: Study 1 Methods	30
Chapter 9: Study 1 Results.....	39
Chapter 10: Study 1 Discussion.....	48
Chapter 11: Study 2 Introduction.....	55
Chapter 12: Study 2 Overview.....	59
Chapter 13: Study 2 Methods	61
Chapter 14: Study 2 Results.....	65
Chapter 15: Study 2 Discussion.....	73
Chapter 16: Conclusions and Future Directions	79
Appendix A: Post-Trial Questionnaire	95
Appendix B: Post-Deliberation Questionnaire	100
Appendix C: Kuhn, Weinstock & Flaton (1994) Coding Sheet Revised	104
Appendix D: Kuhn, Weinstock & Flaton (1994) Coding Scheme Revised	106
Appendix E: Minimal Groups Questionnaire (including Tajfel Matrices).....	109
References.....	112

List of Tables

Table 1: Study 1 guilty verdict descriptive statistics	89
Table 2: Study 1 coding criteria descriptive statistics	90
Table 3: Study 2 Pre-deliberation guilty verdict descriptive statistics	91
Table 4: Study 2 Post-deliberation guilty verdicts descriptive statistics	92
Table 5: Study 2 coding criteria descriptive statistics	93
Table 6: Study 2 jury-level guilty verdict descriptive statistics.....	94

Chapter 1: Introduction

Numerous jury decisions have been second-guessed based on possibilities that racial bias entered into some level of the jury decisions. One early example was the case of *Norris v. Alabama* in which the Supreme Court heard the appeal of one of the nine “Scottsboro Boys” convicted by an all-White jury of raping two White women. The jury convicted all nine boys despite convincing evidence pointing towards their innocence, raising questions about the lack of diversity on the jury and what impact that might have had on the decision-making process. Particularly concerning was the fact that no African-American had ever sat on a jury in Scottsboro, Alabama. Eventually, the conviction was reversed, and it became clear that the Supreme Court would no longer tolerate the total exclusion of African-Americans from serving on juries. The *Norris* case was reversed in 1935, so we may be tempted to believe that problems stemming from a lack of diversity on juries are a part of this country’s past. A more recent case confirms the persistence of issues with jury diversity. Allen Snyder was accused of killing a White man in 1995. Although five out of thirty-six prospective jurors were African-American, he was tried by an all-White jury. All five African-Americans were removed from the jury by the prosecution via peremptory challenges. However, when Snyder appealed his conviction in 2008, the Court ruled that he had not sufficiently demonstrated that the prosecution’s use of the peremptory challenges was discriminatory.

More recent controversial jury outcomes have also been attributed to the lack of diversity on juries. Even more recently, George Zimmerman was accused of killing a Black teenager who was in possession of only a candy bar. While Zimmerman admitted to shooting the teenager, he claimed self-defense under Florida’s stand your ground law. An all-female jury with only one non-white member eventually acquitted Zimmerman of murder. The controversial outcome of this trial has been linked in the media to the homogenous nature of the jury that tried

Zimmerman. Outcomes of cases tried by homogenous juries tend to violate people's perceptions of fairness and contribute to outrage over unpopular verdicts (Fairchild & Cowan, 1997; Sommers, 2006).

Although there are disagreements over the correctness of famous verdicts, there seems to be agreement among both the public and the courts that diversity on juries is important and valuable. Chief Justice Marshall noted in *Peters v. Kiff* (1972) "When any large and identifiable segment of the community is excluded from jury service, the effect is to remove from the jury room qualities of human nature and varieties of human experience the range of which is unknown, and perhaps unknowable." The present research is designed to shed light on the effects of racial diversity in juries.

Chapter 2: Racial Bias in the Legal System

Minorities have a long history of mistreatment in the criminal justice system (for a review, see Hunt, 2015; see also Barnes & Kingsnorth, 1996; Williams, Demuth, & Holcomb, 2007). Despite the implementation of various interventions attempting to eliminate ethnic bias in the justice system, minority defendants are still convicted more often, awarded longer sentences, and sentenced to death more often than White defendants (Mustard, 2001; Williams, et al., 2007). These disparities remain even after controlling for factors such as differential arrest rates (Sorensen, Hope & Stemen, 2003). In a study of single drug felonies in California, African-Americans were more likely to receive a prison term than Whites, and those prison terms were longer those received by Whites (Barnes & Kingsnorth, 1996). The inequalities in punishment are most consequential in capital cases. One study of the outcomes in 2,000 murder cases in Georgia found that Black defendants were 4.2 times more likely than White defendants to be sentenced to death (Baldus, Pulaski & Woodworth, 1983). This analysis controlled for 230 variables including prior convictions, yet the disparity in death sentences remained extreme. Another set of analyses of death penalty cases documented this same pattern of racial disparity for Black defendants when the victim of murder was White (Bowers, 1984).

Racial disparities appear when examining exonerations as well. Additionally, the National Registry of Exonerations reports that almost half (46.4%) of exonerees are Black and 60.4% of exoneration cases featured a minority defendant, a percentage that is much higher than the demographic percentage of minorities in the United States (National Registry of Exonerations, 2016). When looking at DNA exoneration cases the overrepresentation of minorities grows even higher, with 70% of exonerated individuals being members of a racial or ethnic minority group (www.innocenceproject.org).

Even within minority groups, defendants can face harsher outcomes based on stereotypes associated with their appearance. Researchers obtained pictures of African-American defendants from real death-eligible cases in Philadelphia that had advanced to the penalty phase and had participants rate how stereotypically Black the faces were. They found that jurors were more likely to sentence to death defendants who appeared more stereotypically Black than defendants who had less phenotypically African facial features (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). These racial and ethnic disparities in punishment persist despite legal rulings intended to increase diversity in juries and reduce racial bias in decision-making.

The clear and prevalent disadvantages experienced by racial minorities in the justice system inspired research examining what role racial bias in jury decisions might play in these punishment disparities. Researchers have examined the problems caused by prejudice and biases, primarily using jury simulation methodologies. In the majority of these studies, participants either read a written summary of trial evidence or watch a videotaped reenactment of a trial and then render verdicts or provide sentencing outcomes for Black (vs. White) defendants (Sweeney & Haney, 1992). Four meta-analyses have been performed to aggregate the psychology research on bias toward outgroup defendants. In many of these studies, bias toward outgroup defendants has been defined as harsher judgments for defendants who are members of a different racial/ethnic group than participants than for defendants who are members of participants' ingroup. The earliest meta-analysis found a small but significant outgroup bias effect in mock juror verdicts and guilt ratings (Sweeney & Haney, 1992). This meta-analysis, however, only examined White mock jurors' guilt judgments for Black (vs. White) defendants. A second meta-analysis found no evidence of bias towards outgroup defendants in mock juror studies (Mazzella & Feingold, 1994). The authors conducting this meta-analysis did not code for participant race

and defined racial bias as bias toward Black defendants. A third meta-analysis found a small but significant effect for bias against outgroup defendants (Mitchell, Haw, Pfeifer, & Meissner, 2005). This meta-analysis also found larger effect sizes for sentencing decisions than for verdicts. The most recent meta-analysis examining individual characteristics that impact legal decisions found an overall outgroup bias effect, but upon examination of moderators, found that it was driven by White jurors discriminating against Hispanic defendants, not Black defendants (Devine & Caughlin, 2014).

Individual studies also provide evidence of bias toward outgroup defendants. When mock jurors were allowed the opportunity to assign punishment for defendants, they awarded longer sentences and awarded the death penalty more frequently to outgroup defendants than to ingroup defendants (Forsterlee, Forsterlee, Horowitz, & King, 2006; Lynch & Haney, 2009). There is a consistent pattern of mock jurors rendering higher guilt ratings for outgroup defendants than ingroup defendants in recent mock juror simulations (Bucolo & Cohn, 2010; Espinoza & Willis-Esqueda, 2008; Hodson, Hooper, Dovidio, & Gaertner, 2005).

Most mock juror studies operationalize outgroup defendant bias as White mock jurors' harsher guilt ratings for Black defendants as compared to White defendants. Recently, however, there has been more of an emphasis on including a greater variety of minority groups in mock jury research. When researchers have tested bias towards other minority groups, the pattern of White mock juror bias towards minority defendants remains (Adams, Bryden & Griffith, 2011; Espinoza & Willis-Esqueda, 2008). White mock jurors' are more likely to convict, award harsher guilt ratings and longer sentences when judging Hispanic, Mexican-American, and Middle Eastern defendants than when they are deciding cases with a White defendant (Adams et al., 2011; Bodenhausen & Lichtenstein, 1987; Espinoza & Willis-Esqueda, 2008). Some researchers

have had minority participants render judgments for White defendants in mock juror studies. The Mitchell et al. (2005) meta-analysis revealed that the outgroup bias effect was larger for minority participants rendering judgments for White defendants, than for White participants passing judgment on minority defendants. Additionally, the most recent meta-analysis examining racial bias found the largest outgroup bias effects for Whites making judgments about Hispanic defendants (Devine & Caughlin, 2014).

Although there is relative consensus that bias towards outgroup defendants, particularly toward minorities, can and does occur throughout the literature on mock juror bias, there are several studies in the literature that have not found bias toward outgroup defendants in mock juror verdicts (Foley & Pigott, 2002; Mazzella & Feingold, 1994). Some studies have even documented “reverse discrimination” effects by which mock jurors are more lenient towards outgroup defendants than ingroup defendants (Foley & Pigott, 2002). There is a substantial literature suggesting that the modern emergence of political correctness as a desirable trait causes people to deliver unbiased responses, particularly when they are aware of the potential for their responses to be perceived as prejudiced (Carver, Glass & Katz, 1978). Although public prejudice has decreased, biases about other racial and ethnic groups continue to lurk under the surface and enter into decision-making unbeknownst to decision-makers (Gaertner & Dovidio, 1986). These more subtle forms of prejudice predict which situations will be more likely to evoke biased decision-making.

The patterns of racial bias in mock jury decisions mirror the study of prejudice in more basic social psychological research. Dovidio (2001) describes the evolution of the study of prejudice as occurring in three distinct periods or “waves.” During the first wave, researchers considered prejudice to be an abnormality or form of erratic behavior or thought, with research

during this period devoted to identifying and evaluating prejudice and determining its causes (Dovidio, 2001; Gilbert, 1951). From this first stage of prejudice research, the field evolved from its view of prejudice as an abnormality, to the view of prejudice as a part of normal thinking (Dovidio, 2001). For example, Allport (1954) viewed prejudice as a result of an ingroup categorization process by which people differentiate between ingroup and outgroup members, with hostility towards outgroups being part of the process of identifying more with one's ingroup. The social identity perspective on prejudice (Tajfel & Turner, 1979) also emphasized the role of outgroup discrimination in establishing one's identity in an ingroup. The third wave of prejudice research was devoted to discovering and measuring each individual's level of implicit bias (Dovidio, 2001). During this phase researchers developed measures of implicit bias such as the Implicit Association Task (IAT), which measured individuals' level of association with White and Black faces and positive and negative stimuli (Greenwald, McGhee, & Schwartz, 1998). When taking the IAT, many White participants responded more quickly to positive traits when they were provided Black primes with the reverse being true for negative traits (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997).

The second "wave" described by Dovidio (2001) may be the most helpful in understanding discrepancies in findings of prejudice. During this period there was a rise in the importance placed on political correctness and egalitarian principles. Prejudice did not disappear, but instead it took a more subtle form (Dovidio & Gaertner, 1998; Gaertner & Dovidio, 1986). The aversive racism perspective is consistent with the notion that although many people considered themselves to be egalitarian, they still harbored negative feelings toward outgroup members. Often White individuals would not self-report prejudice toward Black individuals; however when a situation was more ambiguous, White individuals exhibited patterns of biased

responses toward black individuals (Gaertner & Dovidio, 1986). Aversive racism was first demonstrated by examining helping behaviors. Using the bystander paradigm (Latane & Darley, 1968), researchers developed a method of testing theories on aversive racism. Participants encountered an individual in need of help. In some conditions there were other bystanders present; in others, the participants were alone. Participants expressed no bias toward the Black victim in the less ambiguous conditions when they were the only available helpers; when other people were around, they helped the White victim twice as often as they helped the Black victim (Gaertner & Dovidio, 1977). When the situation was more ambiguous in terms of whether the participant should intervene because of the presence of bystanders, they were more likely to discriminate against the Black victim. Later research extended this theory to hiring recommendations; when a candidate's criteria were a more questionable match for the position, participants favored the White applicant over the Black applicant (Dovidio & Gaertner, 2000).

Concerns about appearing politically correct are consistent with findings of race salience in mock juror research (Sommers & Ellsworth, 2000; 2001). When racial issues are made salient during the course of a trial, mock jurors are less likely to exhibit bias toward outgroup defendants in their verdict decisions (Sommers & Ellsworth, 2000; 2001). The phenomena of race salience suggests that mock jurors, when confronted with a trial simulation that makes issues of race salient, are able to prevent their biases from affecting their interpretation of the evidence. Race salience fits with these modern theories of prejudice, which maintain that people will not express prejudice in situations that alert them to the potential for their actions to be perceived as biased (Gaertner & Dovidio, 1986; Pearson, Dovidio & Gaertner, 2009; Saucier, Miller & Doucet, 2005). Yet when racial issues call their attention to the possibility of expressing bias, people attend to their biases and correct for them.

As the aversive racism perspective would predict, ambiguous situations tend to be more likely to evoke evidence of bias toward outgroup defendants (Gaertner & Dovidio, 1986). For example, when the prosecution brought in evidence later ruled inadmissible by the judge, mock jurors were more likely to exhibit bias towards outgroup defendants (Hodson, et al., 2005). The ability to make race neutral explanations can also increase attorneys' use of racially motivated peremptory challenges (Baldus, Woodworth, Zuckerman, Weiner, & Broffitt, 2001; Clark, Boccaccini, Caillouet, & Chaplin, 2007; Rose, 1999; Sommers & Norton, 2007); attorneys make racially biased decisions in how they use their peremptory challenges, but few mention race when asked to explain these decisions.

Chapter 3: Jury Diversity

The U.S. Courts have a long history of addressing problems with racial and ethnic minorities being denied the ability to serve on juries. As early as 1879, the courts struck down laws restricting jury service to Whites (*Neal v. Delaware*, 1880; *Strauder v. West Virginia*, 1879). Property and income requirements for jury service, however, had the impact of effectively maintaining all-White membership on juries for a number of years. *Swain v. Alabama* (1965) was the first instance of the Court expressly prohibiting the use of racially biased peremptory challenges to remove African-Americans from juries, yet the standard set in this case required demonstrating a pattern of minorities' exclusion from juries from across the entire county. In *Batson v. Kentucky* (1986), the Supreme Court highlighted its increasing concerns about racial bias, reflecting the societal embrace of egalitarian principles. In this ruling, the Court recognized that racially biased jurors had the potential to impact verdicts (*Batson v. Kentucky*, 1986).

Prior to the *Batson* ruling peremptory challenges did not require any justification or approval from a judge. Prosecutors were (and remain) far more likely to exclude Black venire members than are defense attorneys (Baldus, et al., 2001). As a result, prosecutors were removing minority jurors, often when the defendant was Black. The *Batson* Court sought to remedy this problem by prohibiting the use of peremptory challenges to exclude members of ethnic and racial minority groups from juries, and instituting a procedure by which to challenge discriminatory peremptory challenges. Critics of the *Batson* ruling contend that the challenge relies on a flawed two-step process to prevent discrimination. First, an attorney must raise the issue of a racially motivated peremptory challenge. Then, if the judge agrees that the challenge was based on the potential juror's racial group membership, the attorney has the opportunity to provide a race-neutral explanation for his or her challenge. The judge may either accept the race-

neutral justification as legitimate and allow the peremptory challenge, or reject the explanation as prohibit the challenge.

Despite the protections afforded by *Batson*, racial bias remains in the exercise of peremptory challenges (Baldus et al., 2001; Clark et al., 2007; Rose, 1999). Although the *Batson* ruling was intended to curb the use of peremptory challenges to discriminate against minority group members, the procedure introduced in *Batson* does not successfully eliminate racially motivated peremptory use. Archival research on peremptory challenge use after the *Batson* ruling has determined that very few of *Batson* challenges are successful, allowing prosecutors to use peremptory challenges to prevent minorities from serving on juries without consequence (Melilli, 1996). As evidenced by recent studies on peremptory challenge use, neutral explanations for racially motivated decisions often come naturally to decision-makers and are easy to provide (Sommers & Norton, 2007). In fact, warnings about the illegality of using peremptory challenges to remove prospective jurors based on gender actually increased the elaborative nature of these justifications, while failing to eliminate the use of gender as factor for removal (Norton, Sommers & Brauner, 2007). As a result, prosecutors are often able to provide convincing race-neutral reasons for making peremptory challenges, rendering the vast majority of *Batson* challenges unsuccessful (Melilli, 1996).

The *Batson* standard was extended in *Miller-El v. Dretke* (2005), which allowed a petitioner to include factors outside the case when protesting discrimination in a prospective juror removal. The extension of *Batson* was later upheld in *Snyder v. Louisiana* (2008), which involved an appeal reconsidered in light of the *Miller-El* decision. Both court decisions built upon the *Batson* ruling, but neither significantly enhanced the effectiveness of prohibitions on discriminatory peremptory challenges. In a recent case, *Foster v. Chatman* (2016), the Supreme

Court reversed the dismissal of the petitioner's challenge based on evidence of racism in the prosecutor's notes, obtained years after the trial took place.

The rulings in these court cases were intended to eliminate racial bias in peremptory challenge use, but racially biased peremptory challenges remain (Clark et al., 2007; Rose, 1999). It is difficult to prove discrimination based on a single removal, and including factors outside the case does not provide adequate information about an attorney's motivation. As a result of the inability of *Batson* to successfully prevent discrimination in allowing minority venire members to serve on juries, prosecutors continue to possess the ability to remove minority venire members from jury service. For minority defendants, this practice may help exacerbate the disadvantages they suffer in the legal system compared to White defendants.

Empirical research largely supports the courts' assumption that jury diversity mitigates problems with racial bias in juries. Early research found differences in the quality of deliberations as a function of jury composition that varied on race and attitudinal variables (Cowan, Thompson, & Ellsworth, 1984; Fitzgerald & Ellsworth, 1984). In states that allow the death penalty, potential jurors are screened for bias against the death penalty through voir dire. The overwhelming majority of death-qualified jurors tend to be White men (Fitzgerald & Ellsworth, 1984). Therefore it is likely that death-qualified juries are juries with a very high proportion of White men. A qualitative analysis comparing deliberations in mixed (not death-qualified) juries to juries that were death-qualified showed that mock jurors in mixed juries had a more thorough understanding of the case facts and verdict criteria than did those in death-qualified juries (Cowan, et al., 1984). Death-qualified juries were also more likely to convict the defendant than were mixed juries.

Because death-qualified juries are more conviction-prone (Cowan et al., 1984) and also feature a disproportionate number of White men (Fitzgerald & Ellsworth, 1984), it is likely that an increase in White men on a jury could increase the likelihood that the jury is conviction-prone. An evaluation of sentencing decisions by unanimous mock juries found that juries with a high proportion of White men would be more likely to sentence a defendant to death than would juries with a low proportion of White men (Lynch & Haney, 2011). For the White defendant, there was little difference between the death sentences awarded (60% for juries with fewer than two White men vs. 63% for juries with two or more White men). However, in cases with a Black defendant, 68% of juries with fewer than two White men gave a death sentence, whereas 86% of juries with two or more White men sentenced the Black defendant to death (Lynch & Haney, 2011). Thus, the White men's tendency to sentence a defendant to death was restricted to cases with Black defendants. These findings indicate that the addition of minorities to juries may reduce their tendency to punish minority individuals more harshly.

Deliberations in diverse juries are also higher quality than deliberations in homogenous juries (Sommers, 2006). Prior to deliberations, White jurors were less likely to vote guilty for the Black defendant when they anticipated deliberating in a diverse jury than when they would be deliberating in a homogenous jury. The beneficial effects of diversity were mostly explained by higher quality contributions from the White participants in diverse juries. White participants mentioned more relevant case facts, made fewer factual errors, and discussed the possibility of racism to come into play more when deliberating in a diverse jury. The researchers used length of deliberation, number of case facts discussed, number of inaccuracies, number of mentions of possible missing pieces of evidence, and the mention of race to evaluate quality of deliberations (Sommers, 2006).

More basic psychological research finds support for diversity's benefits in that racially heterogeneous groups outside the courtroom outperform homogenous ones on a number of tasks (McLeod, Lobel, & Cox, 1996; Sommers, Warp & Mahoney, 2008). For example, the anticipation of participating in a racially diverse group increased Whites' information-processing levels as compared to when they were told they would be participating in an all-White group (Sommers et al., 2008). The study of diversity has revealed substantial benefits in terms of improved group performance (Phillips, Mannix, Neale & Gruenfeld, 2004), and these advantages are consistent among varied definitions of diversity (Gruenfeld, Mannix, Williams, & Neale, 1996; Stasser, Stewart, & Wittenbaum, 1995; Watson, Kumar, & Michaelson, 1993).

There are some mixed results in terms of diversity's benefits and improved performance for diverse groups. As diversity is difficult to operationalize and investigate, and some reviews have concluded that it is difficult to recommend diversity based on the existing literature (Webber & Donohue, 2001; Kochan et al., 2003). Although it seems logical that diverse groups would benefit from the varied perspectives of the group members (Nemeth, 1986), empirical research does not consistently support this view. Diversity does not benefit a group in terms of group cohesion, and conflict is more likely to occur in diverse groups (De Dreu & Weingart, 2003). Moreover, some research indicates the need for congruence between social and knowledge ties to reap the benefits of diversity. For example, Phillips et al. (2004) found that for diverse groups to be successful each member of the group must have role that is consistent with the rest of the group's expectations. Overall, however, the research on diversity in both group performance and jury diversity, supports the view that diverse juries will perform better than homogenous ones.

To explain the results from Sommers (2006) the researchers hypothesized that two related theories might explain diversity's benefits in terms of White mock jurors' motivation to avoid prejudice. The theory of race salience describes the reduction of prejudice exhibited by White mock jurors when racial issues are brought up in the context of a trial (Sommers & Ellsworth, 2001). The researchers in Sommers (2006) proposed that the presence of minority jurors may have raised concerns about racial issues for White mock jurors, causing them to enter the deliberations expecting to discuss racial issues. The diversity benefit could also be explained by the finding that White individuals are often motivated to engage in higher levels of thinking when they are evaluating targets who are minority group members (Sargent & Bradfield, 2004; White & Harkins, 1994). Researchers have termed this phenomenon the "watchdog effect," describing the majority group individuals as watchdogs guarding against potential prejudice by themselves or others when evaluating a stigmatized target (Fleming, Petty & White, 2005). This benefit in cognitive processing occurred only when participants expected to discuss race-relevant issues, lending further support to the proposition that Whites are acting as watchdogs to guard against the potential for prejudice to enter their decisions (Fleming et al., 2005).

The motivation of White individuals to avoid appearing prejudiced may explain some of diversity's benefits, specifically for why White individuals tend to benefit from diversity. When asked to consider applicants for a job, participants were more sensitive to the strength of a Black applicant's resume than they were to the strength of a White applicant's resume (Fleming et al., 2005). Similarly, people were more convinced by a strong argument from a stigmatized source than a non-stigmatized source but less convinced by a weak argument if the source of the statement was stigmatized (Petty et al., 1999; White & Harkins, 1994), although this effect may be confined to low prejudice participants (Petty et al., 1999). These findings are consistent with

previous research, which also demonstrated that people provide more polarized evaluations of stigmatized individuals (Hass et al., 1991; Linville & Jones, 1980). The watchdog concept has mainly been tested by examining the degree to which evaluations of stigmatized targets are polarized, indicating a more thorough thinking pattern for non-stigmatized individuals (Fleming et al., 2005).

Research has also established the presence of the watchdog effect into the courtroom. Researchers manipulated the strength of alibi evidence in a criminal case and had mock jurors evaluate the strength of the case. Mock jurors were more sensitive to evidence strength when the defendant in the case was Black than when the defendant was White (Sargent & Bradfield, 2004). The watchdog effect also provides an explanation for the findings that making racial issues salient at trial can decrease bias against minority defendants (Sommers & Ellsworth, 2000; 2001). Although the watchdog effect appears more prominently for individuals who are very low in prejudice on several scales, this tendency to protect stigmatized individuals by more thoroughly processing messages may not be confined to egalitarian-minded individuals.

The watchdog effect relies on sensitivity to argument strength as an indicator of level of processing, consistent with the tenets of dual-process models of persuasion. Both the Heuristic-Systematic Model (HSM) and the Elaboration Likelihood Model (ELM) are based on the theory that people process information in different ways depending on their resources and motivation (Chaiken, 1980; Petty & Cacioppo, 1979). Although there are subtle differences between the HSM and the ELM, both propose a dual-process cognitive strategy by which people process messages more or less thoroughly depending on the situation. The two models also concur that people will process a persuasive message more critically and more in-depth when they possess the ability and motivation to do so. The HSM model focuses more on what types of cues people

rely on when they are evaluating an argument. At lower processing levels, people are more likely to rely on heuristic cues to decision-making such as consensus implies correctness, whereas at higher processing levels people will systematically evaluate the quality of the persuasive arguments (Chaiken, 1980). Similarly, the ELM focuses on whether people use a central (or more systematic) processing route in which they pay attention to the quality of the arguments in the message (like the HSM's systematic processing), or whether they evaluate an argument by relying on peripheral cues such as source credibility or attractiveness (Petty & Cacioppo, 1979).

There are a variety of reasons why people might be motivated to take a more thorough approach to evaluating an argument. Early work on persuasion used personal involvement to increase participants' motivation to process a message (Petty & Cacioppo, 1979). For instance, a typical persuasion study had undergraduate students read arguments supporting a proposal for final comprehensive exams being considered at their university, which would be instituted while the participants were still attending the university or once the students had already graduated. When the students read the self-relevant article (i.e., the proposed exams would be instituted while they were still in school) they processed arguments more systematically when making their decision. When the message was less self-relevant, participants took a more peripheral route and relied on more heuristic cues to process the message. When people process information more systematically, they differentiate between strong and weak arguments. According to the Elaboration Likelihood Model (ELM) this occurs because cognitively elaborating on strong arguments strengthens them further, and elaborating on weak arguments weakens them further (Cacioppo et al., 1986). In the case of the watchdog effect, the presence of minorities on the jury may act as a factor that motivates participants to process information systematically.

Chapter 4: Deliberations

To resolve the issues presented by biased jury decisions and ensure fairer outcomes for minority defendants, the courts and researchers have proposed a variety of protections. These prescribed solutions include diverse or representative juries, deliberations, and warnings about fairness and bias. Ideally, deliberations alone would have the ability to prevent bias toward minority defendants in jury verdicts. The U.S. court system relies on jury groups on the assumption that they are better than individuals in catching and correcting errors (*Apodaca et al. v. Oregon*, 1972). In the court decision in *Apodaca*, a case that dealt with the question of the necessity of unanimity in jury decisions, the court maintained that deliberating jury groups would outperform individual jurors through error correction and increased memory.

Some research supports that jury groups will produce more accurate decisions than individual jurors (Clark, Stephenson, & Kniveton, 1990; Patry, 2008; Vollrath, Sheppard, Hinsz, & Davis, 1989). In mock jury research studies, mock jurors who deliberated were less likely to rely on peripheral, superficial cues such as defendant attractiveness when making verdict decisions than those who rendered individual verdicts (Patry, 2008). Additionally, mock jurors evaluating evidence in groups made fewer memory errors and exhibited better memory performance than individuals (Clark et al., 1990; Vollrath et al., 1989). Moreover, items that jurors remembered incorrectly were more likely to be corrected if they were brought up during a jury deliberation (Pritchard & Keenan, 2002). Juries were also more successful in adhering to judicial instructions to exclude inadmissible testimony after deliberations than they were prior to deliberations (Kerwin & Shaffer, 1994). In some research, deliberating juries even delivered less racially biased verdicts than did individual jurors (Adams, et al., 2011; Bothwell, Pigott, Foley, & McFatter, 2006).

Basic social and cognitive research also suggests that deliberations should improve fact-finding. Memory pooling, defined as improvements in memory due to collaborative remembering, suggests that groups remembering events tend to exhibit improved performance on memory tasks compared to individuals (Clark, Hori, Puthnan & Martin, 2000 ; Stephenson, Clark, & Wade, 1986; Weldon & Bellinger, 1997; Yarmey & Morris, 1998). Some research even suggests that the more people involved in the group remembering, the more accurate the group recall (Stephenson et al., 1986; Weldon & Bellinger, 1997).

Conversely, other research suggests that deliberations impair juror decision-making (Kramer, Kerr, & Carroll, 1990; Lynch & Haney, 2009; Ruva, McEvoy, & Bryant, 2007). In pretrial publicity (PTP) research, deliberations have failed to correct for memory errors and sometimes heightened PTP's prejudicial impact (Kramer, et al., 1990; Ruva, et al., 2007). Sometimes this benefit was restricted to more peripheral details, with deliberations failing to improve memory for critical evidence (Pritchard & Keenan, 2002). A more recent study examining instructions designed to eliminate the negative effects of PTP found that deliberations were not effective in eliminating jurors' use of PTP in verdict decisions (Ruva & Guenther, 2015).

Research on collective memory also suggests that group deliberation does not always result in higher quality outcomes. Some memory research suggests that people recalling events together will experience collaborative inhibition, defined as the reduction in recall of a group remembering together compared to combining individuals' own account (Weldon, Blair, & Huebsch, 2000). Social loafing may account for some of the reduction in recall in groups, although it does not account for all of the impairment (Weldon et al., 2000). Stasser (1988) maintained that the reduction in group recall was due to group members' tendency to prefer

recounting shared memories rather than unique memories, lessening each members' unique contributions. Additionally, participation in collective remembering (i.e., the actual conversation in which people discuss the event) can impair individuals' ability to accurately remember the event (Stone, Barnier, Sutton, & Hutton, 2010).

Most important, the majority of the research does not suggest that deliberations will diminish or eliminate bias toward outgroup defendants. In fact, deliberations may cause jurors to become more punitive toward minority defendants (Lynch & Haney, 2009). In one study, participants watched a videotaped reenactment of a death penalty trial, reported pre-deliberation judgments of mock jurors, and then deliberated in juries. The participants were instructed that the defendant had already been convicted and that they were to decide sentencing. Participants sentenced 57% of Black defendants and 52% of White defendants to death before deliberations, a non-significant difference. After deliberations, mock jurors sentenced 71% of Black defendants to death, only awarding the death penalty for 61% of White defendants. The researchers determined that these effects were mediated by the misuse of penalty phase information such as mitigating and aggravating factors (Lynch & Haney, 2009). Thus, jury deliberations may not serve as an adequate safeguard against racial bias and may in fact increase unfair judgments based on defendant race or ethnicity, particularly in homogenous juries.

Chapter 5: Interracial Interactions

Some features of diverse groups may interfere with decision-making. Interracial interactions are often fraught with discomfort, uncertainty, and anxiety (Blascovich, Mendes, Hunter, Lickel, & Kowai-Bell, 2001; Mendes, Blascovich, Lickel, & Hunter, 2002; Trawalter, Richeson & Shelton, 2009). A body of research establishes that interracial situations tend to invoke more stress than same race interactions, for both majority and minority group members (Trawalter, Richeson & Shelton, 2009). The tension associated with interracial interactions can also decrease cognitive capacity, possibly impairing jurors' ability to engage with the trial information (Richeson & Shelton, 2005; Richeson, Trawalter & Shelton, 2005).

Most of the research examining stress in interracial interactions focuses on the experience of majority group member and their concerns about appearing prejudiced as a primary mechanism to explain anxiety. Individuals with more negative racial attitudes tend to experience heightened anxiety at the prospect of interacting with a member of a stigmatized group (Blascovich et al., 2001; Mendes et al., 2002). The tensions can be even higher when issues of race are a topic of discussion during an interaction (Goff, Steele & Davies, 2008). Raising concerns about political correctness has been demonstrated to pre-occupy White individuals, increasing their tendency to engage in avoidance behaviors towards a minority partner (Norton, Sommers, Apfelbaum, Pura, & Ariely, 2006).

Situations that require the suppression of stereotypes can cause decreased cognitive functioning (Baumeister, Muraven, & Tice, 2000; von Hippel, Silver, & Lynch, 2000). When people devote cognitive energy to inhibiting the expression of stereotypes, it impairs their performance on subsequent cognitively demanding tasks (Baumeister, Muraven, & Tice, 2000; Muraven & Baumeister, 2000). Interracial interactions can cause anxiety and stress by requiring

that people engage in a self-regulatory process, which consumes cognitive resources. Moreover, the degree to which resources are depleted depends on how concerned individuals are with self-regulation. Researchers in one study manipulated people's prejudice concerns by providing false feedback on an IAT, and then having participant engage in an interracial interaction, followed by a Stroop task. Those who were told that they were more prejudiced on the IAT performed worse on the Stroop task than those who were told they were less prejudiced, following their interracial interactions (Richeson & Trawalter, 2005). Participants who believed they needed to suppress their stereotypical views of their partner during their interaction performed worse on the Stroop task after the interaction because they were using more executive resources to regulate their behavior. Another study tested White participants' performance on the Stroop task after an interracial interaction. Participants' racial attitudes (measured by their performance on an Implicit Association Task) were related to cognitive interference on the Stroop task after an interracial interaction but not after a same-race interaction (Richeson & Shelton, 2003). Cognitive decrements following interracial interactions have been document in both minority and White individuals (Richeson et al., 2005). Both White and the minority individuals experienced decreases in cognitive capacity following an interracial interaction, the extent of which was predicted by their negative implicit attitudes towards the other racial group (on the IAT task).

Although many researchers have focused on the consequences of interracial interactions for White individuals there is a similar depletion of cognitive resources for minority individuals following an interracial interaction (Richeson et al., 2005). Minorities, however, may have additional drains on their cognitive resources in an interracial setting. Minority individuals' performance on academic tests can be impaired in the presence of White individuals (Steele &

Aronson, 1995). This phenomenon, known as stereotype threat, occurs because members of stigmatized groups recognize that there are stereotypes about their groups' performance in specific areas. Therefore, these individuals experience psychological threat that interferes with performance in these areas, further confirming the stereotypes that exist.

Rather than being concerned with suppressing stereotypes about others, minorities may be concerned that they will confirm stereotypes that others hold about them. Minorities engaging in interracial interactions often reporting feeling concerned about their potential to be a target of prejudice, regardless of the quality of their contributions to the interaction (Doerr, Plant, Kunstman, & Buck, 2011). Minority concerns also include the worry that they will be seen as a stereotypical member of their group, which increases the pressure and anxiety that they feel in interracial situations (Shelton, Richeson, & Vorauer, 2006). In an interracial deliberation setting, minority individuals may be concerned about confirming stereotypes about their intelligence or cognitive abilities, which could consequently impair their ability to consider trial evidence.

The consequences of resource depletion as they relate to interracial interactions are immense. When interacting with minority group individuals some Whites choose to sit further away from them (Goff et al., 2008). Additionally, when Whites were asked to describe individuals in a photograph identification task, they tended to avoid using race as a characteristic when interacting with a Black partner. This resistance to use race as an identifying characteristic not only impeded performance on the task, but it also increased the likelihood that White participants would exhibit unfriendly behaviors towards their minority group partners (Norton et al., 2006). In a jury deliberation setting decreased cognitive capacity due to concerns about appearing prejudiced could lead to inaccurate evidence processing and lower quality decisions, in addition to causing interpersonal problems and group conflict. Researchers have also

documented decreased performance on cognitive tasks such as the Stroop task following interracial interactions, for both majority and minority group members (Richeson et al., 2005; Shelton & Richeson, 2003).

Jury deliberations share features with the conditions under which anxiety about interracial interactions are most pronounced. Jury deliberations do not tend to have a specific script or structure, and judicial instructions about deliberations are intentionally vague, leaving jurors to their own devices to render accurate legal decisions. The unstructured nature of jury deliberations could prove problematic because unstructured interactions increase the amount of stress people experience compared to situations with a script providing instructions regarding the processes (Avery, Richeson, Hebl, & Ambady, 2009). In the current political climate, it is possible that a legal setting might raise concerns about racial issues. Raising concerns about possible racial prejudice has the potential to worsen tensions surrounding interracial interactions for White individuals (Goff et al., 2008; Norton et al., 2006). Research on interracial interactions suggests that diversity would not benefit the quality of deliberations but instead could possibly increase anxiety and diminish cognitive resources (Mendes et al., 2002; Schmader & Johns, 2003).

Chapter 6: Overview of the Proposed Research

Minorities in general, and Black men in particular, are consistently overrepresented in both the U.S. prison and exoneration statistics. Biased jury decision-making plays a role in harsher outcomes for minority defendants (Mitchell et al., 2005). U.S. courts have repeatedly emphasized the importance of representative juries and the presence of minority jurors on juries, particularly in cases with minority defendants as a method for ameliorating racial bias in jury decisions (*Peters v. Kiff*, 1972; *Ballew v. Georgia*, 1978; *Batson v. Kentucky*, 1986). This emphasis led to the creation of a process by which the removal of jurors through peremptory challenges can be appealed through *Batson* challenges. Studies of *Batson* challenges, however, reveal that they are largely unsuccessful (Clark et al., 2007; Rose, 1999). The combination of unsuccessful *Batson* challenges and the tendency for prosecutors to remove minority jurors from the jury room suggests that minority defendants often have their cases decided by all-White, or majority White juries.

What are the consequences of minority defendants being tried by all-White juries? The consequences appear unfairly severe for members of minority groups, particularly when it comes to sentencing decisions (Baldus, et al., 1983; Eberhardt et al., 2006; Hunt, 2015; Mitchell et al., 2005). Statistics reveal that Black defendants receive longer sentences, and are sentenced to death more often than are White defendants (Mustard, 2001). Mock jury research also indicates that juries composed of mostly White jurors deciding death penalty cases are particularly prone to sentencing Black defendants with the death penalty (Lynch & Haney, 2011).

Whether diversity in jury membership succeeds in reducing bias toward minority defendants in criminal trials has been under-examined. The limited studies on jury diversity largely support the view that diversity increases the quality of deliberations and decreases bias

towards minority defendants (Lynch & Haney, 2011; Sommers, 2006). The findings from jury decision-making research are confirmed by research on more basic group decision-making (Stasser et al., 1995; Phillips et al., 2004). However, there is a need for research examining explanatory mechanisms for the diversity benefit in deliberations. One study of jury diversity revealed that the higher quality of deliberations could be explained mostly by improvements in the contributions by White mock jurors in diverse juries (Sommers, 2006), possibly because of a “watchdog effect.” Specifically, a watchdog effect occurs when majority group members concerned about prejudice pay closer attention to evidence when in the presence of minority group members, particularly when the defendant is a member of a minority group (Fleming et al., 2005).

This finding suggests possible mechanisms for improved jury performance in racially heterogeneous juries but leaves unanswered questions about the impact of minority jurors in diverse juries. Do minority individuals benefit from diversity in a jury group? Will minorities display higher quality and greater quantity of contributions to diverse groups than to homogenous groups? Or will the stress involved in engaging in interracial interactions erode the benefits of diversity?

Alternately, some research suggests that the impact of diversity may not always be beneficial. Diverse groups tend to have more conflict and can be less cohesive than non-diverse groups (De Dreu & Weingart, 2003). Interracial interactions can be fraught with discomfort and cause anxiety, which is not conducive to higher quality performance on group tasks (Shelton & Richeson, 2003; Trawalter, et al., 2009). Situations that are open-ended without a cohesive script are also more likely to evoke anxiety and conflict in interracial interactions, both of which are qualities that define jury deliberations (Avery et al., 2009). The anxieties provoked by interracial

interactions can also impair performance on cognitive tasks and inhibit interpersonal relationships (Norton et al., 2006; Shelton & Richeson, 2003).

The goal of the current research was to investigate whether diverse juries engage in higher quality deliberations than do non-diverse juries. I predicted that diverse juries would engage in higher quality deliberations than would non-diverse juries. Most of the previous research investigates the effect of adding minority jurors to a homogenous jury but the current research also addressed the impact of diversity on minority jurors. I predicted that minority jurors would experience the same benefits from diversity that majority group members do. Another goal of this research was to examine the mechanisms underlying the improved jury deliberations in diverse juries. I also explored the effects of diversity when it is operationalized in a different manner than racial diversity.

In the first study, I manipulated evidence strength and jury diversity, to determine whether diverse juries were more sensitive to evidence strength than non-diverse juries. To address questions of both individual levels of processing and quality of deliberations, analyses were conducted at both the jury and juror level both before and after deliberations to determine what impact the diversity of the group has on individual and group decisions. In the second study, I used a minimal groups paradigm (MGP) and assigned participants to high wealth/power and low wealth/power conditions. In the second study, I explored whether the impact of diversity on evidence processing and deliberations can extend to groups with membership based on something other than racial or ethnic group membership, and explore the role of wealth and power differentials in jury deliberations.

Chapter 7: Overview of Study One

The goal in Study One was to test the hypothesis that diverse juries will have higher quality deliberations than homogenous juries. Sommers (2006) found that racial diversity improved the deliberations of juries but that the benefit resulted mostly from higher quality contributions from White mock jurors. Sommers (2006) attributed this finding to the “watchdog effect,” namely that majority group members are motivated to process information from stigmatized groups (Fleming et al., 2005). In this study, Sommers (2006) also manipulated the race salience of the trial, defined as how salient issues of race were during the trial (Sommers & Ellsworth, 2001). Race salience and the watchdog effect operated similarly, and were derived from more modern theories of prejudice suggesting that when majority group members are aware of the potential to express prejudice, they work hard to avoid it (Dovidio & Gaertner, 1998).

There is reason to believe the beneficial impact of diversity might not extend to minority jurors, at least according to the perspective that diversity improves deliberations through the increased vigilance of White jurors. Thus far the watchdog effect is confined to White jurors evaluating minority targets and arguments from minority sources (Fleming et al., 2005; Petty et al., 1999; White & Harkins, 1994). The main tenet of the watchdog hypothesis is Whites’ motivation to avoid prejudice, and minorities are not usually concerned about appearing prejudiced. Minorities are more sensitive to the potential to be rejected by interaction partners based on their race, and therefore may be more focused on observing the behavior of their interaction partners (Mendoza-Denton, Downey, Purdie, Davis, & Pietrzak, 2002). Minorities may also experience decreased cognitive capacity when engaging in interracial interactions with Whites (Richeson et al., 2005). In the current study I wanted to examine the differential benefits of diversity for minority and majority group members.

This study compared diverse juries (e.g., mixed composition of White and minority jurors) to those composed of only White jurors and those composed of all-minority mock jurors. To test the hypothesis that diverse juries will result in greater sensitivity to case strength, I manipulated the strength of the evidence in the mock trial. Sensitivity to variations in evidence strength should reflect more careful, systematic processing of argument quality, according to the Elaboration Likelihood Model (Cacioppo et al., 1986). I predicted that both White and Black mock jurors would experience more systematic evidence processing in diverse juries. Moreover, I predicted that jurors in diverse juries would have greater sensitivity to case strength, as well as higher quality contributions, in terms of reasoning, to the jury deliberations.

Chapter 8: Study 1 Methods

Participants

Four hundred thirty-three jury-eligible Black and White community members were recruited from Craigslist.org. Participants were paid \$25 for participation, and if they were on time for the experimental session they were entered into a raffle to win an additional \$25. Participants completed a screening questionnaire in the form of a Qualtrics survey to ensure that they were jury-eligible (i.e., 18 years of age or older and a U.S. citizen). Collecting participants' demographic information in advance also allowed me to assign them to the appropriate conditions based on their racial group membership. Overall participants were about evenly split in terms of gender (51% men) and race (51% Black and 49% White). They averaged 39.59 years of age ($SD = 14.83$).

Design, Materials, and Manipulations

This study featured a 3 (Jury diversity: non-diverse White vs. non-diverse Black vs. diverse) x 2 (Strength of evidence: strong vs. weak) between subjects factorial design. Participants were randomly assigned to conditions for which they are eligible but could not be randomly assigned to all six conditions based on race (e.g., a White participant could not be assigned to a non-diverse Black jury condition).

Voir dire. Prior to viewing the trial video, participants completed a short voir dire questionnaire. Participants provided their gender, age, citizenship, marital status, level of education, occupation, voter registration, political views, ethnicity, and jury duty history.

Trial stimulus. Mock jurors viewed a mock trial simulation ranging from 28-31 minutes (28 in the strong case, 31 in the weak case) in which the defendant was on trial for first-degree murder. This trial was based on the fictional case *Commonwealth of Massachusetts v. Johnson*

case featured in Hastie, Penrod and Pennington (1978). The fictional case in this book was originally derived from *Commonwealth of Massachusetts v. Jones* (1975).¹ I developed a script based on a condensed version of the facts in the case, reducing the number of witnesses so that there were 3 witnesses for each side. The witnesses for the prosecution were: a detective, a medical examiner, and a witness (the bartender the night the crime took place). The witnesses for the defense were: the defendant, the defendant's friend, and a witness (a waitress at the bar where the fight took place).

The filming for the mock trial took place in a mock courtroom using volunteer actors, with both attorneys played by law students, and the judge played by a psychology faculty member. All versions of the trial video contained pattern judge's instructions, opening and closing statements from the prosecuting and defense attorneys, and direct and cross-examination of the detective, medical examiner, bartender, waitress, defendant's friend, and the defendant. The strength of evidence was manipulated by adding motive for the murder in the strong case, a more confident judgment from the medical examiner, and clearer views for the prosecutions' eyewitness. All other case facts were held constant between the two evidence strength conditions.

I conducted a pilot study to ensure the strong and weak evidence manipulation was successful. A total of 38 community members recruited from craigslist.org from the New York City area completed the study. All participants completed a demographic screening questionnaire prior to coming in for the study and I recruited Black and White participants for the pilot study to match the community member sample I would be using in Study 1. I extensively pilot tested to ensure that the appropriate percentage of guilty verdicts in the strong vs. weak case. The final

¹Learned through personal communication with the second author.

versions of the trial resulted in different conviction rates, 57% in the strong case, and 10% in the weak case. A Chi-square analysis revealed a significant difference in verdicts between the strong and weak conditions, $\chi^2(1, N = 38) = 10.97, p < .01$.

Manipulations

Jury diversity. Diverse juries contained at least 2 mock jurors of each race. Non-diverse juries consisted of only Black (in the Black Non-diverse conditions) or only White (in the White Non-diverse conditions) mock jurors.

Strength of evidence. The evidence strength manipulation is explained above in the trial stimulus section.

Dependent Variables: Pre-Deliberation

Dichotomous verdicts. Participants provided individual dichotomous verdicts (guilty or not guilty) after watching the trial videos but before the deliberations.

Continuous verdict measures. Participants reported their confidence in their individual verdicts and the likelihood of the defendant's guilt from 0-100%. Participants also rated: the strength of the prosecution's case against the defendant (1= very weak, 7= very strong), the strength of the evidence indicating the defendant's guilt (1= very weak, 7= very strong), and the strength of the evidence supporting the defendant's innocence (1= very weak, 7= very strong).

Witness ratings. Participants rated all witnesses on 6 different dimensions on 7-point bipolar adjective pairs. These items included: trustworthy-untrustworthy, honest-dishonest, attractive-unattractive, not believable-believable (reverse-coded), convincing-unconvincing, certain-uncertain. The witnesses that participants rated were: detective, medical examiner, bartender, friend of the defendant, waitress, and the defendant with higher values corresponding with more positive ratings.

Anticipated social interaction. Participants provided their ratings of their expectations about the deliberations. To measure this, participants answered questionnaires with questions such as: “I will feel comfortable during this interaction” on a 9-point scale (1= strongly disagree, 9= strongly agree). Other items were: self-conscious (reverse coded), not comfortable (reverse coded), be myself, feel accepted, feel comfortable sharing my views, feel comfortable disagreeing, not feel comfortable arguing (reverse coded).

Dependent Variables: Post-Deliberation

Jury verdicts. Each jury was instructed to come to a unanimous verdict and complete a form with a dichotomous (guilty/not guilty) option. If the jury did not reach a unanimous verdict after 45 minutes, the experimenter halted the deliberations and asked jurors to take a vote. These juries were considered hung juries.

Post-deliberation verdicts and measures. After deliberating with their jury groups, participants again provided individual responses to reflect their post-deliberation verdicts and recorded continuous verdict measures. These items were the same as the pre-deliberation dichotomous verdicts and continuous guilt measures.

Social interaction ratings. Participants provided ratings of how comfortable they felt during the deliberations after the juries had finished deliberating. These items were the same as the pre-deliberations social interaction items, but they were phrased to reflect participants’ impressions of how comfortable they were during the interaction (i.e., “I felt comfortable disagreeing with other jurors”; 1= strongly disagree, 9= strongly agree).

Trust in the system. For the trust in the system variables I used items adapted from Lecci and Myers (2008) pre-trial juror questionnaire (PJAQ). Specifically, I used items that Lecci and Myers (2008) classified as “confidence in the system” variables. Participants

completed items such as “Out of every 100 people brought to trial, at least 75 are guilty of the crime with which they are charged” on a 7-point scale (1= strongly disagree, 7= strongly agree). Other items included: “If a suspect runs from police, then he probably committed the crime”, “generally the police make an arrest only when they are sure about who committed the crime”, “a prior record of conviction is the best indicator of a person’s guilt in the present case”, “when it is the suspect’s word against the police officer’s, I believe the police”, “if a witness refuses to take a lie detector test, it is because he/she is hiding something.” Higher scores corresponded with more trust in the system.

Memory test. After viewing the trial and deliberating, participants completed a memory test on the trial and trial evidence. The memory test questions were in multiple choice form, and tested the participants’ memories for: the race of the defendant, the murder weapon, the name of the bar, the topic of the medical examiner’s testimony, and the main point of the defense’s case.

Quality of deliberation measures. I videotaped deliberations, and research assistants transcribed every deliberation. To measure quality of deliberations, I combined coding schemes developed in previous research to measure deliberation quality (Sommers, 2006) and juror reasoning measures (Kuhn, Weinstock, & Flaton, 1994).

First, with the help of several research assistants, I divided the deliberations into units, with units defined as one unique argument (Kuhn et al., 1994). Each unit consisted of an assertion, and any accompanying justification. Research assistants and I then coded each unit. Most research assistants were blind to the hypotheses (although some knew that the study was about diversity), and every coder was blind to the condition of both the juries and the jurors (jurors were represented by their participant numbers). Coding was done in transcript form, so

coders were truly blind to diversity conditions. Several research assistants coded the data and I calculated inter-rater reliability scores using kappa.

First, each unit was coded for quality. To be coded as sufficient, units had to meet 2 criteria: first the evidence was either drawn from the case information or constituted a reasonable inference from the evidence in the case, and second the evidence either increased or decreased the probability that the criteria associated with a specific verdict were met. As such, sufficient units required both an opinion, and also some evidence or reasonable inference to support (or discount) that opinion.

Sufficient units were then evaluated for unit type, indicating the purpose of the juror's utterance. The four unit types were: justifying a chosen verdict, discounting an alternative verdict, discounting a chosen verdict, or supporting an alternative verdict. The chosen verdict was the opinion chosen by the juror during the initial vote process during deliberations. If jurors did not express their verdict in the beginning of deliberations, I used their pre-deliberation verdict choices as their chosen verdicts. The chosen verdict did not change throughout deliberations, so if jurors changed their vote, they would begin either to support the alternative verdict or to discount the chosen verdict.

After characterizing the purpose of each unit, we then chose an argument type. For arguments that justified an opinion, they could be coded as 1 of 4 types: factual, narrative, importing, and credibility. Factual arguments referred to evidence drawn from the case information without elaboration to support an opinion. Narrative arguments used evidence to construct a narrative to support an opinion. Importing arguments used real world knowledge to support an opinion. Credibility arguments concluded that evidence was more accurate because of the source of the information. Discounting arguments could be characterized as 1 of 4 different

types: factual, inconsistent-importing, discounting-importing, and credibility. For discounting arguments, factual arguments involved referring to missing pieces of information to make the case inconsistent with the verdict (corresponding to “missing evidence” category from Sommers (2006)). Inconsistent-importing arguments used common sense to argue that evidence was inconsistent with the verdict. Discounting-importing arguments undermined specific pieces of evidence in order to discount the opinion. Discounting-credibility arguments concluded evidence to be less accurate because of the source of the information.

The Sommers (2006) criteria that we used for our coding scheme were: number of units containing case facts, mentions of “missing evidence,” and the length of deliberations. The number of case facts and the mentions of missing evidence were derived from the coding units, with those coded as “Supporting-factual” and “Supporting-narrative” reserved for case facts, and those in the “Discounting-Factual” conditions were reserved for missing evidence.

Procedure

Potential participants responded to an advertisement on Craigslist, which provided a survey link for a Qualtrics survey. The survey featured demographic questions, and allowed researchers to identify eligible participants. Emails were sent to interested jury-eligible Black and White community members, offering them possible times to complete the study. Participants were recruited in groups of 9-15 participants, allowing for each session to contain either 1 or 2 juries (allowing for the failure of some participants to show for their session). Jury groups were made up of 4 to 7 jurors, and jurors were dismissed (if necessary) so that the jury groups were never larger than 7 or smaller than 4. The majority (73%) of juries contained 6 or more mock jurors. Dismissed jurors completed a brief “post-deliberation” questionnaire and were paid and

debriefed while the rest of the session began deliberations. Dismissal of jurors was done randomly (within the confines of maintaining at least 2 jurors of each race for diverse juries).

Participants then entered the mock courtroom in groups of 4-15, with diversity condition assigned by session to avoid suspicion on the part of participants. Participants completed consent forms, which included a request for permission to videotape them during the deliberation phase of the study. Participants were informed verbally that their videos would not be made public. Participants then filled out demographic questionnaires with questions similar to those asked during voir dire. The mock jurors then watched a trial video and completed their post-trial verdict questionnaires. After participants completed the questionnaires, they were instructed not to discuss the case with other participants and offered a short break. The researchers then divided juries into two different jury groups (if 10 or more participants were present). For each of the juries the experimenter reminded participants of the videotaping that would occur during the deliberations and instructed them that they should choose a foreperson in charge of fetching the researcher once the jury reached a unanimous verdict. Researchers then started the video recording and left the room, waiting outside for the duration of the deliberations. Participants were not informed of any timeline or deadline, although researchers stopped the deliberations if they continued for longer than 45 minutes and had participants count their votes. These juries were considered hung juries.

After deliberations, the mock jurors completed the post-deliberation questionnaires. Participants were then paid, debriefed, and thanked for their participation.

Data Analytic Plan

Pre-deliberation analyses. All analyses of dependent variables analyzed prior to deliberation were analyzed using single level analyses. I used binary logistic regression to

analyze jurors' pre-deliberation verdicts. I used Analysis of Variance (ANOVA) to evaluate jurors' pre-deliberation continuous guilt measures, the ratings of witnesses, and anticipated social interaction. All pre-deliberation analyses were performed using SPSS Statistical Software.

Post-deliberation analyses. All analyses of post-deliberation individual dependent measures were no longer independent, violating assumptions of single-level tests. As a result, all post-deliberation individual measures were analyzed using multilevel techniques. The dichotomous juror verdicts were analyzed using a linear mixed effects design in R. The continuous measures were analyzed using nested ANOVAs in SPSS.

Jury level analyses. Because each jury is an independent data point, the analyses for juries did not require multilevel modeling. For jury verdicts we used binary logistic regression and examine classification tables to determine how well diversity of juries and race of the defendant predicted sensitivity to strength of evidence in mock juries. Main effects and interactions were observed. Significant interactions were followed-up with pairwise tests of the simple main effects.

Chapter 9: Study 1 Results

Descriptive Statistics

In total, 433 participants deliberated in 68 juries of 4 to 7 individuals. Of the original 433 participants, 20 were removed prior to deliberations, leaving 413 participants. I also removed 4 juries from analysis, as they did not meet the requirements for the diverse condition by having at least two jurors of each race. These 4 juries were comprised of 25 participants, so our final dataset contained 388 participants, who deliberated in 64 juries.

Pre- Deliberation Analyses

Verdicts. Prior to deliberations, 26% of participants provided guilty verdicts. I used a standard binary logistic regression to determine whether there were pre-deliberation effects of case strength or diversity on jurors' verdicts before deliberations. In the first step I entered 3 main effects: case strength and the two dummy-coded diversity conditions with diverse as the reference group. In the second step, I entered two interactions: case strength by each dummy-coded diversity condition. In the first step with all the main effects entered, the omnibus test was significant, $\chi^2(3, N = 393) = 39.52, p < .01$, Nagelkerke $R^2 = .14$. At the first step there was a significant main effect of case strength on pre-deliberation verdicts, Wald $\chi^2(1, N = 393) = 34.14, p = .00, \text{Exp}(B) = 4.71, 95\% \text{ CI } [2.80, 7.93]$. The ratio of guilty verdicts to not guilty verdicts was 4.71 times higher in the strong case than it was in the weak case. In the strong case condition 39% of participants voted guilty, with only 12% of participants voting guilty in the weak case. In the second step adding interaction terms did not significantly improve the fit of the model, and the interactions were not significant, Wald $\chi^2(1, N = 393) < 1.90, ps > .05, \text{Exp}(B) < 2.19$.

Continuous guilt measures. I then ran ANOVAs examining the continuous measures for the pre-deliberation ratings of the guilt of the defendant. First, I calculated a verdict confidence measure by combining verdict and confidence in verdict. I multiplied not guilty verdicts by -1 to obtain a continuous scale from -100 (extremely confident in not guilty verdict) to +100 (extremely confident in guilty verdict). Consistent with results from the dichotomous judgments, a 3(Diversity) x 2(Case strength) ANOVA on the verdict confidence measure revealed a main effect for case strength, $F(1, 385) = 40.43, p < .01, \eta_p^2 = .10$. Participants in the strong case provided higher verdict confidence scores ($M = -15.52$) than did participants in the weak case ($M = -61.74$). No other main effects or interactions were significant, $ps > .05$.

Witness ratings. After reverse coding the necessary items, I averaged the scales for all witness ratings so that higher scores corresponded with more positive ratings. I wanted to examine the effect of participant race on ratings of the witnesses, so I combined both of the non-diverse conditions to turn diversity into a 2-level variable (diverse vs. non-diverse). I then ran a 2 (Race of participant: Black vs. White) x 2 (Diversity: diverse vs. non-diverse) x 2 (Case strength: strong vs. weak) ANOVA examining the ratings of the detective ($\alpha = .83$). The ANOVA revealed a main effect for race of participant, $F(1, 382) = 11.30, p < .01, \eta_p^2 = .03$. White participants rated the detective ($M = 4.77$) more positively than did Black participants ($M = 4.32$). No other main effects or interactions were significant ($ps > .1$).

A 2 x 2 x 2 ANOVA examining the impact of race, diversity, and case strength on ratings of the defendant ($\alpha = .85$) revealed a significant main effect for case strength, $F(1, 379) = 61.93, p < .01, \eta_p^2 = .14$. Participants in the weak case rated the defendant more positively ($M = 4.90$) than did participants in the strong case ($M = 3.92$). No other main effects or interactions were significant ($ps > .1$).

Anticipated social interaction. I averaged all 8 items on the anticipated social interaction scale ($\alpha = .85$) and recoded all questions so that higher scores reflected feeling more positively about the forthcoming interaction. For the anticipated social interaction analyses, I wanted to examine race as an independent variable, so I ran a 2 (race) x 2 (diversity) x 2 (case strength) ANOVA. The ANOVA revealed no main effects of case strength or diversity condition on participants' views about their upcoming interaction ($ps > .05$). However, there was a significant main effect for race such that Black participants anticipated feeling more comfortable during deliberations ($M = 5.67$) than did White participants ($M = 5.42$), regardless of diversity condition, $F(1, 382) = 4.45, p < .05, \eta_p^2 = .01$.

Post-Deliberation Analyses

Individual juror verdicts. After deliberations, 9% of participants rendered guilty verdicts. After deliberating in a jury group participants' individual verdicts were no longer independent. I used the glmer function in the package "lme4" in R Statistical software to perform a multilevel logistic regression examining jurors' post-deliberation verdicts. First, I conducted an analysis on the null model to determine how much variance was due to group membership. With the null model I added Jury as the random intercept for the model, and found an intra-class correlation (ICC) of .52, demonstrating that 52% of the variance in the model could be explained by group membership. The mean reliability within jury groups was 87%.

I then ran the linear mixed model with Jury as a random intercept. I entered strength of case as a fixed effect and also the two dummy-coded diversity variables with the diverse condition as the reference group. I also included the interactions between case strength and each of the dummy coded diversity variables. None of the main effects or interactions of the main independent variables had significant effects on post-deliberation verdicts ($ps > .73$).

Continuous guilt measures. I ran a nested ANOVA to examine the effects of case strength and diversity on the continuous guilt measures. For this analysis, I included the jury group as a random clustering effect and case strength and diversity as fixed effects.

First, I examined the verdict confidence measure. I calculated this in the same manner as with the pre-deliberation guilt ratings by using confidence in verdict ratings and multiplying these values by -1 for not guilty verdicts. This resulted in a scale from -100 (highly confident in a not guilty verdict) to +100 (highly confident in a guilty verdict). Then, I ran a 3(Diversity) x 2(Case strength) nested ANOVA to account for the nesting of participants in juries. The results of this analysis revealed a fixed main effect of case strength, $F(1, 60.34) = 4.49$, $p < .05$, $\eta_p^2 = .07$. Participants in the strong case condition were less confident in their not guilty verdicts ($M = -61.40$) than were participants in the weak case condition ($M = -84.07$). No other main effects or interactions were significant, $ps > .05$.

Social Interaction Ratings

To calculate ratings of the quality of the social interaction that occurred during the deliberations, I averaged all post deliberation interaction ratings and recoded several items so that higher numbers corresponded with more positive social interactions ($\alpha = .85$). Prior to deliberations, I observed a main effect of race such that White participants reported expecting to feel less comfortable than Black participants. To determine whether these ratings were impacted by deliberations, I conducted a 2 (Diversity) x 2 (Race of participant) x 2 (Case strength) nested ANOVA, with errors nested within jury groups. Ratings of the quality of the social interactions following deliberations revealed no significant main effects or interactions, $ps > .05$.

Trust in system ratings

I averaged all the items on the trust in the system scale ($\alpha = .78$) and performed a nested ANOVA on the scale. I anticipated racial differences on the scale so I conducted a 2 (Race) x 2 (Diversity) x 2 (Case strength) nested ANOVA. The analysis revealed a significant two-way interaction between case strength and race of participants, $F(1, 108.91) = 7.57, p < .01, \eta_p^2 = .07$. In the weak case, there were no racial differences in ratings of trust in the system ($M_s = 2.79$ vs. 2.97), $d = -.12, p = .51, 95\% \text{ CI}[-.48, .24]$. However, in the strong case, White participants rated their trust in the system as higher ($M = 3.04$) than did Black participants ($M = 2.60$), $d = .57, p < .01, 95\% \text{ CI} [.23, .92]$.

Quality of Deliberations Measures

Deliberations coding. The coding of deliberations revealed that there were 6,320 units (made up of an argument and any accompanying justification) in the 64 jury deliberations. The mean number of units per juror was 16.33 ($SD = 17.41$). Each unit was coded for purpose and quality. We had an overall 90% agreement rate between coders, and disagreements were resolved through discussion. Descriptive statistics for the coding criteria for individual jurors are presented in Table 2. Only analyses with significant diversity effects are presented fully in text.

First, I examined the number of units per juror. We conducted a 2 (race) x 2 (diversity) x 2 (case strength) nested ANOVA with jury group as the random clustering effect to determine any main effects or interactions on the number of units per juror. The results of the nested ANOVA revealed a significant main effect of case strength, $F(1, 82.33) = 5.87, p = .02, \eta_p^2 = .07$. Participants in the strong case provided more units ($M = 20.85$) than did participants in the weak case ($M = 13.55$). No other main effects or interactions were significant ($ps > .05$).

The first category used in the coding process was quality. To be sufficient, units had to be drawn from the evidence or constitute a reasonable inference to be drawn from that evidence. I

used the number of sufficient units per juror as my outcome measure, and conducted a 2 x 2 x 2 nested ANOVA to account for the error due to jury groups. There were no main effects for diversity or race but there was a main effect for case strength, $F(1, 87.64) = 6.48, p = .02, \eta_p^2 = .07$. Participants who viewed the strong case provided more sufficient units ($M = 7.82$) than did participants in the weak case ($M = 5.31$). The main effect was qualified by a significant diversity by strength of case interaction, $F(1, 87.19) = 5.23, p = .02, \eta_p^2 = .06$. Follow-up tests revealed that in the weak case conditions there was no significant effect of case strength ($M_s = 5.47$ vs. 4.68), $d = -.59, p = .53, 95\% \text{ CI} [-2.45, 1.25]$. However, for participants who viewed the strong case, those who deliberated in diverse juries delivered more sufficient units ($M = 10.00$) than did participants who deliberated in non-diverse juries ($M = 6.05$), $d = 3.86, p < .01, 95\% \text{ CI} [2.08, 5.64]$.

Units were then evaluated for the type of argument they represented. The four categories were: supporting the chosen, discounting the alternative, discounting the chosen, and supporting the alternative. I only reported categories that revealed significant diversity effects in a nested ANOVA. The number of discounting alternative arguments, displayed both a significant main effect of case strength ($F(1, 92.59) = 5.90, p = .02, \eta_p^2 = .06$) and a significant interaction between diversity and case strength, $F(1, 92.68) = 4.53, p = .04, \eta_p^2 = .05$. Follow-up tests revealed no difference between jurors deliberating diverse compared to non-diverse juries ($M_s = 1.69$ vs. 2.08), $d = -.31, p = .48, 95\% \text{ CI} [-1.17, .55]$ in the weak case, but for the strong case jurors who deliberated in diverse juries provided significantly more discounting arguments ($M = 3.81$) than those deliberating in non-diverse juries ($M = 2.71$), $d = 1.42, p < .01, 95\% \text{ CI} [.59, 2.25]$.

Supporting arguments (either for the chosen opinion or for an alternative opinion), were coded as: factual, narrative, importing, or credibility. Discounting arguments (for either the alternative or chosen opinion) were coded as: factual, importing-inconsistent, importing-discounting, and credibility. The supporting-facts category was combined with the supporting-narrative category and represented the number of units in which case facts were mentioned. In the supporting-facts category, participants stated facts, and in the supporting-narrative category, participants provided facts and elaborated on those facts to create a story. The discounting-facts category was reserved to represent number of times jurors asked for “missing evidence.” I then combined the remaining variables that reflected similar types of thinking. I combined supporting-importing and supporting-credibility into a “Supporting Reasoning” category. I left “Discounting-inconsistent” as its own category; representing jurors’ ability to take real world knowledge to judge evidence as less consistent with verdict criteria. The last category I created was a “Discounting-Evidence” which combined discounting importing and discounting credibility, both of which discounted evidence because of its unreliability. Only categories with significant diversity effects are presented.

For the “Discounting-Evidence” category, there was a significant interaction between diversity and case strength on the number of discounting arguments used to dispute the credibility of the evidence presented, $F(1, 93.57) = 5.09, p = .03, \eta_p^2 = .05$. Follow-up tests revealed that in the weak case there was no significant effect of diversity on disputing credibility ($M_s = .71$ vs. $.97; d = -.17, p = .49, 95\% \text{ CI } [-.65, .31]$), but for those who viewed the strong case, jurors in diverse juries provided more discounting credibility arguments ($M = 1.70$) than jurors deliberating in non-diverse juries ($M = .80, d = .84, p < .01, 95\% \text{ CI } [.37, 1.30]$).

I examined the 2 x 2 x 2 ANOVA on the number of units mentioning case facts, which is similar to the Sommers (2006) coding criteria of number of case facts mentioned. There was a significant diversity by case strength interaction, $F(1, 88.12) = 4.21$, $p = .04$, $\eta_p^2 = .05$. Pairwise comparisons revealed that in the weak case, there was no effect for diversity ($M_s = 1.15$ vs. 1.71), $d = -.45$, $p = .20$, 95% CI $[-1.13, .24]$. In the strong case, however, jurors deliberating in diverse juries provided more units containing case facts ($M = 2.86$) than did jurors deliberating in non-diverse juries ($M = 1.71$), $d = 1.11$, $p < .01$, 95% CI $[.46, 1.77]$.

The other Sommers (2006) criteria I coded for was jurors' mentions of "missing" evidence or requests for more evidence in the case. There were no significant main effects or interactions of case strength, diversity, or race of participant on participants' ratings of pieces of requested evidence, $ps > .05$.

Jury Level Analyses

Verdicts. Out of the total 64 juries, 56 (88%) voted to acquit the defendant, whereas only 5 juries (8%) voted to convict. Three juries (5%) remained hung after the 45 minutes allotted for deliberation. We counted majority votes for each hung jury, and after that 5 (8%) voted to convict the defendant, whereas 59 (92%) voted to acquit the defendant.

I performed a binary logistic regression on the jury verdicts once the vote counts had been calculated. In the first step I added the strength of case predictor, as well as the two dummy-coded diversity conditions. In the second step I added the interactions between case strength and each of the dummy-coded diversity conditions. None of the main effects or interactions were significant at the jury-level.

Length of deliberations. I also coded for the length of deliberations, consistent with the procedure in Sommers (2006). I performed a 2 (Diverse vs. Non-diverse) x 2 (Case Strength) on

the length of deliberations. Main effects and interactions on deliberation length were not significant ($ps > .05$).

Chapter 10: Study 1 Discussion

The main hypothesis in Study 1 was that racially diverse juries would have higher quality deliberations than either type of non-diverse jury, consistent with the findings of Sommers (2006). My first hypothesis was that jurors deliberating in diverse juries would be more sensitive to variations in evidence strength than would jurors deliberating in non-diverse juries. I did not find that jurors were more sensitive to case strength when deliberating in diverse juries, meaning this hypothesis was not supported.

There are several potential explanations for the lack of sensitivity effects. First, the post-deliberation conviction rate was very low, and this may have created a floor effect, with insufficient variation to detect differences between groups. Another possible explanation for the lack of sensitivity is that diversity does not influence verdicts. This is consistent with findings from previous research finding that diversity influenced the quality of jurors' contribution to discussion, but not their verdicts (Sommers, 2006). This is not consistent, however, with other research finding that mock jurors deliberating in diverse juries deliver less racially biased verdicts and sentences than juries mainly composed of White men (Lynch & Haney, 2011). A third possibility for explaining the lack of sensitivity effects is that they may have been overridden by the shift towards acquittal following deliberations. Prior to deliberations, participants were more likely to convict in the strong case than in the weak case. Although this finding indicates that my manipulation of evidence strength was successful, it is notable that the effect disappeared following deliberations. There was a major shift towards acquittal following deliberations, which is possibly indicative of the leniency effect. Prior research has noted that jurors tend to deliver more lenient judgments after deliberating in a jury (Kerr & MacCoun, 2012). Additionally, Kalven and Zeisel (1966) found evidence that jurors are overwhelming

influenced by the majority vote of the jury group. With an overall low conviction rate, this majority would likely shift most juries towards acquittal. Whether a leniency effect or a result of majority influence, either could have overwhelmed my ability to detect diversity effects.

My second main hypothesis was that jurors deliberating in diverse juries would make higher quality contributions to discussion, consistent with the Sommers (2006) findings. This hypothesis was partially supported, in that while diversity did not improve the quality of deliberations overall, diversity did improve the quality of deliberations in the strong case. For those who viewed the strong case jurors deliberating in diverse juries made more high quality contributions to discussion than did jurors deliberating in non-diverse juries. When viewing the strong case jurors deliberating in diverse juries provided more sufficient units, more counterarguments, more arguments used to discount the reliability of evidence, and more units containing case facts when deliberating in diverse juries than they did when deliberating in non-diverse juries. This finding is consistent with previous research finding that diverse juries have higher quality discussion content than non-diverse juries (Sommers, 2006). These effects were likely restricted to the strong case because the strong case was more ambiguous in terms of guilt, and because we can detect differences more easily when stimuli are ambiguous.

Also not all my quality of deliberations measures revealed significant effects for diversity in the strong case. Contrary to the Sommers (2006) criteria, there were no significant effects for the mentions of missing evidence, and at the jury level there were no significant effects on the length of deliberations. This could be due to the lower sample size of juries compared to jurors, making it more difficult to detect differences. Also, the fact that diversity effects were only found for the strong case, means my sample was further restricted because of how easily juries came to a decision in the weak case.

Participant race did not affect the quality of deliberation contributions, suggesting that diversity similarly affected participants of both races. This finding is somewhat different than that of Sommers (2006) and also suggests that diversity's impact goes beyond the watchdog effect (Fleming et al., 2005). Minorities benefitting from diversity does not rule out the possibility of the watchdog effect, but it does suggest that diversity might operate through other mechanisms as well. The results from the current study suggest that diversity might improve deliberations because of the heterogeneity of experiences and perspectives (Nemeth, 1986), rather than from a reduction of prejudice. It is also possible that diversity's benefits operate differently for minority and majority group members. We cannot determine which explanation is correct from the current research, but this is an exciting avenue for future study.

Black participants reported anticipating they would feel more comfortable during the interaction than White participants did. Although interracial interactions have the potential to make Whites concerned about appearing prejudiced (Mendes et al., 2002; Norton et al., 2006), Whites were equally concerned about deliberating with White and Black jurors, as evidenced by the effect of race but not diversity on ratings of anticipated social interaction. Maybe a trial stimulus featuring a Black defendant was enough to make White participants anxious about the upcoming interaction, regardless of the race of the other mock jurors. The premise of the watchdog effect is that low-prejudiced White individuals increase their vigilance in order to protect minorities from prejudice (Fleming et al., 2005), so the anxiety about the upcoming interaction may be attributable to this concern. White individuals tend to experience anxiety when political correctness concerns are heightened (Norton et al., 2006), and it is possible that the Black defendant was sufficient to raise these concerns. Most of the research on race salience

in the courtroom, however, suggests that jurors' awareness of the race of a Black defendant is not a sufficient condition for making racial issues salient (Sommers & Ellsworth, 2000; 2001; 2009).

White participants rated the detective more positively than did Black participants. Because the detective was a White woman, negative ratings for Black participants might indicate outgroup bias towards an outgroup member (Mitchell et al., 2005), but since no such difference was found in ratings of the defendant, this seems unlikely. The more negative ratings of the police officer by Black participants were more likely indicative of a lack of trust in police. Previous research suggests that minorities have less favorable views of the police than do White individuals (Weitzer & Tuch, 1999; 2004; 2005). Similarly, I found that in the strong case Black participants expressed higher trust in the system than White participants did. Both lower ratings of the police officer and lower trust in the system ratings are consistent with research finding that minority individuals experience more injustice and unfair outcomes at the hands of the justice system (Barnes & Kingsnorth, 1996; Mustard, 2001) and often have more negative views of institutions such as the courts and the police (Overby, et al., 2005; Weitzer & Tuch, 2005). Minorities also exhibit lower trust in a number of institutions, including the courts, jury system, and political institutions (Overby, Brown, Bruce, Smith, & Winkle, 2005; Rose, Ellison & Diamond, 2008).

Participants rated the defendant more positively when deciding the weak case rather than the strong case, but no other significant effects were observed on the ratings of the defendant. The defendant was Black in all conditions, meaning I did not find evidence of more bias when the defendant was an outgroup member compared to when he was an ingroup member. This is contrary to most of the work on mock juror judgments, in particular the most recent meta-analyses (Devine & Caughlin, 2014; Mitchell et al., 2005). Not all meta-analyses, however, find

this outgroup bias effect (Mazzella & Feingold, 1994). Even the main racial bias meta-analysis supporting an outgroup bias effect finds that the effect is larger for minorities making judgments about White defendants, a condition not present in the current study (Mitchell et al., 2005). Additionally, some mock juror studies have found little evidence of bias towards outgroup defendants (Foley & Pigott, 2002), or conditions under which bias can be eliminated (Sommers & Ellsworth, 2000; 2001).

Limitations

The largest limitation to the present study was the low conviction rate, particularly in the post deliberations individual verdicts. These low rates may represent floor effects, which likely explain why it was difficult to detect results from the dichotomous post deliberation verdicts and continuous guilt measures. The pre-deliberation conviction rates were relatively low, and following deliberations the guilt rate was even lower due to the leniency effect of deliberations (Kerr & MacCoun, 2012). Relatedly, the fact that post-deliberation guilt rates were so uniformly low meant there was less ambiguity and discussion necessary for jurors to come to a unanimous decision during deliberation in the weak case condition. The weak case was so unambiguous that it did not provide enough opportunity for jurors to engage in productive discussion, as juries often entered the deliberations unanimous, or with a clear majority of not guilty verdicts. As a result of uniform acquittals upon the beginning of deliberations, I found that the case strength actually influenced the quality of deliberation content. Participants in the strong case condition made higher quality contributions to discussion, than did participants in the weak case. Future studies should insure that they have sufficient ambiguity in their trial stimulus to correct for this, and should possibly pilot test their stimuli after deliberations to determine whether they have sufficient ambiguity at the jury level, and not simply the juror level.

The other limitations to the present study concern the nature of the study and its conditions. I stopped deliberations after 45 minutes, and this may have contributed to the lack of diversity's impact on the length of deliberations. Sommers (2006) did find an effect for diversity on the length of deliberations, and the researchers in that study allowed deliberations to occur for 60 minutes instead.

Conclusions

The first study did not support my initial hypothesis that diverse juries would distinguish between strong and weak evidence more effectively than non-diverse juries. However, the quality of deliberations measures did provide some support for diversity's benefit, with jurors in diverse juries providing more units with case facts and more sufficient arguments than jurors in non-diverse juries. Additionally, jurors in diverse juries were more likely to deliver counterarguments and discount evidence lacking in reliability than jurors in non-diverse juries. The finding of higher quality contributions from jurors deliberating in diverse juries is consistent with previous research examining jury deliberation content (Sommers, 2006).

We did not find any race effects on contributions to deliberations, and this builds upon the finding from Sommers (2006) that White jurors made higher quality contributions to deliberations. The significant benefits of diversity, combined with the lack of effects of participant race, suggest that diversity benefits both majority and minority group members. I did, however, find racial differences on the participants' ratings of the detective in the case, and in their trust in the system. These findings may be related, and reveal some racial differences in how individuals view justice system. Overall, the results of this study support previous research in finding that diverse juries will have higher quality discussions than non-diverse juries (Sommers, 2006), but suggest that the benefits of diversity may extend beyond White's vigilance

in the presence of minority members (Sommers, Warp & Mahoney, 2008). Additionally, this research suggests that future research should explore whether trust in the legal system might play a role in how the inclusion of racial minorities on juries might improve deliberations.

Chapter 11: Study 2 Introduction

In the second study, I sought to define diversity in a different manner than in the first study. I used a minimal groups paradigm (MGP) to investigate the impact of wealth and power differentials and their potential impact on deliberations in heterogeneous juries. Although there is no work directly examining the impact of power and its relationship with jury performance, previous studies on organizational and group dynamics have documented that unequal distribution of power and resources can impact group performance, causing minorities to experience identity threat (Alderfer & Smith, 1982; Foldy, Rivard, & Buckley, 2009; Steele, Spencer, & Aronson, 2002). Stereotype content research also establishes that Whites are perceived as high in competence based on their position as a high status group (Fiske, Cuddy, Glick, & Xu, 2002). These power differentials between members of different racial and ethnic groups have the potential to enter the jury room and interfere with jury performance. Moreover, separating the specific stereotypes surrounding racial groups in their relation to crime can help us understand how differences in wealth and power can impact jury deliberations, even when both are irrelevant to the task at hand.

Research on jury deliberations suggests that power differentials play a significant role in the proceedings of jury deliberations. Traditionally, jury research has found that members of higher social status groups (i.e., higher incomes and education levels) tend to participate more in the jury deliberations than jurors of lower social status groups (Hastie et al., 1983; York & Cornwell, 2006). Similarly, women tend to contribute less to jury deliberations than men (Hastie et al., 1983). More recent surveys have found some conflicting results about race and gender. Racial differences emerged about self-reported participation in jury deliberations, but in a counter-intuitive manner (Cornwell & Hans, 2011). Black jurors reported the highest level of

participation followed by White and Hispanic jurors. Socioeconomic status (SES), however, continues to be influential in more recent examinations of self-reported participation in jury deliberations, with high SES individuals reporting more participation in juries and being perceived as more influential than low SES individuals (Cornwell & Hans, 2011; York & Cornwell, 2006).

Work examining intergroup power also finds that high power individuals tend to act differently in groups than do low power individuals (for a review, see Brauer & Bourhis, 2006). One study found that high power groups tended to express their opinions more often, and were less susceptible to feeling the pressure of an intergroup situation (Galinsky, Magee, Gruenfeld, & Whitson, 2008). Social dominance orientation (SDO) can also be relevant to understanding the impact of power dynamics in jury deliberations (Ho et al., 2015; Sidanius & Pratto, 1999). SDO refers to the tendency for high status individuals to believe that they deserve their position at the top of the power hierarchy, and to maintain their position. People high in SDO are more likely to engage in prejudice and discriminatory behavior toward outgroups (Amiot & Bourhis, 2005). Jurors high in SDO tend to punish outgroup members more than do jurors low in SDO, who are far more likely to be egalitarian, and therefore not discriminatory (Kemmelmeier, 2005).

Racial diversity and diversity of perceived wealth and power are often confounded in jury studies that focus on racial diversity. This occurs because White individuals tend to be perceived as a high status group (Fiske, Cuddy, Glick, & Xu, 2002), and therefore their position in the jury deliberation tends to be that of higher power. In terms of legal decisions, however, different racial groups tend to have specific relationships with the legal system in terms of both their perceptions of the system (Overby et al., 2005; Weitzer & Tuch, 2005), and in others' perceptions of racial minorities being associated with crime (Correll, Park, Judd, & Wittenbrink,

2002; Espinoza & Willis-Esqueda, 2008; Payne, 2001). Moreover, there are problems associated with grouping all ethnic minorities into one non-White category, when each ethnic minority group has its own unique set of values and attitudes (van Knippenberg & Schippers, 2007). The benefit of using a MGP is that it can test the impact of diversity while eliminating some crime-specific stereotypes surrounding racial/ethnic groups, and can eliminate issues that arise from grouping members of different minority groups together.

Early work on intergroup discrimination had previously demonstrated that ingroup favoritism and hostility towards outgroups could be created when there were two groups in competition with each other (Sherif, 1966). These associations form because people derive high levels of self-esteem from identifying with and belonging to their ingroup, and develop antagonistic feelings towards competing groups. Although this early work revealed intergroup hostility in competitive settings (Sherif, 1966), the minimal groups paradigm (MGP) emphasized that directly competing for resources was not a necessary requirement for discrimination between groups. Despite the use of trivial criteria (e.g., similar preferences for one painting over another) to create groups, researchers repeatedly found that people had a significant preference for their ingroup (Billig & Tajfel, 1973; Tajfel & Billig, 1974). Typically, participants reflected this preference in the decisions they made when allocating their resources, showing ingroup favoritism and outgroup discrimination (Tajfel et al., 1971). This paradigm was instrumental in developing social identity theory, which demonstrated people's natural tendency to make distinctions between ingroup and outgroup members (Tajfel & Turner, 1979).

The MGP was initially created to demonstrate that people could develop strong feelings of association with ingroup members and hostility toward outgroup members, even when the groups were created using trivial criteria (Tajfel, Billig, Bundy, & Flament, 1971). Previous

research has established methods of manipulating wealth and power within a MGP, using monopoly money to signify wealth and a greater stake in the allocation decisions to simulate power (Harvey & Bourhis, 2012). In this study, wealth was manipulated through the differential distribution of monopoly money to participants depending upon their assignment to a wealth condition. The researchers manipulated power by allowing some groups to have more control over how resources were allocated than others. In the end, the wealthy and powerful group has more money and more power to distribute that money. Previous MGP studies manipulating wealth and power have demonstrated that independent of wealth and power manipulations, people tend to show favoritism and discrimination based on their group membership (Harvey & Bourhis, 2012).

The current study used the procedure in Harvey and Bourhis (2012) to manipulate power and wealth to create two different groups: a high wealth/power group and a low wealth/power group. I manipulated diversity by combining the wealth/power groups in the diverse conditions. I predicted that diverse juries (i.e., juries combining both wealth/power groups) will exhibit more sensitivity to evidence strength and higher quality deliberations than non-diverse groups.

Chapter 12: Study 2 Overview

Study 2 explores the same theoretical principles as Study 1, but addresses them through a more basic ingroup/outgroup paradigm. Study 1 elucidates the effects of racial diversity on how juries and individuals process trial evidence, but does not approach some of the fundamental reasons for racial differences in perceptions of legal issues. Minorities tend to have less trust in the legal system and in actors in the legal system, than majority group members do (Overby et al., 2005; Weitzer & Tuch, 2005). Additionally, people tend to have associations between minorities and different types of crime (Correll, Park, Judd, & Wittenbrink, 2002; Espinoza & Willis-Esqueda, 2008; Payne, 2001). Yet, much of the research on jury deliberations focuses on the social status and power of individual jurors (York & Cornwell, 2006). The purpose of this research is to study diversity when it is separated from the crime-specific stereotypes associated with racial and ethnic groups, and instead captures the wealth and power differential between Whites and other minority groups. This study will use a minimal groups paradigm (MGP) to manipulate participants' wealth and power to simulate power dynamics that may operate in diverse juries. Using a manipulation of wealth and power in the context of a MGP that has been successful in creating discriminatory patterns of behavior in past research (Harvey & Bourhis, 2012), participants will participate in a trial simulation and deliberate in either diverse (mixed wealth/power groups) or homogenous (separate wealth/power groups) juries.

I predicted that diverse juries (as defined by a combination of high and low wealth/power participants) would exhibit more sensitivity to case strength than would non-diverse juries, irrespective of whether the members all had high or low wealth/power juries. I also predicted that high wealth/power individuals would provide more high quality contributions to deliberations

than would low wealth/power individuals, and that diverse juries would exhibit higher levels of reasoning and have longer deliberations with more case facts than would non-diverse juries.

Chapter 13: Study 2 Methods

Participants

Three-hundred-sixty-nine jury-eligible undergraduate psychology students at an urban public university in the northeast completed the study for course credit. Participants were pre-screened for jury-eligible status (over 18 years old and a U.S citizen). Participants were 75.6% female and averaged 20.75 years old ($SD = 4.29$). The sample was ethnically diverse (42.8% Hispanic, 16.8% Black, 16.0% White, 9.8% Other).

Design and Materials

Participants were randomly assigned to a 3 (jury diversity: diverse vs. non-diverse high wealth/power vs. non-diverse low wealth/power) x 2 (evidence strength: strong vs. weak) between subjects factorial design.

Voir dire. Prior to viewing the trial video, participants completed a short voir dire questionnaire. Participants provided their gender, age, citizenship, marital status, level of education, occupation, voter registration, political views, ethnicity, and jury duty history.

Trial stimulus. I used the same trial in the second study as I did in the first. However, the defendant was Hispanic instead of Black in the second study.

Manipulations

Wealth/Power. Wealth and power were confounded in the present study so that I had a high wealth/power condition and a low wealth/power condition. In the high wealth/power condition, I told participants they would receive \$1,000 in monopoly money and their decisions would account for 70% of the overall allocation of the total sum of monopoly money in the study. In the low wealth/power condition participants received \$300 in monopoly money and their decisions accounted for 30% of the overall allocation of the total sum of the monopoly

money. Participants were told that the totals for the study would be calculated at the end of the study. I informed participants that this was the only purpose for their allocation decisions about the monopoly money.

Jury Diversity. Diverse juries contained at least two members of the high-wealth/power group and the low-wealth/power groups. Non-diverse juries consisted of either all low-wealth/power group members or all high-wealth/power group members.

Evidence Strength. The evidence strength manipulation was identical to the manipulation in Study 1.

Procedure

Participants were recruited through the Sona online recruitment system, and entered the study in groups of 4-16. The procedure for the first part of the study was based on the minimal groups paradigm manipulation developed by Harvey and Bourhis (2012). When participants arrived, they were assigned to either the high-wealth/power group or the low wealth/power group. Participants were in either “Group A” or “Group B,” assigned by odd or even participant numbers to emphasize the randomness of the group assignments. I used randomization to determine whether Group A or Group B was high power. Based on the method of simulating wealth by Vohs, Mean & Goode (2006), I used monopoly money to signify wealth. I awarded \$1,000 to each member of the high wealth/power group and \$300 to each member of the low wealth/power group, consistent with earlier research (Harvey & Bourhis, 2012). After distributing the Monopoly money, the experimenter informed participants that the high power group’s decisions about how to distribute the money would count for 70% of the final distribution of the monopoly money, with the low power group’s decisions counting only 30% in the final allocation of money.

Once the groups were assigned the experimenter explained the distribution task and handed out questionnaires for all participants to fill out before completing the task. Participants first completed three Tajfel matrices originally developed for use in minimal groups paradigms (Tajfel, et. al, 1971). The matrices were labeled “Ingroup” for ingroup member on the top row of the matrices, and “Outgroup” for outgroup members on the bottom row of the matrices. For each matrix, one side of the scale represented rewarding ingroup members and punishing outgroup members, whereas the other side of the scale represented rewarding outgroup members and punishing ingroup members. The middle of the scale allowed for more equal distribution of funds among ingroup and outgroup members. Participants received 3 envelopes, one labelled “Group A Member,” one labelled “Group B Member,” and one labelled “Future Studies.” Consistent with the procedure in Harvey & Bourhis (2012) I asked participants to report how much of the money from the whole experiment they would like to give back to the researchers to contribute to future studies. The purpose of asking participants to donate to future studies was to emphasize that resources were finite, since the participants in the low wealth/power groups did not have much money to donate. Participants placed the appropriate amount of money in each envelope (as dictated by their decisions on the matrices and their donation to future studies). Following this first task, participants filled out additional questionnaires with questions about their group membership and their feelings of wealth and power.

After completing the resource allocation tasks, the participants completed the study in the same manner as in Study 1. Participants completed the voir dire questionnaire, watched the trial video, completed post-trial questionnaires, and then were divided into deliberation groups of 4-7 individuals. In Study 2 70% of juries contained 6 or more mock jurors. I instructed the juries to

arrive at a unanimous decision and then retrieve a researcher at the conclusion of the study. Participants then completed post-deliberation questionnaires and were debriefed.

Dependent Variables

This study used all dependent variables from Study 1, with additional dependent variables for the MGP portion of the study.

Reports of wealth/power. Following the allocation task, participants rated, on a scale of 1-7, how wealthy they felt and how powerful they felt. These questions acted as manipulation checks for the power and wealth manipulations.

Social identity questions. Four items were combined for a composite score representing how happy and satisfied participants felt in their group on a 7-point Likert scale. The questions asked were how much they liked being members of their own group, how happy, at ease, satisfied and vulnerable (reverse scored) they felt as members of their group (Harvey & Bourhis, 2012).

Data Analysis

Data from the mock trial portion of the study will be analyzed in the same manner as the previous two studies. The manipulation checks will be analyzed using one-way ANOVAs to test the differences between the two wealth/power groups.

Chapter 14: Study 2 Results

Descriptive Statistics

Overall, 369 participants deliberated in 60 juries of 4 to 7 individuals. Of the 369 participants, 13 were removed prior to deliberations, leaving 356 participants following deliberations.

Manipulation Checks

Participants' responses to the manipulation check questions reflected the fact that my wealth/power manipulation had the intended effect. I conducted a one-way ANOVA on participants' ratings of how wealthy and powerful they felt. Participants in the high wealth/power condition reported feeling wealthier ($M = 4.60$) than did those in the low wealth/power condition ($M = 3.21$), $F(1, 364) = 64.34$, $p < .00$. Participants in the high wealth/power condition also reported feeling more powerful ($M = 4.52$) than did those in the low wealth/power condition ($M = 3.34$), $F(1, 364) = 51.02$, $p < .00$. I also combined all items on the satisfaction with group membership scale ($\alpha = .78$) based on the procedure in Harvey & Bourhis (2012). Participants in the high wealth/power conditions expressed more satisfaction as members of their group ($M = 4.89$) than did participants in the low wealth/power conditions ($M = 4.04$), $F(1, 364) = 45.20$, $p < .00$.

Pre- Deliberation Analyses

Before deliberations, participants were not aware of any differences in the diversity of their groups. Therefore for all pre-deliberation analyses, I used a 2 (Wealth/power: high vs. low) x 2 (Case strength: strong vs. weak) between subjects factorial design.

Verdicts. Before deliberations, 47% of jurors voted to convict the defendant. I used a standard binary logistic regression to determine whether there were pre-deliberation effects of

case strength or wealth/power group membership on jurors' verdicts before deliberations. In the first step I entered the two main effects (power and case strength), and at the second step entered the interaction between the two variables. The omnibus test was significant, $\chi^2(2, N = 365) = 77.83, p < .01$, Nagelkerke $R^2 = .26$ at the first step. There was a significant main effect of case strength on pre-deliberation verdicts, Wald $\chi^2(1, N = 365) = 67.02, p < .00, Exp(B) = 7.46$, 95% CI [4.61, 12.07]. The ratio of guilty verdicts to not guilty verdicts in the strong case condition was 7.46 times higher than the ratio of guilty to not guilty verdicts in the weak case. In the strong case condition 68% of participants voted guilty, with only 22% of participants voting guilty in the weak case. No other main effects or interactions had significant effects on pre-deliberation verdicts.

Continuous guilt measures. After analyzing the dichotomous verdicts, I examined the continuous guilt measures. I calculated a verdict confidence measure to create a continuous scale from -100 (extremely confident in not guilty verdict) to +100 (extremely confident in guilty verdict), as in Study 1. I conducted a 2 (Power: high wealth/power vs. low wealth/power) x 2 (Case strength: high vs. low) ANOVA on the confidence in verdict measure. The 2 x 2 ANOVA revealed a significant main effect for case strength, $F(1, 350) = 99.09, p < .00, \eta_p^2 = .22$. Participants delivered higher guilt ratings in the strong case ($M = 30.09$) compared to in the weak case ($M = -46.81$). Neither the main effect for the power manipulation nor the interaction between case strength and power were significant, $F(1, 350) < 2.00, ps > .05$.

Witness ratings. After reverse coding the necessary items, I combined the scales for all witness ratings so that higher scores corresponded with more positive ratings. First I examined the effects of the independent variables on the ratings of the detective ($\alpha = .85$). A 2 (Power) x 2 (Case strength) ANOVA on participants' ratings of the detective revealed a significant main

effect for case strength such that participants who viewed the strong case provided more positive ratings of the detective ($M = 4.55$) than did participants who viewed the weak case ($M = 4.28$), $F(1, 348) = 4.90, p = .03, \eta_p^2 = .02$. No other main effects or interactions are significant, $ps > .2$.

I also performed the 2 x 2 ANOVA on participants' ratings of the defendant ($\alpha = .85$). Similar to the ratings of the detective, there was a main effect of case strength on the ratings of the defendant, $F(1, 349) = 103.78, p < .01, \eta_p^2 = .23$. Participants who viewed the strong case rated the defendant less positively ($M = 3.57$) than did participants who viewed the weak case ($M = 4.77$). There was no significant main effect for power, and the power by case strength interaction was not significant, $ps > .5$.

Anticipated social interaction ratings. The participants' social interaction ratings measured how comfortable they expected to feel during the upcoming deliberation. I recoded the necessary items so that higher scores on the scale reflected that participants expected to feel more comfortable during the upcoming interaction. A 2 (Case strength) x 2 (Wealth/power) ANOVA on the ratings of anticipated social interaction revealed no main effects of case strength or wealth/power on the ratings of how comfortable participants expected to feel during the deliberations. However, there was an interaction between case strength and wealth/power, $F(1, 363) = 4.06, p < .05, \eta_p^2 = .01$. Pairwise comparisons did not reveal any simple main effects but for low wealth/power individuals, they anticipated feeling more comfortable when they viewed the strong case ($M = 6.77$) than when they viewed the weak case, ($M = 6.51$), $d = .262, p = .27$, 95% CI [- .20, .73]. This trend was reversed for high power individuals with those viewing the strong case feeling less comfortable ($M = 6.40$) than did high power participants in the weak case. ($M = 6.80$), $d = -.41, p = .08$, 95% CI [- .87, .06].

Post-Deliberation Analyses

Individual juror verdicts. After deliberations 36% of participants voted to convict the defendant. To examine the effects of the independent variables and deliberations on the individual verdicts, I conducted a linear mixed model regression with individual juror verdicts as the outcome measure. First, I examined the null model with jury as a random intercept only to determine how much variance was due to jury group membership. The intra-class correlation was .75, indicating that 75% of the variance in the model was due to group membership. The mean reliability estimates within groups was 95%, indicating a strong relationship between jury group members on their post-deliberation verdicts.

I then conducted a linear mixed model regression using the R program lme4, and the function “glmer” with the family binomial analyses regressions with binary outcomes. The results revealed a main effect of case strength, $z = -2.24$, $p < .05$, 95% CI [-9.77, -.66]. Participants in the strong case were more likely to vote guilty (54%) than were participants in the weak case (14%). No other main effects or interactions reached significance in the model, $ps > .05$.

Continuous guilt measures. For the continuous guilt measures I used a nested ANOVA with jury group as the nesting variable to account for jury-level variance in the models. First, I tested the effect of the continuous confidence in verdict measure. The verdict confidence measure ranged from -100 (100% confidence in not guilty verdict) to 100 (100% confidence in guilty verdict). I then performed a nested ANOVA with the error for the model nested within juries. The analysis revealed a significant effect of case strength, $F(1, 61.51) = 20.94$, $p < .01$, $\eta_p^2 = .25$, with participants in the strong case reporting higher guilt-confidence ratings ($M = 12.77$) than did participants in the weak case ($M = -63.70$). No other main effects or interactions were significant, $ps > .05$. Similar patterns were observed for all continuous guilt measures.

Post-interaction ratings. Prior to deliberations, I observed an interaction between case strength and power on participants' ratings of how comfortable they were during the interaction. I combined all items on the post-interaction scale and recoded any so that higher scores corresponded with having felt more comfortable during the deliberations ($\alpha = .86$). After deliberations, there was no longer a significant interaction on ratings of how comfortable participants felt after the interaction. However, there was a significant main effect of case strength, $F(1, 104.56) = 6.60, p = .01, \eta_p^2 = .06$. Participants in the weak case conditions reported feeling more comfortable ($M = 6.20$) than did participants in the strong case ($M = 5.88$). No other main effects or interactions were significant, $ps > .08$.

Trust in system ratings. For the nested ANOVA on trust in the system scale ($\alpha = .77$) there was a significant main effect of case strength, $F(1, 98.89) = 6.12, p = .02, \eta_p^2 = .06$. Participants who viewed the strong case rated their trust in the system as higher ($M = 3.55$) than did those who viewed the weak case ($M = 3.18$). No other main effects or interactions were significant, $ps > .1$.

Quality of Deliberations Measures

Deliberations coding. Although jurors deliberated in 60 juries, I had technical problems and lost 2 of the jury videos, leaving 58 juries. As in Study 1, those juries were transcribed (by a transcription company) and then divided into units. Units consisted of an argument and any accompanying justification. Overall, there were 3,305 units determined from the 58 juries. The mean number of units per juror was 9.50 ($SD = 9.53$). Those units were then coded for purpose and quality. The Cohen's Kappa for agreement between coders was .83, indicating a high level of agreement.

First, I examined the number of units per juror. I conducted a 2 (Power) x 2 (Diversity) x 2 (Strength of case) nested ANOVA to determine any main effects or interactions on the number of units per juror. The results of the nested ANOVA revealed no significant main effects or interactions ($ps > .05$).

I then evaluated the effects of the independent variables on the number of sufficient units per juror. Sufficient units met the requirement of either being derived directly from the evidence, or a reasonable inference from the evidence. I used the number of sufficient units per juror as my outcome measure, and conducted a 2 x 2 x 2 nested ANOVA to account for the error due to jury groups. The nested ANOVA revealed no significant main effects or interactions for any of the predictor variables.

I then evaluated the effects of the independent variables on the argument type contained in the units. Units supported the chosen, discounted the alternative, discounted the chosen, or supported the alternative. I only report the analyses for categories with significant results on diversity or wealth/power.

For discounting the chosen verdict, there was a significant main effect of power, $F(1, 95.67) = 4.99, p = .03, \eta_p^2 = .05$. Participants in the high wealth/power condition provided more units discounting their chosen verdict ($M = .39$) than did participants in the low wealth/power condition ($M = .16$). No other main effects or interactions were significant, $ps > .05$.

For supporting an alternative verdict, there was a significant main effect for diversity, $F(1, 96.13) = 4.32, p = .04, \eta_p^2 = .04$. Participants who deliberated in diverse juries provided more units supporting an alternative verdict ($M = .56$) than did participants deliberating in non-diverse juries ($M = .26$). There was also a power by case strength interaction, $F(1, 96.05) = 5.95, p = .02, \eta_p^2 = .06$. In the strong case, there was no significant effect of wealth/power ($Ms = .29$

vs. .50), $d = -.28$, $p = .08$, 95% CI [-.59, .03], but in the weak case, high wealth/power individuals provided more arguments supporting an alternative verdict ($M = .51$) than did low wealth/power individuals ($M = .18$), $d = .35$, $p = .03$, 95% CI [.03, .66].

I combined the remaining categories in the same manner as in Study 1. Therefore, I created five categories: Supporting-Facts, Supporting-Reasoning, Discounting-Inconsistent, Discounting Evidence, and “missing evidence.”

There was a significant effect of diversity in the Supporting-Reasoning category. In the Supporting-Reasoning category, in which participants engaged in higher level reasoning by bringing in outside knowledge or judging evidence to be more credible, I found a significant main effect of diversity, $F(1, 91.09) = 4.56$, $p < .05$, $\eta_p^2 = .03$. Participants in the diverse juries provided significantly more supporting-reasoning units ($M = 1.42$) than did participants in the non-diverse juries ($M = .88$).

For number of units containing case facts, there was a significant interaction between diversity and case strength found in the 2 x 2 x 2 nested ANOVA, $F(1, 91.39) = 5.23$, $p = .02$, $\eta_p^2 = .05$. Follow up tests revealed that in the strong case, participants in diverse juries provided more units with case facts ($M = .56$) than did participants in non-diverse juries ($M = .36$), $d = -.29$, $p = .02$, 95% CI [-.53, -.05]. In the weak case, this difference was not significant ($M_s = .35$ vs. .53), $d = .21$, $p = .09$, 95% CI [-.45, .03].

Jury-level analyses

Jury verdicts. Participants deliberated in 60 juries. Of those 60 juries, 17 (29%) voted to convict the defendant, 34 (57%) voted to acquit the defendant, and 9 (15%) remained hung after 45 minutes. I asked each hung jury to provide a public vote to determine majority vote for each jury, and after counting majority votes, 38 (63%) voted to acquit the defendant while 22 (37%)

voted to convict the defendant. I performed jury-level analyses on the verdicts using the majority counts.

First, I performed a binary logistic regression examining the effects of the conditions on jury verdicts. I entered the two dummy-coded diversity variables in the first step, along with the strength of case predictor. I then included interactions between strength of case and each of the dummy-coded diversity conditions in the second step. The omnibus test was significant, $\chi^2(3, N = 60) = 25.35, p < .01$, Nagelkerke $R^2 = .48$. In the first step, there was a main effect of case strength, Wald $\chi^2(1, N = 60) = 15.20, p < .01, Exp(B) = 20.51, 95\% CI [.01, .22]$. The ratio of guilty to not guilty verdicts was 20.51 times higher in the strong than the weak case. In the strong case, 19 (63%) juries voted guilty, compared to only 3 (10%) juries in the weak case. Also in the first step, there was a significant main effect for the dummy-coded non-diverse low wealth/power condition, Wald $\chi^2(1, N = 60) = 3.90, p < .05, Exp(B) = .18, 95\% CI [.03, .99]$. In the non-diverse low wealth/power condition the ratio of guilty to not verdicts was .18 times lower in the diverse condition than in the non-diverse high wealth/power condition. In the diverse condition, 4 (20%) of juries voted guilty compared to 9 (47%) in the low wealth/power diversity condition. The main effect for the dummy-coded non-diverse high wealth/power condition compared to the diverse condition approached significance in the same pattern as the non-diverse low power main effect, Wald $\chi^2(1, N = 60) = 3.60, p = .06, Exp(B) = .19, 95\% CI [.04, 1.06]$. For the non-diverse high power juries, 9 (43%) voted to convict the defendant, compared to 4 (20%) in the diverse condition.

Length of deliberations. Additionally, I examined the length of deliberations in Study 2. The 2 (Diversity) x 2 (Strength of case) ANOVA revealed no significant main effects or interactions of either independent variable on the length of jury deliberations, $ps < .1$.

Chapter 15: Study 2 Discussion

My goal for Study 2 was to examine the effects of diversity in wealth and power on jury deliberations. I manipulated the wealth and power (simultaneously) of participants to mimic the power differentials that could occur in deliberations when jurors have different statuses in society. I predicted an interaction between case strength and diversity conditions, which would reflect higher quality evidence evaluation in diverse groups compared to non-diverse ones. There were significant main effects for case strength throughout my analysis, suggesting that the case strength manipulation was successful; however there were no significant interactions between the strength of case and the diversity conditions on either juror or jury verdicts. Similar to the results in Study 1, I did not find evidence to support my primary hypothesis, that diverse juries (and jurors deliberating in diverse juries) would be more sensitive to evidence strength than non-diverse juries and jurors deliberating in non-diverse juries.

There are several possible explanations for diversity's lack of impact on the sensitivity to evidence strength. First of all, while the conviction rate in the second study was not quite as low as it was in the first study, there was an overall low conviction rate, and this may have reduced my ability to detect differences. No previous research has documented diversity's ability to improve sensitivity to evidence strength, so it is possible that diversity does not improve sensitivity to evidence strength. Previous research has documented how diversity improves the quality of deliberations discussions (Sommers, 2006), but this may not translate to more accurate verdicts. The more likely scenario, however, is that the low conviction rate and the shift towards acquittal that occurred following deliberations eliminated my ability to detect differences in verdicts.

Also, in support of the benefits of diversity, jurors deliberating in diverse juries provided more contributions on several of the quality of deliberation reasoning measures than jurors in non-diverse juries. Participants deliberating in diverse juries provided more units containing case facts, were more likely to discount the alternative opinion, and were more likely to use higher level reasoning arguments to support a verdict choice than were jurors deliberating in non-diverse juries. Consistent with Study 1, and with the Sommers (2006) data, jurors deliberating in diverse juries tend to out-perform jurors deliberating in non-diverse juries, reflected in more high quality contributions to jury deliberations.

I did not, however, find evidence of higher quality deliberations on all deliberation quality variables. Unlike in Study 1, I did not find that jurors deliberating in diverse juries provided more sufficient units or more discounting arguments. Since sufficient units were all of sufficiently high quality, the fact that there were no diversity effects on this variable mean that the diversity effect was weaker in the second study than in the first study. Given the nature of the manipulation (i.e., simulating wealth and power with monopoly money), this is not entirely surprising. It does suggest the need for additional research on this topic to determine the true effects of diversity defined as a combination of wealth/power groups.

It is important to note there were jury level effects for the diversity conditions. Juries in the non-diverse low wealth/power condition were more likely to vote to convict the defendant than were juries in the diverse condition. Thus, low wealth/power groups were more punitive than the diverse juries. Additionally, the high wealth/power condition exhibited the same trend (i.e., non-diverse high wealth/power juries were more punitive than diverse juries) so diversity likely caused this leniency in verdicts. Although deliberations have been demonstrated to increase leniency in juror verdicts (Kerr & MacCoun, 2012), it has not been demonstrated

previously that diversity in jury group composition increases leniency overall. Lynch and Haney (2011) did find that diverse juries (i.e., juries containing more minorities and women) were more lenient towards minority defendants. For all conditions in the present study the defendant was Hispanic, so it could be that diverse juries tend to be more lenient towards minority defendants.

The power manipulation also affected participants' contributions to deliberations. Participants who had been assigned to the high wealth/power group made significantly more arguments supporting the alternative verdict and discounting the chosen verdict than participants in the low power group. High wealth/power individuals may have felt more comfortable and powerful than low power individuals, and this may have contributed to their expression of high levels of reasoning. Galinsky et al. (2008) suggests that high power group members tend to feel more comfortable expressing their opinions in group situations, which could explain this finding. Previous research also finds that higher social status (i.e., education levels, income) tends to correspond with more participation in jury deliberation proceedings (Hastie et al., 1983; York & Cornwell, 2006).

In the strong case the low wealth/power individuals anticipated feeling more comfortable about the upcoming jury deliberations than did individuals in high wealth/power condition, and in the weak case this pattern was reversed. Low wealth/power individuals may have identified with the defendant (an "underdog"), and therefore felt more comfortable going into a deliberation in which they felt the defendant was guilty because they saw the defendant as an ingroup member. For high wealth/power individuals, they saw the defendant as an outgroup member, so they may have felt more comfortable entering deliberations in which they were planning to acquit the defendant.

Similar to the results of Study 1, participants' ratings of trust in the system were affected by the strength of the evidence in the case. Participants who viewed the strong case rated their trust in the system as higher than those who viewed the weak case. This finding demonstrates that experiences with the justice system have the potential to impact people's perceptions of the fairness in the justice system, which is supported by research on racial attitudes toward the police and other legal actors (Overby et al., 2005; Weitzer & Tuch, 2005).

Limitations

Study 2 was not without its limitations. Similar to Study 1, there were unexpected strength of case effects on the quality of deliberations and interactions with diversity, due to the weakness of the weak case. In the weak case we did not detect diversity effects for the most part, which likely occurred because the jurors already agreed upon entering deliberations. The ease with which jurors were able to agree on the final verdict in the weak case hindered the development of more meaningful discussion, and diluted some of our effects in terms of quality of deliberations. Also with a low post-deliberation conviction rate, we likely experienced floor effects on post-deliberation verdicts and guilt ratings which made it more difficult to detect differences between groups.

Another limitation to the second study was the manipulation of wealth and power. Although the manipulation check measures suggested that these manipulations did make jurors feel more or less wealthy and powerful, this was a very weak manipulation of wealth and power, compared to the impact of true wealth and power in the real world. Although possessing monopoly money has been demonstrated to increase feelings of wealth (Harvey & Bourhis, 2012), having false money cannot accurately replicate the experience of actual wealth and social status. However, the fact that I found effects of diversity with diversity defined by these trivial

distinctions suggest that the potential for wealth and power to impact jury groups might be immense.

The weakness of some of the diversity effects is an important limitation of this study's findings. In Study 1 there were significant diversity effects on several of the key categories demonstrating high levels of reasoning, but in Study 2 this was not the case. While I observed some diversity effects on the quality of deliberations measures, this was true for fewer categories than in the previous study. Additionally, jurors deliberating in diverse juries did not outperform those in non-diverse juries on the measures of sufficient units, and discounting of alternative arguments, which are two critical categories for establishing higher quality deliberations. We should be hesitant to draw too strong conclusions from this one study, and future research should explore how robust the wealth/power diversity effect truly is.

Another limitation in this study was the fact that my sample was made up of college students, who generally do not possess as much experience with the justice system and trial procedures as community members. This could explain the difference in conviction rates between the students and community members, although these cannot be directly compared as they were two separate studies. Also the students were all psychology majors studying at the same university, which means some of them may have been familiar to each other outside of the study. This is an important limitation to consider, because interactions with familiar individuals might be different than the interactions with strangers you might experience in jury duty. Using a student population also means that although I manipulated diversity of access to power and wealth, the age of participants was not diverse. Additionally, my sample was ethnically diverse with the largest ethnic group being Hispanic. It is possible that minorities are particularly

sensitive to these issues, and this wealth/power manipulation would not be as effective with a less diverse sample.

Conclusions

The second study did not find that jurors in diverse juries were more sensitive to case strength than those in non-diverse juries. The results revealed, however, that diversity did improve the quality of deliberations on some of the measures, although the diversity effects on the quality of deliberation measures were not as strong as in Study 1. Similar to the results in the first study any diversity benefits occurred for both high and low power individuals, indicating that both high and low social status individuals benefit from deliberating in diverse juries. Diverse juries were also less punitive than non-diverse low power juries, which indicates that diverse juries may be more lenient than non-diverse juries, particularly towards a minority defendant. Also, high wealth/power participants provided more units supporting the alternative verdict and discounting the chosen verdict, suggesting that members of high social status groups may contribute more to deliberations than members of low social status groups.

Chapter 16: Conclusions and Future Directions

Overall, the results of these two studies add to the literature finding that diverse jury deliberations are higher in quality than are the deliberations in homogenous ones. This research was designed to test several novel hypotheses. The first goal was to study the impact of diversity on the quality of deliberations defined as sensitivity to evidence strength. The second goal was to investigate the impact of diversity on mock jurors who were members of minority groups. Additionally this research sought to examine diversity defined as members of an experimentally designed group, by experimentally manipulating wealth and power.

In the first study I tested the hypothesis that diversity would improve the quality of deliberations for both minority and majority mock jurors, with diversity defined as sensitivity to the strength of the evidence. I did not find support for my hypothesis that diverse juries would be more sensitive to the strength of evidence than non-diverse juries. I found no interactions between evidence strength and diversity. The main effect of evidence strength, however, was strong and pervasive, possibly indicating that jurors were sensitive to the quality of evidence without the help of diversity. I also had a very low post-deliberation conviction rate, making it difficult to detect differences between groups. Another possible explanation is that the amount of variance explained by jury group membership (within group variance) was large and may have overpowered any effects of diversity on verdicts and guilt measures.

It is also possible that diversity does not improve sensitivity to case strength. Sommers (2006) also did not find evidence of verdict differences based on the composition of the jury, and studies finding that diversity impacts verdicts suggest that most of the effects lie in the increased likelihood of juries composed of a higher percentage of White males to deliver more punitive judgments against minority defendants (Lynch & Haney, 2011). There is also the possibility that

the group deliberation effect is so immense that we cannot detect diversity effects. For instance, Kalven and Zeisel (1966) found that pre-deliberation guilt votes were the greatest predictor of jury verdicts, with most of the juries eventually siding with the initial majority. Group effects may be more prominent than diversity effects for deliberating juries.

In the first study, I also hypothesized that diversity would improve the quality of the deliberation discussion. Using a previously established coding scheme (Kuhn et al., 1994), I examined the content of deliberations along a number of different reasoning dimensions. I hoped to expand upon previous findings that diverse juries deliberated for longer, reported more case facts, and made fewer errors, among other criteria (Sommers, 2006). My hypothesis that jurors deliberating in diverse juries would provide more high quality units than non-diverse juries was largely supported. Jurors who deliberated in diverse juries contributed more sufficient quality units, and more high-level reasoning arguments supporting a verdict, than jurors deliberating in non-diverse juries. Diversity appeared to improve deliberations in a similar manner for both Black and White participants. Diversity's impact was mainly driven by White mock jurors in previous research (Sommers, 2006). In Sommers (2006), however, the racial make-up of the participant sample in the previous study did not allow for a thorough exploration of the effects of deliberating in diverse compared to non-diverse juries for Black participants. The finding in the current study indicates that the mechanisms used to explain the benefits of diversity may need to be expanded. Although concern about prejudice and the watchdog effect (Fleming et al., 2005) has been previously used to explain diversity's benefits, these theories would likely not apply to minority jurors. Minorities are not concerned with appearing prejudiced, and instead would be more likely to focus on the potential to be the target of prejudice (Mendoza-Denton, et al., 2002). As a result, more research is needed to determine the primary mechanism for diversity's benefits.

Diversity may operate through some more basic mechanisms such as encouraging heterogeneous perspectives and encouraging different avenues for thought (Nemeth, 1986), or it is possible that diversity operates differently for different racial/ethnic groups.

In the first study I also found some evidence of racial differences in terms of trust in the system. Black participants rated the detective more negatively than White participants did, and also reported lower trust in the system than White participants, at least in the strong case. These findings are consistent with previous research documenting racial minorities' distrust of the police and other legal institutions (Overby et al., 2005; Weitzer & Tuch, 2005). The anticipation of unfair outcomes in the system might encourage minority individuals to process information more carefully, and may even have the potential to interact with diversity. From the results of the current study, however, we are unable to determine whether trust in the system is related to diversity effects.

I did not find support for the competing hypothesis that participating in diverse jury groups would impair cognitive performance and reduce the quality of decisions made by deliberating jurors (Richeson & Shelton, 2003). Although participants planning to engage in interracial interactions may often experience decreased cognitive performance, I did not find evidence of this performance decrement. I did, however, find that White participants expressed more anxiety about deliberating in a jury group than Black participants did. This effect is consistent with the idea that concern about appearing prejudiced can increase anxiety about upcoming interracial interactions (Goff et al., 2008), although White participants were equally anxious about interacting with a same-race group.

In the second study, I manipulated diversity in a novel way. Specifically, I manipulated how powerful and wealthy participants were using monopoly money in a minimal groups

paradigm (MGP; Harvey & Bourhis, 2012). I predicted that diverse juries made up of high and low wealth/power individuals would exhibit higher quality deliberations than non-diverse juries, a hypothesis that was partially supported. Jurors who deliberated in diverse juries were not more sensitive to case strength than jurors who deliberated in non-diverse juries.

On the quality of deliberation measures, jurors who deliberated in diverse wealth/power juries tended to make contributions to deliberations that reflected a higher level of reasoning than did those who deliberated in non-diverse juries, particularly in the strong case. Previous research has tested the impact of jury diversity, examining both gender and race as diversity characteristics (Lynch & Haney, 2011; Sommers, 2006); however this is the first study to examine the impact of manipulated power dynamics on jury diversity. The benefits of diversity were not as powerful as in Study 1, but they still provide evidence of a diversity benefit. There were also effects of the wealth/power manipulation, such that participants in the high wealth/power groups provided more units supporting the alternative verdict and discounting the chosen verdict than did participants in low wealth/power groups. This finding is consistent with that in previous research finding that high power individuals tend to feel more comfortable expressing their opinions during intergroup interactions than members of low power groups (Galinsky et al., 2008).

Interestingly, when it came to diverse juries, they were more likely to be lenient compared to non-diverse juries, particularly low power juries. Diverse juries were not more sensitive to the quality of evidence, but instead were less punitive towards the defendant. The defendant was a minority (Hispanic) in all conditions, so perhaps diversity increased leniency towards a minority defendant, similar to the results in Lynch and Haney (2011). However, diversity was defined in this study as mixing high and low wealth/power individuals together.

Perhaps adding low power groups however they are defined (e.g., women, minorities, low SES individuals) can increase sympathy for low status defendants in a jury deliberation.

Limitations

The largest limitation to this study was the weakness of the weak case. In the first study, this caused post-deliberation verdicts to be so low in the first study that I could not even detect a significant effect for the case strength. The weakness of the case with the weaker evidence also likely diluted the effects of the beneficial impact of diversity, as I mostly found evidence for diversity's benefits for the strong case but not the weak case. The ease with which jurors decided the weak case also likely contributed to the unexpected main effect for case strength on most quality of deliberation measures, with higher quality deliberations in the strong case in which verdicts for jurors were closer to a 50/50 split. The strength of case by diversity interaction on the quality of deliberation measures can likely be explained by the fact that the strong case was a more ambivalent case, allowing for more movement in either direction. In the weak case the deliberations were shorter with fewer high quality units, likely because it was an easier case to decide.

Another limitation is that although diversity was defined in a unique manner, I did not examine the impact of diversity for members of racial groups other than Black individuals, therefore I cannot draw conclusions about the impact of diversity for other minority groups. Studies examining the impact of diversity on jury deliberations should seek to include other minority groups in these studies, and study both the impact of diversity on these groups, and also the impact of including jurors from these groups in deliberations for White jurors. Black individuals experience a unique amount of prejudice related to criminality (Goff, Purdue,

Eberhardt & Davies, 2008) and mistreatment at the hands of the justice system (Baldus et al, 1983), and it is possible that other groups might have different experiences in jury deliberations.

Another limitation to this study is the fact that I did not manipulate race of the defendant. One of the proposed benefits of increasing diversity in juries is to increase fair outcomes for minority individuals. I cannot determine whether there was bias, or whether it was reduced, because I do not have comparisons among races. This is similar to previous studies on diversity (Sommers, 2006), and perhaps future research can focus on how diversity can reduce bias. However, in the second study diverse juries were more lenient towards a Hispanic defendant than low power non-diverse juries, suggesting a possible lessening of bias towards a minority defendant. With the largest racial/ethnic group in my sample being Hispanic, however, I am hesitant to draw a strong conclusion about the racial bias effect.

Implications

The courts prescribe several protections against the prejudices of legal decision-makers, and the ability of diversity to guard against biased legal decisions is uncontroversial. Many court decisions are based on the assumption that diversity and deliberations act to benefit the quality of deliberations, correcting for biases and prejudices (*Peters v. Kiff*, 1972). Increasing the diversity of juries also has the ability to improve perceptions of the justice system and increase the perceived fairness of racial outcomes. Verdicts delivered by all-White juries tend to be perceived as less fair than verdicts delivered by diverse juries (Ellis & Diamond, 2003). The increased perceived fairness of verdicts mean that there should be an emphasis on including minority jurors in deliberations.

The findings of diversity benefits for minority jurors are important, as few studies directly compare all-minority juries with diverse juries. Most jury diversity research focuses on

the addition of minority group members to all-White juries. This emphasis is logical, as diversity is typically proposed as a method to reduce racial bias towards minority defendants. The current study is one of the first to examine the deliberations of all-minority juries. It is important to note that minority jurors deliberating in diverse juries exhibited higher quality contributions to deliberations than minority jurors deliberating in non-diverse juries. The benefit for minorities deliberating in diverse juries compared to homogenous juries is a novel finding, and one that deserves further exploration. The results of the current study do not clarify the mechanism by which diversity improves deliberations for minority group members, but future research should delve into this further. The only measure for which we observed racial differences was on the trust in the system variables, and it is possible that minority juries are more likely to focus on problems in the justice system than are juries consisting of mostly White jurors.

This research also has important implications for the effects of wealth and power on jury deliberations. With a relatively trivial manipulation of power and wealth, we were able to see more leniency in diverse juries, and also higher quality deliberations on several dimensions when the power and wealth groups deliberated together rather than separately. Additionally, in some conditions high wealth/power individuals offered more high-level reasoning units than low wealth/power individuals, suggesting that power influences how comfortable jurors feel expressing themselves. This is particularly notable in a jury setting, where ones' wealth and power status is irrelevant to the task at hand.

Future Directions

The results of this research provide many avenues for future research. Although the finding that diversity improves deliberations is not a novel one (Lynch & Haney, 2011; Sommers, 2006), this research expands the types of diversity that can improve jurors' decisions.

Although racial diversity is an important aspect of diversity, many different types of diversity have been demonstrated to improve group decision-making and performance (Gruenfeld, Mannix, Williams, & Neale, 1996; Stasser, Stewart, & Wittenbaum, 1995; Watson, Kumar, & Michaelsen, 1993). Researchers should continue to examine the role of power and wealth dynamics in juries, particularly because power and wealth are so difficult to extricate from race.

The next direction to take this research might also be to explore the role of diverse jury group deliberations in reducing racial bias in jury decisions. In one study, jurors were more lenient when they anticipated deliberating in a diverse group than when they deliberated in a non-diverse group (Sommers, 2006). Another study showed that adding minorities and women to jury proceedings improved outcomes for a minority defendant (Lynch & Haney, 2011). Would this effect still occur if the defendant were not the same race as the minority group jurors? The results of the present research suggest that diverse juries might be more lenient towards any defendant, or perhaps any minority defendant.

Another direction to explore jury diversity might be to examine the impact of diversity from an interracial interactions perspective. Perhaps the common goal and the script of a jury deliberation with the goal of reaching a unanimous decision are sufficient factors to reduce the uncertainty and anxiety that characterizes most interracial interactions (Avery et al., 2009). Perhaps there are certain cases that would be more likely to increase anxiety about interacting in diverse groups, such as raising the potential for racial issues to come into play, which might interfere with diversity's benefits (Goff et al., 2008).

The most important direction we should take from this research is to build on the literature finding evidence of diversity's ability to improve deliberations (Lynch & Haney, 2011; Sommers, 2006). With the inclusion of the current study we now have evidence that both Black

and White individuals make higher quality contributions to deliberations when they participate in diverse juries compared to when they participate in homogenous juries. Additionally, previous research documents the ability of the watchdog effect to promote better information-processing in White jurors deliberating in diverse groups, sometimes even prior to deliberations (Sommers, 2006; Sommers, et al., 2008). The current research suggests that a lack of trust in the system, or being vigilant about the potential for injustice, may encourage minority jurors to process evidence more carefully. Minorities may be more sensitive to potential injustice when deliberating in a diverse group, believing that majority group members may be biased against minorities. Future research should explore the mechanism by which minority jurors make more high quality contributions to jury deliberations, and determine whether diversity's benefit operates through information-processing. Perhaps heterogeneity of perspectives and ideas is the mechanism by which diversity improves deliberations, or perhaps minorities and majority group members benefit from diversity through different mechanisms.

Conclusion

The courts often tout the importance of including all jurors in deliberations, and hope that diverse juries will protect minority jurors from unfair outcomes, and most research suggests that this faith is rightly placed (Lynch & Haney, 2011; Sommers, 2006). The current study supports the view that jurors will deliberate better in diverse juries than in non-diverse juries, and this study finds evidence of this among two definitions of diversity. This study suggests that jurors deliberate better in diverse juries whether diversity as racial diversity or wealth and power. The results of the current study suggest that we need to focus on re-defining the mechanism by which diversity impacts juries. Minorities experienced the same benefits for diversity as White

individuals, whereas much research touting diversity has emphasized Whites' high levels of information processing in interracial situations.

Overall, this research supports the courts' assumption that diverse deliberations will serve to improve jury decision-making and reduce bias against low power and minority defendants. The prevention of minorities from serving on juries not only violates individual rights, but is also a disservice to the quality of jury deliberations. This research highlights the fact that failed *Batson* challenges, and the removal of minorities from juries, can severely impact the quality of jury deliberations.

Table 1

Study 1 guilty verdict descriptive statistics

	<i>n</i>	Diverse <i>n</i> (%)		Non-Diverse White <i>n</i> (%)		Non-Diverse Black <i>n</i> (%)	
		Strong	Weak	Strong	Weak	Strong	Weak
Pre-Deliberation		24	7	29	4	25	11
Verdicts	388	(37.5%)	(11.9%)	(42.0%)	(6.5%)	(37.3%)	(16.4%)
Post-Deliberation		12	1	10	0	5	6
Verdicts	388	(18.8%)	(1.7%)	(14.5%)	(0.0%)	(7.5%)	(9.0%)

Note. Descriptive statistics for guilty verdicts only.

Table 2

Study 1 coding criteria descriptive statistics

		Strong Case		Weak Case	
		<i>M (SD)</i>		<i>M (SD)</i>	
	<i>n</i>	Diverse	Non-Diverse	Diverse	Non-Diverse
Coding Criteria					
Units	388	25.91 (29.65)	15.80 (13.19)	13.92 (17.27)	13.22 (10.26)
Sufficient Units	388	10.00^a (12.06)	6.05^a (5.47)	4.68 (4.75)	5.47 (4.95)
Support-Chosen	388	4.64 (7.71)	2.53 (3.13)	2.44 (2.85)	2.81 (2.90)
Discount-Alternative	388	3.81^a (4.58)	2.29^a (2.71)	1.69 (1.98)	2.08 (2.53)
Discount-Chosen	388	.75 (1.18)	.59 (1.37)	.27 (.61)	.39 (.85)
Support-Alternative	388	.81 (1.23)	.82 (1.24)	.31 (.73)	.29 (.61)
Support-Reasoning	388	2.41 (3.15)	1.50 (1.76)	1.27 (1.41)	1.26 (1.64)
Discount-Inconsistent	388	1.97 (2.83)	1.45 (1.86)	.85 (1.36)	.91 (1.38)
Discount-Evidence	388	1.70^a (2.33)	.80^a (1.35)	.71 (1.11)	.97 (1.70)
Support-Case Facts	388	2.86^a (4.33)	1.71^a (2.12)	1.15 (1.73)	1.71 (1.81)
# of “missing evidence”	388	.87 (1.30)	.50 (.97)	.39 (.87)	.51 (.96)

Note: Within dependent measures, means sharing subscripts differ at $p < .05$.

Table 3

Study 2 Pre-deliberation guilty verdict descriptive statistics

		High Wealth/Power		Low Wealth/Power	
		<i>n</i> (%)		<i>n</i> (%)	
	<i>n</i>	Strong	Weak	Strong	Weak
Pre-Deliberation Verdicts	356	55 (56.7%)	18 (21.4%)	50 (52.1%)	5 (6.3%)

* *Note.* Descriptive statistics for guilty verdicts only.

Table 4

Study 2 Post-deliberation guilty verdicts descriptive statistics

		Diverse		Non-Diverse High		Non-Diverse Low	
		<i>n</i>(%)		<i>n</i>(%)		<i>n</i>(%)	
	<i>n</i>	Strong	Weak	Strong	Weak	Strong	Weak
Post-Deliberation Verdicts	356	28 (43.1%)	4 (6.2%)	41 (63.1%)	15 (30.6%)	36 (57.1%)	4 (8.2%)

Note. Descriptive statistics for guilty verdicts only.

Table 5

Study 2 coding criteria descriptive statistics

Strong Case <i>M(SD)</i>					Weak Case <i>M(SD)</i>			
Diverse			Non-Diverse		Diverse		Non-Diverse	
<i>n</i>	High W/P	Low W/P	High W/P	Low W/P	High W/P	Low W/P	High W/P	Low W/P
Coding Criteria								
Units	13.09	12.88	6.50	9.84	10.52	6.87	11.00	7.46
348	(10.27)	(10.18)	(7.71)	(9.19)	(10.69)	(8.81)	(11.32)	(8.57)
Sufficient Units	5.13	5.24	2.93	4.26	5.12	3.40	4.70	3.64
348	(4.53)	(5.31)	(4.79)	(3.92)	(6.10)	(2.72)	(4.99)	(5.22)
Support- Chosen	2.56	2.56	1.08	2.31	7.38	5.00	8.56	5.48
348	(3.22)	(2.83)	(1.94)	(2.36)	(7.24)	(4.56)	(9.42)	(6.03)
Discount- Alternative	2.19	2.50	1.55	1.78	1.91	1.37	1.66	1.72
348	(2.04)	(3.16)	(2.76)	(2.14)	(2.48)	(1.63)	(2.00)	(3.00)
Discount- Chosen	.38	.18	.30	.17	.59	.30	.36	.06
348	(.94)	(.39)	(.77)	(.63)	(1.37)	(.60)	(1.03)	(.24)
Support- Alternative	.48	.82	.18	.31	.68	.20	.40	.16
348	(.85)	(1.11)	(.39)	(.60)	(2.27)	(.55)	(.90)	(.51)
Support- Reasoning	1.44	1.53	.52	.97	1.38	.97	1.40	.86
348	(2.21)	(1.81)	(1.13)	(1.53)	(2.41)	(1.13)	(2.12)	(1.31)
Discount- Inconsistent	1.44	1.38	.98	.98	1.44	.63	.92	.90
348	(1.48)	(1.95)	(2.21)	(1.38)	(2.05)	(.85)	(1.12)	(1.97)
Discount- Evidence	1.00	1.15	.80	.83	.94	.80	.96	.86
348	(1.34)	(1.44)	(1.31)	(1.20)	(1.46)	(1.19)	(1.40)	(1.39)
Support- Case Facts	.53	.59	.22	.50	.44	.27	.50	.56
348	(.86)	(.84)	(.45)	(.73)	(.70)	(.58)	(.76)	(1.13)
# of “missing evidence”	.12	.23	.14	.04	.13	.15	.08	.14
348	(.33)	(.57)	(.35)	(.20)	(.34)	(.44)	(.33)	(.35)

Table 6

Study 2 jury-level guilty verdict descriptive statistics

		Diverse	Non-Diverse High	Non-Diverse Low
		<i>n</i>(%)	<i>n</i>(%)	<i>n</i>(%)
<i>n</i>				
Post-Deliberation		4	9	9
Verdicts	60	(20%)	(42.9%)	(47.4%)

* *Note.* Descriptive statistics for guilty verdicts only.

Appendix A: Post-Trial Questionnaire

POST-TRIAL QUESTIONNAIRE

1. Do you find the defendant, Mr. Frank Johnson guilty or not guilty of first degree murder?
(CIRCLE ONE)

Guilty

Not Guilty

2. On a scale from 0% - 100%, how confident are you in your verdict? _____ % confident
3. Give a percentage between 0% and 100%. What is the probability that the defendant, Mr. Frank Johnson, committed first degree murder? _____ %

Please indicate whether you agree or disagree with the following statements by choosing the number on the scale corresponding to your belief.

1. The defense had a very strong case.

1	2	3	4	5	6	7	8	9
Strongly								Strongly
Disagree								Agree

2. The evidence against the defendant was very strong.

1	2	3	4	5	6	7	8	9
Strongly								Strongly
Disagree								Agree

Next, we would like you to describe your impressions of witnesses that you heard using a series of adjective pairs. The scales are designed so that you can express the degree to which the person that you are rating seems to fit one end of the scale or the other. Which space you check should depend on the degree to which the word describes the person you are rating.

For example, if you thought that Jane was *slightly* tall, you would mark the item as follows:

Tall : _____ : _____ : _____ : X : _____ : _____ : _____ : Short

However, if you thought that Jane was *extremely* short, you should place the "X" next to short:

Tall : _____ : _____ : _____ : _____ : _____ : _____ : X : Short

If you have any questions about this task, you may ask them at this time.

Based on the testimony that you heard, carefully rate your impressions of *Detective Andrea Jones* as best you can on each of the following dimensions:

trustworthy	:_____:	untrustworthy
honest	:_____:	dishonest
attractive	:_____:	unattractive
not believable	:_____:	believable
convincing	:_____:	unconvincing
certain	:_____:	uncertain

Based on the testimony that you heard, carefully rate your impressions of *David Cross (the medical examiner)* as best you can on each of the following dimensions:

trustworthy	:_____:	untrustworthy
honest	:_____:	dishonest
attractive	:_____:	unattractive
not believable	:_____:	believable
convincing	:_____:	unconvincing
certain	:_____:	uncertain

Based on the testimony that you heard, carefully rate your impressions of *Brian Connolly (the bartender)* as best you can on each of the following dimensions:

trustworthy	:_____:	untrustworthy
honest	:_____:	dishonest
attractive	:_____:	unattractive

not believable	:_____:	believable
convincing	:_____:	unconvincing
certain	:_____:	uncertain

Based on the testimony that you heard, carefully rate your impressions of *Jack Healy (the friend of the defendant)* as best you can on each of the following dimensions:

trustworthy	:_____:	untrustworthy
honest	:_____:	dishonest
attractive	:_____:	unattractive
not believable	:_____:	believable
convincing	:_____:	unconvincing
certain	:_____:	uncertain

Based on the testimony that you heard, carefully rate your impressions of *Rebecca Tyler (the waitress)* as best you can on each of the following dimensions:

trustworthy	:_____:	untrustworthy
honest	:_____:	dishonest
attractive	:_____:	unattractive
not believable	:_____:	believable
convincing	:_____:	unconvincing
certain	:_____:	uncertain

Based on the testimony that you heard, carefully rate your impressions of *Frank Johnson (the defendant)* as best you can on each of the following dimensions:

trustworthy	:_____:	untrustworthy
honest	:_____:	dishonest
attractive	:_____:	unattractive

not believable : _____ : _____ : _____ : _____ : _____ : _____ : _____ : believable

convincing : _____ : _____ : _____ : _____ : _____ : _____ : _____ : unconvincing

certain : _____ : _____ : _____ : _____ : _____ : _____ : _____ : uncertain

You will now be deliberating in a jury group. You will be instructed to discuss the evidence in the case, and then come to a unanimous verdict decision. We are interested in your honest thoughts and feelings about how this conversation will go. Please indicate whether you agree or disagree with the following statements by choosing the number on the scale corresponding to your belief.

1. I will feel self-conscious during this interaction.

1 2 3 4 5 6 7
Strongly Strongly
Disagree Agree

2. I am worried I will not be able to talk comfortably.

1 2 3 4 5 6 7
Strongly Strongly
Disagree Agree

3. I will feel comfortable during this interaction.

1 2 3 4 5 6 7
Strongly Strongly
Disagree Agree

4. I will be able to “be myself” during this interaction.

1 2 3 4 5 6 7
Strongly Strongly
Disagree Agree

5. I will feel accepted during this interaction.

1 2 3 4 5 6 7
Strongly Strongly
Disagree Agree

6. I will feel comfortable sharing my views and perspectives about the case with the other jurors.

1	2	3	4	5	6	7
Strongly					Strongly	
Disagree					Agree	

7. I will feel comfortable telling the other jurors when I disagree with their views.

1	2	3	4	5	6	7
Strongly					Strongly	
Disagree					Agree	

8. I will not feel comfortable arguing for the verdict that I believe is correct.

1	2	3	4	5	6	7
Strongly					Strongly	
Disagree					Agree	

Appendix B: Post-Deliberation Questionnaire

POST-DELIBERATION QUESTIONNAIRE

1. Do you find the defendant, Mr. Frank Johnson guilty or not guilty of first degree murder?
(CIRCLE ONE)

Guilty

Not Guilty

2. On a scale from 0% - 100%, how confident are you in your verdict? _____ % confident
3. What is the probability that the defendant (0%-100% probability, Mr. Frank Johnson, committed first degree murder? _____ %

Please indicate whether you agree or disagree with the following statements by choosing the number on the scale corresponding to your belief.

1. The defense had a very strong case.

1	2	3	4	5	6	7	8	9
Strongly Disagree								Strongly Agree

2. The prosecution had a very strong case.

1	2	3	4	5	6	7	8	9
Strongly Disagree								Strongly Agree

Please indicate your agreement with the following statements.

1. Even though I didn't agree with the other jurors, I felt pressure to go along with their verdict.

1	2	3	4	5	6	7	8	9
Strongly Disagree								Strongly Agree

2. I convinced at least one of my fellow jurors to change their verdict.

1	2	3	4	5	6	7	8	9
Strongly Disagree								Strongly Agree

3. If this was an actual trial, I would have fought harder to convince my fellow jurors that the defendant was [guilty/not guilty].

1	2	3	4	5	6	7	8	9
Strongly								Strongly
Disagree								Agree

You have just finished deliberating in a jury group. We are interested in your honest thoughts and feelings about how this deliberation session went. Please state your agreement with the following statements:

1. I felt self-conscious during this deliberation.

1	2	3	4	5	6	7
Strongly						Strongly
Disagree						Agree

2. I felt comfortable during the deliberations.

1	2	3	4	5	6	7
Strongly						Strongly
Disagree						Agree

3. I felt like I could “be myself” during deliberations.

1	2	3	4	5	6	7
Strongly						Strongly
Disagree						Agree

4. I felt accepted during deliberations.

1	2	3	4	5	6	7
Strongly						Strongly
Disagree						Agree

5. I felt like others listened to me during the deliberations.

1	2	3	4	5	6	7
Strongly						Strongly
Disagree						Agree

6. I felt comfortable sharing my views and perspectives about the case with the other jurors.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

7. I felt comfortable telling the other jurors when I disagreed with their views.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

8. I did *not* feel comfortable arguing for the verdict that I believed was correct.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

Please indicate your agreement with the following statements:

1. If a suspect runs from police, then he probably committed the crime.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

2. Out of every 100 people brought to trial, at least 75 are guilty of the crime with which they are charged.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

3. Generally, the police make an arrest only when they are sure about who committed the crime.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

4. A prior record of conviction is the best indicator of a person's guilt in the present case.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

5. When it is the suspect's word against the police officer's I believe the police.

1	2	3	4	5	6	7
Strongly						Strongly
Disagree						Agree

6. If a witness refuses to take a lie detector test, it is because he/she is hiding something.

1	2	3	4	5	6	7
Strongly						Strongly
Disagree						Agree

Please answer the following questions to the best of your ability:

1. What was the race of the defendant?
 - a. African-American
 - b. White
 - c. Hispanic
 - d. Other
 - e. Don't remember
2. What was the weapon used by the defendant in this case?
 - a. Hammer
 - b. Knife
 - c. Gun
 - d. No Weapon
3. What was the name of the bar where the crime occurred?
 - a. O'Reilly's
 - b. O'Connor's
 - c. Gleason's
 - d. Smith's
4. Which of the following did the expert testify about:
 - a. The time of death
 - b. The bullet pattern
 - c. The direction of the knife
 - d. Whether the victim had a heart attack
5. Which of the following characterizes the defense's case:
 - a. Insanity Plea
 - b. Self-Defense
 - c. Crime of passion
 - d. Innocence (someone else committed the crime)

Appendix C: Kuhn, Weinstock & Flaton (1994) Coding Sheet Revised

1. Participant Number _____
2. Jury Number _____ Unit Number _____
3. Opinion reached by the juror
1 = Guilty
0 = Not Guilty

Function: What is the purpose of the unit?

4. Jusverd: The function of the unit was to *justify the chosen opinion*.
1 = yes (If 1, enter 0 for 5-7)
0 = no (If 0, go to 5)
5. Disaltv: The function of the unit was to *discount an alternative opinion*
1 = yes (If 1, enter 0 for 6-7)
0 = no (If 0, go to 6)
6. disver: The function of the unit was to *discount the chosen opinion*
1 = yes (If 1, enter 0 for 7)
0 = no (If 0, go to 7)
7. supalt: The function of the unit was to support an alternative opinion
1 = support an alternative opinion
0 = DID NOT support an alternative opinion

Quality

8. Legit
1 = Legitimate
0 = Illegitimate
9. Relevanc
1 = Relevant
0 = Not Relevant
10. Suff
1 = Sufficient – scored a 1 on legit AND relevanc
0 = Deficient – scored a 0 on EITHER legit OR relevanc

SUPPORTING

Nonjudgmental

11. Factual (Snjfact)
1 = yes; 0 = no
12. Narrative (Snjnarr)
1 = yes; 0 = no

Judgmental

13. Importing (sjimpor)
1 = yes; 0 = no
14. Credibility (sjcred)
1 = yes; 0 = no

DISCOUNTING

Inconsistent

15. Nonjudgmental – factual (dincnj)
1 = yes; 0 = no

16. Judgmental – importing (dincjud)

1 = yes; 0 = no

Discounting of evidence

17. Judgmental – importing (Dunjimp)

1 = yes; 0 = no

18. Judgmental – credibility (Dunjcre)

1 = yes; 0 = no

Appendix D: Kuhn, Weinstock & Flaton (1994) Coding Scheme Revised

Coding Sheet for Units

We are coding three things:

1. the function of the unit (the purpose it was intended to serve; 4-7)
2. the quality of the unit (legitimate, relevance, sufficient; 8-10)
3. the type of the argument in the unit (supporting vs. discounting, 11-18)

Also, before you begin coding, read through the expert report to make sure you know what the final opinion is (to help you code for supporting vs. discounting arguments).

19. Participant: Participant Number
20. Opinion: Opinion reached by the juror
 - a. 1 = Guilty
 - b. 0 = Not guilty

Function: What is the purpose of the unit?

21. Jusverd: The function of the unit was to *justify the chosen opinion*.
 - a. 1 = yes
 - b. 0 = no
 - i. If 1, enter 0 for 5-7
 - ii. If 0, go to 5
22. Disaltv: The function of the unit was to *discount an alternative opinion*
 - a. 1 = yes
 - b. 0 = no
 - i. If 1, enter 0 for 6-7
 - ii. If 0, go to 6
23. disver: The function of the unit was to *discount the chosen opinion*
 - a. 1 = yes
 - b. 0 = no
 - i. If 1, enter 0 for 7
 - ii. If 0, go to 7
24. supalt: The function of the unit was to support an alternative opinion
 - a. 1 = The function of the unit was to support an alternative opinion
 - b. 0 = The function of the unit was not to support an alternative opinion

Quality

25. Legit: Legitimate vs. Illegitimate
 - a. 1 = Legitimate – the evidence is either drawn directly from the trial or consists of a reasonable inference drawn from the information

- b. 0 = Illegitimate – the evidence is not either drawn directly from trial or constitutes a reasonable inference drawn from testimony
- 26. Relevanc: Relevant v. Not Relevant
 - a. 1 = Relevant – the evidence either increases or decreases the probability that the criteria associated with the opinion are met
 - b. 0 = Not Relevant – the evidence does not either increases or decreases the probability that the criteria associated with the opinion are met
- 27. Suff: Sufficient vs. deficient. **NOTE: to be sufficient, the statement must also be legitimate and relevant**
 - a. 1 = Sufficient – scored a 1 on legit AND relevanc
 - b. 0 = Deficient – scored a 0 on EITHER legit OR relevanc

ONLY THOSE ITEMS THAT ARE SUFFICIENT ARE FURTHER CATEGORIZED

Each unit should be placed in ONE of the following 8 categories. That is, only one of the following 8 categories should receive a one for a particular unit—the rest should receive zeros.

SUPPORTING

Nonjudgmental

- 28. Factual (Snjfact: 1 = yes; 0 = no)
The expert refers to pieces of evidence drawn directly from trial without elaboration to support the opinion choice
- 29. Narrative (Snjnarr: 1 = yes; 0 = no)
The expert uses evidence to construct a possible narrative consistent with evidence but elaborating on it

Judgmental

- 30. Importing (sjimpor: 1 = yes; 0 = no)
The expert compares the evidence with real world knowledge or ‘common sense’, thereby making the evidence more supportive of the opinion choice
- 31. Credibility (sjcred: 1 = yes; 0 = no)
The expert evaluates the source of the evidence and concludes that the evidence is more accurate because of the source (hence supportive of the opinion choice).

DISCOUNTING

Inconsistent

32. Nonjudgmental – factual (dincnj: 1 = yes; 0 = no)

The expert refers to pieces of direct testimony alleged to be inconsistent with the opinion being discounted (refers to evidence “missing from the case”)

33. Judgmental – importing (dincjud: 1 = yes; 0 = no)

The expert uses real world knowledge or ‘common sense’ to support the argument that the particular evidence is inconsistent with the verdict choice

Discounting of evidence

34. Judgmental – importing (Dunjimp: 1 = yes; 0 = no)

The expert compares the evidence to real-world knowledge or ‘common sense’, with the outcome that the evidence is judged to be less supportive of verdict choice

35. Judgmental – credibility (Dunjcre: 1 = yes; 0 = no)

The expert evaluates the source of the evidence with the outcome that this evidence is less likely to be accurate (thus less supportive of verdict choice)

Appendix E: Minimal Groups Questionnaire (including Tajfel Matrices)

You will now be asked to distribute your resources to others. You are required to do this by using a matrix, and deciding which strategy you will use. You are not allowed to give money to yourself, instead you will send the money to members of your group and to members of the other group. You will then place the amount of money in the envelopes given to you by the researcher. If you are in Group A, for example

Here is an example:

In-group Member	10	20	30	40
Out-group Member	40	30	20	10

If you chose the second matrix, you would place \$20 in the envelope for the member of your group and \$30 in the envelope for the other group member. Please remember to circle the matrix.

Please circle the column representing your distribution choice:

Matrix 1

In-group Member	50	45	40	35	30	25	20	15
Out-group Member	15	20	25	30	35	40	45	50

Matrix 2

In-group Member	25	30	35	40	45	50	55	60
Out-group Member	15	20	25	30	35	40	45	50

Matrix 3

In-group Member	10	20	30	40	50	60	70	80
Out-group Member	15	20	25	30	35	40	45	50

You will now have money leftover. This money can be either used for future research studies, or can be returned to us. Please indicate how much money you would like to be used in future studies and place it in the “future studies” envelope.

Please circle the choice indicating your agreement with the following statements:

1. How wealthy do you feel?

1	2	3	4	5	6	7
Not at all						Extremely
Wealthy						Wealthy

2. How powerful do you feel?

1	2	3	4	5	6	7
Not at all						Extremely
Powerful						Powerful

3. How happy do you feel as a member of your group?

1	2	3	4	5	6	7
Not at all						Extremely
Happy						Happy

4. How at ease do you feel as a member of your group?

1	2	3	4	5	6	7
Not at all						Extremely
At ease						At ease

5. How satisfied do you feel as a member of your group?

1	2	3	4	5	6	7
Not at all						Extremely
Satisfied						Satisfied

6. How vulnerable do you feel as a member of your group?

1	2	3	4	5	6	7
Not at all						Extremely
Vulnerable						Vulnerable

Please answer the following questions by circling the number indicating your level of agreement with the following statements:

1. Overall, my Group (A or B) has very little to do with how I feel about myself.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

2. The Group (A or B) I belong to is an important reflection of myself.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

3. The Group (A or B) I belong to is unimportant to my sense of what kind of person I am.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

4. In general, belonging to my Group (A or B) is an important part of my self-image.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

References

- Adams, L. T., Bryden, M. W., & Griffith, J. D. (2011). Middle Eastern racial bias and the impact of jury deliberation. *American Journal of Forensic Psychology*, 29(3), 41-59.
- Alderfer, C. P., & Smith, K. K. (1982). Studying intergroup relations embedded in organizations. *Administrative Science Quarterly*, 27(1), 35-65. doi:10.2307/2392545
- Allport, G. W. (1954). *The nature of prejudice*. Oxford, England: Addison-Wesley.
- Amiot, C. E., & Bourhis, R. Y. (2005). Ideological beliefs as determinants of discrimination in positive and negative outcome distributions. *European Journal of Social Psychology*, 35(5), 581-598. doi:10.1002/ejsp.238
- Apodaca v. Oregon*, 406 U.S. 404 (1972).
- Avery, D. R., Richeson, J. A., Hebl, M. R., & Ambady, N. (2009). It does not have to be uncomfortable: The role of behavioral scripts in Black–White interracial interactions. *Journal of Applied Psychology*, 94(6), 1382-1393. doi:10.1037/a0016208
- Baldus, D. C., Woodworth, G. G., Zuckerman, D., Weiner, N. A., & Broffitt, B. (2001). The use of peremptory challenges in capital murder trials: A legal and empirical analysis. *University of Pennsylvania Journal of Constitutional Law*, 3, 3–169.
- Baldus, D. C., Pulaski, C., & Woodworth, G. (1983). Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience. *Journal of Criminal Law & Criminology*, 74(3), 661-753.
- Ballew v. Georgia*, 435 U.S. 223 (1978).
- Barnes, C.W. & Kingsnorth, R. (1996). Race, drug, and criminal sentencing: Hidden effects of the criminal law. *Journal of Criminal Justice*, 24(1), 39-55. doi:10.1016/0047-2352(95)00051-8

Batson v. Kentucky, 476 U.S. 79 (1986).

Baumeister, R. F., Muraven, M., & Tice, D. M. (2000). Ego depletion: A resource model of volition, self-regulation, and controlled processing. *Social Cognition*, 18(2), 130-150. doi:10.1521/soco.2000.18.2.130

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, 3(1), 27-52. doi:10.1002/ejsp.2420030103

Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social Psychology*, 80(2), 253-267. doi:10.1037/0022-3514.80.2.253

Bodenhausen, G. V., & Lichtenstein, M. (1987). Social stereotypes and information-processing strategies: The impact of task complexity. *Journal of Personality and Social Psychology*, 52(5), 871-880. doi:10.1037/0022-3514.52.5.871

Bothwell, R. K., Pigott, M. A., Foley, L. A., & McFatter, R. M. (2006). Racial Bias in Juridic Judgment at Private and Public Levels. *Journal Of Applied Social Psychology*, 36(9), 2134-2149. doi:10.1111/j.0021-9029.2006.00008.x

Bourhis, R. Y., Sachdev, I., & Gagnon, A. (1994). Intergroup research with the Tajfel matrices: Methodological notes. In M. P. Zanna, J. M. Olson (Eds.). *The psychology of prejudice: The Ontario Symposium*, 7, 209-232.

Bowers, W.C. (1984). *Legal homicide: Death as punishment in America, 1864-1982*. Boston, MA: Northeastern University Press.

Brauer, M., & Bourhis, R. Y. (2006). Social power. *European Journal of Social Psychology*, 36(4), 601-616. doi:10.1002/ejsp.355

- Bucolo, D. O., & Cohn, E. S. (2010). Playing the race card: Making race salient in defence opening and closing statements. *Legal and Criminological Psychology*, 15(2), 293-303. doi:10.1348/135532508X400824
- Cacioppo, J. T., Petty, R. E., Kao, C., & Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology*, 51(5), 1032-1043. doi:10.1037/0022-3514.51.5.1032
- Carver, C. S., Glass, D. C., & Katz, I. (1978). Favorable evaluations of Blacks and the handicapped: Positive prejudice, unconscious denial, or social desirability?. *Journal of Applied Social Psychology*, 8(2), 97-106. doi:10.1111/j.1559-1816.1978.tb00768.x
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5), 752-766. doi:10.1037/0022-3514.39.5.752
- Clark, J., Boccaccini, M. T., Caillouet, B., & Chaplin, W. F. (2007). Five factor model personality traits, jury selection, and case outcomes in criminal and civil cases. *Criminal Justice and Behavior*, 34, 641-660. doi: 10.1037/10140-006.
- Clark, S. E., Hori, A., Putnam, A., & Martin, T. P. (2000). Group collaboration in recognition memory. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, 26(6), 1578-1588. doi:10.1037/0278-7393.26.6.1578
- Clark, N. K., Stephenson, G. M., & Kniveton, B. H. (1990). Social remembering: Quantitative aspects of individual and collaborative remembering by police officers and students. *British Journal of Psychology*, 81(1), 73-94. doi:10.1111/j.2044-8295.1990.tb02347.x

- Cohn, E. S., Bucolo, D., & Sommers, S. R. (2012). Race and racism. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 279-301). Washington, DC US: American Psychological Association. doi:10.1037/13085-013
- Commonwealth of Massachusetts vs. Jones*. 366 Mass. 805 (1975).
- Cornwell, E. Y., & Hans, V. P. (2011). Representation through participation: A multilevel analysis of jury deliberations. *Law & Society Review*, 45(3), 667-698.
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83(6), 1314-1329. doi:10.1037/0022-3514.83.6.1314
- Cowan, C. L., Thompson, W. C., & Ellsworth, P. C. (1984). The effects of death qualification on jurors' predisposition to convict and on the quality of deliberation. *Law and Human Behavior*, 8(1-2), 53-79. doi:10.1007/BF01044351
- De Dreu, C. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology*, 88(4), 741-749. doi:10.1037/0021-9010.88.4.741
- Devine, D. J., & Caughlin, D. E. (2014). Do they matter? A meta-analytic investigation of individual characteristics and guilt judgments. *Psychology, Public Policy, and Law*, 20(2), 109-134. doi:10.1037/law0000006
- DeSantis, A., & Kayson, W. A. (1997). Defendants' characteristics of attractiveness, race, and sex and sentencing decisions. *Psychological Reports*, 81(2), 679-683. doi:10.2466/PRO.81.6.679-683.

- Doerr, C., Plant, E. A., Kunstman, J. W., & Buck, D. (2011). Interactions in Black and White: Racial differences and similarities in response to interracial interactions. *Group Processes & Intergroup Relations*, 14(1), 31-43. doi:10.1177/1368430210375250
- Dovidio, J. F. (2001). On the nature of contemporary prejudice: The third wave. *Journal of Social Issues*, 57(4), 829-849. doi:10.1111/0022-4537.00244
- Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, 11(4), 315-319. doi:10.1111/1467-9280.00262
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, 33(5), 510-540. doi:10.1006/jesp.1997.1331
- Dovidio, J. F., Smith, J. K., Donnell, A., & Gaertner, S. L. (1997). Racial attitudes and the death penalty. *Journal of Applied Social Psychology*, 27(16), 1468-1487. doi:10.1111/j.1559-1816.1997.tb01609.x
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. (2006). Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes. *Psychological Science*, 17(5), 383-386. doi:10.1111/j.1467-9280.2006.01716.x
- Ellis, L. & Diamond, S. S. (2003). Race, Diversity, and Jury Composition: Battering and Bolstering Legitimacy, 78 *Chi.-Kent. L. Rev.* 1033.
- Espinoza, R. E., & Willis-Esqueda, C. (2008). Defendant and defense attorney characteristics and their effects on juror decision making and prejudice against Mexican Americans. *Cultural Diversity and Ethnic Minority Psychology*, 14(4), 364-371. doi:10.1037/a0012767

- Esqueda, C., Espinoza, R. E., & Culhane, S. E. (2008). The effects of ethnicity, SES, and crime status on juror decision making: A cross-cultural examination of European American and Mexican American mock jurors. *Hispanic Journal of Behavioral Sciences*, 30(2), 181-199. doi:10.1177/0739986308315319
- Fairchild, H. H., & Cowan, G. (1997). The O. J. Simpson trial: Challenges to science and society. *Journal of Social Issues*, 53(3), 583-591. doi:10.1111/0022-4537.00036
- Fitzgerald, R., & Ellsworth, P. C. (1984). Due process vs. crime control: Death qualification and jury attitudes. *Law and Human Behavior*, 8(1-2), 31-51. doi:10.1007/BF01044350
- Fleming, M., Petty, R., & White, P. (2005). Stigmatized targets and evaluation: Prejudice as a determinant of attribute scrutiny and polarization. *Personality and Social Psychology Bulletin*, 31(4), 496-507. doi:10.1177/0146167204271585
- Fiske, S. T., Cuddy, A. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902. doi:10.1037/0022-3514.82.6.878
- Foldy, E., Rivard, P., & Buckley, T. R. (2009). Power, safety, and learning in racially diverse groups. *Academy of Management Learning & Education*, 8(1), 25-41. doi:10.5465/AMLE.2009.37012177
- Foley, L. A., & Pigott, M. A. (2002). Race, self presentation and reverse discrimination in jury decisions. *American Journal of Forensic Psychology*, 20(3), 37-52. doi: 20022002-17734-003
- Foster v. Chatman*. 578 U.S. ____ (2016)
- ForsterLee, R., ForsterLee, L., Horowitz, I. A., & King, E. (2006). The Effects of Defendant

- Race, Victim Race, and Juror Gender on Evidence Processing in a Murder Trial.
Behavioral Sciences & the Law, 24(2), 179-198. doi:10.1002/bsl.675
- Gaertner, S. L., & Dovidio, J. F. (2005). Understanding and Addressing Contemporary Racism: From Aversive Racism to the common Ingroup Identity Model. *Journal of Social Issues*, 61(3), 615-639. doi:10.1111/j.1540-4560.2005.00424.x
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio, S. L. Gaertner (Eds.). *Prejudice, Discrimination, and Racism* (pp. 61-89). San Diego, CA US: Academic Press.
- Gaertner, S. L., & Dovidio, J. F. (1977). The subtlety of White racism, arousal, and helping behavior. *Journal of Personality and Social Psychology*, 35(10), 691-707.
doi:10.1037/0022-3514.35.10.691
- Galinsky, A. D., Magee, J. C., Gruenfeld, D. H., Whitson, J. A., & Liljenquist, K. A. (2008). Power reduces the press of the situation: Implications for creativity, conformity, and dissonance. *Journal of Personality and Social Psychology*, 95(6), 1450-1466.
doi:10.1037/a0012633
- Gilbert, G. M. (1951). Stereotype persistence and change among college students. *The Journal of Abnormal and Social Psychology*, 46(2), 245-254. doi:10.1037/h0053696
- Goff, P. A., Eberhardt, J. L., Williams, M. J., & Jackson, M. C. (2008). Not yet human: Implicit knowledge, historical dehumanization, and contemporary consequences. *Journal Of Personality and Social Psychology*, 94(2), 292-306. doi:10.1037/0022-3514.94.2.292
- Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: Stereotype threat and distance in interracial contexts. *Journal of Personality and Social Psychology*, 94(1), 91-107. doi:10.1037/0022-3514.94.1.91

- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4-27. doi:10.1037/0033-295X.102.1.4
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480. doi:10.1037/0022-3514.74.6.1464
- Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, 67(1), 1-15. doi:10.1006/obhd.1996.0061
- Harvey, S., & Bourhis, R. Y. (2012). Discrimination in wealth and power intergroup structures. *Group Processes & Intergroup Relations*, 15(1), 21-38. doi:10.1177/1368430211408364
- Hass, R., Katz, I., Rizzo, N., & Bailey, J. (1991). Cross-racial appraisal as related to attitude ambivalence and cognitive complexity. *Personality and Social Psychology Bulletin*, 17(1), 83-92. doi:10.1177/0146167291171013
- Hastie, R., Penrod, S., & Pennington, N. (1983). Inside the jury. Cambridge, MA: Harvard University Press. <http://dx.doi.org/10.4159/harvard.9780674865945>
- von Hippel, W., Silver, L. A., & Lynch, M. E. (2000). Stereotyping against your will: The role of inhibitory ability in stereotyping and prejudice among the elderly. *Personality and Social Psychology Bulletin*, 26(5), 523-532. doi:10.1177/0146167200267001
- Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., & ... Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO₇ scale. *Journal of Personality and Social Psychology*, 109(6), 1003-1028. doi:10.1037/pspi0000033

- Hodson, G., Hooper, H., Dovidio, J. F., & Gaertner, S. L. (2005). Aversive Racism in Britain: The Use of Inadmissible Evidence in Legal Decisions. *European Journal of Social Psychology*, 35(4), 437-448. doi:10.1002/ejsp.261
- Hunt, J. S. (2015). Race in the justice system. In B. L. Cutler, P. A. Zapf, B. L. Cutler, P. A. Zapf (Eds.), *APA handbook of forensic psychology, Vol. 2: Criminal investigation, adjudication, and sentencing outcomes* (pp. 125-161). Washington, DC, US: American Psychological Association. doi:10.1037/14462-005
- Innocence Project (2016). <http://www.innocenceproject.org/>. Downloaded July 25, 2016.
- Kalven, H., & Zeisel, H. (1966). *The American jury*. Boston: Little, Brown.
- Kimmelmeier, M. (2005). The Effects of Race and Social Dominance Orientation in Simulated Juror Decision Making. *Journal of Applied Social Psychology*, 35(5), 1030-1045. doi:10.1111/j.1559-1816.2005.tb02158.x
- Kerwin, J., & Shaffer, D. R. (1994). Mock jurors versus mock juries: The role of deliberations in reactions to inadmissible testimony. *Personality and Social Psychology Bulletin*, 20(2), 153-162. doi:10.1177/0146167294202002
- Kochan, T., Bezrukova, K., Ely, R., Jackson, S., Joshi, A., Jehn, K., & Thomas, D. (2003). The effects of diversity on business performance: Report of the Diversity Research Network. *Human Resource Management*, 42(1), 3-21. doi:10.1002/hrm.1006
- Kramer, G. P., Kerr, N. L., & Carroll, J. S. (1990). Pretrial publicity, judicial remedies, and jury bias. *Law and Human Behavior*, 14(5), 409-438. doi:10.1007/BF01044220
- Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science*, 5(5), 289-296. doi:10.1111/j.1467-9280.1994.tb00628.x

- Latane, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, 10(3), 215-221. doi:10.1037/h0026570
- Lecci, L., & Myers, B. (2008). Individual differences in attitudes relevant to juror decision making: Development and validation of the Pretrial Juror Attitude Questionnaire (PJAQ). *Journal of Applied Social Psychology*, 38(8), 2010-2038. doi:10.1111/j.1559-1816.2008.00378.x
- Linville, P. W., & Jones, E. E. (1980). Polarized appraisals of outgroup members. *Journal of Personality and Social Psychology*, 38(5), 689-703. doi:10.1037/0022-3514.38.5.689
- Lynch, M., & Haney, C. (2009). Capital jury deliberation: Effects on death sentencing, comprehension, and discrimination. *Law and Human Behavior*, 33(6), 481-496. doi:10.1007/s10979-008-9168-2
- Lynch, M., & Haney, C. (2011). Mapping the racial bias of the white male capital juror: Jury composition and the “empathic divide”. *Law & Society Review*, 45(1), 69-102. doi:10.1111/j.1540-5893.2011.00428.x
- Kerr, N.L. & MacCoun, R.J. (2012). Is the leniency asymmetry really dead? Misinterpreting asymmetry effects in criminal jury deliberation. *Group Processes and Intergroup Relations*, 15(5), 585-602. doi: 10.1177/1368430212441639
- van Knippenberg, D., & Schippers, M. C. (2007). Work Group Diversity. *Annual Review of Psychology*, 58, 515-541. doi:10.1146/annurev.psych.58.110405.085546
- Mallett, R. K., Wilson, T. D., & Gilbert, D. T. (2008). Expect the unexpected: Failure to anticipate similarities leads to an intergroup forecasting error. *Journal of Personality and Social Psychology*, 94(2), 265-277. doi:10.1037/0022-3514.94.2.265
- 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(15), 1315-1344. doi:10.1111/j.1559-1816.1994.tb01552.x
- McLeod, P., Lobel, S., & Cox, T. H. (1996). Ethnic diversity and creativity in small groups. *Small Group Research*, 27(2), 248-264. doi:10.1177/1046496496272003
- Melilli, K. (1996). Batson in Practice: What we have learned about Batson and peremptory challenges. *Notre Dame Law Review*, 71, 1996.
- Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interaction with white and black men. *Personality and Social Psychology Bulletin*, 28(7), 939-952. doi:10.1177/01467202028007007
- Mendes, W. B., Blascovich, J., Major, B., & Seery, M. (2001). Challenge and threat responses during downward and upward social comparisons. *European Journal of Social Psychology*, 31(5), 477-497. doi:10.1002/ejsp.80
- Mendoza-Denton, R., Downey, G., Purdie, V. J., Davis, A., & Pietrzak, J. (2002). Sensitivity to status-based rejection: Implications for African American students' college experience. *Journal of Personality and Social Psychology*, 83(4), 896-918. doi:10.1037/0022-3514.83.4.896
- Miller-El v. Dretke*, 545 U.S. 231 (2005).
- 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(6), 621-637. doi:10.1007/s10979-005-8122-9

- Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126(2), 247-259.
doi:10.1037/0033-2909.126.2.247
- Mustard, D.B. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the U.S. federal courts. *Journal of Law and Economics*, 64, 285-314. doi: 10.1086/320276
- National Registry of Exonerations (2016). Retrieved August 4, 2016, from,
<http://www.law.umich.edu/special/exoneration/Pages/ExonerationsRaceByCrime.aspx>
- Navarrete, C., McDonald, M. M., Asher, B. D., Kerr, N. L., Yokota, K., Olsson, A., & Sidanius, J. (2012). Fear is readily associated with an outgroup face in a minimal group context. *Evolution and Human Behavior*, 33(5), 590-593. doi:10.1016/j.evolhumbehav.2012.02.007
- Neal v. Delaware*, 103 U.S. 370 (1880).
- Nemeth, C. J. (1986). Differential contributions of majority and minority influence. *Psychological Review*, 93(1), 23-32. doi:10.1037/0033-295X.93.1.23
- Norris v. Alabama*, 294 U.S. 587 (1935).
- Norton, M. I., Sommers, S. R., Apfelbaum, E. P., Pura, N., & Ariely, D. (2006). Color Blindness and Interracial Interaction: Playing the Political Correctness Game. *Psychological Science*, 17(11), 949-953. doi:10.1111/j.1467-9280.2006.01810.x
- Padilla, J. B., Miller, M. K., & Broadus, A. D. (2008). Analysis of Hispanic representation and conceptualization in psychology and law research. *Behavioral Sciences & the Law*, 26(5), 655-670. doi:10.1002/bsl.818
- Patry, M. W. (2008). Attractive but guilty: Deliberation and the physical attractiveness bias. *Psychological Reports*, 102(3), 727-733. doi:10.2466/PRO.102.3.727-733
- Payne, B. (2001). Prejudice and perception: The role of automatic and controlled processes in

- misperceiving a weapon. *Journal of Personality and Social Psychology*, 81(2), 181-192.
doi:10.1037/0022-3514.81.2.181
- Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary prejudice: Insights from aversive racism. *Social and Personality Psychology Compass*, 3(3), 314-338. doi:10.1111/j.1751-9004.2009.00183.x
- Peters v. Kiff*, 407 U.S. 493 (1972)
- Petty, R. E., & Cacioppo, J. T. (1979). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology*, 37(10), 1915-1926. doi:10.1037/0022-3514.37.10.1915
- Petty, R., Cacioppo, J., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, 41(5), 847-855. doi:10.1037/0022-3514.41.5.847
- Petty, R., Fleming, M., & White, P. (1999). Stigmatized sources and persuasion: Prejudice as a determinant of argument scrutiny. *Journal of Personality and Social Psychology*, 76(1), 19-34. doi:10.1037/0022-3514.76.1.19
- Pfeifer, J. E., & Ogloff, J. R. (1991). Ambiguity and guilt determinations: A modern racism perspective. *Journal of Applied Social Psychology*, 21(21), 1713-1725.
doi:10.1111/j.1559-1816.1991.tb00500.x
- Phillips, K. W., Mannix, E. A., Neale, M. A., & Gruenfeld, D. H. (2004). Diverse groups and information sharing: The effects of congruent ties. *Journal of Experimental Social Psychology*, 40(4), 497-510. doi:10.1016/j.jesp.2003.10.003
- Pritchard, M. E., & Keenan, J. M. (2002). Does jury deliberation really improve jurors' memories? *Applied Cognitive Psychology*, 16(5), 589-601. doi:10.1002/acp.816

- Richeson, J. A., & Shelton, J. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, 14(3), 287-290. doi:10.1111/1467-9280.03437
- Richeson, J. A., & Trawalter, S. (2005). Why Do Interracial Interactions Impair Executive Function? A Resource Depletion Account. *Journal of Personality and Social Psychology*, 88(6), 934-947. doi:10.1037/0022-3514.88.6.934
- Richeson, J. A., Trawalter, S., & Shelton, J. (2005). African Americans' Implicit Racial Attitudes and the Depletion of Executive Function after Interracial Interactions. *Social Cognition*, 23(4), 336-352. doi:10.1521/soco.2005.23.4.336
- Rose, M. R. (1999). The peremptory challenge accused of race or gender discrimination? Some data from one county. *Law and Human Behavior*, 23(6), 695-702. doi:10.1023/A:1022393506784
- Rose, M. R., Ellison, C., & Diamond, S. S. (2008). Preferences for juries over judges across racial and ethnic groups. *Social Science Quarterly*, 89(2), 372-391. doi:10.1111/j.1540-6237.2008.00537.x
- Ruva, C. L., & Guenther, C. C. (2015). From the shadows into the light: How pretrial publicity and deliberation affect mock jurors' decisions, impressions, and memory. *Law and Human Behavior*, 39(3), 294-310. doi:10.1037/lhb0000117
- Ruva, C., McEvoy, C., & Bryant, J. (2007). Effects of Pre-Trial Publicity and Jury Deliberation on Juror Bias and Source Memory Errors. *Applied Cognitive Psychology*, 21(1), 45-67. doi:10.1002/acp.1254
- Sargent, M. J., & Bradfield, A. L. (2004). Race and Information Processing in Criminal Trials:

- Does the Defendant's Race Affect How the Facts Are Evaluated? *Personality and Social Psychology Bulletin*, 30(8), 995-1008. doi:10.1177/0146167204265741
- Saucier, D. A., Miller, C. T., & Doucet, N. (2005). Differences in Helping Whites and Blacks: A Meta-Analysis. *Personality and Social Psychology Review*, 9(1), 2-16.
doi:10.1207/s15327957pspr0901_1
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440-452.
doi:10.1037/0022-3514.85.3.440
- 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(4), 320-328.
doi:10.1007/s10979-008-9153-9
- Shelton, J. N., & Richeson, J. A. (2015). Interacting across racial lines. In M. Mikulincer, P. R. Shaver, J. F. Dovidio, J. A. Simpson, M. Mikulincer, P. R. Shaver, J. A. Simpson (Eds.). *APA handbook of personality and social psychology, Volume 2: Group processes* (pp. 395-422). Washington, DC, US: American Psychological Association.
doi:10.1037/14342-015
- Shelton, J. N., Richeson, J. A., & Vorauer, J. D. (2006). Threatened identities and interethnic interactions. *European Review Of Social Psychology*, 17321-358.
doi:10.1080/10463280601095240
- Sherif, M. (1966). *The psychology of social norms*. Oxford England: Harper Torchbooks.
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY, US: Cambridge University Press.
- Snyder v. Louisiana*, 552 U.S. 472 (2008).

- Sommers, S. R. (2006). On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology*, 90(4), 597-612. doi:10.1037/0022-3514.90.4.597
- Sommers, S. R. (2007). Race and the decision making of juries. *Legal and Criminological Psychology*, 12(2), 171-187. doi:10.1348/135532507X189687
- Sommers, S. R., & Ellsworth, P. C. (2009). 'Race salience' in juror decision-making: Misconceptions, clarifications, and unanswered questions. *Behavioral Sciences & the Law*, 27(4), 599-609. doi:10.1002/bsl.877
- Sommers, S. R., & Ellsworth, P. C. (2001). White juror bias: An investigation of prejudice against Black defendants in the American courtroom. *Psychology, Public Policy, and Law*, 7(1), 201-229. doi:10.1037/1076-8971.7.1.201
- Sommers, S. R., & Ellsworth, P. C. (2000). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin*, 26(11), 1367-1379. doi:10.1177/0146167200263005
- Sommers, S. R., & Norton, M. I. (2007). Race-based judgments, race-neutral justifications: Experimental examination of peremptory use and the Batson Challenge procedure. *Law and Human Behavior*, 31(3), 261-273. doi:10.1007/s10979-006-9048-6
- Sommers, S. R., Warp, L. S., & Mahoney, C. C. (2008). Cognitive effects of racial diversity: White individuals' information processing in heterogeneous groups. *Journal of Experimental Social Psychology*, 44(4), 1129-1136. doi:10.1016/j.jesp.2008.01.003
- Sorensen, J., Hope, R., & Stemen, D. (2003). Racial disproportionality in state prison admissions: Can regional variation be explained by differential arrest rates? *Journal of Criminal Justice*, 31(1), 73-84. doi: 10.1016/S0047-2352(02)00200-3

- Stasser, G. (1988). Computer simulation as a research tool: The DISCUSS model of group decision making. *Journal of Experimental Social Psychology*, 24, 393-422. doi: 10.1016/0022-1031(88)90028-5
- Stasser, G., Stewart, D. D., & Wittenbaum, G. M. (1995). Expert roles and information exchange during discussion: The importance of knowing who knows what. *Journal of Experimental Social Psychology*, 31(3), 244-265. doi:10.1006/jesp.1995.1012
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811. doi:10.1037/0022-3514.69.5.797
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.) *Advances in Experimental Social Psychology*, 34 (379-440). San Diego, CA US: Academic Press.
- Stephenson, G. M., Clark, N. K., & Wade, G. S. (1986). Meetings make evidence? An experimental study of collaborative and individual recall of a simulated police interrogation. *Journal Of Personality And Social Psychology*, 50(6), 1113-1122. doi:10.1037/0022-3514.50.6.1113
- Stone, C. B., Barnier, A. J., Sutton, J., & Hirst, W. (2010). Building consensus about the past: Schema consistency and convergence in socially shared retrieval-induced forgetting. *Memory*, 18, 170 -184. doi: 10.1080/09658210903159003
- Strauder v. West Virginia*, 100 U.S. 303 (1879).
- Swain v. Alabama*, 380 U.S. 202 (1965).
- Sweeney, L. T., & Haney, C. (1992). The influence of race on sentencing: A meta-analytic

- review of experimental studies. *Behavioral Sciences & the Law*, 10(2), 179-195.
doi:10.1002/bsl.2370100204
- Tajfel, H., & Billig, M. (1974). Familiarity and categorization in intergroup behavior. *Journal of Experimental Social Psychology*, 10(2), 159-170. doi:10.1016/0022-1031(74)90064-X
- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2), 149-178.
doi:10.1002/ejsp.2420010202
- Trawalter, S., Richeson, J. A., & Shelton, J. N. (2009). Predicting behavior during interracial interactions: A stress and coping approach. *Personality and Social Psychology Review*, 13(4), 243-268. doi:10.1177/1088868309345850
- Turner, J. C., Brown, R. J., & Tajfel, H. (1979). Social comparison and group interest in ingroup favouritism. *European Journal of Social Psychology*, 9(2), 187-204.
doi:10.1002/ejsp.2420090207
- Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science*, 314(5802), 1154-1156. doi:10.1126/science.1132491
- Vollrath, D. A., Sheppard, B. H., Hinsz, V. B., & Davis, J. H. (1989). Memory performance by decision-making groups and individuals. *Organizational Behavior And Human Decision Processes*, 43(3), 289-300. doi:10.1016/0749-5978(89)90040-X
- Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal*, 36(3), 590-602. doi:10.2307/256593

- Webber, S., & Donahue, L. M. (2001). Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis. *Journal of Management*, 27(2), 141-162. doi:10.1016/S0149-2063(00)00093-3
- Weitzer, R., & Tuch, S. A. (2005). Racially Biased Policing: Determinants of Citizen Perceptions. *Social Forces*, 83(3), 1009-1030. doi:10.1353/sof.2005.0050
- Weitzer, R., & Tuch, S. A. (2004). Race and Perceptions of Police Misconduct. *Social Problems*, 51(3), 305-325. doi:10.1525/sp.2004.51.3.305
- Weitzer, R., & Tuch, S. A. (1999). Race, class, and perceptions of discrimination by the police. *Crime & Delinquency*, 45(4), 494-507. doi:10.1177/0011128799045004006
- Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, 23(5), 1160-1175. doi:10.1037/0278-7393.23.5.1160
- Weldon, M. S., Blair, C., & Huebsch, P. D. (2000). Group remembering: Does social loafing underlie collaborative inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1568-1577. doi: 10.1037/0278-7393.26.6.1568
- White, P., & Harkins, S. (1994). Race of source effects in the Elaboration Likelihood Model. *Journal of Personality and Social Psychology*, 67(5), 790-807. doi:10.1037/0022-3514.67.5.790
- Williams, M. R., Demuth, S., & Holcomb, J. E. (2007). Understanding the influence of victim gender in death penalty cases: The importance of victim race, sex-related victimization, and jury decision making. *Criminology: An Interdisciplinary Journal*, 45(4), 865-891. doi:10.1111/j.1745-9125.2007.00095.x

- Yarmey, A. D. & Morris, S. (1998). The effects of discussion on eyewitness memory. *Journal of Applied Social Psychology*, 28(17), 1637-1648. doi:10.1111/j.1559-1816.1998.tb01693.x
- York, E. & Cornwell, B. (2006). Status on trial: Social characteristics and influence in the jury room. *Social Forces*, 85(1), 455-477.