

# UC Berkeley

## UC Berkeley Electronic Theses and Dissertations

### Title

The Promise of Accountability: Countering Racial Bias in Decision-Making

### Permalink

<https://escholarship.org/uc/item/1zr709pp>

### Author

Carbone, Christina Stevens

### Publication Date

2014

Peer reviewed|Thesis/dissertation

**The Promise of Accountability:  
Countering Racial Bias in Decision-Making**

By

Christina Stevens Carbone

A dissertation submitted in partial satisfaction of the  
requirements for the degree of  
Doctor of Philosophy  
in  
Jurisprudence and Social Policy  
in the  
Graduate Division  
of the  
University of California, Berkeley

Committee in charge:

Professor Victoria C. Plaut, Chair  
Professor Catherine Albiston  
Professor Jack Glaser  
Professor David Sklansky

Fall 2014

Copyright © 2014 Christina Stevens Carbone  
All Rights Reserved

## Abstract

### The Promise of Accountability: Countering Racial Bias in Decision-Making

by

Christina Stevens Carbone

Doctor of Philosophy in Jurisprudence & Social Policy

University of California, Berkeley

Professor Victoria Plaut, Chair

Evidence of racial disparities across a number of important domains, combined with documentation of persisting prejudice and stereotypes disfavoring racial minorities, suggests that decision processes are not immune from the influence of racial bias. The operation of such bias threatens public values of fairness and equality, particularly in contexts where matters of life and liberty are at stake, such as the criminal justice system. Accountability—the requirement of having to justify oneself through the giving of reasons—is a procedural device commonly used by organizations of all kinds (e.g., police departments, prosecutor offices, courts) to improve the quality of decisions.

While scholars and organizations turn to accountability as a strategy to reduce racial bias, the existing evidence justifying this reliance is tenuous at best. As my review of the literature shows, whereas previous research suggests accountability can effectively address a range of cognitive biases under the right conditions, studies focusing on the effects of accountability on intergroup biases have produced mixed results. In some cases, accountability can operate to actually bolster intergroup biases. Further, no known study has directly examined the effect of accountability on implicit (unconscious) racial bias. To avoid the adoption of ineffective or potentially harmful practices, further research is needed on accountability’s ability to reduce the influence of racial bias in the decision process.

The current project presents results from three experimental studies testing accountability’s capacity to attenuate the effects of explicit and implicit racial bias. Using a case file paradigm, online participants reviewed either a White or racial minority criminal suspect and rendered judgments about the seriousness of the crime, suspect guilt, and the appropriate punishment. Study 1 compared the relative effects of holding decision-makers accountable either for the decision outcome itself (outcome accountability) or for the process used to reach the decision (process accountability). In this study, college students reviewed either a White or Black suspect in a drug possession case. Only limited evidence was found for accountability’s ability to address racial bias, with the process accountability manipulation faring marginally better than outcome accountability.

Focusing on different aspects of the justification process, Study 2 examined the effects of having decision-makers respond to a general prompt to provide reasons for their judgments (undirected accountability) versus responding to a series of targeted questions designed to focus attention on the criteria used and the weighing of factors within the decision process (directed accountability). This latter form of accountability more closely simulates filling out a

standardized form, which many organizations use when evaluating cases in a variety of contexts, such as hiring. In this study, a national sample of adults recruited from Amazon Mechanical Turk reviewed either a White or Black suspect across three different criminal files: drug possession, vandalism, and robbery. Contrary to expectations, those in the control condition did not show the baseline pattern of racial bias. While control participants tended to give harsher outcomes to the White suspect, this same pattern was not found in the two accountability conditions. Limited evidence of racial bias was found in each of the accountability conditions, at least for one of the case files.

Study 3 also examined undirected versus directed accountability, with MTurk workers reviewing either a White or Hispanic suspect across two case files: vandalism and battery. In addition to the measures of explicit racial bias included in the first two studies, measures of participants' implicit racial bias were also examined. While racial disparities in outcomes did not emerge for any of the conditions, the control and two accountability conditions all showed some evidence that suspect race was related to their decisions, at least some of the time.

Keeping in mind the limitations stemming from the experiment-based nature of the project, the collective findings from these studies caution against conclusions that accountability effectively addresses the influence of racial bias in the decision-making process.

## **DEDICATION**

To my loving family, who has supported me throughout my journey to this point.

To my husband, my friend and partner in life.

To my puppies, who bring joy to my heart and provided me with companionship while writing this dissertation.

## **ACKNOWLEDGEMENTS**

I am most grateful to my advisor, Victoria Plaut, for her crucial encouragement, her insightful advice, and her wonderful example of how to be an exceptional scholar, mentor, and person. I would like to thank the other members of my dissertation committee, Catherine Albiston, David Sklansky, and Jack Glaser, for their valuable feedback and support for this project. I am grateful for the financial and academic contributions to this project given by the Culture, Diversity, & Intergroup Relations Lab at Berkeley Law, Berkeley's Center for Race & Gender, and the Department of Jurisprudence and Social Policy. I would also like to acknowledge my fellow doctoral students who have impressed, inspired, and befriended me throughout my time in the JSP program.

## TABLE OF CONTENTS

<b>CHAPTER 1: WHY ACCOUNTABILITY?</b> .....	1
What is Accountability? .....	1
Accountability within the Legal System .....	3
Accountability as a Bias-reducing Strategy .....	5
A Cautionary Tale .....	8
The Current Project .....	10
<b>CHAPTER 2: LITERATURE REVIEW</b> .....	12
The Bias-Reducing Effect of Accountability .....	12
The Operation of Intergroup Biases .....	15
Accountability Studies on Intergroup Bias .....	17
Accountability's Mechanism .....	20
Unanswered Questions: The Need for Further Research .....	23
<b>CHAPTER 3: STUDY 1</b> .....	31
Research Questions and Hypotheses .....	32
Methods .....	33
Results .....	37
Summary and Discussion .....	44
<b>CHAPTER 4: STUDY 2</b> .....	46
Research Questions and Hypotheses .....	47
Methods .....	47
Results .....	51
Drug Case .....	51
Vandalism Case .....	56
Robbery Case.....	58
Summary and Discussion .....	62
<b>CHAPTER 5: STUDY 3</b> .....	64
Research Questions and Hypotheses .....	64
Methods .....	65
Results .....	68
Vandalism Case .....	69
Battery Case.....	74



Summary and Discussion .....	81
<b>CHAPTER 6: GENERAL DISCUSSION &amp; CONCLUSION .....</b>	<b>84</b>
Overview of Studies 1, 2, & 3 .....	84
Limitations .....	86
Conclusion.....	88
<b>REFERENCES.....</b>	<b>89</b>
<b>APPENDIX A: CASE FILE NARRATIVES.....</b>	<b>100</b>
<b>APPENDIX B: LEGAL INSTRUCTIONS .....</b>	<b>103</b>
<b>APPENDIX C: MEANS AND STANDARD DEVIATION TABLES .....</b>	<b>105</b>

## CHAPTER 1: WHY ACCOUNTABILITY?

In 2013, the San Jose Police Department in California began phasing in a new “curb sitting” policy in response to community concerns about possible racial profiling during stops and detentions (Salonga, 2013). Aimed at increasing transparency and accountability, the new policy requires officers to document data about each vehicle and pedestrian stop whenever there is a temporary detention involving handcuffing or making someone sit on a street curb or in a police car. Such data includes the race of the suspect, the reason for the stop, the type of detention, the reason for the detention, whether a search was conducted, and the result of the stop. San Jose is just one of many police departments across the country working to implement similar data-gathering and accountability policies (Center for Policing Equity, 2013). The response by the San Jose Police Department to allegations of racial profiling is also typical whenever systematic problems are identified—greater accountability. Indeed, “accountability” often gets invoked as a solution to all sorts of political, economic, legal, and social crises. Yet, inquiries into the effectiveness of accountability have been limited in important ways and have not yielded consistent results. This is particularly true for accountability strategies aimed at addressing instances of racial bias in decision-making.

At its core, this project examines the extent to which accountability attenuates the influence of racial bias in decision-making. While this project is predominantly empirical, that is, it seeks to understand the psychological phenomena of accountability as it relates to racial bias, it is also motivated by a set of normative concerns and values. Indeed, empirical inquiries are almost always embedded within a normative framework which provides a justification for why the particular empirical question is worthy of our time, attention, and resources. The purpose of this chapter is to articulate the normative reasons behind undertaking this project and why legal scholars and practitioners should care about accountability and its effects on racial bias.

The thrust of my argument is that the legal system already relies on accountability as a tool to improve decision-making across a number of contexts, and, therefore, we ought to evaluate whether such a tool actually serves its intended purpose. As illustrated through the opening example, police departments as well as other legal organizations (e.g., courts, prosecutor offices, probation departments) have or will adopt some form of accountability process with an aim to improve decision making. Accountability is a politically attractive strategy that has been offered to address persisting social inequalities stemming from individual-level biases, including implicit or unconscious bias (Kang, 2008; Bartlett, 2009; Mitchell, 2009; Heilman & Haynes, 2008). The literature on symbolic structures (Edelman, 1992; Edelman, Erlanger, & Lande, 1993) and “false consciousness” critiques of procedural justice (MacCoun, 2005), however, send a cautionary note for implementing strategies that appear at face-value to obtain fair outcomes but have not been sufficiently tested in terms of producing substantive change. Thus, careful attention should be given to the empirical evidence of accountability’s effectiveness in reducing racial bias, particularly since some studies have shown that accountability can actually *bolster* cognitive biases (e.g., Lambert et al., 1996; Gorden, Roxelle, & Baxter, 1988).

### What is Accountability?

Before elaborating further on these arguments, attention must first be given to what is meant by “accountability.” Indeed, scholars have identified multiple, related faces of the concept of accountability. When the criminal offender is held accountable for his crimes by facing

punishment, accountability operates in the sense of settling accounts, that is, of getting what one deserves or has coming to him. When corporations are held accountable to their shareholders or when the government is held accountable to the general public, accountability stands in for the concept of transparency. When individuals are held accountable within organizational contexts, accountability can be understood as oversight, reporting on progress, or tying performance to particular desired outcomes (Green & Kalev, 2008; Kalev, Dobbin, & Kelly, 2006.) Depending on the context, then, accountability can be synonymous with the concepts of liability, answerability, responsibility, responsiveness, obligation, obedience, fidelity, and amenability (Dubnick & Justice, 2004). Likewise, accountability can function as a mechanism to fulfill a range of different goals, including promoting democracy, justice, and ethical behavior, enhancing performance, and engendering trust or legitimacy (see Dubnick, 2005).

Within the current project, accountability is understood as answerability or "the implicit or explicit expectation that one may be called on to justify one's beliefs, feelings, and actions to others" (Lerner & Tetlock, 1999, p. 255).<sup>1</sup> I chose to examine this particular definition of accountability because it focuses directly on the decision process and various instantiations of this reason-giving function can be found across several different domains and types of organizations. This definition makes a subtle and important distinction that implicates accountability's intended function. At first glance, this form of accountability is closely tied to the concept of transparency in the sense that one is anticipating revealing, or making transparent, the criteria and analysis underlying one's judgment. The set of proffered reasons can be evaluated by a third party as either adequate or not and, if not, appropriate remedies can be activated. Accountability can shed light on abuses of discretion in the same way that, in the words of Justice Louis Brandeis, "[s]unlight is said to be the best of disinfectants; electric light the most efficient policeman" (Brandeis, 1914). But while the purpose of transparency is to allow for meaningful review by another, the psychological literature on accountability flips the focus around to examine the effect of accountability on the decision-maker himself. The function of interest, then, is how accountability can affect the performance of the decision-maker prior to the time of actual review. From this approach, the alleged psychological power stems from the *mere anticipation* of being held accountable.

A final preliminary point about accountability is that in most contexts it is best conceptualized as a recurring system or cycle. As depicted in Figure 1, a system of accountability is made up of several different phases (Schlenker & Weigold, 1989). Keeping these phases in mind is important because factors at play in each phase can potentially impact the effectiveness of accountability in reducing bias. The first phase, which will receive extensive discussion in the next chapter, focuses on the characteristics of the particular decision task as well as the psychological factors at play during the decision process. Task characteristics that may affect the decision-making process include the clarity of task prescriptions, task complexity, the amount of cognitive and time resources available to carry out the task, and the degree of discretion allowed to the decision-maker. The relevant psychological factors include presumed audience expectations, knowledge of institutional or task norms, the familiarity or expertise the decision-maker has with the particular task, and internal and external motivations towards task

---

<sup>1</sup> While this definition focuses on justification to others, some scholars have argued for a broader understanding of accountability that includes being answerable to oneself (Schlenker & Weigold, 1989). However, almost all the psychological studies examining accountability have focused on accountability to others, where the 'other' is the experimenter, someone with knowledge in the particular field, or an unspecified person.

performance. All of these factors feed into a decision process that culminates in phase 2 of accountability, the rendering of a final judgment or choice.

The decision-maker is then faced in phase 3 with having to justify his or her choices, a process that can take on different forms depending on the parameters of the specific task. Phase 4 consists of a third party reviewing the decision and justifications. Here, accountability's effectiveness may depend on the authority or legitimacy of the reviewer, the frequency with which review takes places, and the degree of scrutiny undertaken by the reviewer. Finally, phase 5 consists of feedback provided by the reviewer to the decision-maker and any tangible or intangible consequences (including rewards and punishments) that may be imposed. A factor potentially at play during phase 5 is the degree to which the feedback and consequences are made public or kept private.

Oftentimes, the cycle of accountability may not go through all five phases, but rather go directly from justification for one decision (phase 3) back to a new decision task (phase 1). This may be particularly true for decision-makers who routinely make a large number of decisions as part of their daily jobs (e.g., police, prosecutors, judges, managers, medical professionals, and other frontline workers). Review and feedback on every single decision may not be practically feasible from an organizational and resource perspective. Thus, phases 4 and 5 may occur at an aggregate level or only for a subset of decisions, yet another factor that may affect accountability's effectiveness.

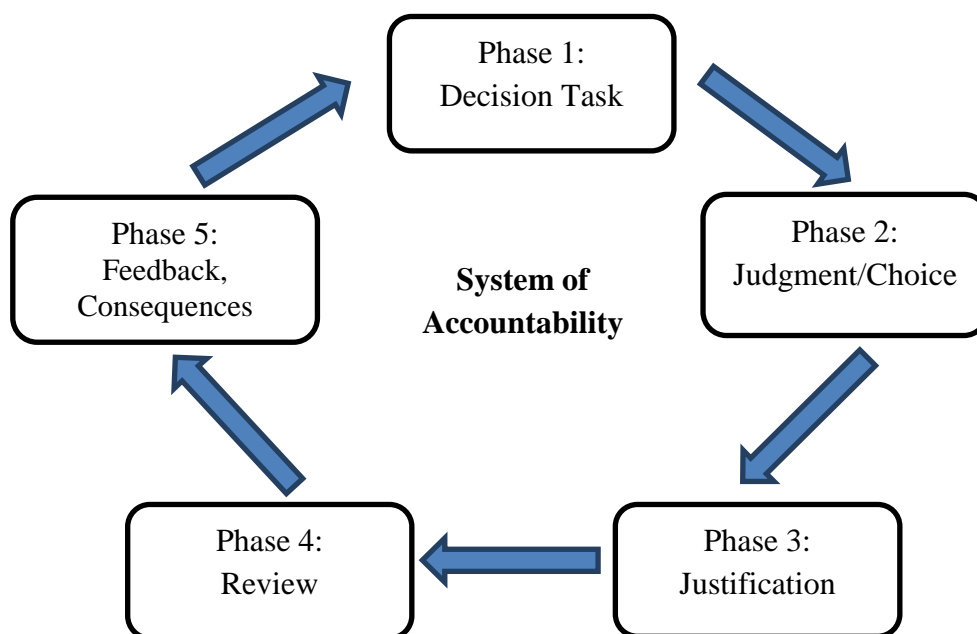


Figure 1: Phases of Accountability.

### Accountability within the Legal System

The extent to which the law incorporates accountability into its logic and practice reflects accountability's perceived importance as a decision-enhancing tool. Most frequently, courts and legislatures have turned to accountability as a strategy to counter or place a check on the exercise

of discretion. On the one hand, discretion is necessary to remain flexible in novel or unanticipated circumstances, to allow individualization of treatment when justice requires it, to resolve conflicts between rules, and to fill in gaps remaining from articulated rules. On the other hand, discretion permits the use of illegitimate criteria and the inconsistent application of criteria, and is a form of power that is susceptible to corruption (Hawkins, 1992). Discretion, then, functions as a double-edged sword because the very qualities that enable positive outcomes in terms of flexibility and responsiveness are the same qualities that create opportunities for unfair differential treatment (Gelsthorpe & Padfield, 2003; Hawkins, 1992). Even when formal legal rules apparently constrain the decision-making process, they often take the form of general standards which inherently require some degree of subjective judgment, for example, “reasonable articulable suspicion” for police detentions and “probable cause” to arrest (*Terry v. Ohio* (1968); *Draper v. United States* (1959)).

Drawing on accountability to reign in this discretion, courts have exercised their rule-making powers to require decision-makers to give a statement of reasons for their decisions in a number of contexts. The United States Supreme Court, for example, created the requirement that juvenile courts must provide a statement of reasons when deciding to waive jurisdiction over a juvenile (*U.S. v. Kent*, 1966), and that administrative agencies must state the grounds for their actions (*Securities and Exchange Commission v. Chenery Corp.*, 1943). The California courts have followed suit by requiring: administrative agencies to set forth the basis for their findings (*Topanga Assn. for a Scenic Community v. County of Los Angeles*, 1974); judges to explain the reasons when a sentence diverges from the determinate sentencing law (*People v. Martin*, 1986); judges to give a statement of reasons when denying bail pending appeal (*In re Podesto*, 1976); the Adult Authority to support all its denials of parole with a written, definitive statement of its reasons (*In re Sturm*, 1974); the Narcotic Addict Evaluation Authority to provide a written statement of reasons for revoking outpatient status for those in the civil addict program (*In re Bye*, 1974); and judges to expressly state their reasons when exercising discretion to impose a consecutive, as opposed to concurrent, sentence (*In re Spears*, 1984). Legislatures, too, have imposed accountability conditions through statute by, for example, requiring judges to state on the record the reasons for requiring lifetime sex offender registration (California Penal Code section 290.006).

Most often, courts adopt these accountability mechanisms in order to ensure meaningful judicial review. But the rationale goes beyond a transparency function. Courts have imposed these accountability measures because "a requirement of articulated reasons acts as an inherent guard against the careless decision, insuring that the judge himself analyzes the problem and recognizes the grounds for his decision" (*In re Podesto*, 1976, p. 937). Thus, accountability is intended to improve decision-making performance: "Requiring a written statement of reasons imposes an intellectual discipline that may lead to better reasoned decisions" (*People v. Martin*, 1986, p. 450).

Within the criminal justice context, some legal doctrines specifically concerned with the operation of bias incorporate the structure of accountability. One example is found in the Supreme Court's standard for stop-and-frisk decisions by police officers as articulated in *Terry v. Ohio* (1968). In order to justify a temporary detention, the officer must be able "to point to specific and articulable facts" that lead a reasonable officer to believe that a crime has, is, or will occur. On its face, *Terry* holds police officers accountable by requiring they articulate the specific reasons for making the stop. Presumably, an expectation of being held accountable in this way leads officers to evaluate “suspicious behavior” more carefully and cautiously, thus

avoiding unwarranted stops based on hunches or bias. This standard of "reasonable articulable suspicion," however, has received many criticisms, some of which reflect potential problems with accountability more generally. Because a court reviewing a detention decision under *Terry* uses an objective standard, the officer's subjective intent at the time is irrelevant. Thus, an officer can generate additional reasons post-hoc to justify his actions, even if these factors were not necessarily motivating the officer's behavior. Richman (1998) goes a step further to argue that at times the subsequent articulation of the grounds for the stop and frisk comes not from the officer, but from the prosecutor who argues the permissibility of the detention in court. This possibility for post-hoc rationalization is one reason to think critically about the structure of accountability systems and examine the contexts under which this kind of response is likely to occur.

A second example where accountability is embedded within legal doctrine is the burden-shifting scheme triggered by a challenge under *Batson v. Kentucky* (1986), which enforces the Fourteenth Amendment's prohibition of striking a juror from the panel on the basis of his race, ethnicity, gender, or religion. The challenging party must first make out a prima facie case for discrimination. The burden then shifts to the opposing party who must articulate his reasons for excusing the potential juror from the panel. A *Batson* challenge will be sustained if the court determines the proffered reason is merely pretextual and the attorney acted with purposeful discrimination. This doctrinal scheme presumes that instances of purposeful discrimination can be unearthed by a having a judge determine the credibility and legitimacy of reasons for the strike proffered by the attorney. Though several critics have challenged the efficacy of *Batson* as an enforcement mechanism (Page, 2005), that is, the aspect of accountability focused on meaningful review, little research has addressed the flip side of accountability by examining whether the specter of a *Batson* challenge alters the behavior or decision process of attorneys who may be self-monitoring to avoid a challenge from being raised. On the one hand, the anticipation of having to provide a neutral reason and having that reason scrutinized by an authoritative third party may very well result in more carefully self-screened uses of peremptory challenges on the part of the attorneys. On the other hand, the same post-hoc rationalization that is of concern in the *Terry* context may operate here as well to some unknown extent.

### **Accountability as a Bias-reducing Strategy**

While accountability has been considered within legal doctrine as a tool to improve the decision-making process generally, scholars have also pointed to accountability as a potential strategy to address inter-group bias specifically. The search for such bias-reducing strategies comes in response to the persisting group-based inequalities that have been documented across several critical domains, including the distribution of resources (e.g., employment, education, wealth, housing, health care), life trajectories (e.g., probability of incarceration), and the distribution of status and power (e.g., representation in leadership positions, class divisions and mobility) (Plaut, 2010; Pettit & Western, 2004). Nowhere is evidence of continuing racial disparities more apparent than in the criminal justice system, with a 1 in 15 incarceration rate for Black adults compared to an incarceration rate of 1 in 106 for White adults (Pew Center on the States, 2008). While African Americans represent 13% of the general population, they make up 28% of all arrests, 40% of all inmates in prisons and jails, and 42% of inmates on death row (Hartney & Vuong, 2009). Available data show that other racial minorities are also overrepresented relative to Whites within the criminal justice system.

These racial disparities demand our attention because they raise a challenge to fundamental moral and public values like fairness and equality. They also implicate important, tangible costs for those who are disadvantaged within these social systems. Within the criminal justice context, the stakes encompass both direct costs to those caught up in the system (e.g., deprivation of liberty, exposure to risky conditions of incarceration, the disadvantages and stigma attached to a criminal record, and limited access to public goods) (Pager & Quillian, 2005) as well as indirect costs falling on those connected with the offender (e.g., family stability, increased turnover in communities of color, emotional and material consequences for children of incarcerated parents) (Hagan & Dinovitzer, 1999). More broadly, the perception that the system is unfair because it treats people differently on the basis of race threatens the legitimacy of the system and, in turn, people's willingness to comply and cooperate with the system (Tyler, 2006).

Several factors contribute to these ongoing disparities, including the legacy of both formal (legal) and informal historical practices, contemporary structural or institutional practices, different behavior patterns across groups, and individual bias. These factors often operate in tandem, interact with each other, and can be cumulative in nature. While it is difficult to measure how much of the racial disparity can be attributed to psychologically-driven biases held by decision-makers, there is little doubt that racial stereotypes and prejudice are contributing factors at least some of the time. For example, evidence of racial disparities in the criminal justice system remains even after controlling for other relevant factors (Mustard, 2001; Mitchell, 2005; Baldus, et al., 1998), suggesting that individual-level bias may be influencing these outcomes. More directly, several studies have shown that race can influence perceptions of crime (Eberhardt et al., 2004; Payne, 2001), perceptions of offenders (Bridges & Steen, 1998; Graham & Lowery, 2004), decisions about whether to shoot armed or unarmed targets (Correll et al. 2002; Plant & Peruche, 2005; Peruche & Plant, 2006), the likelihood of receiving the death penalty (Eberhardt et al., 2006; Goff et al., 2008), perceptions of juvenile culpability (Rattan et al., 2012), support for capital punishment, mandatory sentencing, and other deterrent measures (Gilliam & Iyengar, 2000), perceptions of neighborhood crime (Quillian & Pager, 2001), and estimations of risk of crime victimization (Quillian & Pager, 2010).

Given the connection between racial bias and disparate outcomes, the question for some time has been what can be done to attenuate or eliminate the operation of individual bias in the decision-making process? Not surprisingly, several scholars have argued that accountability mechanisms can function to reduce both explicit and implicit (unconscious) biases in decision-making processes (see, e.g., Kang, 2008; Bartlett, 2009; Heilman & Haynes, 2008; Mitchell, 2009). This proffered solution is an attractive one for at least three reasons.

First, accountability is a tool that operates at the nexus between individual decision-makers and the organizational context. It is, therefore, capable of shaping and guiding discretion in desired ways. Organizational sociologists and those studying discretion emphasize the way in which various factors, structures, and norms present within the decision context moderate the role of bias. Discretionary judgments do not simply permit arbitrariness or reflect the whims of the individual decision maker. Rather, a set of external factors operate to guide discretion in systematic ways (Gelsthorpe & Padiel, 2003; Hawkins, 1992; Feldman, 1992; Baumgartner, 1992). Baron & Pfeffer (1994), for example, draw attention to the "'micro-macro' connections--the links between social structures, institutions, and organizations, on the one hand, and, on the other, cognitions, perceptions, interests, and behaviors at the individual or small-group level" (p. 191). Because of an organization's capacity to create and control "opportunity structures" (Petersen & Saporta, 2004) that facilitate the influence of bias, the "*proximate cause* of most

discrimination is whether and how personnel practices in work organizations constrain the biasing effects of these automatic cognitive processes" (Reskin, 2000, p. 320; see also Bielby, 2003). Situating individual actors within the broader organizational context and appreciating the interaction between the two opens up avenues for potential intervention. Accountability seems particularly fit to step into this gulf: "accountability serves as a critical rule and norm enforcement mechanism--the social psychological link between individual decision-makers on the one hand and social systems on the other" (Tetlock, 1992, p. 337). It is an institutional practice that organizations can adopt to shape and constrain the workings of individual bias and, therefore, lies at the center of this micro-macro nexus.

The second reason accountability is an attractive solution is that it offers some relative advantages in addressing discriminatory outcomes. Unlike alternative strategies such as litigation, accountability as an organizational practice focuses on preventing harms from occurring rather than remedying harm already done and it has the potential to respond to discrimination on a large scale. It is also a top-down approach that avoids the difficulties of relying on victims to recognize their injury, of identifying the responsible parties, and going through the arduous process of bringing a claim to court (Felstiner, Abel, & Sarat, 1980). This strategy may be more successful than bringing discrimination claims directly since courts seem reluctant to rely on social scientific evidence of implicit bias (see *Wal-Mart v. Dukes*, 2011; *Pippen v. Iowa*, 2012), struggle to fit instances of systematic bias or implicit bias within the intent requirement of the Equal Protection clause (see Krieger, 1995; *McCleskey v. Kemp*, 1987), and are reluctant to question discretion exercised by legal actors (see *McCleskey v. Kemp*, 1985). For example, the lower court in *McCleskey* found that the statistical studies relied upon failed to support a *prima facie* case that the death penalty was imposed because of the defendant's race, noting that, "The very exercise of discretion means that persons exercising discretion may reach different results from exact duplicates. Assuming each result is within the range of discretion, all are correct in the eyes of the law" (*McCleskey v. Kemp*, 1985, p. 898). The Supreme Court affirmed the lower court's ruling on the basis that, even accepting the conclusions of the statistical studies showing systematic bias, the defendant "offers no evidence specific to his own case that would support an inference that racial considerations played a part in his sentence" (*McCleskey v. Kemp*, 1987, p. 292-93).

A third reason that accountability is an attractive policy solution is that, relative to other methods known to reduce implicit bias, accountability is arguably more compatible with organizational operations and is viewed as less problematic in terms of respecting individual preferences and beliefs. For some time, psychologists have been studying ways to counter implicit bias in particular, with some success in demonstrating that implicit bias is not uncontrollable or inevitable despite being unconscious (see Blair, 2002 for a review). Both implicit and explicit bias can be reduced through consensus information regarding stereotypes (Sechrist & Stangor, 2001; Stangor et al., 2001), sufficient training in the negation of stereotypic associations (Kawakami et al., 1999; Kawakami et al., 2000), perceptual training in individuating the faces of out-group members (Lebrecht et al., 2009), the use of mental imagery to produce counterstereotypes (Blair, Ma, & Lenton, 2001), exposure to atypical exemplars of out-group members (Dasgupta & Greenwald, 2001), positive intergroup contact (Allport, 1954; Pettigrew & Tropp, 2000), and perspective-taking strategies (Galinsky & Moskowitz, 2000). Many have criticized some of these strategies as not readily translatable to "real-world" contexts. By contrast, accountability is a mechanism already familiar to various types of organizations and can be inserted into the decision-making context with relative ease. Rather than standing out as a



special type of intervention, accountability is a mechanism that can be layered upon existing organizational processes.

Additionally, whereas some of the other bias-reducing strategies mentioned above seek to undue the actual stereotypic associations themselves, accountability operates to simply buffer one's bias from decision judgments and outcomes. The distinction here, then, is between targeting the stereotypes themselves versus preventing the stereotypes from influencing particular decisions. Though many would argue that holding racial bias is morally repugnant, accountability does not impose this normative view on its targets. Rather, it allows individuals to hold any beliefs they choose in their private lives while constraining the expression of that bias in the workplace organizational context.

### **A Cautionary Tale**

Before accountability can be touted as a 'best practice' approach for legal institutions, it must be carefully vetted and tested in terms of actually producing desired outcomes. Failure to do this adequately could lead to several negative outcomes. Of most concern is a situation where accountability practices lend legitimacy to the decision-making process but in fact do little to address the intended problems. Such a possibility is raised by work done on symbolic structures (Edelman, 1992; Edelman, Erlanger & Lande, 1993; Edelman, Uggen, & Erlanger, 1999). Through the implementation of certain procedural structures, traditionally disadvantaged groups are perceived as being treated fairly by the organization. For example, Edelman and colleagues have documented that while the role of law extended into workplace organizations through the adoption of internal dispute resolution procedures, these structures became imbued with managerial norms in ways that deemphasize workplace discrimination (Edelman, Erlanger, & Lande, 1993). Importantly, adoption of these and similar process structures lends legitimacy to organizations, irrespective of their effectiveness in adequately addressing discriminatory harms. As a result, when a discrimination claim is brought before the courts, judges are more likely to infer nondiscrimination based on the mere presence of these structures (Edelman et al., 2011).

Similar effects of symbolic structures were found in a series of experimental studies done by Cheryl Kaiser and her colleagues (2013). Among high-status group members (e.g., Whites, men) the presence of diversity structures created an illusion of fairness, that is, these organizations were viewed as more procedurally fair for low-status groups (i.e., racial minorities, women), regardless of whether actual outcomes were equal among groups or not. This illusion of fairness created by the very presence of diversity structures was shown to have important consequences on people's perceptions of discrimination within the organization and the treatment of those who claimed discrimination. First, despite concrete evidence that an organization discriminated against women, participants were less supportive of sex-based discrimination litigation against a company that had diversity structures in place. Second, participants were less likely to say that discrimination was occurring in a company that had diversity structures in place, despite evidence showing salary inequalities for men and women within the company. Third, when diversity structures existed, participants not only viewed the claims of discrimination litigants as less valid, but they also viewed the litigants themselves more negatively and as complainers. These researchers look with skepticism at organizational procedures that are seemingly based on legal principles and given deference without closer scrutiny as to whether they achieve substantive outcomes.

A similar note of caution comes from the literature pertaining to procedural justice. The key finding from procedural justice research is that instituting fair procedures increases perceptions of legitimacy and, in turn, increases acceptance of outcomes--even when the outcome is unfavorable (Tyler & Huo, 2002; Tyler, 1990). Critics have expressed concerns about engendering a sense of "false consciousness" and creating a "procedural justice trap" whereby a focus on process distracts from substantive outcomes (Fox, 1993). MacCoun (2005), for example, describes procedural fairness as a double edged sword, citing arguments that fair processes can be used in a manipulative fashion to induce compliance from people, to legitimize positions of authority, or maintain the status quo. When form is exalted over substance, seemingly fair procedures can legitimize decision systems while at the same time mask unjust outcomes (Fox, 1993). Indeed, Kaiser et al. (2013) found that many of the insulating effects of diversity structures for companies were mediated by participants' perceptions of procedural fairness.

Translating these concerns into the context of accountability, the mere exercise of giving reasons can increase perceptions of fairness and the acceptance of outcomes.<sup>2</sup> Indeed, courts have explicitly viewed accountability as relevant to public perceptions and serving the ends of legitimacy. In *People v. Martin* (1986), for example, the California Supreme Court expressed that "a statement of reasons serves to preserve public confidence in the sentencing process" (p. 450.) Similarly, in *In re Podesto* (1976), the Court stated that "articulated reasons aid in preserving public confidence in the decision-making process 'by helping to persuade the parties [and the public] that . . . decision-making is careful, reasoned and equitable'" (p. 938, citations omitted; see also *United States v. Capriola*, 1976, p. 321: "the preservation of the appearance of judicial integrity and impartiality requires that the sentencing judge record an explanation."). Those advocating the adoption of accountability as a bias-reducing strategy ought to first scrutinize whether adequate empirical evidence supports its effectiveness in the particular context contemplated.

Implementing accountability structures also comes at a cost to the organization itself, and so, to the extent organizations are genuinely committed to improving their decision processes, it is in their interest to adopt tested and proven strategies. Though accountability structures can be systematically integrated into organizational processes, doing so involves several steps that require a significant investment in time and resources. For example, an organization might research and pilot the most appropriate form of accountability, train personnel about the goals and logistics of these new procedures, and set up some type of review or monitoring body to ensure that the procedures are being properly followed and analyze feedback at the individual, departmental, or organizational level. Apart from up-front set up costs, the ongoing operation of accountability procedures would foreseeably add time to the initial decision process and also require a time commitment from the persons reviewing decision outcomes and providing

---

<sup>2</sup> Though not a study designed to examine either accountability or procedural justice, an experiment by Langer, Chanowitz, & Blank (1978) nicely demonstrates the intended point. These researchers demonstrated that compliance was achieved when a reason-giving form was followed, irrespective of whether the reason was a valid one. In this experiment, people standing in line to use a copy machine were asked one of three requests by someone trying to cut in line: 1) "Excuse me. I have 5 pages. May I use the Xerox machine?;" 2) "Excuse me. I have 5 pages. May I use the Xerox machine because I'm in a rush?;" or 3) "Excuse me. I have 5 pages. May I use the Xerox machine because I have to make some copies?" When no reason was provided, only 60% of people complied with the request. But when the request was accompanied by a reason (even a "placebic reason"), the compliance rate jumped to 94% and 93%.

feedback. These time and resource expenditures would be rendered somewhat futile if accountability structures fail, in fact, to improve the decision process in a meaningful way.

As described in later chapters, the body of psychological research on accountability shows that it does indeed have the potential to counteract a variety of cognitive biases. A small set of field studies within organizations have likewise documented the beneficial effects of accountability in promoting equal outcomes among traditionally disadvantaged groups (Kalev, Dobbin, & Kelly, 2006; Castilla, 2013). But the literature also shows that accountability's effects are highly contextually dependent and, at times, can actually have perverse effects by increasing bias (Tetlock, Skitka, & Boettger, 1989; Lerner & Tetlock, 1999). Scholars putting accountability forward as a potential remedy to address implicit biases in particular frequently do so without adequately acknowledging its nuances and limitations (Kang, 2008; Heilman & Haynes, 2008; Mitchell, 2009). Nor do they cite a single study which included an implicit measure of bias and directly demonstrated accountability's effectiveness in combatting this type of bias. The lesson to be learned from the literatures on symbolic structures and procedural justice is that process-based strategies to address discriminatory outcomes need to be evidence-based.<sup>3</sup> Thus, consistent with the goals of the current project, researchers should give attention to the complexities of accountability as a psychological phenomenon and its interaction with the operation of inter-group biases.

### **The Current Project**

As argued above, different kinds of organizations—police departments, probation departments, prosecutors, courts, and businesses—turn to accountability as a strategy to improve the decision-making process and reduce racial bias. Accountability is an attractive solution for a number of practical and policy reasons. Yet, several questions remain about the efficacy of implementing such a strategy. I argue that further empirical study is needed before claims about accountability's ability to reduce racial bias are adequately supported.

The current dissertation project represents a step towards building greater knowledge about the bias-reducing effects of accountability. Across three experimental studies, I test the effectiveness of three different forms of accountability in attenuating the influence of race in the decision-making process. Using a case file paradigm, participants in each study are asked to make decisions pertaining to either a White or racial minority criminal suspect, while either expecting to have to justify their decisions or not. To test the impact of accountability, racial disparities in outcomes are compared across conditions as well as the extent to which participants' explicit and implicit racial biases are related to case decisions.

The remainder of the dissertation is organized as follows. Chapter 2 reviews the existing psychological literature on accountability, with a particular focus on studies that have examined accountability's effects on inter-group bias. To understand how accountability might attenuate the influence of racial bias in particular, I also provide an overview of how inter-group bias operates as well as the theorized mechanism driving accountability's effects on decision-making. I then identify several questions left unanswered by the existing research, several of which are taken up by the current set of studies.

---

<sup>3</sup> By evidence-based, I refer to any type of systematic empirical (qualitative or quantitative) study, including those examining the basic psychological phenomenon in general as well as those conducted within a specific organizational setting.

Chapter 3 presents the results from the first experimental study. Study 1 examined the effectiveness of process accountability and outcome accountability relative to a control condition. Process accountability is where a decision-maker expects having to justify the process he/she used in reaching the decision whereas outcome accountability is where the expectation is having to justify the merits of the outcome of the decision itself. Taking on the role of a prosecutor, participants read about either a White or Black suspect in a drug possession case and considered issues of guilt, case seriousness, and punishment. In addition to explicit measures of racial bias, participants' motivation to control prejudice and social anxiety were also examined.

Chapter 4 presents the results from the second experimental study, which examined the effectiveness of process/(undirected) accountability and directed accountability relative to a control condition. Whereas undirected accountability involves articulating a justification for one's decision process in an open-ended format, directed accountability requires the decision-maker to respond to specific, targeted questions about how they reached their decision and weighed different factors. This latter form of accountability is more akin to standardized decision forms frequently used within organizations. Study 2 used the same basic paradigm as the first study, but asked participants to review three separate case files so that the effectiveness of accountability could be examined across multiple decision points. Whereas college students participated in the first study, this second study recruited a more diverse sample from workers on Amazon Mechanical Turk.

Chapter 5 presents the results from the third experimental study, which also examined the effects of undirected and directed accountability. Examining bias towards a different minority group, participants were asked to review two criminal case files, this time describing either a White or Hispanic suspect. In addition to the measures included in the previous studies, Study 3 asked participants to evaluate the suspect's character and included participants' implicit racial bias towards Hispanics.

Chapter 6 presents a general discussion of the results from all three studies as well as the studies' limitations. Directions for future research are indicated.

## CHAPTER 2: LITERATURE REVIEW

As argued in the previous chapter, scholars and policy makers often turn to accountability to improve decision making processes, including to combat the influence of explicit and implicit racial bias. Indeed, accountability structures can already be found embedded within legal doctrines and institutions. To avoid blind reliance on such structures, however, greater care and attention should be given to the critical question: To what extent is accountability actually effective in attenuating racial bias?

In this chapter, I first review previous research on the general bias-reducing effects of accountability. Then, because of the current project's focus on racial bias, I briefly review some of the psychological literature on intergroup bias. Finally, I review the psychological literature on accountability with the following questions in mind: Under what conditions is accountability effective at attenuating bias in the decision-making process? Through what mechanism does accountability work? What areas of conceptual and practical interest remain under-researched within this literature?

### **The Bias-Reducing Effect of Accountability**

A commonly held belief by scholars, leaders, and lay persons alike is that holding individuals accountable will lead to better decisions. In general, this claim refers to decisions that are more accurate, well-reasoned, and free from the influence of cognitive biases, including stereotypes. The empirical question arises as to whether this belief is justified, that is, can accountability successfully counteract the influence of cognitive biases in the decision process? The short answer is sometimes. Evidence shows that accountability can attenuate a wide range of cognitive biases, though its effectiveness is highly contextually dependent. At times, accountability can actually produce the opposite effect by bolstering biases in the decision process.

Several studies have demonstrated the effectiveness of accountability in addressing a range of cognitive bias, including among others: numerical anchoring, the tendency for numerical estimates to be tied to initial values without sufficient adjustment to new information (Kruglanski & Freund, 1983); overconfidence, the tendency to be more confident in one's judgment than is warranted by actual rates of being correct (Tetlock & Kim, 1987); ordering effects, such as the tendency to rely more heavily on information obtained earlier in time (Kruglanski & Freund, 1983; Webster et al., 1996; Schadewald & Limber, 1992); and the fundamental attribution error, the tendency to attribute positive outcomes to dispositional variables (e.g., personal traits) and to attribute negative outcomes to situational variables (e.g., environment) (Wells et al., 1977; Lerner et al., 1998). These studies generally show that individuals expecting to be held accountable tend to exhibit less evidence of cognitive bias in their decisions. This general finding, however, is a highly qualified one. Some studies have failed to find any effect for accountability (e.g., Simonson & Nye, 1992). More importantly, a number of studies have demonstrated that accountability can amplify cognitive biases (Tetlock & Boettger, 1994; Simonson, 1989; Taylor, 1995). In a review of research on accountability, Lerner & Tetlock (1999) concluded that accountability is likely to attenuate bias only when certain conditions are in place.

First, bias tends to be reduced when the decision-maker learns they will be held accountable prior to forming any opinions (Lerner & Tetlock, 1999). This is because when a

person has already committed to a judgment they will tend to engage in defensive bolstering and rationalize their opinions rather than take a self-critical approach. Once committed to a judgment, individuals are motivated to reduce cognitive dissonance, that is, a sense of discomfort stemming from inconsistencies between one's beliefs and one's behavior (Festinger, 1957). Similarly, individuals may also be motivated by a desire to manage the impressions others have of them (Tedeschi, Schlenker, & Bonoma, 1971). In order to appear consistent and maintain their credibility and legitimacy, decision-makers may try to generate as many reasons as they can in support of their judgment. But these reasons tend to represent a post-hoc reconstruction of the decision process rather than an accurate reflection of what generated the final judgment.

Second, accountability is more likely to attenuate cognitive bias when the views of the anticipated audience are unknown (Lerner & Tetlock, 1999). When the views of the audience are known in advance, the decision-maker is likely to simply shift their judgments to fall in line with those views. This shift towards anticipated views leads to a decision process that is judgment-driven rather than one characterized by thorough and careful consideration of all relevant task information. Again, the goal of impression management is in play here. Since people are motivated to present a positive image of themselves to others, they can preemptively avoid criticism by catering to the preferences of the audience. When the views of the anticipated audience are unknown, decision-makers cannot tailor their responses in such a way; rather, they are motivated to consider counterfactuals, weigh trade-offs, and generate more thorough justifications.

In some situations, an audience can signal certain expectations that do facilitate accountability's effectiveness. Specifically, studies have shown that when an audience is interested in obtaining accuracy, individuals are better at processing task information in an objective manner. Thompson et al. (1994), for example, found that accountability coupled with an accuracy instruction decreased people's tendency to rely on salient traits when forming impressions of another person. In this study, participants were first subtly exposed to positive and negative trait terms by having them create sentences from scrambled words, and subsequently asked to form an impression of an ambiguously described person. Results showed that while participants in the control condition used greater prime-consistent than prime-inconsistent traits in their descriptions, those in the accountability/accuracy condition showed the reverse pattern. Additionally, those in the accountability/accuracy condition spent more time reading information about the person than did their control counterparts, indicating attempts to form more accurate impressions.

Third, accountability has proven more effective when the focus of justification is placed on the process used by the decision-maker rather than on the outcome itself (Lerner & Tetlock, 1999). Process accountability exists when "evaluation is based solely on the quality of the *procedure* that a judge or decision maker uses in arriving at a response, regardless of the quality of the outcome of that response" (Siegel-Jacobs & Yates, 1996, p. 2). Conversely, outcome accountability focuses solely on the quality of the outcome itself. In general, process accountability improves individuals' task performance in several ways. Siegel-Jacobs & Yates (1996) found that process accountability improved calibration (the correlation between one's accuracy and one's confidence in their judgments) as well as judgment consistency. In a study by Brtek & Motowidlo (2002), process accountability increased the accuracy of leadership potential ratings of managers, as compared to actual ratings of the managers by supervisors on the job. Davis, Mero, & Goodman (2007) showed that participants engaged in a complex radar simulation task showed improved performance during the second of two trial phases when held

accountable for the procedures used in making decisions. The positive effects of process accountability are thought to derive from participants seeking out more information, being more attentive during the task, and engaging in more complex modes of processing.

The evidence on outcome accountability is more mixed, with some studies finding beneficial effects (e.g., Davis, Mero, & Goodman, 2007; Mero & Motowidlo, 1995), while others finding that outcome accountability actually worsened performance relative to a no-accountability condition (Brtek & Motowidlo, 2002; Siegel-Jacobs & Yates, 1996). To explain the difference between the two types of accountability, some have noted that process accountability provides a suggestion for enhancing performance that is not dependent on the uncertainties associated with outcomes (Siegel-Jacobs & Yates, 1996). Though further evidence is needed, outcome accountability may produce high levels of stress induced by the demands of finding a suitable solution, which could lead to a more simplified decision process.

Fourth, Lerner & Tetlock (1999) suggest that the anticipated audience should possess certain characteristics in order for accountability to be most effective. The audience should be perceived as having adequate domain-specific knowledge and competence to be able to properly evaluate the justifications given by the decision-maker. Additionally, the audience should have a legitimate reason for inquiring into the reasons behind the person's judgments. If accountability is seen as illegitimate, controlling, or intrusive, it may produce undesired effects, including backlash and increased stress.

Finally, while not emphasized by Lerner & Tetlock (1999), another factor contributing to accountability's effectiveness is whether decision-makers have sufficient cognitive resources to engage in more integratively complex thought.<sup>4</sup> Cognitive resources include sufficient time to complete the task and sufficient processing capacity, i.e., mental focus and attention dedicated to the present task. Thompson et al. (1994) showed that accountability coupled with an accuracy instruction successfully reduced the influence of primed traits on an impression formation task. But accountability's effectiveness disappeared when participants were asked to rehearse a nine-digit number as they completed the task. With their attentional resources simultaneously devoted to a second task, these participants were placed under a cognitive load and did not have the necessary processing capacity to reinterpret the ambiguous target information apart from the primed traits. Similarly, Roets, Van Hiel, & Kruglanski (2013) demonstrated that when decision-makers have high cognitive capacity, accountability facilitated the use of relevant information to modify baseline judgments, but accountability actually impaired this ability when decision-makers had low cognitive capacity. Studies have also shown that accountable decision-makers do not show improved performance when under time constraints (Kruglanski & Freund, 1983). Thus, the effectiveness of accountability may be undermined when individuals are operating under time pressure or are multi-tasking.

Even taking these moderating factors into account, Lerner & Tetlock (1999) note the persistent variability in outcomes across studies. To further account for this variability, the authors offer a flexible, nuanced model of accountability. First, accountability is likely to attenuate bias when the bias results from a lack of either effort or self-critical awareness of one's judgment process. As discussed further below, in these situations accountability provides the

---

<sup>4</sup> Tetlock & Kim (1987) describe integratively complex thought in terms of two dimensions: conceptual differentiation and information integration. Conceptual differentiation is "a function of the number of alternative interpretations or perspectives that a person draws upon in understanding a problem" (p. 701). Information integration is "a function of the number of causal or conceptual connections that a person identifies between (among) differentiated perspectives" (p. 701).

motivational impetus for the decision-maker to engage more deeply and thoroughly with the task than would otherwise be the case. Second, accountability is likely to have no effect when the person simply lacks the requisite knowledge to solve the problem. No amount of motivation can overcome an inability to do the task better. Third, accountability is likely to amplify bias during choice tasks (e.g., hiring decisions) when the biased choice also happens to be the most justifiable. Here, accountability encourages a preference to pick the option that is the most easily justifiable because this option is a safe bet, thus addressing any self-presentation concerns the decision-maker may hold. Accountability can amplify bias during judgment tasks (e.g., levels of competence or dangerousness) by leading people to use all informational cues, including irrelevant ones, in their decision. In trying to build up arguments to justify their judgment, accountable individuals may utilize the information at hand in an indiscriminate manner, relying on both judgment-relevant and judgment-irrelevant cues.

### **The Operation of Intergroup Biases**

Lerner & Tetlock's (1999) review of the accountability literature considered a wide range of cognitive biases. The current project focuses on intergroup bias in particular. Intergroup bias has several different dimensions, but generally refers to an "evaluative, emotional, cognitive, or behavioral response toward another group in ways that devalue or disadvantage the other group and its members either directly or indirectly by valuing or privileging members of one's own group" (Dovidio & Gaertner, 2010). Intergroup bias thus hinges upon perceptions of social groups defined by characteristics such as race, gender, religion, or sexual orientation. Intergroup bias can involve not only out-group derogation, that is, more negative valuations and treatment of members from other social groups, but also in-group favoritism, that is, more positive valuations and treatment of one's own social group. Indeed, giving the benefit of the doubt or extra opportunities to members of one's own group can be just as pernicious in producing discriminatory outcomes as penalizing members from other groups (Gaertner & Dovidio, 2000).

Bias is a multi-dimensional phenomenon, involving cognitive, affective, and motivational components. Stereotypes, for example, are cognitive associations between particular attributes (e.g., nurturing, aggressive) and members of a defined group (e.g., women, Blacks). As I show in Figure 2, these stereotype associations can shape impressions of others and the decision process in ways leading to discriminatory judgments and outcomes. Stereotypes are not only the product of certain psychological processes (e.g., illusory correlation), but are also socially and historically constructed since they are associations learned through frequent and early exposure to cultural cues and messages (e.g., media, accounts transmitted through one's network) (1) (see, e.g., Orbe & Harris, 2006). As part of a socialization process, stereotypes continually get reinforced by exposure to these cultural cues and social interactions throughout our lives. For example, research has shown that local news coverage of crime over-represents the amount of violent crime committed by African-Americans and Hispanics and underrepresents violent crime committed by Whites (Gilliam, Iyengar, Simon, & Wright, 1996).

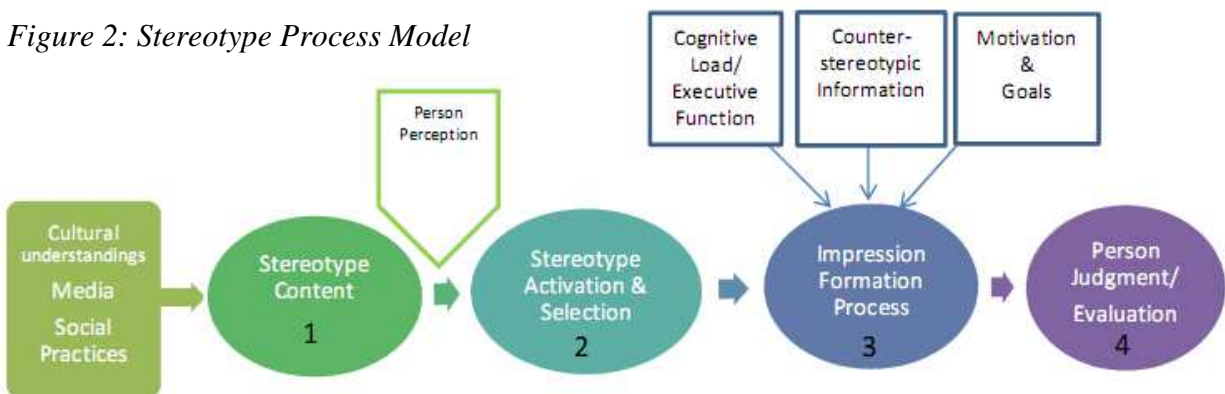
Upon perceiving a person, one or more stereotypes can potentially become activated (2). Several robust stereotypes relevant to the criminal justice context have been documented, including an association between Blacks and aggressiveness, dangerousness, and crime (Eberhardt et al., 2004; Devine, 1989; Payne, 2001; see also Hugenberg & Bodenhausen, 2003, 2004). Additionally, associations between Blacks and apes has been correlated with differential death penalty outcomes (Goff et al., 2008). Selection between competing stereotypes is



determined in part by a particular category's salience and chronic accessibility (Macrae & Bodenhausen, 2000). For example, a job candidate's gender might become relatively more salient when applying for certain job fields (e.g., construction worker, auto mechanic, nanny, nurse), while a person's race might become more salient if he or she is the only racial minority in an otherwise all-White group.

Once activated, stereotypes can influence evaluations of a person as well as decision outcomes relevant to that person (3, 4). Specifically, stereotypes provide expectations about the world and guide the way we attend to, process, and remember information (e.g., Lenton, Blair, & Hastie, 2001; Walsh, Banaji, & Greenwald, 1995). For example, research has shown that perceivers have better memory for stereotype-consistent information relative to stereotype-inconsistent or neutral information (Fyock & Stangor, 1994; Bodenhausen, 1988; Stangor & McMillan, 1992; Neuberg, 1994). People are particularly likely to process information in stereotype-consistent ways when the situation is ambiguous or subjective decision processes are involved (Gaertner & Dovidio, 2000; Hodson, Dovidio, & Gaertner, 2002). For example, ambiguously aggressive behaviors are likely to be perceived as more mean and threatening when the perpetrator is Black as opposed to White because of an existing stereotype associating Blacks with violence and aggression (Sagar & Schofield, 1980). Whether stereotypes become activated and/or influence the impression formation process is dependent on several factors, including the amount of time and cognitive capacity available, the presence of counter-stereotypic information, and the motivation and goals held by the decision maker (Macrae & Bodenhausen, 2000). Thus, the context in which evaluations and decisions are made is of great importance and, as discussed further below, accountability structures help shape the decision context in particular ways.

Figure 2: Stereotype Process Model



It is important to note that the bias literature makes a distinction between explicit and implicit forms of bias. Explicit bias operates at a conscious level and is characterized as overt, intentional, and subject to control. The expression of explicit bias is commonly regarded as an instance of “old-fashioned” racism—the “bad apple” bigot who openly holds and endorses animus towards members of other groups. Explicit bias is generally measured through self-report instruments which assume the person has access to and can articulate their feelings and beliefs on a particular issue. Based on these self-report measures, research shows that levels of explicit bias have steadily declined over time (Schuman et al., 2001; Bobo, 2011). While overt forms of prejudice and discrimination can still easily be found (e.g., the continued existence of hate groups and hate crimes), explicit bias often cannot explain the biased beliefs and behaviors of

those who endorsed egalitarian values. Pursuing other explanations, psychologists began studying more implicit forms of bias.

Implicit bias describes the more subtle, often unconscious, operation of stereotypes and prejudice. Studies have shown that implicit bias stems from automatic and difficult-to-control mental systems which can operate outside a person's awareness (Blair & Banaji, 1996). All people—even those who personally endorse and hold egalitarian values—are susceptible to expressing implicit bias since such bias stems, in part, from basic psychological processes (Devine, 1989; Gaertner & Dovidio, 2000). In their research on aversive racism, Gaertner & Dovidio (2000) have shown that some people can simultaneously endorse egalitarian values and also hold negative racial feelings and beliefs. In order to dissociate these negative racial sentiments from one's self-image as being unbiased, aversive racists will tend to discriminate only when their behavior can be justified on race-neutral grounds or will otherwise be perceived as socially acceptable.

Similarly, implicit bias measures (e.g., the Implicit Association Test), which are designed to bypass a person's awareness and control, demonstrate a tendency among self-described unprejudiced people to automatically associate groups and characteristics in stereotype-consistent ways (Greenwald, McGhee, & Schwartz, 1998). A large body of research now shows that implicit bias exists, is fairly pervasive, and can be uniquely predictive of people's subsequent behaviors (Greenwald et al., 2009; Green et al., 2007; Rudman & Glick, 2001; Rooth, 2007). For example, Green et al. (2007) asked physicians to review a clinical vignette of either a White or Black patient presenting with acute coronary syndrome and then measured physicians' levels of both explicit and implicit bias. Whereas levels of explicit bias were unrelated to physicians' decisions, their levels of implicit racial bias were significantly correlated with their recommendations to treat White patients and not treat Black patients with thrombolytic therapy for a diagnosis of coronary artery disease.

The explicit/implicit distinction is not limited to the bias literature, but rather fits within the framework for a more general dual process theory (Petty & Cacioppo, 1986; Chaiken & Trope, 1999; Kahneman, 2011). This theory posits that human thought and behavior is guided by two conceptually distinct processing systems that are differentially engaged depending on one's mental capacity and motivation. One system (System 1) is characterized by rapid, low effort, and associative thought while the other system (System 2) is characterized by careful, effortful, and deliberate thought. An example to illustrate the difference between these systems is learning how to drive a car. When learning for the first time, one must devote a lot of attention and effort into coordinating all the novel movements and tasks involved in driving. The beginner driver is relying on System 2 processing to perform this task. But after years of driving, the movements and tasks involved have been mastered and the person no longer has to consciously attend to a large part of what they are doing. System 1 processing is engaged and the mechanics of driving have become habitual. As with all things that are habitual, we often just do them without having to think about it. In the same way, stereotypes and prejudice can be thought of as learned habits that can operate even when we are not aware of it (Devine, Forscher, Austin, & Cox, 2012).

### **Accountability Studies on Intergroup Bias**

Of the subset of studies that have specifically examined accountability with respect to inter-group bias, the results are quite mixed. Some have found accountability to be successful in

decreasing bias and the use of stereotypes. For example, Kruglanski & Freund (1983) found that accountability moderated ethnic stereotyping. In evaluating a composition written by an ethnically-identifiable student (either an Ashkenazi or Sephardi Jew in Isarel), accountable participants<sup>5</sup> were significantly less likely to render stereotype-consistent judgments than those in the control group, at least when these participants were not faced with time pressure. Ford et al. (2004) also successfully demonstrated that accountability could reduce the expression of racial bias with White sales managers in a hiring task. Participants were presented with a job description and resumes for two job candidates. While the first resume was the same for all participants (describing a White applicant), the researchers manipulated race for the second resume, with half the participants reviewing another White applicant and the other half reviewing a Black applicant. As to the second resume, those in the control condition exhibited a racial bias by evaluating the White applicant significantly more positively than the Black candidate. But there was no difference in the evaluation of candidates in the high-accountability condition.<sup>6</sup>

Although not specifically dealing with inter-group bias, Thompson et al.'s (1994) study on accountability is instructive because it examined the same basic cognitive process underlying the operation of stereotypes. When accountability was coupled with an accuracy instruction, participants were less likely to evaluate an ambiguous target in ways consistent with previously primed trait constructs. Specifically, after reading a paragraph describing a particular person, participants in the non-accountability condition were more likely to describe the individual's personality in ways consistent with positive and negative traits they had been exposed to during an earlier task where they were asked to make sentences using a scrambled set of words. By contrast, those in the accountability/accuracy condition tended to describe the individual in more prime-inconsistent ways.

Some studies show mixed evidence of accountability's effectiveness. For example, in Hattrup & Ford's (1995) study, participants were given the opportunity to gather information about eight hypothetical co-workers before rating the extent to which they would prefer working on a task with each person. In the test condition, the co-workers were labeled by their occupational roles, and participants could opt to view up to a total of 9 attributes for each co-worker that were either consistent or inconsistent with the given occupation. Accountable participants<sup>7</sup> engaged in more exhaustive information seeking and spent more time reviewing that information compared to unaccountable participants. However, a labeling effect still

---

<sup>5</sup> In the accountability condition, participants were told that the research was intended to assess the teachers' evaluative ability prior to their graduation and that they would have to explain their grade assignment to other members of the group. They were also told they could measure their evaluative ability by comparing their grade assignment with that made by a team of experienced teachers.

<sup>6</sup> The accountability instruction included, in part: "in the business environment, whenever people evaluate something, they often must be able to account for the reasons for their evaluations. Since the result of evaluating job applicants for the company is crucial, you will be asked to discuss your views and justify the basis for your evaluation to your supervisor, who may have different views on this subject." It should be noted, however, that the researchers may have manipulated more than accountability alone. In the low-accountability condition, the instructions stated that the researchers were specifically interested in participants' judgments based on their instincts and gut feelings. In the high-accountability condition, by contrast, the instructions stated that the researchers were interested in participants' judgments based on careful deliberation. Thus, it is not clear whether these additional instructions may explain some of the differing results across conditions.

<sup>7</sup> Accountable participants were informed they would be asked to write a short one- to two-paragraph explanation of each of their final target ratings and were shown a copy of a form they would use to provide these written explanations. They were further told that there was a chance that they may be interviewed by one of the researchers to determine how they arrived at their final ratings.

persisted, suggesting that accountability was not sufficient to eliminate participants' reliance on stereotypes during information seeking, even in the face of stereotype-inconsistent information.

A study by Bodenhausen, Kramer, and Susser (1994) showed that accountability was sufficient to undermine the biased tendency of happy people to make more stereotypic judgments. Taking on the role of a peer disciplinary review panel member, participants reported the likelihood an accused student was guilty of either assault or cheating after reading a description containing ambiguous evidence. When motivated by an accountability instruction,<sup>8</sup> happy-induced participants failed to show greater bias in their judgments towards the stereotyped target. It is possible, though, that holding participants accountable may have interfered with the emotion manipulation in a different way than anticipated, e.g., made these participants less happy. A comparison between accountable and unaccountable participants in the neutral condition, for instance, shows that accountability actually increased judgments of guilt in a stereotype-consistent way.<sup>9</sup>

Several studies have shown that accountability can actually increase or bolster the influence of stereotypes in the decision process. Gordon, Roxelle, and Baxter (1988) found that participants held accountable<sup>10</sup> for their evaluations of a job candidate during an interview evaluated older candidates more negatively than those in the unaccountable condition. Accountable participants also recalled significantly fewer pieces of information about the candidates than those in the control condition, suggesting these participants formed more simplified impressions. The authors suggest that the increased bias effects are due to the similarity between the accountability manipulation and transmission effects, that is, "when individuals are informed that they will be involved in transmitting the impressions they form of a given stimulus person to others (transmission set) their cognitive representation of the stimulus person tends to be more organized, rigid, and polarized" (p. 21). In essence, when anticipating having to summarize information to others, the person tends to focus on the "take away" points and perhaps relies more on stereotypes as an organizing framework.

Lambert and colleagues (1996) also found evidence that anticipated public settings bolstered previously held racial attitudes when evaluating a Black target. Using an ambiguous description of a Black-identified person, these researchers found that participants' judgments were more consistent with their explicit racial attitudes when anticipating having to discuss their views with others. These results held even when participants were given information about the racial attitudes of the audience, leading the researchers to conclude that anticipating discussing their views led participants to focus on and solidify their previously formed private racial attitudes due to the implicit pressure to defend one's views. In a subsequent study (2003, Study 1), the researchers found that these effects were largely limited to those participants high in social anxiety. Additionally, anticipated public settings were shown to lead to increased expression of stereotypes due to an impaired ability to engage in controlled processing, as demonstrated by lowered performance on a response time task. Specifically, those anticipating a

---

<sup>8</sup> The accountability instruction read: "Bear in mind that you will be held accountable for your judgments, just as if you were a judge on a real peer discipline panel. That is, you will have to be able to justify the decisions that you make about the case you read."

<sup>9</sup> Because the main focus of this study was on the effect of happiness on stereotyping, the authors did not provide specific tests regarding accountable and non-accountable participants in the neutral mood condition, though the reported means suggest a bolstering effect of accountability (Lambert et al., 1996).

<sup>10</sup> Participants in the accountability condition were told that following completion of the task they would meet with two representatives from a personnel association who would be interested in hearing their description of the job applicant and the basis on which they believed they formed their impressions.

public setting made more identification errors when a Black prime preceded an image of a tool than when it preceded an image of a gun. Again, levels of anxiety moderated these effects, with the ironic consequence that prejudice was exacerbated among those presumably most worried about doing the wrong thing in public.

A study by Biernat & Fuegen (2001) examining gender stereotypes likewise found that accountability<sup>11</sup> actually decreased the likelihood that a female job candidate would be chosen in a hiring task. Participants reviewed 14 resumes and selected three candidates to be placed on a short-list. After reading letters of recommendations for these three candidates, participants then selected one applicant to hire. Although accountability did not affect short-listing decisions, participants in the accountability condition were somewhat less likely to hire a female than those in the no-accountability condition (41% versus 58%). This pattern is consistent with existing stereotypes that women are less competent than men, though the authors did not report whether the proportion of females hired in the accountability condition significantly differed from chance.<sup>12</sup>

In sum, the existing literature provides mixed answers about whether accountability can effectively reduce the influence of intergroup biases. Particularly concerning for those who offer accountability as a solution to reduce bias, it is not merely a case of occasional ineffectiveness. In some instances (the exact triggering factors remaining unclear), accountability may actually produce more harm than good by bolstering intergroup biases.

### **Accountability's Mechanism**

We know from at least some studies that accountability can attenuate cognitive biases that would otherwise lead to undesirable outcomes. What accounts for accountability's effectiveness in these cases? The proposed theoretical mechanism is that accountability leads people to increase their cognitive complexity, that is, to engage with the decision task more deeply and thoroughly and become more "vigilant information processors" (Tetlock, 1983). Accountable individuals are more likely to consider counterfactuals and weigh possible alternative outcomes so they are more prepared to justify their decision and counter criticism when needed. They also demonstrate other earmarks of high quality decision-making, including greater investments of time, greater tolerance for inconsistency, more awareness of informational determinants, and being receptive to new evidence (Tetlock, 1983). To use the framework of dual process theory, accountability encourages more System 2 processing (e.g., effortful, deliberate, and analytic thinking) rather than System 1 processing (e.g., fast, habitual, intuitive thinking). This more systematic thinking means that accountable individuals are less likely to rely on stereotypes or other types of heuristics.

Studies using different kinds of measures support the claim that accountability's effectiveness lies in its ability to prompt more integratively complex thought. Tetlock (1983), for example, asked participants to report their thoughts about three controversial issues and then coded these responses according to their structural properties. Participants accountable to an audience with unknown views were significantly more likely to reason in ways that

---

<sup>11</sup> Participants in the accountability condition were told that the campus department would make recommendations based on their evaluations and that they should be able to justify any decision they made.

<sup>12</sup> Although there was no significant interaction between accountability and participant gender, the researchers noted that male participants consistently tended to over-hire female job candidates, perhaps because they were motivated to appear nonsexist, except when they were held accountable.

differentiated multiple aspects of the issue and integrated these different dimensions with one another. As a proxy for the amount of cognitive elaboration, both Ford & Weldon (1981) and Thompson et al. (1994) found that those in the accountability condition spent significantly more time reading the task information than those in the no-accountability condition.

Mero & Motowidlo (1995) found that accountable participants took more and better notes and had higher levels of attentiveness as assessed through behavioral cues than non-accountable participants. Additionally, accountable participants self-reported as being more engaged during the task, including spending more time thinking about the challenges presented, discussing the task more with others, and debating more between alternative responses. Subsequent analysis of these data and a partial replication of the study suggested that attentiveness and note-taking helped to explain the effects of accountability on rater accuracy (Mero, Motowidlo, & Anna, 2003). The researchers concluded that accountability led participants to alter their process strategies in ways that better prepared them for making those ratings.

Although one consequence of accountability is more cognitively complex thought, the mechanism is essentially a motivational one. Operating in the background is a concern of how one's performance will reflect on back on him/herself.<sup>13</sup> Schlenker & Weigold (1989) argue that accountability necessarily implicates identity concerns because "identity is constructed via the layering of judgments that emerge from being accountable for particular experiences over time" (p. 23). Thus, the qualities that a person wishes to project to others (e.g., competent, thorough, reliable, fair) become more salient, as do the prescriptions or standards pertaining to the task and the presumed or known expectations of the audience. Anticipating the reactions from others, accountability motivates one to self-regulate by monitoring and controlling their behavior and evaluating their own conduct (Schlenker & Weigold, 1989). When held accountable, concerns increase about committing oneself to a particular outcome or judgment that is seen by others as illegitimate, poorly justified, or mistaken. This "fear of invalidity" (Kruglanski & Freund, 1982) renders an individual more cautious before reaching a judgment or choice. Because the anticipated evaluation of one's decision implicates a person's image, this preemptive defensive strategy can sometimes be accompanied by higher levels of stress which can actually impair performance (Schlenker & Weigold, 1989; Lambert et al. 2003).

Though the literature has mostly focused on accountability as triggering a defensive reaction, some scholars have argued that accountability can be viewed as posing a challenge rather than a threat (Schlenker & Weigold, 1989). When people have high outcome expectations, they respond proactively to the challenging situation. This "acquisitive style" includes more personal investment in the task and a greater expenditure of effort and persistence in the face of difficulties or obstacles. Thus, when there is a reasonable likelihood of success, accountability may positively drive people to step up to the challenge rather than focus on simply avoiding negative evaluations.

Accountability's ability to motivate improved cognitive processing is generally supported by the larger literature on motivated cognition, which shows that motivation and cognition are closely related and interactive systems. The ways in which we seek out information and interpret the world are shaped by whatever sets of goals are activated at the time. Indeed, other motivational concerns have been shown to moderate the activation of stereotypes in several

---

<sup>13</sup> The link between greater cognitive complexity and the relevance of one's identity is further supported by the literature on persuasion and attitude change. Specifically, when the personal relevance of a message is increased, people tend to scrutinize the evidence more carefully and are more responsive to the strength of the message content (Petty & Brinol, 2010; Petty & Cacioppo, 1990; Fleming and Petty, 2000).

contexts. For example, Spencer et al. (1998) showed that either threatening or enhancing one's self-image through negative or positive feedback on an intelligence test moderated differences in whether automatic stereotypes of Blacks and Asians were exhibited during subsequent word completion tasks (see also Fein & Spencer, 1997). Similarly, Sinclair & Kunda (1999) showed that receiving negative feedback from a Black doctor led to greater activation of negative racial stereotypes and lesser activation of conflicting positive professional stereotypes, particularly for those high in prejudice. The reverse pattern emerged for those participants receiving positive feedback from the Black doctor. Participants receiving the negative feedback were supposedly inclined to view the doctor as incompetent in order to counter the blow from the feedback, thus triggering negative stereotypes. This self-serving mechanism did not emerge when the feedback was about someone else. Not surprisingly, the motivation to control prejudice is another motivational goal that affects stereotype activation and application. Social desirability concerns and the need to be seen by others as not being prejudiced have been offered as explanations for the finding that participants show less prejudice in the presence of a Black experimenter rather than a White experimenter (see Blair, 2002).<sup>14</sup>

The general literature on bias reduction points to another aspect of the relationship between motivation and cognition, which is that both are necessary components to successfully address bias. Individuals must have both the will and the ability to do so (Devine et al., 2012). The studies by Kruglanski & Freund (1982) and Thompson et al. (1994) provide examples of this principle. In both cases, accountability reduced reliance on stereotypes or primed traits, but this effect disappeared when accountable individuals were placed under time constraints. In light of the strain placed on cognitive resources by this time pressure, the motivational force of the accountability manipulation was not a sufficient condition by itself to produce reductions in bias. Thus, accountability's motivational mechanism is limited by the person's cognitive processing capacity.

As a final note on the mechanism of accountability, some critics have noted the psychological limitations inherent in asking people to justify their actions (MacCoun, 2006). First, requiring a justification presumes that individuals are capable of introspection and have accurate knowledge of the reasons driving their decisions. But research suggests that people are, in fact, quite poor at accurately explaining their behavior (Nisbett & Wilson, 1977). After reviewing several literatures on the accuracy of subjective reports about higher mental processes, Nisbett and Wilson (1977) reached the following conclusion:

People often cannot report accurately on the effects of particular stimuli on higher order, inference-based responses. Indeed, sometimes they cannot report on the existence of critical stimuli, sometimes cannot even report that an inferential process of any kind has occurred. The accuracy of subjective reports is so poor as to suggest that any introspective access that may exist is not sufficient to produce generally correct or reliable reports (p. 233).

---

<sup>14</sup> As discussed further below, studies examining motivations to regulate prejudice and strategies to avoid race have yielded mixed results in terms of their effects. Apfelbaum et al. (2008), for example, provide an example of how avoidance of race can backfire to produce more prejudicial behavior. In their study, White participants who strategically engaged in colorblindness during an interracial interaction tended to exhibit more negative nonverbal behavior towards their partner, an effect attributed to a decreased ability to exert inhibitory control. Similarly, Goff et al. (2008) found that the threat of appearing racist caused Whites to engage in distancing behavior during interactions with Black partners, a result attributed to stereotype threat (e.g., concerns about being judged or treated on the basis of a "White racist" stereotype).

When asked why they took a particular course of action, people often search for plausible reasons that explain and legitimize their behavior. For example, moral psychologist Jonathan Haidt challenges the rationalist model of moral judgment, whereby judgments flow directly from a process of reasoning and reflection (Haidt, 2001). Haidt suggests instead that an intuitionist model is a more accurate one, with moral intuitions guiding judgments. Moral reasoning is generated in a post hoc fashion to be consistent with and support the already-made judgment. Thus, people are better described as lawyers building a persuasive case rather than scientists searching for the truth.

While accountability may encourage greater awareness of the information one is using in the decision process, existing evidence suggests there are introspective limits to this. The proffered reasons may simply not reflect the actual determinants of the decision. Additionally, demanding a justification from someone may focus attention on easily articulated or salient goals as opposed to other factors that are or should be at play. This narrowing of focus can potentially lead to suboptimal decisions in terms of simultaneously satisfying multiple goals (MacCoun, 2006).

In sum, existing research suggests that accountability is an external motivational force that encourages the use of particular cognitive strategies during a decision task, including seeking out more information, being more aware of informational determinants, considering alternatives, and distinguishing and integrating information in more complex ways. In terms of addressing racial bias specifically, this more deliberative and effortful approach should reduce reliance on cognitive heuristics like stereotypes. However, the nature of the mechanism at play raises at least three possible concerns. First, while the prospect of being evaluated can result in more cautious and deliberative thinking, concerns about performance and the consequences to one's identity can also trigger anxiety and stress that can actually impair decision performance. Second, the effectiveness of accountability's motivational force is constrained by the cognitive capacity and resources (e.g., time, adequate focus) available to the decision-maker during the task. Third, existing research suggests there may be limits to people's ability to provide accurate introspective knowledge about the reasons for their decisions, raising the possibility that requiring an articulation of reasons will simply result in post-hoc justifications.

### **Unanswered Questions: The Need for Further Research**

The literature reviewed above suggests that accountability's effectiveness in attenuating bias is highly contextually dependent. At times, the expectation of having to justify one's decision to others motivates people to engage more deeply with the decision task, leading to a more cognitively complex process that minimizes reliance on cognitive shortcuts like stereotypes. At other times, however, holding people accountable can either produce no effect or actually bolster the influence of cognitive biases on decision outcomes. Several factors have been identified in the literature that seem to moderate accountability's effects, yet predicting when accountability will produce the desired effect remains difficult even after taking these factors into consideration. As described in the following sections, there are several additional aspects of accountability that remain under-examined. In particular, further research is needed on testing accountability's effectiveness on intergroup bias specifically, affectively-driven biases, and implicit biases, on different motivational manipulations, on accountability as a recurring cycle, and on different forms of accountability.



## Further Examination of Intergroup Bias Specifically

The handful of existing studies relevant to the operation of intergroup bias yield very mixed results, suggesting a complex interplay between this type of bias and accountability. Because social categories such as race, gender, and sexual orientation lie at the center of important historical and contemporary issues, intergroup biases arguably operate in ways relatively unique from other forms of cognitive bias. Existing within the social consciousness are debates about the normative significance of these social categories, the extent to which discrimination is ongoing, and the concern of being perceived as biased against a particular group based on one's actions or decisions. When considering the role of race in contemporary society, for example, one simultaneously confronts messages about a colorblind America (Bonilla-Silva, 2010), evidence of ongoing racial inequality (Plaut, 2010), debates about affirmative action (*Fisher v. University of Texas*, 2013; *Schuetz v. Coalition to Defend Affirmative Action*, 2013), and concerns about reverse discrimination (*Ricci v. DeStefano*, 2009). Such contentious narratives can hardly be said to exist in reference to something like an overconfidence bias. Even within the broad category of intergroup bias there is variation in the significance and meaning attached to biases based on different social groups. For example, within the employment context some social groups are recognized as legally-protected categories (e.g., race & gender) while others do not receive special protection under the law (e.g., weight).

Relatedly, intergroup biases are more salient and visible. While few people are likely aware of what numerical anchoring is or whether they may be susceptible to this type of bias in a given situation, many people are attuned to issues of intergroup bias and are aware of its normative valuation in society. The same cues that permit categorization into group membership (e.g., appearance, name characteristics, language accents, etc.) can also place a decision-maker on notice about the potential influence of intergroup bias, a signal in the decision process that may not be available for other types of cognitive bias.

The implications of these unique features of intergroup bias on accountability should be explored further. For example, intergroup biases may trigger several other motivational goals, such as impression management, that may undermine cognitive complexity during the decision task. Across a series of studies, Michael Norton and colleagues demonstrated that in order to mask potentially biased judgments, decision-makers engage in casuistry, that is, specious reasoning in the service of justifying questionable behavior (Norton, Vandello, & Darley, 2004; Norton, Sommers, Vandello, & Darley, 2006; Norton, Vandello, Biga, & Darley, 2008). Specifically, decision-makers structure and change their decision criteria in ways that support their preferred outcome, thus making their decision appear legitimate and grounded in objective reasons (Uhlmann & Cohen, 2005). Such a strategy may be used not only to maintain a non-biased appearance for others, but also to maintain an illusion of objectivity for oneself.

In one such study, for example, the researchers made political correctness norms salient during a college admission task (Norton, Vandello, & Darley, 2004). Decision-makers choosing between a White and Black candidate overwhelmingly selected the Black candidate in order to demonstrate that they were not biased against minority group members. Critically, the reported importance of different decision criteria (i.e., GPA versus number of advanced placement courses) changed depending on which of these criterion happened to support selection of the Black candidate. Thus, the justifications given by decision-makers simply changed while outcomes were still strongly influenced by the race of the applicant. Holding decision-makers accountable did not change either the selection or justification bias resulting from casuistry.

Rather than engaging in more integratively complex thought, accountable decision-makers continued to restructure their reasons in ways that supported their preferred choice—a choice that fulfilled their motivation to appear unbiased.

A related factor that may interact in unintended ways with intergroup bias is that accountability is most effective when the views of the anticipated audience are unknown. In the context of intergroup bias, however, prevailing social and legal norms provide an indication of what an audience's views are likely to be. In an organizational context where non-discrimination norms are salient, a decision-maker can preemptively avoid potential criticisms about who to hire or promote by choosing a minority candidate. This process, however, may result in an over-compensation effect, that is, minority groups being treated systematically more favorably than traditionally high-status groups. From a normative perspective, this may be a desirable effect in a context where the goal is to compensate for structural and historical inequalities that accumulate prior to the particular decision point. Accountability structures incorporated into affirmative action programs, for example, have been somewhat successful in increasing minority representation within organizations (Kalev, Dobbin, & Kelly, 2006). Yet this type of overcorrection may not be appropriate in other contexts, such as the criminal justice system.

### **Testing Accountability's Effectiveness on Implicit Bias**

Most studies on accountability do not include any measures of individual difference in bias, and when they are included they are self-report measures of explicit bias (e.g., Lambert, 2003). Despite the lack of supporting evidence, several scholars have suggested that accountability can effectively combat implicit bias (Kang, 2008; Bartlett, 2009; Heilman & Haynes, 2008; Mitchell, 2009). Because the processes underlying the operation of implicit and explicit bias differ considerably, one needs to question whether the same intervention strategies can adequately address both forms at once. While accountability can lead decision-makers to pay greater attention to the cues they use and gain greater awareness of their cognitive processes, does this extend to cues that generally operate automatically and unconsciously?

One possibility is that greater attention to and processing of task information triggered by accountability will override intuitive or gut feelings about the target, producing a "sober second thought" (Mitchell, 2009). The bias reduction literature has shown that implicit biases are malleable and can be changed through simple associative interventions. For example, Kawakami and colleagues showed across a number of experiments that extensive counterstereotype or stereotype negation training could reduce stereotype activation and the expression of bias (Kawakami et al., 2000; Kawakami, Dovidio, & van Kamp, 2005; Kawakami, Dovidio, van Kamp, 2007). Attempts to reduce implicit bias through more deliberative strategies have been met with more mixed results. On one hand, Wallaert, Ward, & Mann (2010) found that a simple instruction to "be careful not to stereotype" was sufficient to reduce pro-White bias on an Implicit Association Test. Efforts to direct thoughts in particular ways can also affect levels of implicit bias. For example, Blair, Ma, and Lenton (2001) demonstrated that those who engaged in counterstereotypic mental imagery (e.g., imagining the characteristics of a strong woman) had reduced levels of implicit stereotypes relative to a control group. On the other hand, the same study also found that telling participants the task is a measure of gender stereotypes and they should try to suppress such stereotypes did not change levels of implicit bias. Similarly, Payne, Lambert, & Jacoby (2002) found that telling participants to avoid the influence of racial cues during a task did not significantly reduce errors on a race-object identification task. If anything,

motivating participants to avoid the use of race increased the extent to which they made more stereotype-congruent versus incongruent errors. A study by Legault, Gutsell, & Inzlicht (2011) examined whether levels of prejudice could be changed by motivating people to control their prejudice for either autonomous (self-determined) or external reasons. They found that while self-determined motivation reduced prejudice on both explicit and implicit bias measures relative to the control group, external motivation to reduce prejudice produced a backfire effect such that participants actually showed increased levels of prejudice relative to the control group.

The mixed evidence cited above suggests there is reason to be skeptical that the process of overriding implicit bias will generally occur in response to accountability manipulations. Indeed, something more than just greater motivation may be necessary to counteract implicit bias. For example, a successful bias-reduction program undertaken by Devine and colleagues suggests that implicit bias should be approached in the same way as breaking a bad habit (Devine et al., 2012). This requires that the person (1) be aware of their biases, (2) be motivated to eliminate their bias, (3) have the appropriate tools and strategies to combat bias, and (4) have knowledge of the situations in which such bias is likely to be triggered. Although the goal of accountability is not necessarily to eliminate a person's bias altogether, but rather to buffer the decision process from existing biases, one might expect that a similar combination of factors will still be required. The effect of accountability on implicit bias is certainly one area that needs greater attention.

## **Cognitive versus Affective Biases**

The psychological literature on accountability has mainly focused on different types of *cognitive* biases. The literature on intergroup bias, however, makes a distinction between stereotypes and prejudice. While stereotypes are cognitive associations, prejudice refers to the emotional or evaluative orientation towards a person because of their group membership. In the context of judging a criminal defendant, for example, people's decision-making process may be influenced by the prevailing stereotype linking minorities with crime, but the process may also be influenced by a more general disliking of minorities as outgroup members. While this cognitive-affective distinction does not represent a strict boundary between mental systems (Madva & Brownstein, 2013), it raises the question about whether accountability can attenuate affectively-driven bias. In other words, accountable decision-makers are motivated to engage in greater cognitive complexity during the decision task, but how does this enhanced cognitive work impact initial emotional responses?

A clear answer to this question may be difficult to reach given debates within the emotions literature about the relationship between cognitive and emotional processing (see Zajonc, 1980, 1984; Lazarus, 1982, 1984).<sup>15</sup> Some scholars have noted that affective evaluations have been shown to occur quite rapidly and automatically such that they are thought to be a part of perception (Zajonc, 1980). Zajonc (1980) has argued that affective reactions not only can precede cognitive processing, but are also more difficult to control and less susceptible to change than cognitive judgments. Thus, cognitive overriding of initial attitudinal reactions might prove difficult to achieve. It is one thing for accountability to motivate greater cognitive effort when

---

<sup>15</sup> To some extent, this debate parallels the general automatic versus controlled distinction made within the bias literature. In other words, the elicitation and effect of emotions on behavior can occur at both higher (e.g., conceptual, controlled) and lower (e.g., pre-attentive, automatic) levels of processing.

the decision-maker otherwise lacks motivation. But it is a different situation when the decision-maker is already motivated by a particular emotional response generated within the context of a decision task. Bodenhausen, Sheppard, & Kramer (1994), for example, found that particular emotions have different kinds of effects on cognitive processing. Specifically, participants induced to feel anger (but not sadness) made greater use of racial stereotypes in judging the guilt of others. In such cases, the motivational force of accountability is potentially competing with the emotion-based motivation. How might accountability's effectiveness be diminished when the decision maker is experiencing anger or threat, the source of which is an outgroup member about whom the decision must be made?

The existing literature on accountability has yet to explore this kind of question or address the affective side of intergroup bias in any kind of systematic way. One of the rare exceptions is a study by Lerner, Goldberg, & Tetlock (1998). As part of two ostensibly unrelated tasks, participants first viewed either an anger-inducing or neutral video and then assessed responsibility in a fictional tort case. For both tasks, participants were either held accountable or not for their judgments. The researchers found that accountability decreased the tendency of angry participants to make more punitive attributions in the tort cases.<sup>16</sup> Supporting the conclusion that accountability triggered greater systematic thought, accountable participants' judgments were related to their perceptions of free will for the defendants (suggesting they paid attention to mitigating information in the cases), but this relationship was not observed for unaccountable participants. Additionally, accountable participants were relatively less influenced by their pre-existing level of anger compared to unaccountable participants. While this study provides some evidence that accountability can effectively attenuate emotional responses, further research is needed to examine this effect across different decision contexts and with different types of emotional responses.

## **Examining Different Motivational Prescriptions**

Schlenker and Weigold (1989) define accountability as "being answerable for conducting oneself in a manner that is consistent with relevant prescriptions for how things should be" (p. 24). While most studies of accountability focus on its ability to increase motivation on the part of a decision maker, few studies examine how this motivation might be shaped and directed in particular ways. The typical accountability instruction puts people on alert, but provides very little guidance as to how they should be deploying their efforts. Some work suggests that accountable people, rather than engaging in integratively complex thought, are likely to conform to the known (or easily guessed) views of an audience (Tetlock, Skitka, & Boettger, 1989; Tetlock, 1983). Conformity in this case is the easiest (or cognitively "lazy") way to reach what would be considered an acceptable solution to the reviewing audience. However, specific goals, like accuracy, have been shown to preserve the kind of depth of processing that is usually sought after (Thompson et al., 1994). Studies just looking at accuracy motivation alone have also found it to have a number of beneficial outcomes, including encouraging individuating rather than category-based processes (Neuberg & Fiske, 1987) and overcoming negative expectancies when forming impressions about others (Neuberg, 1989).

---

<sup>16</sup> Interestingly, accountability reduced levels of punitiveness even among those who viewed the neutral-emotion video.

Another motivational goal which seems particularly relevant in attempting to attenuate intergroup bias is the explicit goal of equal treatment regardless of a person's group characteristics. In the psychological literature, there is mixed evidence regarding the effects of a directive to be non-prejudiced or avoid the use of race in judgments. Payne, Lambert, & Jacoby (2002), for example, found that calling attention to race increased the accessibility of stereotypes, and Legault, Gutsell, & Inzlicht (2011) found that using external reasons to motivate prejudice regulation resulted in higher levels of explicit and implicit bias. By contrast, Wallaert, Ward, & Mann (2010) found that a simple instruction to "be careful not to stereotype" was sufficient to reduce pro-White bias on an IAT test. It may also be possible that making a non-discrimination goal salient might result in an overcompensation effect, whereby, in an effort to appear non-prejudiced, decision-makers actually give preferential treatment to minority targets. Depending on the context, other goals such as fairness or impartiality might be appropriate to signal along with accountability.

### **Examining Accountability as a Recurring Cycle**

As described in the previous chapter, rather than thinking about accountability as a single event, in many contexts it is better modeled as a recurring cycle consisting of multiple phases. Thus, a decision-maker must confront a number of psychological factors during the decision task itself, render a judgment or choice, justify his/her actions, be subject to review by a third party, and receive feedback and/or consequences stemming from the review process. In many organizational contexts, multiple rounds of decision-making and justification may occur before the review process. Particularly when individuals render a large number of similar types of judgments as part of their jobs, the review process may occur for only a subset of the total decisions or at an aggregate level.

Perhaps not surprisingly, the psychological literature on accountability has focused almost entirely on the psychological factors faced by an individual during the decision task. Effects are examined after creating the *mere expectation* that one will be held accountable. As a result, a large portion of studies do not actually have participants go through the justification process. Similarly, there are almost no studies that examine the effect of accountability across multiple decision points. Thus, effects that might derive from actually writing down or otherwise articulating one's reasoning have been largely unexplored. Whether the experiential aspect of justification either adds to or detracts from accountability's effectiveness may depend on the form of accountability at issue (see next section).

### **Examining Different Forms of Accountability**

With one exception, the existing literature has not systematically examined the effectiveness of different forms of accountability. The exception consists of studies that have compared *process* versus *outcome* accountability, that is, the extent to which one is expecting to be held accountable for either the process they used in reaching the decision or the decision outcome itself (see, e.g., Davis, Mero, & Goodman, 2007; Siegel-Jacobs & Yates, 1996; Brtek & Motowidlo, 2002; Mero & Motowidlo, 1995). Yet, this distinction between process and outcome barely scratches the surface in terms of how accountability is likely to be operationalized in various organizational settings. For example, accounting for one's actions through the keeping of records is a form of accountability likely to be used by organizations because of its relatively low

cost, minimal intrusiveness as an intervention, and its potential to also serve multiple functions, such as creating aggregate data to be used in other ways. Indeed, this is the type of accountability that is already being implemented in police departments across the country.

Another distinction to be explored is whether the justification happens face-to-face with the reviewer or in writing. Even when justifying a decision in writing, there is great variation in the form accountability can take. It could consist of simply checking a box from a list of pre-constructed reasons, as police officers have been doing in New York City (Jones-Brown, Gill, & Trone, 2010). It can consist of an open-ended directive to the decision-maker to state the reasons for their decisions, as judges are often required to do on the record. Or it may require the decision-maker to complete a form, directing them to respond specifically to different aspects of the decision task, much like hiring evaluation forms that are often used by organizations. Each of these variations might impact accountability's effectiveness, particularly as they correlate with different degrees of how rigorous or demanding the justification process is. One can imagine, for example, that an inoculation effect might emerge if the justification process is too easy or vague. In this situation, the prospect of being held accountable no longer prompts the same amount of motivation because the decision-maker knows through previous experience that the justification process is relatively non-demanding and can be satisfied even with low effort. Such possibilities warrant further examination so that organizations can better consider the most effective way to operationalize the justification and review processes.

### **Summary**

The above review takes stock of our current state of knowledge regarding accountability's effectiveness in attenuating biases. The research to date has gone far in identifying specific factors that moderate accountability's effects. We have known for some time that any calls to implement accountability as a bias-reducing strategy must be approached with caution since only highly specific forms of accountability are likely to be effective. When the bias stems from a lack of critical attention or effort, informing decision-makers in advance that they will be accountable for their judgment process to an audience with unknown views is most likely to increase deep, systematic thought. Importantly, motivation is the key mechanism behind accountability's effectiveness. In the face of being reviewed and judged by another, decision-makers are motivated to devote more effort to the task, more thoroughly attend to and integrate available information, and consider alternative outcomes. Decision-makers must also have adequate cognitive resources available to invest in the task at hand, since no amount of motivation can overcome a lack of ability.

Organizations implementing a system of accountability would do well to keep these moderating factors in mind. Yet, many organizations today—corporations, police agencies, courts—are looking for strategies to address the instances of racial bias in particular. Many of the factors identified within the general literature on accountability may very well apply to the context of racial bias. However, there are relatively few accountability studies that have directly examined intergroup bias, let alone racial bias. These studies have produced mixed results, even after taking into account many of the factors identified within the general accountability literature. Racial bias may present unique challenges to accountability for several reasons, including its normative status and salience within society, its affective components, and its often implicit nature. Further research is needed to explore these aspects more closely.

Additionally, the accountability literature could be pushed further in terms of examining different forms of accountability likely to be adopted within different organizational settings. While the ideal situation would be to test the effectiveness of a particular form of accountability within the specific organizational setting, such opportunities are often not available for a variety of reasons. Experimental studies, while admittedly limited in terms of external validity, are appropriate in isolating and identifying general factors that can be applied across a range of contexts. Working within this paradigm, studies should expand to examine accountability as a continuous cycle of decision-making, to test the effects of different motivational goals, and to compare different iterations of the justification process.

In the set of experimental studies described in the next few chapters, I aim to address some of these gaps in the accountability literature by examining the extent to which different forms of accountability can effectively attenuate the influence of explicit and implicit racial bias in the decision-making context.

## CHAPTER 3: STUDY 1

Racial bias in decision-making raises important issues of fairness in several domains, including employment, housing, health care, and the criminal justice system. Implementing a system of accountability is one potential way to reduce bias and increase the legitimacy of decisions within organizations. While procedural fairness is a cornerstone of the American legal system, the adoption of any procedural practice should be preceded by sufficient evidence validating its efficacy in actually creating fairer outcomes. The literature on accountability reviewed in the previous chapter suggests that further research is needed to examine this complicated and nuanced intervention strategy. In the studies presented in this and the following chapters, I attempt to broaden our knowledge about accountability in the hopes of informing policy decisions around the adoption of accountability structures. As a starting point, this first study takes up the comparison between process and outcome accountability and their potential for reducing intergroup bias.

Process accountability exists when “evaluation is based solely on the quality of the *procedure* that a judge or decision maker uses in arriving at a response, regardless of the quality of the outcome of that response” (Siegel-Jacobs & Yates, 1996, p. 2). For example, was the decision well-reasoned and based on the proper criteria? Conversely, outcome accountability focuses solely on the quality of the outcome itself, without regard to the nature or quality of the procedure used to reach the decision. For example, did the decision-maker end up with the ‘right’ or most optimal choice? While both types of accountability may increase motivation, process accountability also provides a suggestion for enhancing performance that is not dependent on the uncertainties associated with outcomes (Siegel-Jacobs & Yates, 1996).

In general, process accountability has been shown to improve accuracy in making probability judgments (Siegel-Jacobs & Yates, 1996), increase the correspondence between evaluations of workers based on an interview and supervisors’ ratings of actual job performance (Brtek & Motowidlo, 2002), and improve performance on a tactical Navy simulation program testing accuracy in classification and shooting decisions (Davis, Mero, & Goodman, 2007). These effects are thought to derive from participants seeking out more information, being more attentive during the task, and engaging in more complex modes of processing (Siegel-Jacobs & Yates, 1996). The evidence on outcome accountability is more mixed, with some studies finding beneficial effects (e.g., Davis, Mero, & Goodman, 2007; Mero & Mitowidlo, 1995), but others finding that outcome accountability actually worsened performance relative to a no-accountability condition (Brtek & Motowidlo, 2002; Siegel-Jacobs & Yates, 1996). Some have suggested these negative effects result from the high levels of stress induced by the demands of reaching the “right” outcome in contexts of uncertainty, which are detrimental to various judgment and decision processes (Siegel-Jacobs & Yates, 1996).

The current study builds on the existing literature in several ways. First, it allows me to test not only the relative effects of process and outcome accountability, but also whether the demands of outcome accountability differentially affect those most prone to stress (i.e., those high in social anxiety). Second, it asks participants to render judgments within the criminal justice context, a context which has seldom been examined by other studies (for exceptions see Bodenhausen et al., 1994 and Lambert et al., 2003 (Study 2)). From a policy perspective, the criminal context is an important one to study given the high stakes involved and the well-documented racial disparities existing within the criminal system. Additionally, the type of



deliberative judgments made by participants in this study are somewhat analogous to the types of decisions made not only by jurors (with respect to judgments of guilt), but also prosecutors and judges (with regard to charging and sentencing decisions).

Third, the current study is one of only a handful of accountability studies that have focused on intergroup bias specifically. Here, judgments are compared for White and Black targets since White-Black disparities within the criminal context are the largest. Additionally, studies have documented prevailing stereotypes linking Blacks with crime-related traits, e.g., aggressiveness and dangerousness (Devine, 1989; Correll et al., 2002), thereby making it more likely for us to see initial evidence of racial bias.

Fourth, unlike most previous studies, the current study includes individual difference measures of explicit bias. These measures tap into not only cognitive stereotypes linking racial groups with crime, but also general attitudes towards different racial groups. By including both types of measures, we can potentially evaluate whether accountability is equally effective in addressing affect-driven and cognitive biases. The criminal context is one where we are likely to see stronger emotional reactions from participants since crime evokes moral judgments and can be seen as posing a threat to the community.

### **Research Questions and Hypotheses**

The primary research question driving this dissertation project is to what extent does holding someone accountable for their decisions reduce racial bias in the decision-making process. For this first study, I am interested in testing the relative effects of different types of accountability, specifically whether it is better to hold someone accountable for the process they use in reaching their decision or for the outcome itself. Although the existing literature examining this distinction is somewhat mixed, the bulk of the evidence would seem to suggest that process accountability is more likely to exhibit the desired effect in reducing bias. Within the context of intergroup bias, it is also possible that outcome accountability may trigger an overcompensation effect since people may focus on reaching an outcome which is less likely to reveal them as being prejudiced (e.g., less harsh outcomes for the Black suspect). Therefore, I hypothesize that the process accountability manipulation will more effectively attenuate the influence of racial biases in participants' decisions.

The main research question will be approached in two ways. First, I will examine whether there is evidence of racial bias in decision outcomes and, if so, can accountability effectively reduce the disparity in these outcomes. I hypothesize that evidence of racial disparities will be found in the control condition, but not in the two accountability conditions.

Two measures of individual differences—motivation to control prejudice and social anxiety—will be examined in relation to this first line of inquiry. I focus on these factors in particular because previous research has shown that a person's motivation to control prejudice can affect the expression of bias (Devine et al., 2002) and that social anxiety can affect accountability's effectiveness in combatting bias (Lambert et al., 2003). I expect that participants low in motivation to control prejudice will show the largest racial disparities. I also expect to find that the accountability manipulations will be less effective among participants high in social anxiety.

The second approach to the research question focuses on whether participants' racial biases are related to their judgments and, if so, can accountability effectively decouple this

relationship. I hypothesize that participants' racial biases will be related to their decisions in stereotype-consistent ways in the control condition, but not in the two accountability conditions.

Within this study, I also examine the potential mechanism behind accountability's effectiveness by testing whether accountability increases levels of cognitive complexity among participants. I expect to find greater task engagement and effort among those participants receiving the two accountability manipulations.

## Methods

### Participants

A sample of 453 undergraduates was recruited through the psychology department using a brief description of the study posted to the research website.<sup>17</sup> Interested students completed the study online at their own convenience, receiving partial course credit for doing so. Thirty-three people (7.3%) were excluded from the analysis based on their responses to the suspicion check question because they indicated the study had something to do with race, bias, or discrimination. Demographics for the remaining sample ( $N = 420$ ) are presented in Table 1.

### Procedure

Participants were asked to take on the role of a prosecutor and review materials for a drug-related criminal case, describing either a White, Black, or racially non-identified suspect. Participants were randomly assigned to one of three conditions: process accountability, outcome accountability, or a no accountability control condition. This 3 (accountability)  $\times$  3 (race of target) between-subjects design resulted in a total of 9 experimental conditions. Participants completed a shortened practice case to become familiar with the task and the types of decisions they would be asked to make. After the practice case, reminder instructions were given to participants to reinforce the accountability manipulations. The target case file consisted of a police report and a description of the relevant law. Participants were asked their perceptions of the case and how they would handle it with regard to filing, determination of guilt, and sentencing. Participants in all conditions were asked to provide the reasoning behind the decisions they made. At the end, they completed a set of questionnaires, manipulation and suspicion check questions, and demographic questions. Participants were then debriefed about the purpose of the study.

Because it is preferable to measure potential moderating variables prior to the experimental task, I included the individual difference measures in a pre-screening packet distributed and completed by a substantial subset of participants. I recruited as much as possible from those who completed the pre-screening packet, but needed to expand recruitment beyond this group. Those who did not complete the pre-screening packet completed the individual difference measures after the experimental task. Analyses showed that, with the exception of the measure for social anxiety, the timing of when participants completed the individual difference measures did not influence scores. Levels of social anxiety had significantly lower ratings after the experimental task than those who completed these measures before the experimental task ( $t(414) = 2.11, p < .05, M$  before = 3.14,  $M$  after = 2.98).

---

<sup>17</sup> The description read: "We are interested in learning about people's ability to understand and apply the law. Specifically, we are interested in how ordinary people think and reason about legal problems. You will be asked to take on the role of a prosecutor, review a criminal case file, and then make some decisions about how to handle the case."

*Table 1: Study 1 Sample Demographics*

Characteristic	<i>n</i>	%
<b>Gender</b>		
Male	119	28.5
Female	298	71.5
<b>Race</b>		
Caucasian/White	126	30.2
African-American	13	3.1
Latino/Hispanic	42	10.1
Asian	204	48.9
American Indian/Alaskan Native	3	0.7
Other	29	7.0
<b>Political Views</b>		
Very liberal	29	7.0
Liberal	178	42.9
Moderate	169	40.7
Conservative	37	8.9
Very conservative	2	0.5
<b>Age</b>		
18	95	23
19	85	21
20	87	21
21	90	22
22+	58	14
<b>Combined income of parents</b>		
Below \$20,000	38	9.3
\$20,000-\$60,000	90	22.1
\$60,000-\$80,000	62	15.2
\$80,000-\$100,000	57	14.0
\$100,000-\$150,000	73	17.9
Greater than \$150,000	88	21.6
<b>Social Class</b>		
Poor	20	4.8
Working class	50	12.1
Lower middle class	45	10.9
Middle class	139	33.7
Upper middle class	133	32.2
Lower upper class	15	3.6
Upper class	11	2.7

**Materials**

**Instructions/accountability manipulation.** All participants were given a set of instructions explaining their role and the nature of the upcoming task.<sup>18</sup> Those randomly assigned

<sup>18</sup> The instructions read: “We are interested in how ordinary people think and reason about legal problems. For this study, we ask that you take on the role of a criminal prosecutor, that is, the person responsible for initiating criminal

to the process accountability condition received an additional instruction focusing evaluation on the process used by participants to reach their decisions, adapted from Davis, Mero, & Goodman (2007). The instruction read: “Bear in mind that you will be held accountable for the process you used to reach your judgment in this case, just as if you were a prosecutor in a real district attorney’s office. Therefore, on completion of the task, you may be asked to justify to us the process you used to reach your decisions.” Those randomly assigned to the outcome accountability condition received an additional instruction focusing evaluation on the ultimate decision made by participants, again adapted from Davis, Mero, & Goodman (2007). This instruction read: “Bear in mind that you will be held accountable for your judgment in this case, just as if you were a prosecutor in a real district attorney’s office. Therefore, on completion of the task, your decisions will be compared to outcomes in similar cases, and you will be asked to justify to us differences between your decisions and that of others.” Those assigned to the control condition received no additional instructions.

**Case report/race manipulation.** Participants were presented a police report containing basic background information about the suspect, an incident summary, and the suspect’s criminal and employment histories (see Appendix A). The narrative described an arrest for possession of cocaine, with the evidence designed to be somewhat ambiguous. The race of the suspect (Black or White) was manipulated in two ways. First, the suspect was identified using a racially stereotypical name (DeShawn Williams or Colin Miller). Second, the race of the suspect was provided as part of the background information on the police report. For the race control condition, the suspect’s name and racial background were excluded.

The practice case completed first also involved simple possession of a controlled substance (methamphetamine). For this practice round, only a brief narrative description of the facts was provided, excluding all background information about the suspect.

**Law instructions.** Participants were told the legal elements required to convict someone for possession of a controlled substance, adapted from California’s jury instructions (see CALCRIM No. 2304 (2014); Appendix B).

## Measures

**Dependent measures.** On the decision form, participants indicated on a 7-point scale how serious they thought the conduct was in this case and how strongly the facts implicated the suspect’s guilt. Participants next decided whether to file charges as either a misdemeanor or felony. Finally, participants were asked to make sentencing recommendations by writing in the appropriate monetary fine, number of community service hours, and days spent incarcerated. For ease of presentation, all means and standard deviations for each of the outcome measures are shown in Appendix C.

**Measures of individual difference.** Participants completed several measures of individual difference, including levels of explicit racial bias, motivation to control prejudice, social anxiety, and general punitiveness.

**Explicit racial bias.** Three measures of explicit racial bias were included. The first was an affective measure of bias using a warmth thermometer. Participants indicated on a sliding

---

cases and handling them as they make their way through the criminal justice system. As an assistant district attorney, you have been given the responsibility of handling all drug-related cases. Your job is to review a case file from an actual criminal case and the relevant California law and answer questions about how you would handle the case as prosecutor.” Participants were then told the type of decisions they would have to make and that they would have the opportunity to go through the process with an abbreviated, hypothetical case.

scale from 0 to 100 how warmly they feel towards several social groups, including Whites and Blacks. For the analyses, scores were calculated by subtracting warmth towards Blacks from warmth towards Whites, so that higher scores signify greater preference for Whites over Blacks. The second and third measures, adapted from Correll et al. (2002), tapped into the cognitive component of bias by measuring crime stereotypes. Participants were asked about their *personal endorsement* of race-crime associations. They were specifically asked to estimate the percentage of Blacks who are dangerous, aggressive, violent, and engage in criminal activities based on their personal beliefs ( $\alpha = .96$ ), and the same for Whites ( $\alpha = .94$ ). Participants were also asked about their *knowledge/awareness* of race-crime associations. Specifically, they were asked to estimate the percentage of Blacks who are dangerous, aggressive, violent, and engage in criminal activities based on what most White Americans would say ( $\alpha = .98$ ), and the same for Whites ( $\alpha = .96$ ). For the analyses, each of these measures was calculated by subtracting scores pertaining to Whites from scores pertaining to Blacks, so that higher scores correspond with the prevailing race-crime stereotype that Blacks are more closely associated with crime.

**Motivation to control prejudice.** Because accountability is thought to influence task performance via motivational channels, it is theoretically important to know participants' level and source of motivation to control prejudice. I used Plant & Devine's (1998) 10-item Internal and External Motivation to Respond Without Prejudice Scales (IMS and EMS), which measure the degree to which people's reasons for avoiding prejudice is driven by either external or internal forces.<sup>19</sup> Those scoring high on the EMS measure are primarily motivated to avoid prejudice out of concern for the negative reactions of others. By contrast, those scoring high on the IMS measure are primarily motivated to avoid prejudice because it is seen as inconsistent with internalized standards that are personally important to the individual. Both scales showed high levels of reliability in the current study (EMS:  $\alpha = .82$ , IMS:  $\alpha = .84$ ).

**Social anxiety.** Levels of social anxiety have been shown in previous research to moderate the effects of accountability. Accountable participants high in social anxiety exhibited a tendency to rely more heavily on previously held racial attitudes and were less able to engage in controlled processing (Lambert et al., 2003). I include the social anxiety subscale of the Self-Consciousness Scale developed by Fenigstein, Scheier, & Buss (1975) ( $\alpha = .78$ ).<sup>20</sup>

**Measure of cognitive complexity.** As a measure of cognitive complexity, I include the 3-item measure used by Mero & Motowidlo (1995), which asks participants to report how engaged they were during the task, how much time and effort they spent thinking about their decisions, and the extent to which they debated between alternative responses ( $\alpha = .70$ ).

**Suspicion and manipulation check questions.** As a suspicion check, participants responded in an open-ended format what they thought the purpose of research was. As a check on the accountability manipulation, participants were asked the extent to which they thought they would have to explain the outcomes they chose and the process they followed to make their decisions (adapted from Brtek & Motowidlo, 2002). As a check on the race manipulation, participants were asked to identify the racial/ethnic group the suspect belonged to from a list of options.

---

<sup>19</sup> Sample items from the IMS scale include, "Being non-prejudiced toward racial minorities is important to my self-concept" and "Because of my personal values, I believe that using stereotypes about racial minorities is wrong." Sample items from the EMS scale include, "I try to act non-prejudiced toward racial minorities because of pressure from others" and "If I acted prejudiced toward people from minority racial groups, I would be concerned that others would be angry with me."

<sup>20</sup> Sample items include, "It takes me time to overcome my shyness in new situations" and "I feel anxious when I speak in front of a group."

**Demographic information.** Participants were asked to provide information about their personal background, including their age, gender, year in college, political views (very liberal to very conservative), race/ethnic identification, the highest level of education achieved by their father and mother, the combined yearly income of their parents, their social class background (from “poor” to “upper class”), and how many years they have lived in the United States.

## **Data Analysis**

The data analysis strategy was modeled off of Green et al. (2007), another study that examined the relationship between bias and another type of deliberative decision (i.e., the diagnosis and proposed treatment of patients). In a two-pronged approach, I first looked at whether outcomes differed by suspect race as predicted in the control condition, with the Black suspect receiving harsher outcomes than the White suspect. To evaluate whether accountability changes this pattern of outcomes, means for the White and Black suspect were compared within each of the accountability conditions. The second phase of the inquiry entailed looking at whether participants’ explicit biases influenced their case decisions using moderated regression analysis. Thus, regardless of whether overall means differed by race condition, I examined whether racial bias influenced these decisions and whether this relationship was changed in the two accountability conditions.

Independent sample t-tests and analysis of variance (ANOVA) tests were used for the continuous measures to compare the means between the White and Black suspects within each condition. Prior to these analyses being conducted, the measures for recommended fine amount, number of community service hours, and days spent incarcerated were log transformed to compensate for skewed data. Extreme outliers for these measures were identified by looking at values beyond three times the interquartile range. Rather than excluding these data, they were winsorized, that is, replaced by values at the extreme outlier boundary. The measure of whether to file charges as a misdemeanor or felony was analyzed using binary logistic regression.

## **Results**

For ease of presentation, results for the non-identified race condition are not shown. While the non-identified race condition was included as a benchmark to see whether any observed racial disparities were due to in-group favoritism or out-group derogation, the primary research focus is on the differences between the White and Black suspects.

### **Manipulation Checks**

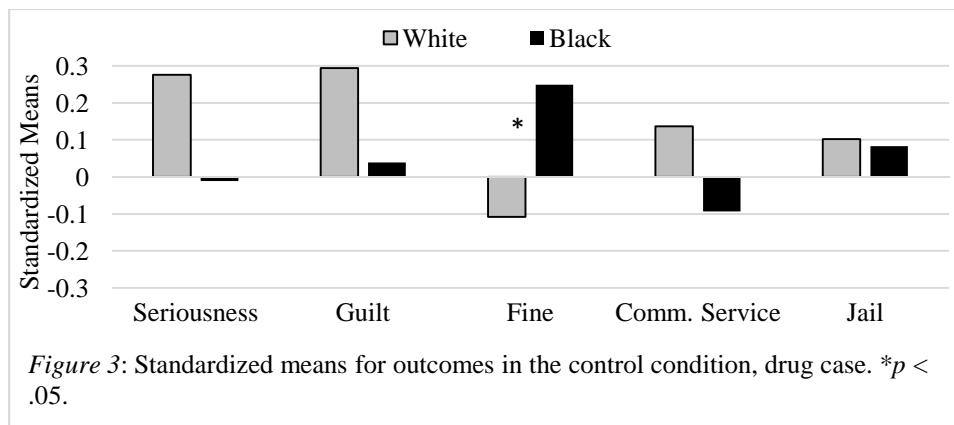
Participants were asked to identify the racial background of the suspect. For those reviewing the White suspect, 62% correctly identified the suspect as White, 20% incorrectly identified the racial background, and 19% reported they could not recall this information. For those reviewing the Black suspect, 58% correctly identified the suspect as Black, 29% incorrectly identified the racial background, and 13% could not recall this information.

To check for the robustness of the accountability manipulation, participants were asked to indicate the extent to which they thought they might have to explain the outcomes they chose as well as the extent to which they thought they might have to explain the process they followed when making their decisions. Analyses showed that both measures had a significant effect by accountability condition. For the outcome question, those in the process condition had

significantly higher expectations of having to explain the outcomes they reached compared to the other two conditions ( $F(2,412) = 16.58, p < .01$ ;  $M$  Control = 4.79,  $M$  Process = 5.68,  $M$  Outcome = 4.75). Similarly, only those in the process condition reported significantly higher expectations of having to explain the process they used in making their decisions ( $F(2,412) = 14.93, p < .01$ ;  $M$  Control = 4.45,  $M$  Process = 5.26,  $M$  Outcome = 4.38). These results indicate that while the process accountability manipulation had its intended effect, the outcome accountability manipulation increased outcome accountability expectations in the process condition but not the outcome condition.

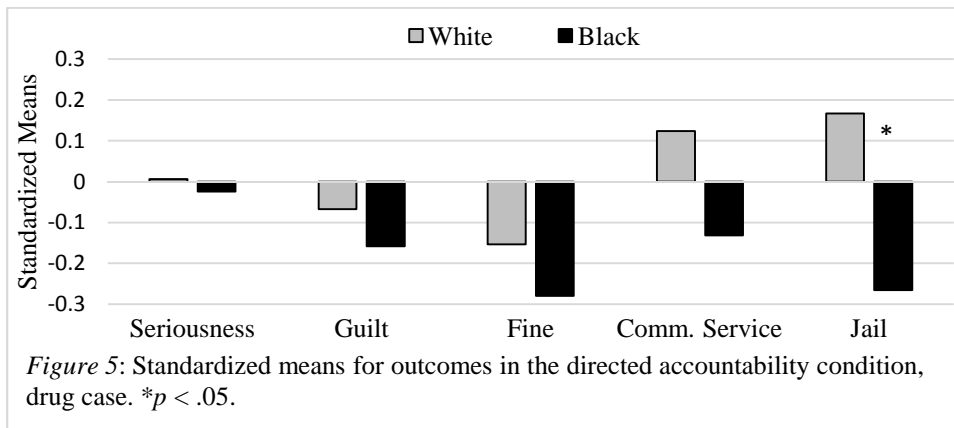
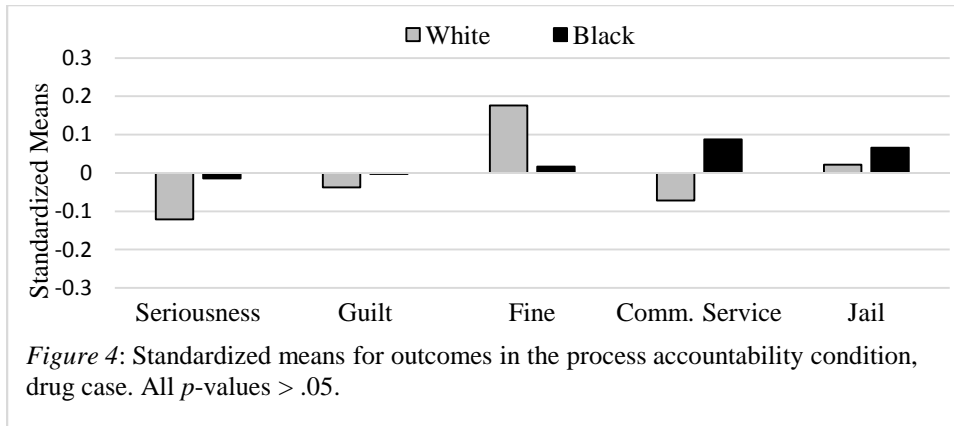
## Analyses of Case-related Outcomes

**Case ratings and decisions.** Overall, participants rated the case as being of moderate seriousness ( $M = 4.78$ ). As intended, the facts surrounding the suspect's guilt were viewed with ambiguity ( $M = 4.94$ ). The standardized means for the continuous outcome measures are shown in Figure 3 for the control condition. In general, we do not observe the expected pattern of the Black suspect receiving harsher outcomes relative to the White suspect. The one exception to this is for the recommended fine amount. Here, consistent with the operation of racial bias, the Black suspect did receive a significantly higher fine amount than the White suspect ( $t(97) = -2.02, p = .05$ ;  $M$  Black suspect = 749.75,  $M$  White suspect = 633.19). This difference remains even after controlling for key participant demographics,  $F(1,81) = 5.81, p = .02$ .<sup>21</sup>



The standardized means for the process and outcome accountability conditions are presented in Figures 4 and 5. Importantly, unlike in the control condition, we do not see harsher fine recommendations for the Black suspect. This pattern is consistent with the predicted effect of accountability in reducing racial bias. In general, the differences by suspect race are even smaller in the process condition, with no significant effects across any of the outcomes. In the outcome condition, though, we do have potential evidence of overcompensation effects. Specifically, participants in this condition recommended significantly *fewer* days of incarceration for the Black suspect relative to the White suspect ( $t(84) = 2.02, p = .05$ ;  $M$  Black suspect = 93.09,  $M$  White suspect = 132.82). This difference persists after controlling for participant demographics,  $F(1,71) = 4.23, p < .05$ .

<sup>21</sup> Demographic measures included as covariates in the model are age, gender, political views, race/ethnic identification, the combined yearly income of their parents, and social class background.



Participants were told the difference between a misdemeanor and felony charge and asked to decide how they would classify the current case. Table 2 shows the percentage of participants from each condition who opted to file the case as a felony. Logistic regression analysis showed no significant differences in treatment between the White and Black suspects within any of the conditions.

Table 2: Misdemeanor vs. Felony Filing Decisions

Accountability Condition	Suspect Race	
	White % Felony (N)	Black % Felony (N)
Control	38% (18)	36% (21)
Process	24% (11)	34% (15)
Outcome	38% (18)	24% (10)

### Analyses of Case-related Outcomes Using Individual Difference Measures

**Motivation to control prejudice.** So far, with a few exceptions, we have not observed differences by suspect race in any of the conditions. One possibility is that differences by suspect race depend on participants' motivations to control prejudice. For example, we may be more likely to find evidence of racial disparity among those who are relatively low in their motivations to control prejudice. Similarly, as suggested by Lambert et al. (2003), participants' levels of



social anxiety may interact with accountability so that those relatively high in social anxiety are more likely to exhibit racial bias. To examine these possibilities, moderated regression analyses were conducted to see whether IMS, EMS, or social anxiety scores interacted with suspect race for the outcomes of interest.

Participants' EMS scores did not moderate the effects of suspect race for outcomes in any of the conditions. Participants' IMS scores did moderate the effects of suspect race on recommended fine amounts in the outcome accountability condition,  $\beta = -.38, t = -2.81, p < .01$ . As shown in Figure 6, those low in IMS tended to give harsher outcomes to the Black suspect, consistent with the expected anti-Black bias. By contrast, those high in IMS actually treated the Black suspect significantly more leniently, consistent with an overcompensation effect. IMS scores also moderated the effect of suspect race on recommended incarceration time in the process accountability condition. Showing a similar pattern, those low in IMS were significantly more likely to recommend longer time for the Black suspect,  $\beta = .33, t = 1.98, p = .05$ . But there is a non-significant trend in the opposite direction for participants high in IMS,  $\beta = -.21, t = -1.33, n.s.$

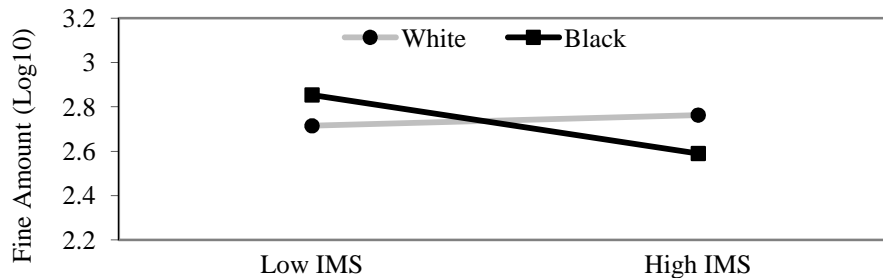


Figure 6: Interaction between IMS and suspect race for recommended fine amount in the outcome accountability condition. Simple effects: White:  $\beta = .09, t = .69, n.s.$ ; Black:  $\beta = -.52, t = -3.02, p < .01$ ; Race difference at 1 *SD* below:  $\beta = .27, t = 1.79, p = .08$ ; Race difference at 1 *SD* above:  $\beta = -.34, t = -2.23, p = .03$ .

As in previous studies on the moderating role of motivations to control prejudice (Devine et al., 2002), the interaction between IMS and EMS scores was also examined. High- and low-motivation groups were constructed based on median splits of IMS (*Mdn* = 5.80, high IMS, *M* = 6.48; low IMS, *M* = 4.92) and EMS (*Mdn* = 4.00, high EMS, *M* = 5.09; low EMS, *M* = 3.05). No significant three-way interactions between IMS, EMS, and suspect race emerged for any of the accountability conditions.

**Social anxiety.**<sup>22</sup> Although work in previous studies has suggested that stress induced by being held accountable can lead those prone to anxiety to actually exhibit greater bias, no evidence of this was found in the present study. Across all conditions, participants' social anxiety scores did not moderate the relationship between suspect race and decision outcomes.

### Relationship between Bias Measures & Participants' Decisions

Thus far we have examined whether outcome differences by suspect race emerged within the three conditions. The second part of the inquiry is whether, regardless of mean differences,

<sup>22</sup> Although social anxiety scores for participants differed by whether the scores were measured prior to or after the task, these scores did not vary by either accountability or race conditions.

participants' racial bias was nonetheless related to their decisions. Means and correlations for the stereotype and attitudinal measures are shown in Table 3. Comparing means for the stereotype measures, participants were more likely to associate Blacks with crime than Whites with crime based on their own personal beliefs,  $t(393) = 6.11, p < .01$ . Similarly, participants were more likely to associate Blacks with crime than Whites with crime based on what they thought most White Americans would say,  $t(394) = 25.72, p < .01$ . Finally, participants reported that in general they felt significantly more warmly towards Whites than Blacks as a group,  $t(398) = -4.64, p < .01$ . All of these differences are consistent with the anti-Black racial bias we would expect to generally observe.

*Table 3: Means and Zero-Order Correlations among Racial Bias Measures*

Measure	Mean (SD)	1	2	3	4	5
1. Black Stereotype Endorsement	28.92 (21.45)	—				
2. White Stereotype Endorsement	24.61 (16.68)	.76**	—			
3. Black Stereotype Awareness	48.12 (24.99)	.67**	.71**	—		
4. White Stereotype Awareness	18.45 (15.40)	.59**	.63**	.44**	—	
5. Warmth, Blacks	65.28 (21.63)	-.41**	-.22**	-.14**	-.15**	—
6. Warmth, Whites	69.71 (20.73)	-.16**	-.22**	-.13*	-.11*	.60**

*N* for each measure: Black Stereotype Endorsement—394; White Stereotype Endorsement—394; Black Stereotype Awareness—395; White Stereotype Awareness—395; Warmth, Blacks—399; Warmth, Whites—399; \*\* $p < .01$ ; \* $p < .05$ .

Moderated regression analyses were conducted to examine whether each of the bias measures influenced decision outcomes. Consistent with prevailing cultural biases, we would expect to see participants' bias scores related to harsher outcomes for the Black suspect (or, alternatively, more lenient outcomes for the White suspect) in the control condition. If accountability is effective in attenuating the influence of racial bias, then we would expect to find no relationship between participants' bias measures and suspect race in the process and outcome conditions. Results are presented by condition and organized by the relative bias measures constructed by taking the difference of the scores for the two groups (e.g., Black-crime association minus White-crime association). In all graphs, values for continuous measures are shown at 1 *SD* above and below the mean. All results hold controlling for participants' demographics unless stated otherwise.

**Control condition.** Looking first at the control condition, personal endorsement of the race-crime stereotype was unrelated to participants' case decisions. For ratings of stereotype awareness, we find a significant interaction between bias scores and suspect race,  $\beta = -.72, t = -3.38, p < .01$ . As shown in Figure 7, while stereotype awareness was unrelated to fine amount for the Black suspect, it was positively related to fine amount for the White suspect. Counterintuitively, this means that the more participants are aware of the Black-crime association relative to the White-crime association, the larger the fine amount for the White suspect. A similar interaction pattern emerges for the number of recommended community

service hours,  $\beta = -.43$ ,  $t = -2.68$ ,  $p < .01$ . As before, stereotype awareness was significantly related to recommended hours for the White suspect,  $\beta = .43$ ,  $t = 2.73$ ,  $p < .01$ , but not for the Black suspect,  $\beta = -.10$ ,  $t = -.85$ , n.s.

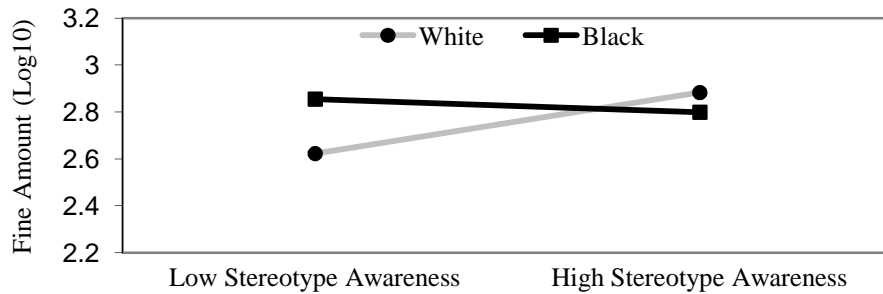


Figure 7: Interaction between stereotype awareness and suspect race for recommended fine amount in the control condition. Simple effects: White:  $\beta = .56$ ,  $t = 3.45$ ,  $p < .01$ ; Black:  $\beta = -.12$ ,  $t = -1.02$ , n.s.

In the control condition, warmth measures did not moderate the relationship between suspect race and any outcome measures. No other outcome measures apart from the ones reported above showed moderation patterns.

**Process accountability condition.** Contrary to expectations, personal endorsement of the race-crime stereotype moderated the relationship between suspect race and guilt ratings,  $\beta = -.37$ ,  $t = -2.31$ ,  $p < .05$ . As shown in Figure 8, a curious pattern emerges, with stereotype endorsement marginally significantly related to guilt ratings for the White suspect, but not significantly related to guilt ratings for the Black suspect. Thus, the more participants associated Blacks with crime than Whites with crime, the more likely they were to believe the White suspect was guilty.

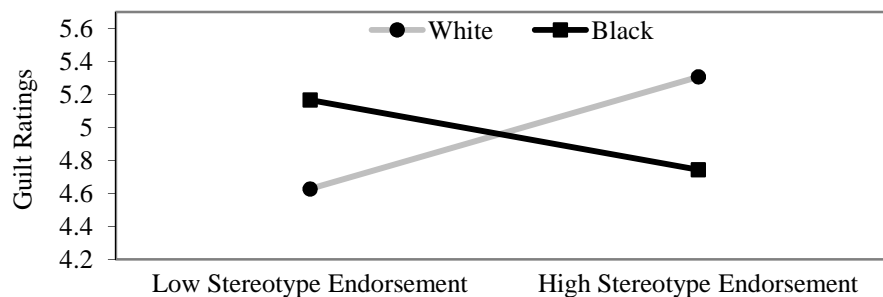


Figure 8: Interaction between stereotype endorsement and suspect race for guilt ratings in the process condition. Simple effects: White:  $\beta = .31$ ,  $t = 1.95$ ,  $p = .06$ ; Black:  $\beta = -.19$ ,  $t = -1.30$ , n.s.

Endorsement of the race-crime stereotype also moderated the relationship between suspect race and the decision of whether to file the case as a misdemeanor or felony,  $B = -1.32$ ,  $Wald = 5.19$ ,  $Exp(B) = .27$ ,  $p < .05$ . Again, endorsement was related to more felony classification for the White suspect,  $B = 1.02$ ,  $Wald = 4.92$ ,  $Exp(B) = 2.77$ ,  $p < .05$ , but not significantly related to felony classification for the Black suspect,  $B = .47$ ,  $Wald = .78$ ,  $Exp(B) = 1.60$ , n.s.

Warmth measures did not moderate outcomes in the process condition, nor were the explicit bias measures significantly related to any outcomes not reported above.

**Outcome accountability condition.** Contrary to hypotheses, personal endorsement of the race-crime stereotype moderated the relationship between suspect race and recommended fine amounts,  $\beta = .30$ ,  $t = 2.45$ ,  $p < .05$ . As shown in Figure 9, personal endorsement corresponded with higher fine amounts for the Black suspect, but not the White suspect.

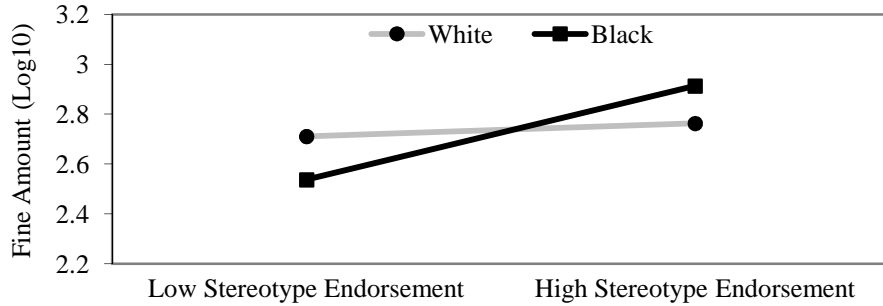


Figure 9: Interaction between stereotype endorsement and suspect race for recommended fine amount for the outcome condition. Simple effects: White:  $\beta = .10$ ,  $t = .86$ , n.s.; Black:  $\beta = .75$ ,  $t = 3.19$ ,  $p < .01$ .

In the outcome condition we also find that awareness of the race-crime stereotype moderated the relationship between suspect race and recommended fine amount,  $\beta = .47$ ,  $t = 2.31$ ,  $p < .05$ . The pattern, though, is opposite from that found for stereotype endorsement. Here, stereotype awareness was negatively related to outcomes for the White suspect, but unrelated to outcomes for the Black suspect (see Figure 10).

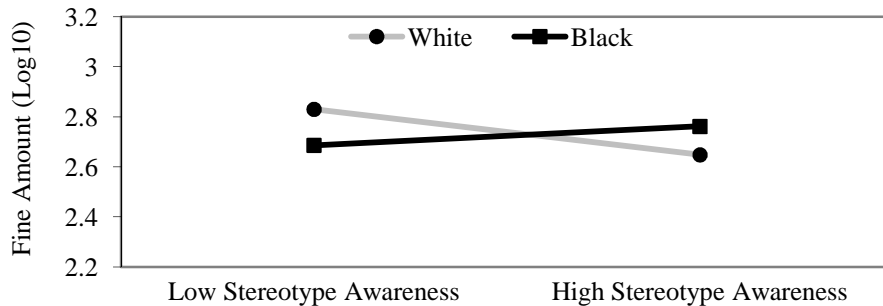


Figure 10: Interaction between stereotype awareness and suspect race for recommended fine amount in the outcome condition. Simple effects: White:  $\beta = -.36$ ,  $t = -2.54$ ,  $p = .01$ ; Black:  $\beta = .15$ ,  $t = .89$ , n.s.

As with the other two conditions, warmth measures did not moderate the relationship between suspect race and participants' decisions. The bias measures did not shown moderation patterns for any other outcomes.

### Cognitive Complexity

Contrary to expectations, an analysis of the 3-item measure of cognitive complexity showed no significant effects by experimental condition ( $F < 1$ ). Thus, there is no evidence that,

based on self-report, accountable participants spent more time and effort on the task, were more engaged, or found themselves debating between alternative responses.

## Summary and Discussion

The data analysis was driven by two primary inquiries: the first was whether outcomes reflect differences by suspect race and the second was whether participants' racial bias was related to decision outcomes. As to the first question, differences by suspect race were found in the control condition for only one of the dependent measures, recommended fine amount. Consistent with an anti-Black bias, participants recommended larger fine amounts for the Black suspect relative to the White suspect. A question raised by this finding is why fine amount was significant as opposed to some of the other outcome measures? One possible explanation is that it is due to ordering effects. In the set of questions asking participants to fill in the most appropriate sentence, the question for fine amount was first. It may be the case that any greater punitiveness towards Blacks was expressed on this first question and not the others that followed. Another possibility is that fines might translate into punishment differently depending on how easily one can afford to pay the fine. Participants may have adjusted their recommended fine amount to the extent they assumed that members from particular racial groups had a greater or lesser ability to pay. However, this does not seem a likely explanation for the observed pattern because, if anything, it is more likely that participants would presume the White suspect was economically better situated, thus justifying a higher fine amount for the White suspect. Since Blacks are more closely associated with being poor (Devine, 1989), it does not seem likely that this same logic could explain the higher fine amounts for the Black suspect.

It is somewhat surprising, however, that we do not observe a more robust pattern across outcomes. One possible explanation for this is that the race manipulation was relatively weak, at least based on participants' responses to the manipulation check question. Yet studies examining race have routinely used stereotypical names as a manipulation with great success (e.g., Bodenhausen et al., 1994; Bertrand & Mullainathan, 2004). In the present study, I also indicated the suspect's race through a demographic checkbox presented as part of his background information. Additionally, looking only at those participants who correctly identified the suspect's race, the pattern of means across the conditions essentially stays the same.

Consistent with accountability's theorized ability to reduce the influence of bias, this racial difference for fine amount was not found in either of the process or outcome accountability conditions. While no significant effects were found in the process condition, there was evidence of reverse bias in the outcome condition, with the Black suspect receiving fewer days of incarceration than the White suspect. One explanation for this overcompensation effect is that being held accountable for outcomes focused participants' attention on appearing non-prejudiced. Thus, they may have been more cautious in assigning higher penalties to the Black suspect. Although recommended incarceration time was the only measure to reach significance, there was a clear pattern across the board with the Black suspect receiving more lenient outcomes.

Additional differences in outcomes by suspect race emerged when participants' internal motivation to control prejudice was taken into account. In both the process and outcome accountability conditions, those low in IMS tended to exhibit an anti-Black bias whereas those high in IMS tended to treat the Black suspect more leniently than the White suspect. On the one hand, this pattern makes sense since those high in IMS have internalized the goal of being

egalitarian and are less likely to exhibit racial bias (Devine et al., 2002). On the other hand, it is interesting that we see a moderation pattern for IMS but not for EMS, since accountability serves as more of an external motivational force. Therefore, we would have expected to see a difference emerge among high- versus low- EMS people in the accountability conditions.

The second question regarding the relationship between racial bias and decision outcomes is more difficult to answer. The observed moderation effects are difficult to interpret because they fail to show a consistent pattern across outcomes. In the control condition, awareness of the race-crime stereotype moderated the effect of race on recommended fine amount and community service hours. Here, the greater the relative awareness of the Black-crime association, the greater the recommended fine and number of community service hours for the White suspect. This pattern is somewhat puzzling because it is inconsistent with the operation of prevailing anti-Black stereotypes. Yet we see this same pattern in the process condition. In the process condition, stereotype endorsement moderated ratings of guilt and the decision of whether to classify the case as a felony. In both instances, endorsement was unrelated to outcomes for the Black suspect, but positively related to outcomes for the White suspect. Again, the more closely participants associate Blacks with crime, the harsher the outcomes for the White suspect.

In the outcome condition, stereotype endorsement was related to fine recommendations in a way consistent with the prevailing race-crime stereotype. Specifically, scores were unrelated to outcomes for the White suspect, but positively predicted fine amounts for the Black suspect. The more people personally associated Blacks with crime (relative to Whites with crime), the higher the fines imposed for the Black suspect. Stereotype awareness was also related to fine recommendations, again in a stereotype-consistent way. Here, while scores were unrelated to outcomes for the Black suspect, they were negatively related to fines for the White suspect. Thus, the more participants associate Whites with crime (relative to Blacks with crime), the harsher the outcomes for the White suspect.

Although all of the conditions showed some evidence of bias at some point in the analyses, it might be said that the process accountability condition fared a little better than the outcome accountability condition. Unlike the outcome accountability condition, there were no significance differences in outcomes by suspect race in the process condition, except for those low in IMS. While the stereotype bias measures did moderate results in both the process and outcome accountability conditions, the pattern of influence was much clearer in the outcome condition. In the process condition, the moderation effects were counterintuitive and difficult to interpret. In the outcome condition, by contrast, both endorsement and awareness of stereotypes moderated the effects of suspect race in ways theoretically consistent with the operation of stereotypes. When evaluating the outcome condition, however, it is important to keep in mind that the accountability manipulation here may not have been as effective, as indicated by responses to the manipulation check question (which were no different than the control group).

## CHAPTER 4: STUDY 2

This second study seeks to partially test the robustness of findings from the first study and examine two additional aspects of accountability: (1) the relative effectiveness of directed and undirected accountability, and (2) the effectiveness of accountability across multiple cases. The previous study utilized what I call undirected accountability, that is, participants were told they would be held accountable and were asked to justify their decisions using open-ended questions that ask for their overall reasoning. This type of accountability is compared to what I call directed accountability, which asks people to justify their decisions in response to pointed questions (e.g., What factors did you take into account when reaching this decision? How did you weigh these factors against one another?). Responding to targeted questions in this way may help facilitate more cognitively complex thought by guiding and prompting the decision-maker to be more aware of his decision process and weigh alternatives. This latter form of accountability more closely simulates filling out a standardized form, which is something organizations of many kinds (e.g., police, prosecutors, managers, front-line workers) frequently use when evaluating case files. Therefore, it is useful to test the effectiveness of this form of accountability directly.

This second study also examines the effectiveness of accountability over time, across multiple cases. Because the effects of accountability are theorized to be triggered by creating merely the *expectation* of being held accountable, no study to my knowledge has examined multiple decisions over time. By doing so in this study, I am treating accountability as a dynamic process rather than a one-time interaction. Considering this dynamic is important when thinking about a system of accountability within an organization and the types of factors that need to be present in order to maintain the effectiveness of the intervention. One concern in particular is that over time the accountability function will become routinized and lose effectiveness as people learn to generate standardized responses. By adding two additional cases to the paradigm used in Study 1, I can note different effects of accountability after participants have already had to justify their decisions on previous occasions.

This second study permits further examination of racial bias within different crime contexts. It is possible that there is something unique to the drug case used in the first study that dampened the expected bias effects in the control condition, especially given shifting attitudes about drug policy in the United States (Pew Research Center for the People & the Press, 2014; Malinowska-Sempruch, 2014). The added case files in this second study—vandalism and robbery—are distinct from the drug case in terms of the harm done (e.g., property and person, respectively).

The expected bias effects may also have been dampened in the first study due to the particular characteristics of the sampled population. Berkeley students in general tend to be more liberal and politically active, and some of the psychology students recruited for the first study may well have studied racial bias as part of their coursework. This second study broadens the population sampled by recruiting participants from the Amazon Mechanical Turk worksite. This will enable me to examine the effects of interest with people from a wider range of ages, geographic locations, political ideology, and social class.

## Research Questions and Hypotheses

As before, the primary research question is to what extent does holding someone accountable reduce evidence of racial bias in the decision-making process. I hypothesize that racial differences in outcomes will be found in the control condition, but not in the two accountability conditions. Taking participants' motivation to control prejudice into account, I expect that larger racial disparities will emerge among low-IMS participants in particular. I also expect that accountability will be most effective in reducing racial disparities among high-EMS participants. I further hypothesize that participants' racial biases will be related to their case decisions in the control condition, but not in the two accountability conditions. A third question of interest for this second study is whether the effects of accountability (if any) are consistent across multiple case decisions. Because the accountability manipulations are essentially twofold—the first being the initial instructions and the second being the actual experience of justifying oneself, we may find larger differences between the control and accountability conditions emerging after the first case file depending on how difficult participants find the justification process. On the one hand, if participants find the justification process relatively easy, they may not feel the need to put forth additional effort on subsequent cases. On the other hand, if participants find the justification process relatively demanding, they are put on notice that they need to expend more effort to adequately address the questions put to them. Because the justification questions in the directed accountability conditions are designed to promote greater awareness of the decision process and consideration of alternatives, I expect the directed accountability condition to be relatively more effective.

## Methods

### Design Overview

As in Study 1, participants were asked to take on the role of a prosecutor and, after reviewing a criminal case file and the relevant law, make a number of decisions about how to handle the case. Rather than a single target case file, participants reviewed three case files: the same drug case used in Study 1, a vandalism/trespass case, and a robbery case. Suspect race was manipulated between-subjects, describing either a White or Black suspect in each case. Participants were randomly assigned to one of three conditions: control, undirected (process) accountability, or directed accountability. This 3 (accountability) x 2 (race of suspect) between-subjects design resulted in a total of 6 experimental conditions.

### Participants

A sample of 361 adults was recruited through Amazon Mechanical Turk, an online marketplace for work. A brief description of the study was posted to the website and interested parties signed up to complete the study at their own convenience in exchange for \$2.25. Twenty-one people (6%) were excluded from the analysis based on their responses to the suspicion check question because they indicated the study had something to do with the race, leaving a sample of 340. Table 4 contains the demographic information for the sample.



*Table 4: Study 2 Sample Demographics*

Characteristic	<i>n</i>	%
<b>Gender</b>		
Male	149	47
Female	171	53
<b>Race</b>		
Caucasian/White	251	79
African-American	30	9
Latino/Hispanic	13	4
Asian	20	6
American Indian/Alaskan Native	2	0.6
Other	3	1
<b>Political Views</b>		
Very liberal	34	11
Liberal	110	34
Moderate	99	31
Conservative	63	20
Very conservative	14	4
<b>Political Party</b>		
Republican	61	19
Democrat	142	44
Libertarian	14	4
Independent	92	29
Other	11	3
<b>Age</b>		
18-25	85	27
26-35	109	34
36-45	67	21
46-55	27	8
56+	30	9
<b>Education</b>		
Did not finish high school	1	0.3
High school diploma	45	14
Some college	124	39
College degree	116	36
Professional/graduate degree	34	11
<b>Household Income</b>		
Below \$20,000	49	15
\$20,000-\$60,000	164	51
\$60,000-\$80,000	50	16
\$80,000-\$100,000	20	6
\$100,000-\$150,000	32	10
Greater than \$150,000	4	1

Note: Depending on the question, demographic information was available for 317-320 participants. Percentages were calculated based on the total number that answered each question.

## Procedure

The same procedures were followed as in Study 1, with a few modifications. In addition to the shortened practice case and the target drug case, participants subsequently viewed a vandalism/trespass case and a robbery case. Only participants in the undirected and directed accountability conditions (not the control condition) were asked to provide the reasoning behind the decisions they made after each case. After reviewing the three target cases, participants completed the suspicion and manipulation checks questions, a memory task, individual difference measures, and demographic background questions. They were then debriefed about the purpose of the study.

## Materials

**Accountability manipulation.** Those randomly assigned to the process accountability condition received the same instruction as the process accountability condition from Study 1, informing them that they would be asked to justify the process used to reach their decisions. Those assigned to the directed accountability condition received the following instruction: “Bear in mind that you will be held accountable for your judgments in these cases, just as if you were a prosecutor in a real district attorney’s office. Therefore, on completion of each task, you will be asked to justify the reasons for your decisions. Specifically, you will be asked to articulate what factors you took into consideration and how you weighed those different factors.” Those in the control condition received no additional instructions. Instruction reminders to reinforce the accountability manipulations were given following the practice task. At the end of each case, those in the undirected condition were asked to respond to three open-ended questions regarding their reasoning behind the suspect’s guilt, whether to charge the case as a misdemeanor or felony, and the most appropriate sentence. Those in the directed condition were asked about the same three subject areas, but were asked to respond specifically to prompts about the factors they used in reaching their decision and how they weighed those factors.<sup>23</sup>

**Race manipulation.** The race of the suspect (White or Black) was manipulated between-subjects through racially stereotypical names and a checked demographic box on the police report. The names Colin Miller and DeShawn Williams were used for the drug case, Andrew Rogers and Darnel Rogers for the vandalism case, and Scott Hill and Tyrrell Hill for the robbery case.

**Case reports.** The first case file, the drug case, was the same one used in Study 1. The additional vandalism/trespass and robbery case files were adapted from vignettes used in previous studies on prosecutors (Frederick & Stemen, 2012; Jacoby et al., 1982; Appendix B). The facts in each case were designed to leave room for ambiguity as to the suspect’s guilt since previous research has shown that racial bias is most likely to operate under conditions of

---

<sup>23</sup> Regarding your decision about the likelihood of the suspect’s guilt in this most recent case . . . How do the key facts of the case support the suspect’s guilt or innocence? What are the strengths and weaknesses of the evidence in this case? Regarding your decision about whether to file the case as a misdemeanor or felony . . . What specific factors did you consider in making this decision? Were there any aggravating or mitigating circumstances present (i.e., factors that either increase or decrease the severity of the crime)? How did you evaluate these factors and weigh them against one another? Regarding your decision about the sentence you think is most appropriate . . . How was the type of sentence you chose more appropriate than the other available options? What factors did you consider when making each of the sentence recommendations in this case?

ambiguity (Gaertner & Dovidio, 2000; Hodson, Dovidio, & Gaertner, 2002; Sagar & Schofield, 1980).

## Measures

**Dependent measures.** Outcome measures included the seriousness of the conduct (7-point scale), the suspect's likely guilt (7-point scale), whether to file charges as a misdemeanor or felony, and the appropriate sentencing recommendations for monetary fine, community service, and time spent incarcerated. Because the latter three measures all related to punishment, a punishment composite measure was constructed by averaging the standardized scores (Drug case  $\alpha = .75$ , Vandalism case  $\alpha = .52$ , Robbery case  $\alpha = .62$ ). For ease of presentation, all means and standard deviations for each of the outcome measures are shown in Appendix C.

**Measures of individual difference.** Participants completed warmth measures towards Whites and Blacks, the Internal & External Motivation to Respond without Prejudice Scales (IMS:  $\alpha = .89$ ; EMS:  $\alpha = .89$ ), personal endorsement of the Black-crime stereotype ( $\alpha = .97$ ) and White-crime stereotype ( $\alpha = .96$ ), and awareness of the cultural Black-crime stereotype ( $\alpha = .98$ ) and White-crime stereotype ( $\alpha = .97$ ). For the analyses, difference scores were constructed for the warmth and stereotype measures to represent the relative bias of participants towards Whites versus Blacks.

**Participant demographics.** Participants were asked to report their age, gender, political views (1 = very liberal, 5 = very conservative), political party, highest education level, household income, and racial/ethnic background.

**Cognitive complexity.** Cognitive complexity was measured in two ways. First, participants completed the same 3-item self-report measure of task engagement used in Study 1 and adapted from Mero & Motowidlo (1995). Added to this scale was a fourth item, which specifically asked how important it was to the participant to appropriately understand and apply the law in each case. Because a scale reliability analysis showed that reliability would be improved by removing the question about debating between alternatives ( $\alpha = .71$ ), the debating question was analyzed separately. Second, participants completed a memory task where they wrote down as many facts as they could recall from each of the case files. To the extent participants processed and engaged with the materials more deeply, the facts from the cases should be better encoded in memory and be easier to subsequently recall.

**Manipulation and suspicion check questions.** To check for suspicion, participants were asked what they thought the research purpose of the task was. They were also asked which racial/ethnic group the suspects belonged to and to what extent they thought they might have to explain their decision process to the researchers.

## Data Analysis

As in Study 1, I first examined outcome differences using independent sample t-test analyses, comparing means for the White and Black suspect within each of the accountability conditions. To look at the relationship between bias measures and outcomes, I use moderated regression analyses to see whether there was an interaction between these measures within the different accountability conditions. To examine the effect of accountability over multiple case decisions, I used a mixed model ANOVA with the case file outcomes as the within-subjects factor and suspect race and accountability conditions as the between-subjects factors. Before

conducting these analyses, the measures for recommended fine amount, number of community service hours, and days incarcerated were log transformed to adjust for skewness in the data. Additionally, extreme outliers (more than three times the interquartile range) were winsorized to avoid undue influence from these data.

## Results

### Manipulation Checks

Participants were asked to indicate the racial background of the suspects. Of those in the White condition, 82% correctly identified the suspect's racial background, 16% incorrectly identified race, and 2% could not recall. The manipulation for the Black suspect was relatively less successful. Only 51% of participants correctly identified the race of the Black suspect, 47% incorrectly identified the racial background, and 1% could not recall.<sup>24</sup>

The accountability manipulation check question showed a significant main effect, with participants in the undirected and directed conditions reporting significantly higher expectations of having to justify the process they used in reaching their decisions ( $F(2,314) = 22.56, p < .01$ ; control:  $M = 4.83$ , undirected:  $M = 6.01$ , directed:  $M = 6.10$ ). Thus, the accountability manipulation had its intended effect.

### Drug Case

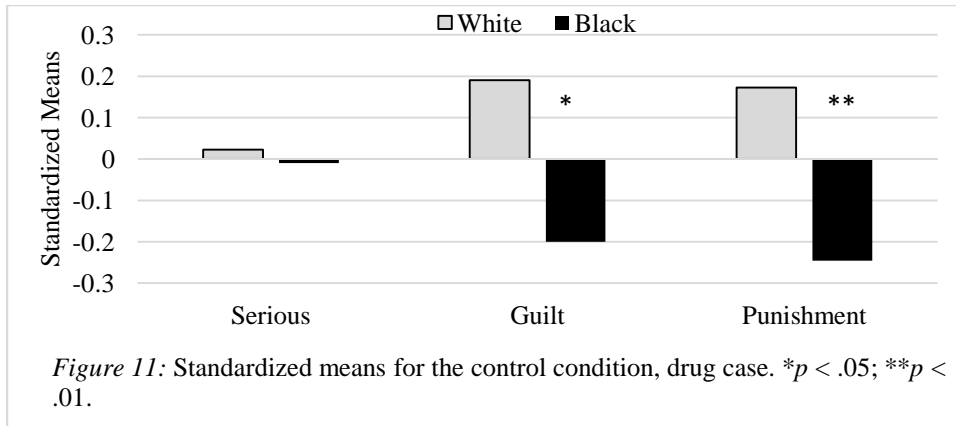
Overall, participants rated the drug case as being of moderate seriousness ( $M = 4.59$ ). They also saw the facts of the case as fairly ambiguous, as indicated by mean ratings of suspect guilt ( $M = 5.04$ ).

### Analyses of Case-related Outcomes

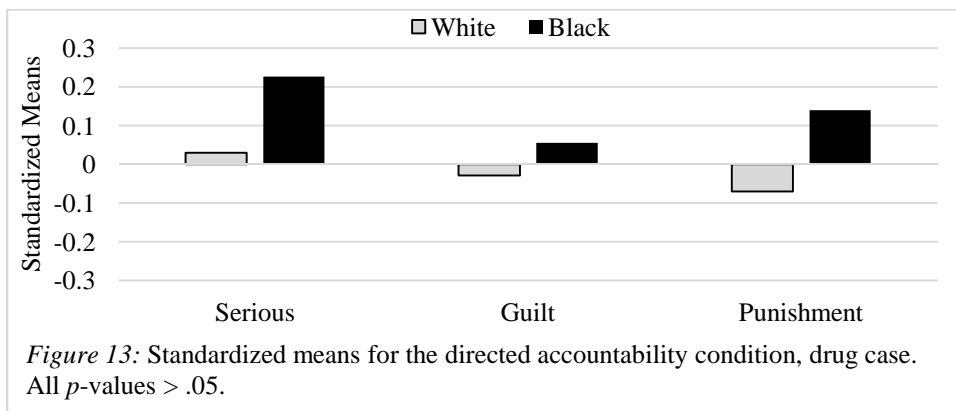
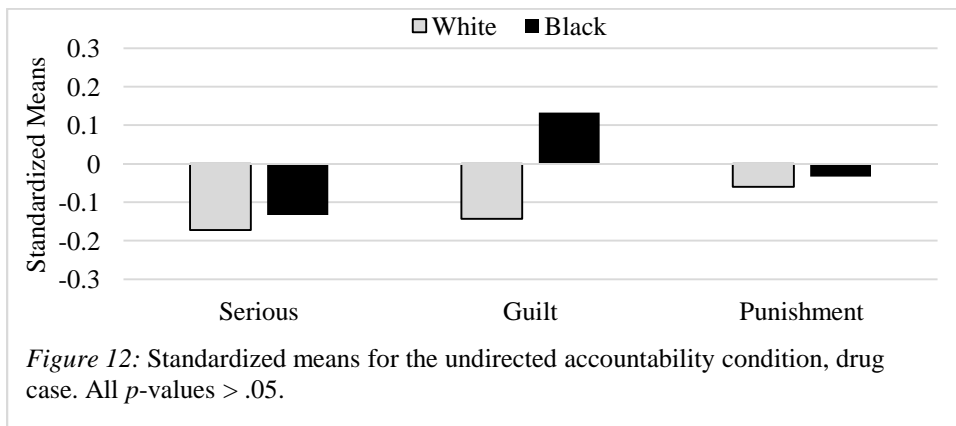
**Case ratings and decisions.** Figure 11 shows the standardized means for the continuous outcome variables in the control condition for the drug case. Significant differences emerged for two of the three measures: guilt ratings,  $t(97) = 2.05, p = .04$ , and punishment,  $t(109) = 2.67, p < .01$ . After controlling for participant demographic variables, these differences remained significant. In both instances, the White suspect received harsher outcomes than the Black suspect. Thus, the racial bias appears to flow in the opposite direction from the expected pattern.

---

<sup>24</sup> A full set of analyses could not feasibly be done with the sample size remaining after excluding participants who incorrectly identified the suspect race. However, a comparison of means between the full sample and one excluding those who identified the suspect race incorrectly shows the same pattern of results for all three case files.



The standardized means for the undirected and directed accountability conditions are presented in Figures 12 and 13. Unlike in the control condition, we do not see any significant differences between the White and Black suspect across outcomes. Thus, creating the expectation of having to justify one's responses seems to have the predicted effect of reducing discrepancies in how suspects from different racial groups are treated.



Participants were also asked whether they would file the case as either a felony or misdemeanor. The percentages of participants deciding to file the case as a felony in each

condition are presented in Table 5. Analyses showed no significant differences by suspect race in any of the accountability conditions.

*Table 5: Misdemeanor vs. Felony Filing Decisions, Drug Case*

Accountability Condition	Suspect Race	
	White % <i>Felony</i> (N)	Black % <i>Felony</i> (N)
Control	26% (15)	33% (17)
Undirected	30% (17)	25% (12)
Directed	32% (21)	44% (26)

In sum, two of the four dependent measures showed evidence of the predicted accountability effect. For suspect guilt and punishment, a racial discrepancy between the White and Black suspects was found in the control condition. Although the racial discrepancy was in the opposite direction, it disappeared in both accountability conditions.

### Analyses of Case-related Outcomes Using Individual Difference Measures

**Motivation to control prejudice.**<sup>25</sup> For the entire sample, the mean for internal motivation to control prejudice (IMS) was 5.73 ( $SD = 1.21$ ) and the mean for external motivation to control prejudice (EMS) was 3.36 ( $SD = 1.55$ ). To examine whether differences by suspect race depended on participants' motivation to control prejudice, moderated regression analyses were conducted using participants' IMS and EMS scores by condition. In the analyses for all the case files, differences by race were examined at 1 SD above and below the variable mean.

In the control condition, IMS scores significantly interacted with suspect race for recommended punishment,  $\beta = -.30$ ,  $t = -2.44$ ,  $p < .05$ . While there is no difference by suspect race among those low in IMS, participants high in IMS gave significantly harsher punishment recommendations for the White suspect relative to the Black suspect (see Figure 14). This pattern is consistent with an overcompensation effect, with high IMS participants attempting to not be prejudiced by giving harsher recommendations to the White suspect and more lenient recommendations to the Black suspect. It should be noted, however, that this interaction effect becomes non-significant after controlling for participant demographic variables. EMS scores did not interact with suspect race for any of the outcomes in the control condition.

<sup>25</sup> While the interaction between participants' IMS and EMS scores are of theoretical interest (Devine et al., 2002), such analyses would have been underpowered given the small number of participants in some of the cells (e.g., five). Therefore, I was unable to examine the interaction of these scores as a potential moderator.

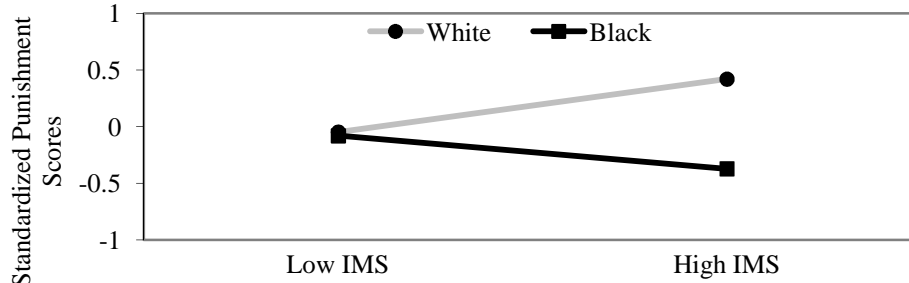


Figure 14: Interaction between IMS scores and suspect race for punishment in the control condition, drug case. Simple effects: White:  $\beta = .28, t = 2.23, p < .05$ ; Black:  $\beta = -.17, t = -1.27, n.s.$ ; Race diff. at 1 *SD* below:  $\beta = -.02, t = -.14, n.s.$ ; Race diff. at 1 *SD* above:  $\beta = -.47, t = -3.63, p < .01$ .

In the undirected condition, participants' IMS scores did not interact with suspect race for any of the outcomes. There was a significant interaction between EMS scores and suspect race for recommended punishment,  $\beta = -.29, t = -2.24, p < .05$ . Among those low in EMS, the Black suspect tended to receive higher ratings than the White suspect, though the difference was only marginally significant,  $\beta = .24, t = 1.73, p = .09$ . Although the opposite pattern is found among high EMS participants, the difference is not significant,  $\beta = -.21, t = -1.47, n.s.$  This pattern is consistent with accountability exerting an external motivational force, with participants high in EMS responding to this motivational force by showing reduced racial disparities. EMS scores, however, did not interact with suspect race for any other outcome variable.

In the directed condition, neither participants' IMS nor EMS scores interacted with suspect race for any of the outcomes.

In sum, we see evidence of an overcompensation or social desirability effect in the control condition, as indicated by harsher outcomes for the White suspect among high IMS participants. Those who are most internally motivated to control their prejudice may have been more cautious in recommending harsher sentences to the Black suspect in order to avoid being prejudiced. By contrast, we see that EMS scores, rather than IMS scores, moderate results in the undirected accountability condition. Supporting the claim that accountability counters the influence of bias, high EMS participants—those most likely to be responsive to accountability's motivational forces—showed no evidence of racial disparities in their decision outcomes. Those low in EMS, however, tended to give harsher recommendations to the Black suspect.

### Relationship between bias measures and participants' decisions.

The correlations for the racial bias measures analyzed in this study are presented in Table 6. Consistent with prevailing cultural stereotypes, participants associated Blacks with crime more than Whites, both in terms of their own personal beliefs,  $t(317) = 4.02, p < .01$ , and their estimate of what most White Americans would think,  $t(315) = 20.61, p < .01$ . Similarly, participants felt more warmly towards Whites than Blacks as a group,  $t(308) = 6.46, p < .01$ .

*Table 6: Zero-Order Correlations among Racial Bias Measures*

Measure	Mean (SD)	1	2	3	4	5
1. Black Stereotype Endorsement	29.25 (22.23)	—				
2. White Stereotype Endorsement	26.06 (18.50)	.78**	—			
3. Black Stereotype Awareness	49.83 (25.59)	.70**	.70**	—		
4. White Stereotype Awareness	24.10 (17.58)	.57**	.71**	.53**	—	
5. Warmth, Blacks	67.36 (22.89)	-.36**	-.12*	-.15**	-.05	—
6. Warmth, Whites	75.12 (19.59)	.18**	.08	.05	.14*	.53**

Note: *N* for each measure: Black Stereotype Endorsement = 317; White Stereotype Endorsement = 317; Black Stereotype Awareness = 316; White Stereotype Awareness = 316; Warmth, Blacks = 308; Warmth, Whites = 309. \*\**p* < .01, \**p* < .05.

Moderated regression analyses were conducted to examine whether any of the bias measures influenced decision outcomes. For these analyses, participants' racial bias was tested using difference scores calculated from the stereotype endorsement, stereotype awareness, and warmth measures.<sup>26</sup> While we would expect participants' racial bias to be related to decisions in the control condition in stereotype-consistent ways, we would expect to find no relationship between the bias measures and outcomes in the two accountability conditions. Results are presented by condition and organized by explicit bias measure. In all graphs, values for continuous measures are shown at 1 SD above and below the mean.

**Control condition.** In the control condition, all three racial bias measures were unrelated to participants' case decisions.

**Undirected accountability condition.** In the undirected condition, stereotype endorsement did not moderate the relationship between suspect race and decision outcomes. There was a significant interaction between stereotype awareness and suspect race for recommended punishment,  $\beta = .33$ ,  $t = 2.18$ ,  $p < .05$ . Specifically, stereotype awareness was positively related to fine amounts for the Black suspect, but negatively related to fine amounts for the White suspect (see Figure 15). This pattern is consistent with the operation of prevailing biases since greater awareness of the Black-crime association meant harsher outcomes for the

<sup>26</sup> For stereotype endorsement and awareness, scores signifying the association of Whites with crime were subtracted from scores signifying the association of Blacks with crime. For the warmth measure, scores for Blacks were subtracted from scores for Whites, so that higher values signify greater warmth towards Whites.



Black suspect, and vice versa. Although the simple slope effects are non-significant, this interaction effect remained significant after controlling for participant demographics.

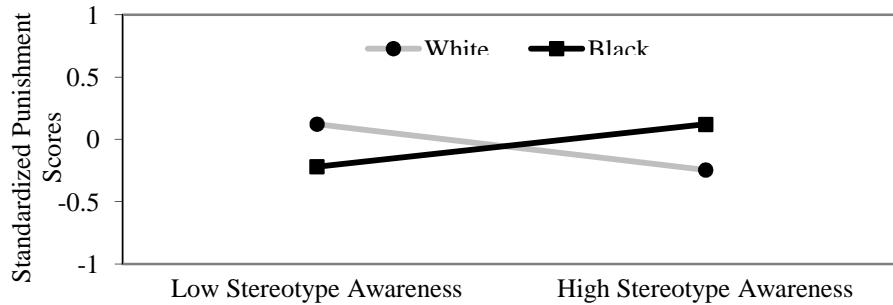


Figure 15: Interaction between stereotype awareness and suspect race for recommended punishment in the process condition, drug case. Simple effects: White:  $\beta = -.23$ ,  $t = -1.51$ , n.s.; Black:  $\beta = -.22$ ,  $t = -1.39$ , n.s.

Participant warmth ratings did not interact with suspect race for any outcomes.

**Directed accountability condition.** Consistent with hypotheses, none of the racial bias measures moderated the relationship between suspect race and outcomes in the directed accountability condition.

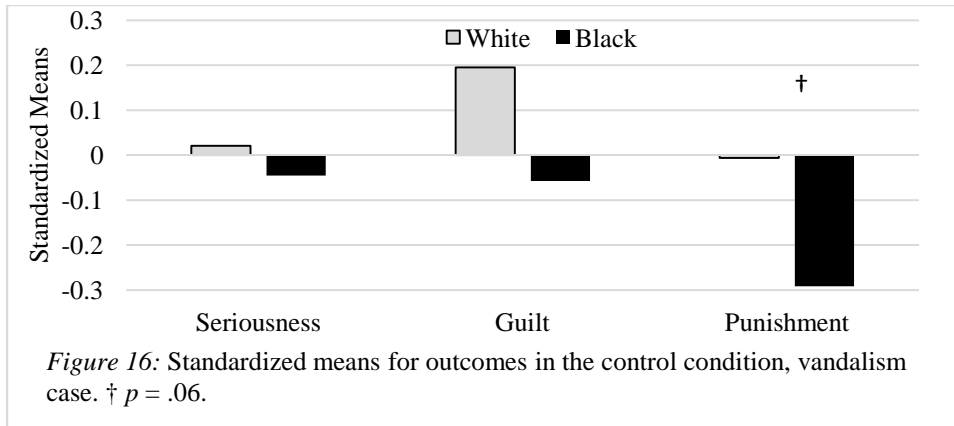
In sum, it is surprising that we do not see significant relationships between the bias measures and suspect race for any outcomes in the control condition. The only significant interaction we see is between stereotype awareness and suspect race for recommended punishment in the undirected condition. The observed pattern of results is consistent with the operation of racial bias, suggesting that the undirected accountability manipulation was not fully successful in insulating decision outcomes from racial bias. It should be kept in mind, however, that the simple slope effects were non-significant. As in the control condition, there were no significant interactions for the bias measures in the directed accountability condition.

### Vandalism Case

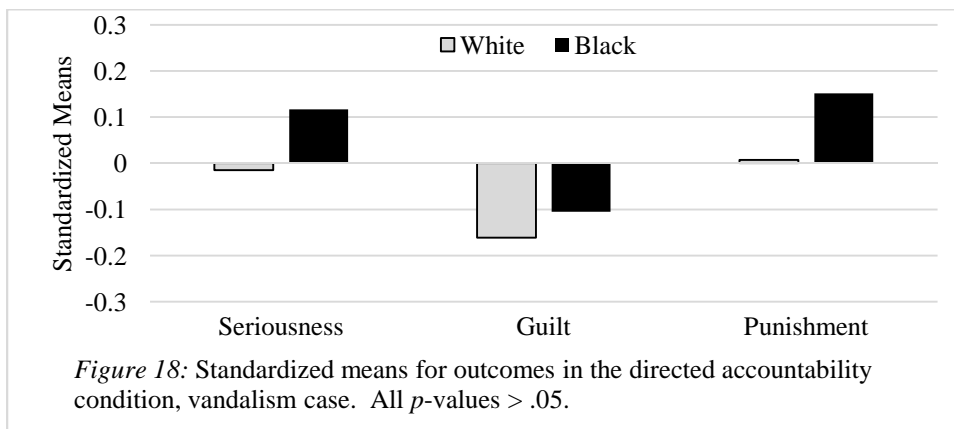
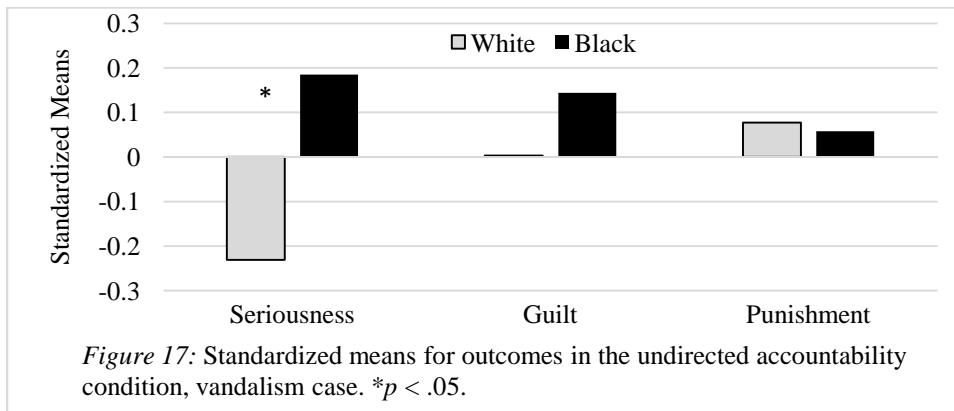
Overall, participants rated the vandalism case as slightly more serious than the drug case ( $M = 4.85$ ,  $SD = 1.25$ ). As designed, participants tended to view the facts of the case as fairly ambiguous, as indicated by guilt ratings ( $M = 5.52$ ,  $SD = 1.39$ ).

### Analyses of Case-related Outcomes

**Case ratings and decisions.** Figure 16 presents the standardized means for the continuous outcome variables in the control condition. The difference by suspect race is marginally significant for recommended punishment,  $t(110) = 1.89$ ,  $p = .06$ . As with the first case file, participants tended to give more lenient punishments to the Black suspect than the White suspect. This difference becomes significant after controlling for participant demographics,  $F(1,100) = 4.05$ ,  $p = .05$ .



The standardized means for the undirected and directed accountability conditions are presented in Figures 17 and 18. In the undirected condition, participants rated the crime as significantly more serious for the Black suspect than the White suspect,  $t(98) = -2.04, p = .04$ . This difference, however, is only marginally significant after controlling for participants' demographic characteristics,  $F(1,81) = 3.43, p = .07$ . No significant differences by race were found in the directed accountability condition.



The percentages of participants deciding to file the case as a felony in each condition are presented in Table 7. Results from logistic regression analyses showed no difference between the

White and Black suspects in either the control or undirected accountability conditions. In the directed accountability condition, however, the Black suspect was significantly more likely to have the case filed as a felony,  $B = .98$ ,  $Wald = 5.37$ ,  $Exp(B) = 2.65$ ,  $p < .05$ .

*Table 7: Misdemeanor vs. Felony Filing Decisions, Vandalism Case*

Accountability Condition	Suspect Race	
	White % Felony (N)	Black % Felony (N)
Control	14% (8)	22% (12)
Undirected	33% (18)	30% (14)
Directed	22% (13)	42% (22)

In sum, we see two different patterns between the control and undirected accountability conditions: a pro-Black/anti-White tendency in the former and a pro-White/anti-Black tendency in the latter. While the racial difference for punishment in the control condition is robust when covariates are included, the difference in the undirected accountability condition for seriousness is non-significant once participant demographics are taken into consideration. We also see evidence of a pro-White/anti-Black bias in the directed accountability when it comes to choices of filing the case as a felony.

### **Analyses of Case-related Outcomes Using Individual Difference Measures**

**Motivation to control prejudice.** Outcome measures were examined in light of participants' IMS and EMS scores to see whether further differences by racial suspect emerged. However, neither measure significantly interacted with any of the outcome measures.

### **Relationship between bias measures and participants' decisions.**

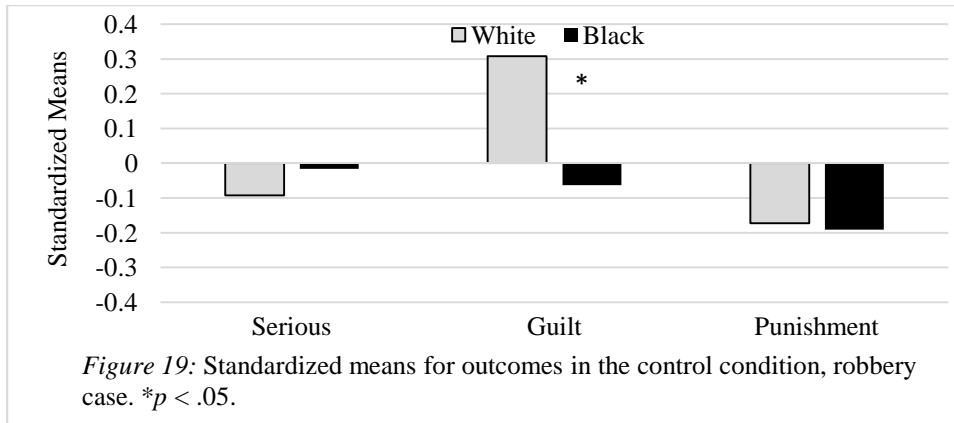
Examining the relationship between participants' bias measures and their decisions in the vandalism case, we find no significant interactions for the two stereotype or warmth measures.

### **Robbery Case**

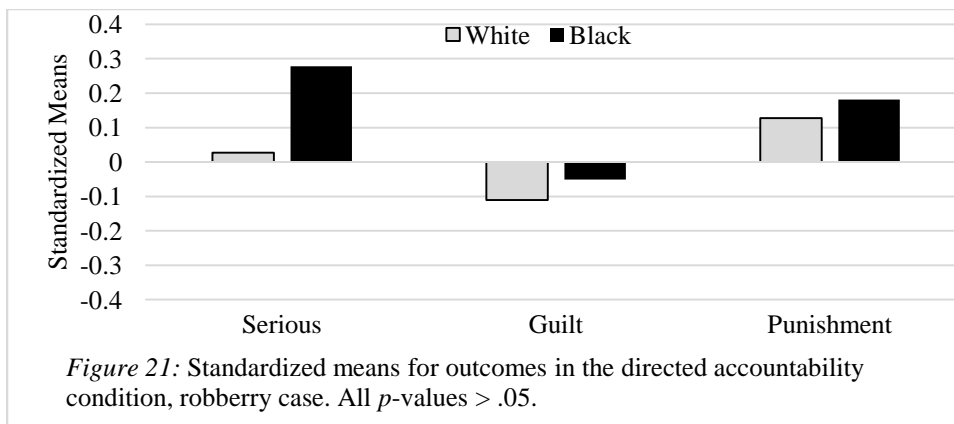
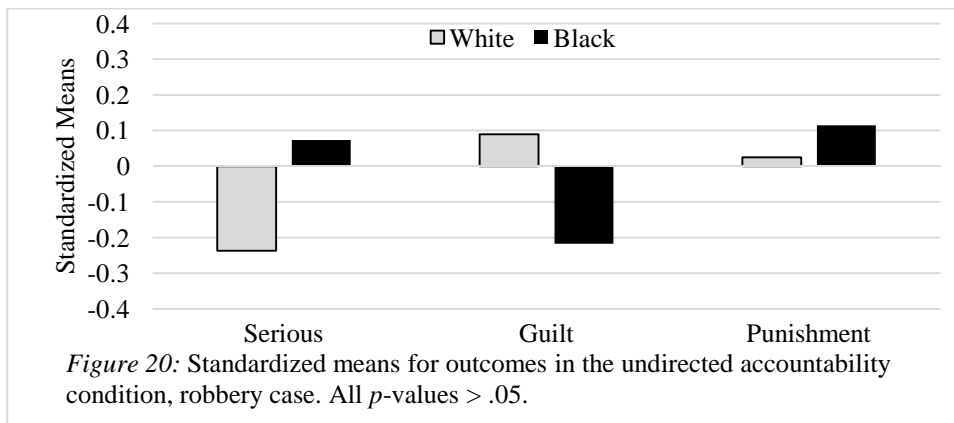
Overall, participants rated the robbery case as more serious than either the drug or vandalism case ( $M = 6.24$ ,  $SD = .84$ ). As with the other two cases, participants tended to view the facts of the case as fairly ambiguous, as indicated by average guilt ratings ( $M = 5.00$ ,  $SD = 1.59$ ).

### **Analyses of Case-related Outcomes**

**Case ratings and decisions.** Figure 19 presents the standardized means for the continuous outcomes in the control condition. The only significant difference by suspect race is for ratings of guilt,  $t(107) = 2.18$ ,  $p = .03$ . Here, participants rated the White suspect as more likely to be guilty than the Black suspect. This difference remained significant after controlling for participants' demographic information.



Figures 20 and 21 present the standardized means for the undirected and directed accountability conditions in the robbery case. Although the undirected accountability condition showed the same pattern for guilt ratings as the control condition, the racial difference was not significant. In fact, none of the racial differences were significant. In the directed accountability condition, the difference in guilt ratings between the Black and White suspects were relatively smaller than in the other two conditions and non-significant. Likewise, there were no significant differences by suspect race for any other outcomes in the directed condition.



The percentages of participants choosing to file the case as a felony are presented in Table 8. Analyses showed no significant differences by suspect race in any of the conditions.

*Table 8: Misdemeanor vs. Felony Filing Decisions, Robbery Case*

Accountability Condition	Suspect Race	
	White % Felony (N)	Black % Felony (N)
Control	90% (51)	83% (45)
Undirected	83% (45)	83% (39)
Directed	86% (50)	86% (44)

In sum, we continue to see evidence of overcompensation or social desirability effects in the control condition, as indicated by higher ratings of guilt for the White suspect relative to the Black suspect. Consistent with accountability attenuating racial bias, we also see that this racial difference disappears in both accountability conditions.

### Analyses of Case-related Outcomes Using Individual Difference Measures

**Motivation to control prejudice.** Analyses examining outcome differences while taking into account participants' IMS and EMS scores revealed no additional significant effects.

### Relationship between Bias Measures and Participants' Decisions

As with the other case files, moderated regression analyses were conducted to see whether participants' racial biases were related to decision outcomes. While we would expect to find racial biases related to outcomes in the control condition, we would not expect to find evidence of these relationships in the two accountability conditions.

**Control condition.** Contrary to expectations, neither stereotype endorsement, stereotype awareness, nor warmth ratings moderated the relationship between suspect race and any of the outcomes.

**Undirected condition.** In the undirected condition, neither stereotype endorsement nor stereotype awareness moderated the relationship between suspect race and any of the outcomes. However, there was a significant interaction between warmth ratings and suspect race for the decision of whether to classify the case as a felony,  $B = 1.51$ ,  $Wald = 5.12$ ,  $Exp(B) = 4.55$ ,  $p < .05$ . Warmth difference ratings were negatively related to the White suspect,  $B = -.61$ ,  $Wald = 2.75$ ,  $Exp(B) = .54$ ,  $p = .10$ , and positively related to the Black suspect,  $B = .91$ ,  $Wald = 2.61$ ,  $Exp(B) = 2.47$ , n.s. While this pattern is consistent with the operation of prevailing biases, the simple slope effects did not reach conventional levels of significance for either suspect.

**Directed condition.** In the directed condition, stereotype endorsement did not moderate the relationship between suspect race and any outcomes. A significant interaction emerged between stereotype awareness and suspect race for the decision of whether to classify the case as a felony,  $B = -1.58$ ,  $Wald = 6.50$ ,  $Exp(B) = .21$ ,  $p = .01$ . While there was no relationship between stereotype awareness and felony classification for the White suspect,  $B = .63$ ,  $Wald = 2.01$ ,

Exp(B) = 1.88, n.s., there was a significant negative relationship for the Black suspect,  $B = -.95$ , Wald = 4.83, Exp(B) = .387,  $p < .05$ . Thus, contrary to the direction of prevailing biases, greater awareness of the Black-crime association was related to a reduced likelihood of classifying the case as a felony for the Black suspect.

Warmth ratings did not moderate the relationship between suspect race and case decisions in the directed condition, nor were moderation patterns found for any outcomes not reported above.

In sum, we did not see the hypothesized relationship between participants' bias measures and case outcomes in the control condition. While there was a significant interaction for warmth ratings and suspect race for felony classification decisions in the undirected condition, the simple slope effects were non-significant. In the directed condition, we found that stereotype awareness was related to suspect race felony classification decisions. Here, though, the pattern is the opposite from what one would expect from the operation of prevailing biases since greater awareness of the Black-crime association was related to a reduced likelihood of felony classification for the Black suspect.

### **Examining Decisions across Case Files**

An additional question of interest was whether the effects of accountability differed across multiple case decisions. In particular, I examined whether differences by suspect race changed across case files using a mixed model ANOVA analysis. With suspect race and accountability condition as the between-subjects factor and the three case file decisions as the within-subjects factor, there were no significant three-way interactions. This suggests that the accountability manipulations did not change the discrepancies between the White and Black suspects across case files.

### **Cognitive Complexity**

The theorized mechanism behind accountability's effectiveness is increased cognitive complexity. However, both self-report and indirect measures of this construct failed to yield any significant effects. An analysis of the 3-item self-report measure for cognitive complexity showed no significant effects across experimental conditions ( $M_s$ : control = 5.87, undirected = 5.86, directed = 5.89;  $F < 1$ ). Likewise, the single item self-report measure for the extent to which participants found themselves debating between alternatives did not significantly differ by accountability condition ( $M_s$ : control = 5.54, undirected = 5.41, directed = 5.61,  $F < 1$ ).

As an indirect measure of cognitive complexity, participants completed a memory task at the end of the experiment where they were asked to write down all the facts they could recall from each of the three case files. I hypothesized that participants who engaged with the materials more deeply and thoroughly would be able to retain more case facts in memory. Two raters, blind to condition, scored the memory task responses by giving a point for each correct case fact. A percentage agreement assessment conducted for each case revealed sufficient reliability between coders: 86.5% for the drug case, 89.5% for the vandalism case, and 89.2% for the robbery case. In none of the cases was there a significant difference in the total number of facts recalled across accountability conditions ( $M_s$ : control = 27.29, undirected = 25.85, directed = 25.71).

## Summary and Discussion

For the drug case, we saw evidence of racial disparities in outcomes in the control condition. But the racial disparities were in the opposite direction from what would be expected given prevailing biases since the White suspect tended to receive harsher outcomes relative to the Black suspect. We tended to see this overcompensation effect particularly among high-IMS participants. Surprisingly, bias measures were unrelated to suspect race for all outcomes in the control condition.

This relative harshness towards the White suspect was not observed in either the undirected or directed accountability conditions. In fact, no significant racial disparities were observed in these conditions except when participants' EMS scores are taken into account in the undirected condition. In this condition, EMS moderated the relationship between suspect race and recommended punishment. While we tended to see a bias-consistent racial disparity among those low in EMS, we saw no such racial gap among those high in EMS. One interpretation for these results is that accountability, as an external motivational force, was effective in reducing bias among those most responsive to external factors for controlling prejudice. In terms of the relationship between the bias measures and case decisions, there was a significant interaction for stereotype awareness within the undirected accountability condition. However, the simple slopes for both the Black and White suspect were non-significant. Bias measures did not moderate any results in the directed condition.

For the vandalism case, we again saw some evidence of an overcompensation effect in the control condition, as indicated by more lenient punishment recommendations for the Black suspect relative to the White suspect. Contrary to expectations, we did not see any moderation patterns for participants' motivation to control prejudice or any relationship between bias measures and case decisions in the control condition. In the process condition, the opposite pattern emerges in terms of racial disparities, with the crime perceived as more serious for the Black suspect than for the White suspect. In the directed condition, we also see a pattern consistent with prevailing racial bias, with the crime more likely to be classified as a felony for the Black suspect. As with the control condition, there was no moderation pattern for motivation to control prejudice or any relationship between bias measures and case decisions.

For the robbery case, we continued to see evidence of an overcompensation effect in the control condition, with the White suspect seen as more likely guilty than the Black suspect. Contrary to expectations, there was no moderation pattern for motivation to control prejudice or any relationship between bias measures and case decisions. In the undirected and directed accountability conditions, we do not find any significant racial disparities in case outcomes, suggesting that the manipulations were successful in this regard. In the directed accountability condition, however, stereotype awareness moderated the relationship between suspect race and felony classification. The pattern for this interaction was inconsistent from what one would expect with prevailing racial biases. Nonetheless, this suggests that the directed accountability manipulation was less effective in decoupling the relationship between participants' racial bias and their case decisions.

Based on the results from the drug and robbery cases, we might conclude that accountability was effective in reducing racial disparities in outcomes. Case outcomes from the vandalism case, however, caution against this conclusion because we do see some evidence of racial disparities consistent with prevailing biases in both accountability conditions. Further

caution is warranted because stereotype awareness was related to suspect race for one of the case decisions in the directed condition in the robbery case.

One of the unexpected findings from all three case files is that the racial disparities observed in the control condition were not in the expected direction. One possible explanation is that participants may have been concerned with appearing or acting prejudiced and, therefore, erred on the side of caution by making decisions that were relatively more favorable to the Black suspect. All of the analyses were run excluding participants who indicated on the suspicion check question that the study had something to do with race, yet it is possible that participants suspected the study's nature on an implicit level or simply failed to report their suspicions. Participants' concern with being politically correct might also partially account for the large difference we see between those who correctly identified the suspect as being White (82%) and those who correctly identified the suspect as being Black (51%). Perhaps those viewing the Black suspect were more reticent to label the suspect as Black because doing so would raise concerns about race.

Regardless of the reason for the unexpected pattern of results, the nature of the bias that accountability attenuated in these cases is somewhat different than the kind of racial bias the study intended to target. One previous study examining people's use of race when political correctness norms were made salient found that accountability did not counteract the tendency of decision-makers to prefer a Black over a White candidate (Norton, Vandello, & Darley, 2004, Study 5). By contrast, there was evidence found in the present study that accountability did counteract decision-makers' tendency to give harsher outcomes to the White suspect in the control condition.

Contrary to the hypothesis regarding the mechanism behind accountability, there was no indication on either direct or indirect measures that accountable participants engaged with the task more than non-accountable participants. Also contrary to hypotheses, there was no indication that the accountability manipulations affected the magnitude of the race disparities across multiple case files.



## CHAPTER 5: STUDY 3

Study 2 examined the relative effects of undirected and directed accountability across three criminal case files. Contrary to expectations, we found that participants in the control group tended to give harsher outcomes to the White suspect on several measures. There was mixed evidence that accountability effectively countered the influence of race on participants' case decisions.

This third study again examines the effect of undirected and directed accountability, but with several modifications. First, whereas the first two studies compared a White versus a Black suspect, the present study compares a White versus a Hispanic/Latino suspect. This was done, in part, because the Black-White disparity within the criminal justice context is more widely known and, therefore, presenting a Black suspect may have triggered social desirability concerns among participants not necessarily captured by the included suspicion-check question. Equally important, though, including a Hispanic suspect in this study begins to fill a long-standing gap in the research, which has been largely defined by a Black-White dichotomy. Relatively few studies within the bias literature have focused on Hispanics as a target group. Though to a lesser extent than Blacks, Hispanics are also overrepresented within the criminal justice system and face discrimination in other social domains as well. Examining Hispanics as a target group is important given the increasing representation of this group within the U.S. population. For example, from 2000 to 2012, the Hispanic population increased by 50%, bringing the total number to 53 million people (Brown, 2014).

Second, extensive pre-testing of criminal case files resulted in the identification of two factual scenarios which most consistently exhibited the baseline pattern of racial bias we would expect to see in the control condition, that is, with the Hispanic suspect receiving harsher outcomes relative to the White suspect. Using these two case files, I could directly test the effect of accountability on racial bias *per se* rather than racial disparities arising out of social desirability or avoidance-of-prejudice concerns.

A third change from the previous study is that the suspect information provided to participants was simplified to include the person's name, age, and photo. Fourth, as explained in more detail below, changes were made to the dependent measures to constrain potential outliers and include evaluations of the suspect's character. Finally, the first two studies only examined explicit measures of bias. But claims about accountability's effectiveness have encompassed implicit (unconscious) forms of bias as well. Therefore, two measures of implicit bias were included in the present study to be able to examine these claims.

### Research Questions and Hypotheses

To answer the question of whether accountability can reduce racial bias in the decision-making process, I followed the same two lines of inquiry used in the first two studies: 1) an examination of racial disparities in case outcomes, and 2) an examination of whether participants' racial biases were related to their decisions. I hypothesized that the Hispanic suspect would receive harsher outcomes relative to the White suspect in the control condition, but not in the two accountability conditions. As before, I also examined outcomes in light of participants' internal and external motivation to control prejudice in order to see whether further racial disparities emerge after taking these individual differences into account. I further hypothesized

that we would see participants' explicit and implicit racial biases related to case decisions in the control condition, but not in the two accountability conditions.

As in the previous study, I examined whether accountable participants approached the task with greater effort and cognitive complexity, and whether the effectiveness of accountability changed over the course of the two case files.

## Methods

### Participants

Participants were recruited through Amazon Mechanical Turk using the same study description as before. While 589 participants evaluated the two criminal case files, only 576 participants completed the measures of legal attitudes and explicit bias. Forty-four participants (7.5%) were excluded from all analyses because their responses to the suspicion check question indicated they believed the study had something to do with race. Due to technical difficulties, only 381 participants had complete data on the two sets of implicit bias measures using the Go/No-Go Association Task (GNAT). Of these, 64 participants had at least one of their implicit scores dropped from the analysis because they had a  $d'$  score of 0 or less, indicating the participant was either unable to discriminate any signal from noise or was not performing the task as instructed (Nosek & Banaji, 2003). The overall sample profile was consistent with the previous study (58% female, 79% White,  $M$  age = 34, 40% with college degree, 35% with some college, 63% identifying with the liberal side of political spectrum, 40% with household income between \$20K-\$50K).

### Procedure

After reading the initial instructions<sup>27</sup> and familiarizing themselves with the type of decisions they would be making, participants reviewed either a White or Hispanic suspect across two criminal case files. After reading the relevant law and completing the decision form for each case file, those in the two accountability conditions also provided justifications for their decisions either through general open-ended questions (undirected condition) or a set of targeted questions (directed condition). Participants then completed the suspicion-check question, the manipulation check questions, measures of legal attitudes, a memory task, measures of explicit bias, and measures of implicit bias. Demographic information was collected at the very beginning of the study, with the exception of yearly income and racial/ethnic identity, which were measured after the completion of the explicit bias measures. Participants were thanked and debriefed at the end.

---

<sup>27</sup> The prosecutor's instructions for this study vary slightly from the previous two studies: "For this study, you will take on the role of a criminal prosecutor. Prosecutors initiate criminal cases and handle them as they progress through the criminal justice system. As a prosecutor, your duty is to protect the community and advocate for victims by enforcing the law and ensuring justice is done. Your job will be to review two case files and answer questions about how you would handle each case. All your responses will remain anonymous." Additionally, to prime the race-crime stereotype, each instruction page contained crime-related images taken from the Internet, some of which specifically depicted a Hispanic or minority target. Unlike the first two studies, participants did not initially review a practice case.

## Materials

**Race manipulation.** Suspect race was manipulated in two ways. First, the suspect's name was stereotypical of their racial group, specifically "Charles Roderick" and "Michael Hamilton" for the White suspects and "Carlos Rodriguez" and "Miguel Hernandez" for the Hispanic suspects. Additionally, headshot photos of the suspects were presented against a neutral background, resembling a booking photo. Headshots for the White suspects were taken from a pre-existing database of photos used in previous research.<sup>28</sup> Headshots for the Hispanic suspects were edited from images available on the Internet. For each case file, the photos of the White and Hispanic suspect were paired based on independent ratings of perceived age and stereotypicality.

**Case reports.** Through extensive pre-testing, two case files were identified as most consistent in producing the expected racial disparity in the control condition. The first case file described a version of the same vandalism case used in Study 2, with four males breaking into a school and destroying property. The second case file was adapted from Sommers & Ellsworth (2000) and described a simple battery in a bar between the suspect and a female victim (see Appendix A).

**Legal instructions.** Participants read about the elements of each crime through jury instructions currently used in California for vandalism and battery (see CALCRIM Nos. 2900 and 960 (2014), respectively; Appendix B).

**Accountability manipulation.** No changes were made to the accountability manipulations from Study 2. Those in the two accountability conditions received the instruction: "Bear in mind that you will be held accountable for your judgments in these cases, just as if you were a prosecutor in a real district attorney's office." Those in the undirected accountability condition received the additional instruction, "Therefore, on completion of each task, you will be asked to justify the process you used to reach your decision." Those in the directed accountability condition were given the additional instruction, "Therefore, on completion of each task, you will be asked to justify the reasons for your decisions. Specifically, you will be asked to articulate what factors you took into consideration and how you weighed those different factors." Those in the control condition received no additional instructions.

**Demographic Information.** Participants were asked to report their gender, age, highest level of education obtained, political view (very liberal to very conservative), estimated yearly income, and racial/ethnic background.

## Measures

**Dependent measures.** Participants were asked to rate the seriousness of the conduct in this particular case (1=not serious at all; 7=very serious), how strongly the facts implicated the suspect's guilt (1=not at all; 7=very much so), the extent to which they believe the suspect is an aggressive or violent person and the extent to which they believe the suspect has a poor moral character (1=not at all; 7=very much), whether to file the case as a felony or misdemeanor, how severe the punishment should be in this case (1=not at all severe; 7=very severe), what amount of monetary fine is an appropriate punishment (9-point scale ranging from "less than \$250" to "greater than \$2000"), and what amount of time spent in jail is an appropriate punishment (9-point scale ranging from "less than 15 days" to "greater than 180 days"). Because aggressiveness and poor moral character both reflected attributions of the suspect's character, these two items

---

<sup>28</sup> This database was compiled and made available by Dr. Jennifer Eberhardt.

were combined into a single composite (vandalism:  $r = .67$ ; battery:  $r = .79$ ). Similarly, a composite punishment measure was constructed by averaging punishment severity, fine amount, and jail time (vandalism  $\alpha = .67$ ; battery  $\alpha = .84$ ). For ease of presentation, all means and standard deviations for the outcome measures are shown in Appendix C.

**Explicit measures.** Two explicit attitudinal measures were included in this study. The first is a warmth thermometer measure, with participants indicating how warmly they feel towards seven different social groups, including Whites/European Americans and Latino Americans/Hispanics (0=very cold; 100=very warm). Participants' relative bias was used in the analyses by subtracting warmth towards Whites from warmth towards Hispanics. For the second measure, participants indicated the extent to which they prefer White Americans versus Latino/Hispanic Americans (1=somewhat prefer Whites over Latinos/Hispanics; 5=somewhat prefer Latinos/Hispanics over Whites).

To measure awareness of cultural stereotypes between different racial/ethnic groups and crime, participants estimated the percentage of Latinos/Hispanics and White Americans who are aggressive, violent, engage in criminal activities, and break the law based on their perception of what most White Americans would estimate (see Correll et al., 2002). Responses to these four items were averaged together for each target group (Hispanics:  $\alpha = .95$ , Whites:  $\alpha = .94$ ). Participants' relative awareness of the prevailing race-crime stereotype was used in the analyses by subtracting estimates for Whites from estimates for Hispanics.

The last measure of explicit bias was the racial bias subscale of the Pretrial Juror Attitude Questionnaire ("RB-PJAQ"; Lecci & Myers, 2008). The RB-PJAQ consists of 4 items and has been shown in previous research to predict overall verdict tendencies (Lecci & Myers, 2008).<sup>29</sup> One item was dropped from the scale in order to increase reliability of the measure ( $\alpha = .75$ ).<sup>30</sup>

To measure participants' motivation to control prejudice, the Internal & External Motivation to Respond without Prejudice Scales (IMS/EMS, Devine et al., 2002) were used. (IMS:  $\alpha = .84$ ; EMS:  $\alpha = .85$ ).

**Implicit measures.** The Go/No-go Association Task (GNAT, Nosek & Banaji, 2001) was used to assess participants' attitudes toward Whites and Hispanics as well as the extent to which they stereotypically associate each target group with crime. As with other measures of implicit social cognition, the GNAT is designed to assess automatically activated associations without the person's conscious awareness or control. In a series of tasks, participants identified words belonging to a particular category as quickly as possible. Both target and distractor words are shown one at a time in the middle of the computer screen and participant decided whether the word (e.g., "felon") belonged to the target category (e.g., "criminal").<sup>31</sup> Participants hit the space bar when the word matched the target category, but did nothing when there was no match. Feedback was given on each trial telling participants whether they correctly identified the word. For the practice rounds, participants matched words to a single target category in order to familiarize themselves with the task procedures and the words comprising that particular category. During the test rounds, two categories were presented at the top of the screen and participants identified the words belonging to either category. Thus, for the race-crime stereotype

---

<sup>29</sup> Sample items include, "The large number of African Americans currently in prison is an example of the innate criminality of that subgroup" and "Minority suspects are likely to be guilty, more often than not."

<sup>30</sup> The item dropped from the scale was "The defendant is often a victim of his own bad reputation."

<sup>31</sup> For the race-crime GNAT, the target words were: felon, criminal, offender, suspect, and convict. For both GNATs, the distractor words were: table, door, month, contents, context, seat, village, paper, moment, and errand. Target names for the Hispanic category were: Juan, Javier, Pedro, Manuel, Alejandro, Pablo, Fernando, and Ricardo. Target names for the White category were: Curtis, Frank, Brad, Paul, Jay, Donald, Steve, and Ted.

GNAT, the categories “Criminal” and “White” were paired together for one set of trials while the categories “Criminal” and “Hispanic” were paired together for another set of trials. For the attitudinal GNAT, the categories “Good” and “Bad” were each paired with the categories “White” and “Hispanic” for different test rounds. Each test round consisted of 60 trials and participants had 700 milliseconds to indicate whether the word matched the category. The presentation order of the test rounds were randomized across participants.

The GNAT is based on signal detection theory and uses participants’ errors to determine differences in sensitivity to different pairing conditions. Participants’ ability to discriminate signal (target words) from noise (distractor words) should be facilitated when the two presented categories are more strongly associated. Thus, the extent to which the category “Hispanic” is associated with “good” versus “bad” should be reflected in the relative ease of discriminating Hispanic names when also identifying one versus the other evaluative attribute. For the analyses, participants’ biases were calculated by subtracting scores for the White category from scores for the Hispanic category for both the good/bad and stereotype GNATs.

**Suspicion and manipulation check questions.** As a suspicion check, participants were asked to write in an open-ended format their understanding of the study’s purpose. As a check on the accountability manipulation, participants were asked to report the extent to which they thought they might have to explain the process they used in reaching their decisions to the researchers (1=definitely did not believe I would have to explain; 7=definitely believed I would have to explain). As a check on the race manipulation, participants were asked to identify the racial/ethnic group the suspects belonged to.

**Cognitive complexity measures.** Two measures of cognitive complexity were included in this study. The first is the same 3-item scale ( $\alpha = .77$ ) used in Study 2 asking participants about their level of engagement during the task, how much time and effort they spent thinking about their decisions, and how important it was to them to reach the best possible decision in each case. A single item was also included asking participants the extent to which they found themselves debating between alternative responses.<sup>32</sup> A second, indirect measure of cognitive complexity was a memory task where participants were asked to sort true and false items from a list of 12 statements about each case file. The number of correct responses for each case file was summed for each participant. The number of correct responses served as a proxy for the extent to which participants attended to and engaged with the task.

## Results

### Manipulation Checks

The manipulation of suspect race was largely effective, with only 9% of the sample incorrectly identifying the race of the suspects they had reviewed. For the accountability manipulation check question, there was a significant effect by accountability condition,  $F(2,537) = 45.77, p < .01$ . Those in the control condition were significantly less likely to believe they would have to explain the process used to reach their decisions than participants in the undirected or directed accountability conditions ( $M_s$ : Control = 4.08, Undirected = 5.65, Directed = 5.48).

---

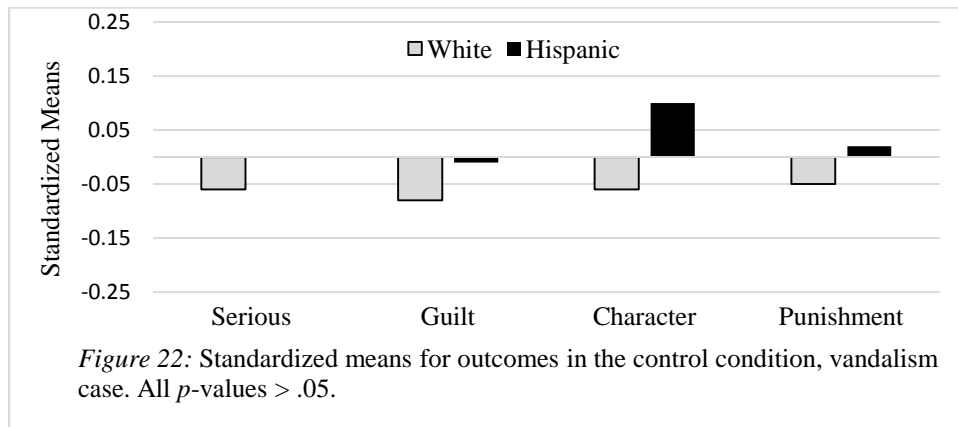
<sup>32</sup> This item was examined separately because it did not reliably fit as part of the three other self-report items for cognitive complexity.

## Vandalism Case

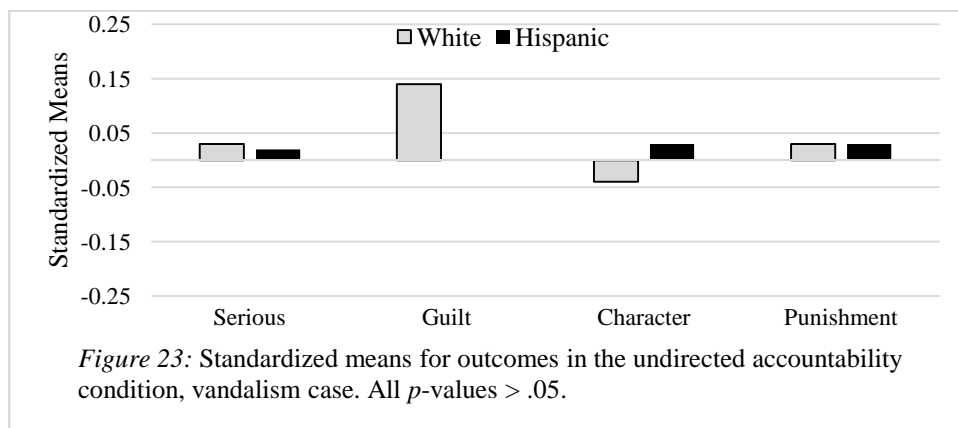
### Analyses of Case-related Outcomes

Overall, participants rated the vandalism case as being of moderate seriousness ( $M = 4.39$ ) and guilt ratings indicated that the presented facts were viewed as fairly ambiguous ( $M = 4.78$ ).

**Case ratings and decisions.** The standardized means for the continuous outcome measures in the control condition are shown in Figure 22. Contrary to expectations based on the pre-testing, all differences by suspect race were small and non-significant.



The differences between the two suspects remained small and non-significant within the two accountability conditions (see Figures 23 & 24).



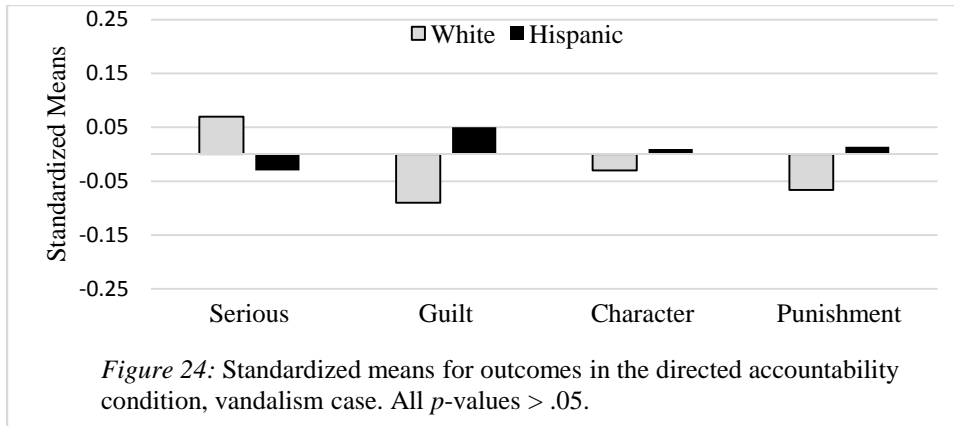


Table 9 provides the descriptive data for participants' choice of whether to file the case as a misdemeanor or felony offense. Analyses showed no significant differences by suspect race in any of the accountability conditions.

Table 9: Misdemeanor vs. Felony Filing Decisions, Vandalism Case

Accountability Condition	Suspect Race	
	White % Felony (N)	Hispanic % Felony (N)
Control	25% (27)	21% (20)
Undirected	14% (13)	22% (19)
Directed	17% (13)	19% (16)

### Analyses of Case-related Outcomes Using Individual Difference Measures

**Motivation to control prejudice.**<sup>33</sup> For the total sample, the mean IMS score was 5.53 ( $SD = 1.18$ ) and the mean EMS score was 3.25 ( $SD = 1.43$ ). To examine whether differences by suspect race depend on participants' motivation to control prejudice, moderated regression analyses were conducted using participants' IMS and EMS scores. In the analyses, differences by race were examined at 1  $SD$  above and below the variable mean.

There was no interaction between EMS and suspect race for any of the outcomes in any of the conditions. However, there was one significant interaction between IMS and suspect race for perceived crime seriousness in the undirected accountability condition,  $\beta = -.27$ ,  $t = -2.55$ ,  $p = .01$ . As shown in Figure 25, while those low in IMS tended to perceive the crime as more serious for the Hispanic suspect, those high in IMS perceived the crime as more serious for the White suspect. This pattern is consistent with those high in IMS being concerned about acting in a prejudiced manner, perhaps leading them to be relatively more lenient towards the Hispanic suspect. IMS did not show a moderation pattern for any other outcomes or in any other condition.

<sup>33</sup> While the interaction between participants' IMS and EMS scores are of theoretical interest (Devine et al., 2002), such analyses would have been underpowered given the small number of participants in some of the cells (e.g., five). Therefore, I was unable to examine the interaction of these scores as a potential moderator.

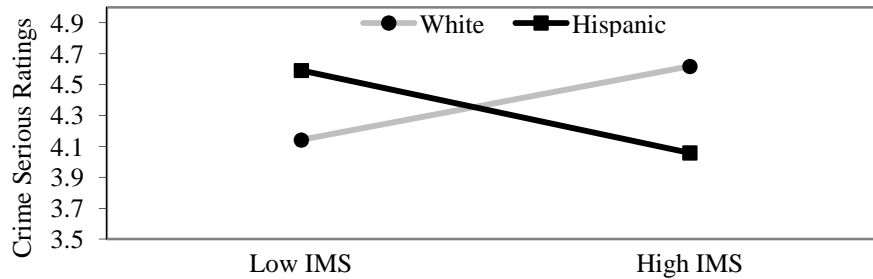


Figure 25: Interaction between IMS scores and suspect race for crime seriousness in undirected accountability condition, vandalism case. Simple effects: White:  $\beta = .21, t = 1.99, p = .05$ ; Hispanic:  $\beta = -.24, t = -1.69, p = .09$ ; Race diff. at 1 *SD* below:  $\beta = -.20, t = -1.61, n.s.$ ; Race diff. at 1 *SD* above:  $\beta = .25, t = 2.07, p < .05$ .

### Relationship between Bias Measures and Participants' Decisions

Although mean differences were generally not found by suspect race, a key question is whether participants' racial biases were nonetheless related to their decisions. Means and correlations for explicit and implicit attitudinal and stereotype measures are shown in Table 10. Looking at the means for the explicit stereotype measures, participants associated Hispanics with crime significantly more than Whites with crime,  $t(507) = -19.82, p < .01$ . Participants reported having significantly warmer feelings towards Whites than Hispanics,  $t(456) = 9.39, p < .01$ . They also expressed a slight preference for Whites over Hispanics on the group preference measure.

For the implicit bias measures, participants were more likely to associate Whites with "good" ( $d' = 2.54$ ) than to associate Whites with "bad" ( $d' = 2.26$ ),  $t(347) = 5.68, p < .01$ . Conversely, participants were more likely to associate Hispanics with "bad" ( $d' = 2.75$ ) than to associate Hispanics with "good" ( $d' = 2.55$ ),  $t(347) = -3.92, p < .01$ . After subtracting the target group's association with "bad" from its association with "good" for each participant, results showed that overall participants had more positive implicit attitudes towards Whites than Hispanics,  $t(346) = 6.38, p < .01$ . The race-crime stereotype GNAT also yield the predicted results, with participants more likely to associate Hispanics with crime ( $d' = 2.35$ ) than Whites with crime ( $d' = 1.80$ ),  $t(332) = 11.44, p < .01$ .

To examine whether participants' biases were related to their decisions, moderated regression analyses were conducted with the interaction between the bias measures and suspect race predicting participant decisions. While we would expect participants' racial bias to be related to decisions in stereotype-consistent ways in the control condition, we would expect to find no relationship between the bias measures and outcomes in the two accountability conditions. Results are presented by condition and organized by bias measure. In all graphs, values for continuous measures are shown at 1 *SD* above and below the mean.



*Table 10: Means and Zero-order Correlations among Racial Bias Measures*

Measure	Mean (SD)	1	2	3	4	5	6	7	8	9
1. Hispanic-Crime Stereotype	51.72 (23.33)	—								
2. White-Crime Stereotype	32.97 (20.54)	.53**	—							
3. Warmth, Hispanics	71.98 (20.09)	-.03	-.05	—						
4. Warmth, Whites	79.18 (18.19)	-.03	-.03	.67**	—					
5. Group Preference	2.45 (.90)	.06	-.01	.24**	-.22**	—				
6. Racial Bias (PJAQ)	2.11 (0.97)	.06	.23**	-.12*	.13**	-.35**	—			
7. Good/Bad GNAT, Hispanics	-.19 (.94)	-.03	-.03	.12*	-.02	.18**	-.20**	—		
8. Good/Bad GNAT, Whites	.28 (.92)	-.01	.05	.03	.06	-.06	.04	-.18**	—	
9. Hispanic-Crime GNAT	2.35 (.97)	-.03	-.11	-.04	-.01	-.02	-.11*	.02	.08	—
10. White-Crime GNAT	1.80 (.92)	.05	-.07	.01	.06	.02	-.02	.09	-.01	.54**

\* $p \leq .05$ , \*\* $p \leq .01$ , N for each measure: Hispanic-Crime Stereotype—512, White-Crime Stereotype—510, Warmth, Hispanics—465, Warmth, Whites—476, Group Preference—535, Racial Bias (PJAQ)—540, Good/Bad GNAT, Hispanics—355, Good/Bad GNAT, Whites—352, Hispanic-Crime GNAT—333, White-Crime GNAT—333.

### Control condition

**Explicit measures.** Looking first at the relationship between explicit attitudinal measures and outcomes in the control condition, warmth ratings were unrelated to participants' case decisions. Participants' group preference ratings were significantly related to suspect race for recommended punishment,  $\beta = -.19, t = -2.01, p < .05$ . As shown in Figure 26, while group preference ratings were unrelated to punishment for the White suspect, they were significantly related to punishment for the Hispanic suspect. Consistent with the operation of prevailing biases, participants who prefer Whites over Hispanics tended to recommend harsher punishment for the Hispanic suspect.

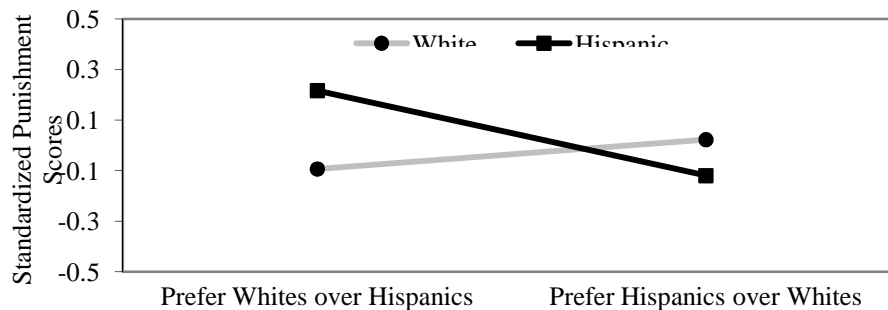


Figure 26: Interaction between group preference ratings and suspect race for punishment in the control condition, vandalism case. Simple effects: White:  $\beta = .08, t = .79, n.s.$ ; Hispanic:  $\beta = -.21, t = -1.97, p = .05$ .

Participants' awareness of the cultural race-crime stereotype and RB-PJAQ scores were unrelated to case decisions in the control condition.

**Implicit measures.** Participants' scores on both the good/bad and race-crime stereotype GNATs were unrelated to case decisions in the control condition.

### Undirected accountability condition.

**Explicit measures.** Participants' explicit bias measures did not significantly interact with suspect race for outcomes in the undirected condition, with one exception. Group preference ratings significantly interacted with suspect race for guilt,  $\beta = .22, t = 2.09, p < .05$ . Contrary to prevailing biases, greater preference for Whites was marginally related to higher guilt ratings for the White suspect,  $\beta = -.20, t = -1.85, p = .07$ . Group preference ratings were unrelated to guilt for the Hispanic suspect,  $\beta = .12, t = 1.10, n.s.$  This interaction becomes marginally significant after controlling for participant demographics,  $\beta = .22, t = 1.93, p = .06$ .

**Implicit measures.** Participants' scores on the good/bad GNAT were unrelated to case decisions. However, scores on the race-crime stereotype GNAT significantly interacted with suspect race for the decision of whether to file the case as a felony or misdemeanor,  $B = -1.49, Wald = 6.34, Exp(B) = .23, p = .01$ . Contrary to expectations, implicit awareness of the Hispanic-crime association was related to less felony classification for the Hispanic suspect only,  $B = -.60, Wald = 3.59, Exp(B) = .55, p = .06$ . When participant demographics were controlled for, the interaction remained significant and the simple slope effect for the Hispanic suspect became significant,  $B = -1.40, Wald = 6.59, Exp(B) = .25, p = .01$ .

**Directed accountability condition.** None of the explicit or implicit bias measures significantly interacted with suspect race for any of the outcomes in the directed condition.

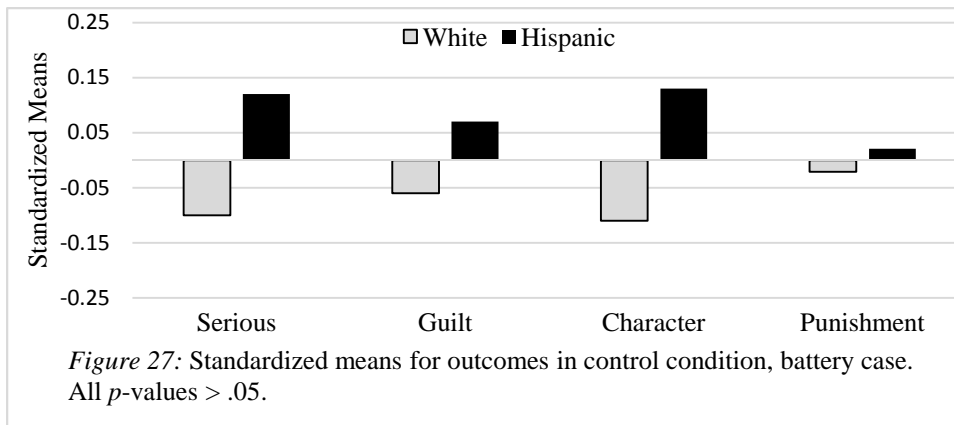
In sum, consistent with the expectations, we see some (limited) evidence in the control condition of a relationship between bias measures and participants' decisions. Specifically, the more participants expressed an explicit preference for Whites, the harsher the punishment recommended for the Hispanic suspect. In the undirected accountability condition, there was no significant interaction between participants' explicit bias measures and suspect race for any of the outcomes. For the implicit measures, there was a significant relationship between participants' race-crime stereotype scores and suspect race for one outcome measure: felony classification decisions. The direction of this effect, however, is counter to expectations since relatively stronger Hispanic-crime associations were associated with fewer felony classifications for the Hispanic suspect. Consistent with the predicted effect of accountability, there was no significant interaction between bias measures and suspect race for any of the outcomes in the directed accountability condition.

### Battery Case

#### Analyses of Case-related Outcomes

While participants viewed the battery case as being slightly more serious than the vandalism case ( $M = 4.93$ ), they viewed the presented facts as equally ambiguous ( $M = 4.78$ ).

**Case ratings and decisions.** Figures 27-29 display the standardized means for the continuous outcomes across the three conditions. The control condition generally showed the expected pattern, with the Hispanic suspect receiving harsher treatment. But the differences by suspect race are small and non-significant. No significant differences were found in either of the accountability conditions.



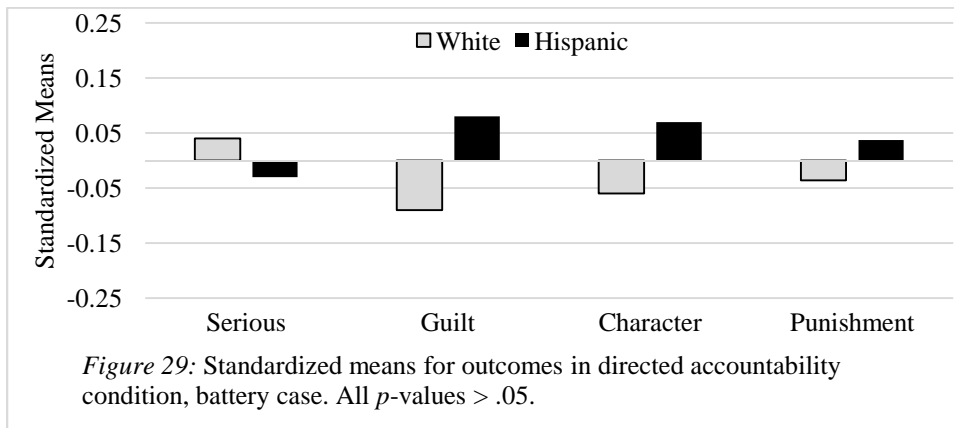
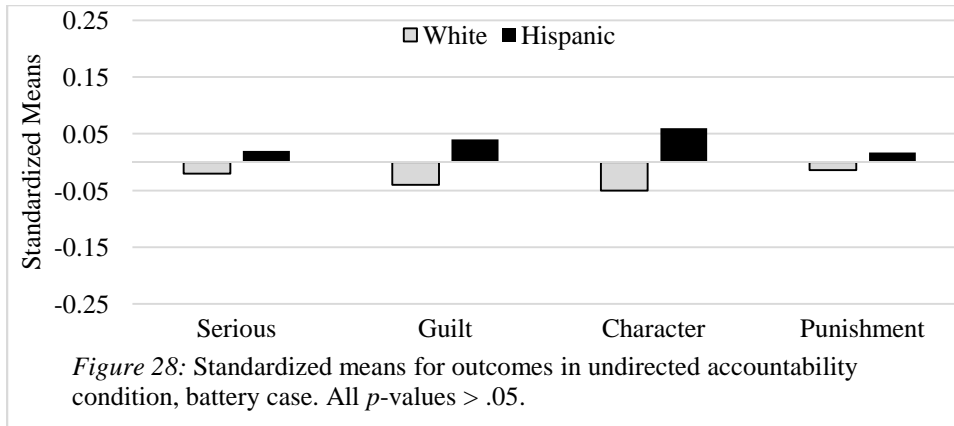


Table 11 reports the frequency with which participants treated the case as a felony (rather than a misdemeanor). As with the other outcome measures, analyses showed no significant differences by suspect race across the three conditions.

Table 11: Misdemeanor vs. Felony Filing Decisions, Battery Case

Accountability Condition	Suspect Race	
	White % Felony (N)	Hispanic % Felony (N)
Control	31% (33)	39% (37)
Undirected	31% (27)	27% (24)
Directed	30% (23)	26% (22)

### Analyses of Case-related Outcomes Using Individual Difference Measures

**Motivation to control prejudice.** As before, the moderating role of IMS/EMS on the effect of suspect race on outcomes was examined. In the control condition, IMS did not moderate the relationship between suspect race and any outcomes. However, EMS did moderate the relationship between suspect race and perceived crime seriousness, suspect character, and recommended punishment. Figure 30 depicts the significant interaction for crime seriousness,  $\beta = .27, t = 2.39, p < .05$ . Whereas there was no difference by suspect race among low EMS participants, the Hispanic suspect received significantly higher ratings than the White suspect among those high in EMS. This same pattern holds for ratings of suspect character and

recommended punishment.<sup>34</sup> This pattern makes some sense since there are no external motivational factors at play in the control condition. When not pressured to control their prejudices, those who are externally motivated tended to exhibit bias towards the Hispanic suspect.

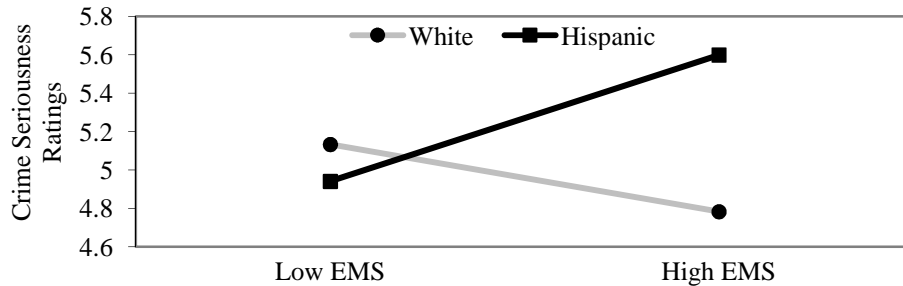


Figure 30: Interaction between EMS scores and suspect race for crime seriousness in control condition, battery case. Simple effects: White:  $\beta = -.13, t = -1.18, n.s.$ ; Hispanic:  $\beta = .25, t = 2.20, p < .05$ ; Race difference at 1 SD below:  $\beta = .07, t = .64, n.s.$ ; Race difference at 1 SD above:  $\beta = -.31, t = -2.75, p < .01$ .

In the undirected accountability condition, IMS moderated the relationship between suspect race and ratings of crime seriousness,  $\beta = -.28, t = -2.72, p < .01$ . As shown in Figure 31, whereas those low in IMS rated the crime as more serious for the Hispanic suspect, those high in IMS rated the crime as more serious for the White suspect. This pattern is consistent with participants high in IMS being more concerned about acting prejudiced.

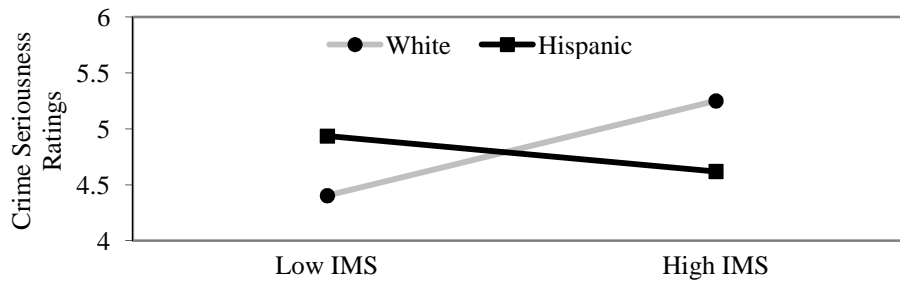


Figure 31: Interaction between IMS and suspect race for crime seriousness in undirected accountability condition, battery case. Simple effects: White:  $\beta = .34, t = 3.28, p < .01$ ; Hispanic:  $\beta = -.13, t = -.93, n.s.$ ; Race difference at 1 SD below:  $\beta = -.22, t = -1.77, p = .08$ ; Race difference at 1 SD above:  $\beta = .26, t = 2.16, p < .05$ .

EMS also moderated the relationship between suspect race and ratings of crime seriousness, recommended punishment, and the felony classification decision in the undirected accountability condition. For all of these outcomes, EMS scores were positively related to outcomes for the Hispanic suspect, but negatively related to outcomes for the White suspect.

<sup>34</sup> Suspect character—Interaction effect:  $\beta = .27, t = 2.39, p < .05$ ; Simple effects: White:  $\beta = -.17, t = -1.49, n.s.$ ; Hispanic:  $\beta = .22, t = 1.89, p = .06$ ; Race diff. at 1 SD below:  $\beta = .08, t = .71, n.s.$ ; Race diff. at 1 SD above:  $\beta = -.30, t = -2.67, p < .01$ . Recommended punishment—Interaction effect:  $\beta = .26, t = 2.31, p < .05$ ; Simple effects: Hispanic:  $\beta = .11, t = .96, n.s.$ ; White:  $\beta = -.26, t = -2.31, p < .05$ ; Race difference at 1 SD below mean:  $\beta = .15, t = 1.34, n.s.$ ; Race difference at 1 SD above mean:  $\beta = -.22, t = -1.92, p = .06$ .

Whether these relationships translate into significant differences by suspect race depends on the particular outcome examined. Ratings of crime seriousness and recommended punishment follow the pattern shown in Figure 32.<sup>35</sup> As shown for ratings of crime seriousness, the White suspect received significantly higher ratings than the Hispanic suspect among those low in EMS, but the Hispanic suspect received significantly higher ratings among those high in EMS. For felony classification, there is a significant race difference among low EMS participants, but the difference is not significant among those high in EMS.<sup>36</sup>

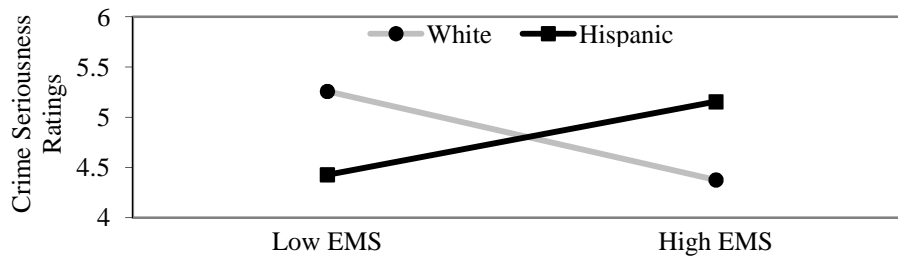


Figure 32: Interaction between EMS and suspect race for crime seriousness in undirected accountability condition, battery case. Interaction effect:  $\beta = .46, t = 3.99, p < .01$ ; Simple effects: White:  $\beta = -.36, t = -3.11, p < .01$ ; Hispanic:  $\beta = .30, t = 2.54, p = .01$ ; Race difference at 1 SD below:  $\beta = .34, t = 2.92, p < .01$ ; Race difference at 1 SD above:  $\beta = -.32, t = -2.74, p < .01$ .

In the directed accountability condition, IMS moderated the relationship between suspect race and guilt ratings,  $\beta = -.41, t = -3.07, p < .01$ . As shown in Figure 33, there was a significant race difference among those low in IMS, with the Hispanic suspect receiving higher guilt ratings. This race difference disappeared among those high in IMS, again consistent with high-IMS participants being more concerned about acting in a prejudiced manner.

EMS scores did not moderate the relationship between suspect race and any outcomes in the directed accountability condition.

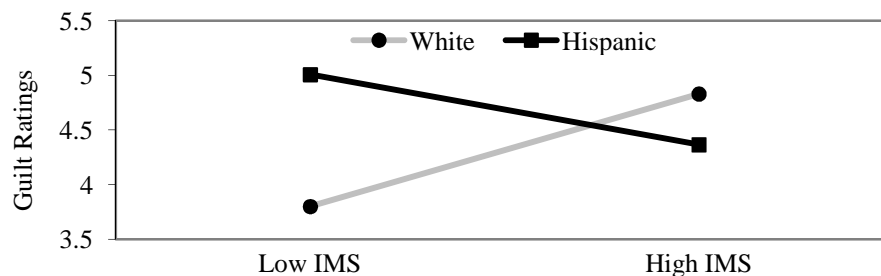


Figure 33: Interaction between IMS and suspect race for guilt ratings in directed accountability condition, battery case. Simple effects: White:  $\beta = .33, t = 2.45, p < .05$ ; Hispanic:  $\beta = -.21, t = -1.86, p = .07$ ; Race difference at 1 SD below:  $\beta = -.39, t = -3.14, p < .01$ ; Race difference at 1 SD above:  $\beta = .15, t = 1.24, n.s.$

<sup>35</sup> Interaction effect for recommended punishment:  $\beta = .35, t = 3.02, p < .05$ ; Simple effects: Hispanic:  $\beta = .33, t = 2.80, p < .01$ ; White:  $\beta = -.15, t = -.146, n.s.$ ; Race difference at 1 SD below mean:  $\beta = .24, t = 2.06, p < .05$ ; Race difference at 1 SD above mean:  $\beta = -.26, t = -2.22, p < .05$ .

<sup>36</sup> Interaction effect for felony classification decision:  $B = 1.24, Wald = 7.48, Exp(B) = 3.47, p < .01$ ; Simple effects: White:  $B = -.16, Wald = .36, Exp(B) = .85, n.s.$ ; Hispanic:  $B = 1.08, Wald = 8.80, Exp(B) = 2.95, p < .01$ ; Race difference at 1 SD below:  $B = 1.77, Wald = 6.25, Exp(B) = 5.88, p = .01$ ; Race difference at 1 SD above:  $B = -.72, Wald = 1.76, Exp(B) = .49, n.s.$

In sum, differences by suspect race emerged after taking into account participants' IMS and EMS scores. In the control condition, high EMS participants gave the Hispanic suspect significantly harsher outcomes than the White suspect. This pattern makes sense given the lack of external factors motivating participants to be non-prejudiced in the control condition. In the undirected accountability condition, we see a mixed pattern of results. Consistent with having concerns about being prejudiced, those high in IMS perceived the crime as more serious for the White suspect than the Hispanic suspect. But EMS tended to be related to harsher outcomes for the Hispanic suspect and to more lenient outcomes for the White suspect, sometimes resulting in significant racial differences among both low- and high-EMS participants across different outcome measures. This is contrary to expectations that the undirected accountability manipulation would be a source of external motivation and reduce bias among those most responsive to this type of pressure, i.e., those high in EMS. In the directed accountability condition, we see the expected pattern with respect to IMS and no moderation patterns for EMS.

### Relationship between Bias Measures & Participants' Decisions

#### Control condition.

**Explicit measures.** Participants' warmth and group preference ratings as well as their awareness of the cultural race-crime stereotype did not significantly interact with suspect race for case decisions in the control condition. Participants' racial bias scores on the RB-PJAQ subscale did significantly interact with suspect race for recommended punishment,  $\beta = .20, t = 1.98, p = .05$ . As shown in Figure 34, racial bias scores were positively related to punishment for the Hispanic suspect only. This pattern is consistent with the expected operation of prevailing bias.

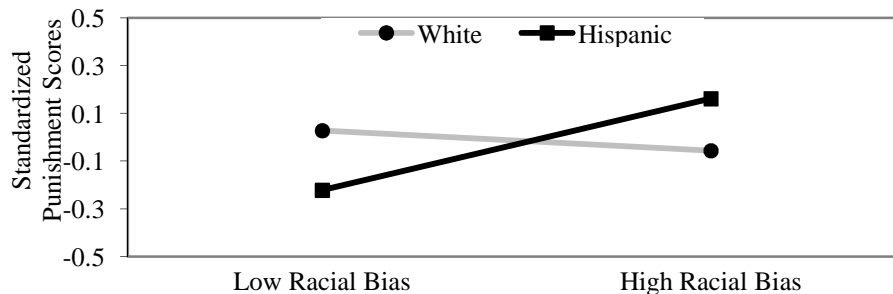


Figure 34: Interaction between RB-PJAQ scores and suspect race for recommended punishment in the control condition, battery case. Simple effects: Hispanic:  $\beta = .23, t = 2.33, p < .05$ ; White:  $\beta = -.05, t = -.49, n.s.$

**Implicit measures.** Participants' scores on the good/bad and race-crime stereotype GNATs were unrelated to suspect race across all outcomes measures.

#### Undirected Accountability Condition.

**Explicit measures.** Participants' warmth ratings significantly interacted with suspect race for perceived crime seriousness,  $\beta = -.29, t = -2.68, p < .01$ . As shown in Figure 35, while warmth ratings were unrelated to seriousness ratings for the Hispanic suspect, they were

significantly related to ratings for the White suspect. Consistent with prevailing biases, the more warmly participants feel towards Whites, the less serious they rated the crime.

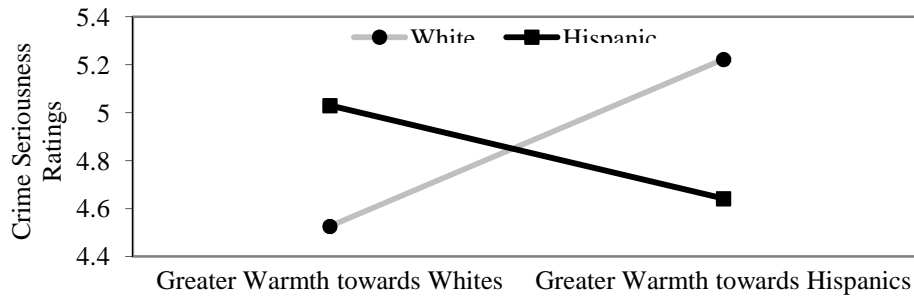


Figure 35: Interaction between warmth ratings and suspect race for crime seriousness in undirected condition, battery case. Simple effects: White:  $\beta = .28, t = 2.50, p = .01$ ; Hispanic:  $\beta = -.15, t = -1.32, n.s.$

In the undirected accountability condition, group preference ratings significantly interacted with suspect race for perceived crime seriousness in the same way as warmth ratings,  $\beta = -.21, t = -1.99, p < .05$ . The more participants preferred Whites over Hispanics, the less serious they rated the crime for the White suspect. Group preference ratings were unrelated to perceived seriousness for the Hispanic suspect. Group preference ratings did not moderate results for any other outcome.

Participants' scores on the racial bias subscale of the PJAQ significantly interacted with suspect race for perceived crime seriousness,  $\beta = .27, t = 2.69, p < .01$ , and felony classification decisions,  $B = .67, Wald = 4.07, Exp(B) = 1.96, p < .05$ . As shown in Figure 36, for crime seriousness, RB-PJAQ scores were positively related to seriousness ratings for the Hispanic suspect, but unrelated to ratings for the White suspect. The same pattern was found for the felony classification decision.<sup>37</sup>

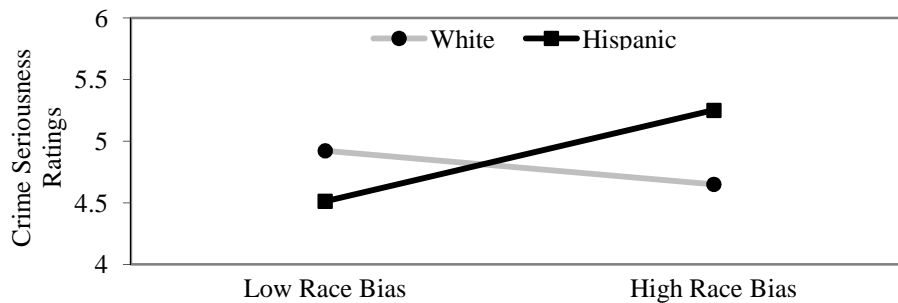


Figure 36: Interaction between RB-PJAQ scores and suspect race for crime seriousness in undirected condition, battery case. Simple effects: White:  $\beta = -.11, t = -1.06, n.s.$ ; Hispanic:  $\beta = .29, t = 2.68, p < .01$ .

<sup>37</sup> Simple effects for the felony classification decisions: White suspect:  $B = -.22, Wald = .62, Exp(B) = .80, n.s.$ ; Hispanic suspect:  $B = .73, Wald = 5.06, Exp(B) = 2.08, p < .05$ . The simple effect for the Hispanic suspect becomes marginally significant after controlling for participants' demographics,  $B = .78, Wald = 3.60, Exp(B) = 2.19, p = .06$ .



Participants' awareness of the cultural race-crime stereotype was unrelated to participants' case decisions in the undirected accountability condition.

**Implicit measures.** In the undirected condition, scores from the race-crime stereotype GNAT did not interact with suspect race for participant decisions. However, scores on the Hispanic Good/Bad GNAT significantly interacted with suspect race for perceived crime seriousness,  $\beta = -.31$ ,  $t = -2.44$ ,  $p < .05$ . As shown in Figure 37, for the Hispanic suspect, greater implicit preference for Whites was associated with greater perceived seriousness, and greater implicit preference for Hispanics was associated with less perceived seriousness. This pattern is consistent with the operation of prevailing racial biases. Good/Bad GNAT scores were not significantly related to any other outcomes.

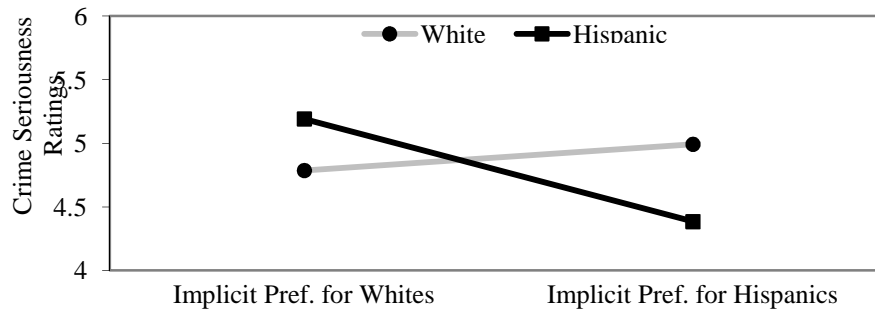


Figure 37: Interaction between Hispanic-White GNAT scores and suspect race for crime seriousness in undirected accountability condition, battery case. Simple effects: White:  $\beta = .13$ ,  $t = 1.04$ , n.s.; Hispanic:  $\beta = -.32$ ,  $t = -2.38$ ,  $p < .05$ .

### Directed condition.

**Explicit measures.** Participants' warmth ratings and RB-PJAQ scores were unrelated to case decisions in the directed accountability condition. Although awareness of the cultural race-crime stereotype significantly interacted with suspect race for guilt ratings,  $\beta = -.30$ ,  $t = -2.28$ ,  $p < .05$ , the simple slope effects for both the White and Hispanic suspect were non-significant. Group preference ratings significantly interacted with suspect race for the felony classification decision only,  $B = -1.08$ ,  $Wald = 4.54$ ,  $Exp(B) = .34$ ,  $p < .05$ . While group preferences were unrelated to felony decisions for the Hispanic suspect,  $B = -.23$ ,  $Wald = .83$ ,  $Exp(B) = .80$ , n.s., they were positively related to felony decisions for the White suspect,  $B = .86$ ,  $Wald = 3.74$ ,  $Exp(B) = 2.36$ ,  $p = .05$ . Consistent with prevailing biases, the more participants prefer Whites over Hispanics, the more likely the case was treated as a misdemeanor.

**Implicit measures.** Participants' scores on the Good/Bad GNAT were unrelated to case decisions in the directed condition. Scores for the race-crime stereotype GNAT significantly interacted with suspect race for perceived crime seriousness only,  $\beta = .33$ ,  $t = 2.06$ ,  $p < .05$ . While implicit stereotypes were unrelated to crime seriousness for the Hispanic suspect,  $\beta = .11$ ,  $t = .87$ , n.s., stronger implicit White-crime associations tended to relate to increased seriousness ratings for the White suspect, though this relationship was only marginally significant,  $\beta = -.31$ ,  $t = -1.94$ ,  $p = .06$ .

In sum, no condition was completely immune from having participants' racial biases related to case decisions. In the control condition, consistent with expectations, participants' RB-PJAQ scores positively predicted harsher punishment for the Hispanic suspect. Yet, all other bias

measures were unrelated to suspect race for the control group. In the undirected accountability condition, participants' warmth ratings, group preference ratings, RB-PJAQ scores, and Good/Bad GNAT scores were related to suspect race for at least some outcomes measures. The observed relationship between the bias measures and suspect race were all in the direction one would expect for the operation of prevailing cultural biases. Thus, undirected accountability was relatively ineffective in attenuating the relationship between racial bias and case outcomes. In the directed accountability condition, there was more limited evidence of racial biases relating to suspect race, with one significant interaction for group preferences and one significant interaction for the race-crime stereotype GNAT. In both instances, the relationships with suspect race were consistent with the prevailing cultural biases.

### **Examining Decisions across Case Files**

To test whether accountability's effectiveness changes across multiple decision points, I examined whether differences by suspect race changed across case files using a mixed model ANOVA. With suspect race and accountability condition as the between-subjects factor and the two case file decisions as the within-subjects factor, there were no significant three-way interactions. This suggests that the accountability manipulations did not change the discrepancies between the White and Hispanic suspects across case files.

### **Cognitive Complexity**

For the self-report measure of cognitive complexity, there was a significant effect by accountability condition,  $F(2,538) = 5.97, p < .01$ . Participants in the control condition had significantly lower ratings for task engagement, effort, and importance than those in the directed accountability condition ( $M$ s: 6.09 vs. 6.36,  $t(355) = -3.03, p < .01$ ) and in the undirected accountability condition ( $M = 6.34, t(378) = -2.78, p < .01$ ). The ratings between the two accountability conditions did not significantly differ from each other. For the item measuring the extent of debate among alternatives, the effect for accountability was only marginally significant,  $F(2,535) = 2.33, p < .10$ . Here, participants in the directed condition reported debating more ( $M = 5.40$ ) than those in the control ( $M = 5.07$ ) or undirected ( $M = 5.08$ ) conditions.

For the memory task, there were no significant differences between the groups in terms of correct items recalled across the two cases ( $M$ s: Control—17.42, Undirected—17.69, Directed—17.89).

## **Summary and Discussion**

For the vandalism case, we saw no statistically significant racial disparities in outcomes in the control condition. Additionally, we saw no racial disparities in the control condition after taking into account participants' motivation to control prejudice. The lack of significant racial disparity in outcomes is surprising in light of previous rounds of pre-testing. Turning to the question of whether participants' racial biases were related to case decisions in the control condition, the answer is by and large no. Only group preference ratings were significantly related to punishment recommendations for the Hispanic suspect. While this relationship was consistent with the operation of prevailing cultural biases, no other bias measures moderated the relationship between suspect race and case decisions. This result is contrary to the hypothesis that participants' racial biases would be related to case decisions in the control condition.

Does accountability change this pattern of results? Looking first at the undirected accountability condition, we also saw no differences by suspect race based on the outcome means. When participants' motivation to control prejudice was taken into account, the White suspect received significantly higher crime seriousness ratings than the Hispanic suspect among high IMS participants. Though this effect only emerged for one outcome measure, the pattern is consistent with high IMS participants being relatively more concerned with acting prejudiced. The undirected accountability manipulation was only moderately successful in insulating case decisions from racial bias. Inconsistent with prevailing biases, greater preference for Whites was related to higher guilt ratings for the White suspect, and stronger implicit awareness of the Hispanic-crime association tended to relate to less felony classification for the Hispanic suspect. Thus, for the vandalism case, the undirected accountability manipulation was not completely successful in insulating racial bias from the participants' case decisions, even though the bias ran in a counterintuitive direction.

For the directed accountability condition, a comparison of outcome means showed no significant differences by suspect race, nor did differences emerge after taking into account participants' motivation to control prejudice. In terms of directed accountability's effectiveness at insulating racial bias from case decisions, the manipulation appeared successful since no bias measure significantly interacted with suspect race.

For the battery case, we continued to see no race differences in the control condition based on outcome means, once again a departure from pre-testing. When participants' motivation to control prejudice was taken into account, significant differences by suspect emerged among those high in EMS. Specifically, for crime seriousness ratings and judgments of suspect character, the Hispanic suspect received harsher outcomes relative to the White suspect. This pattern is consistent with the expected operation of prevailing racial bias. In terms of the interactions between racial bias and case decisions, we generally do not see any significant relationships between bias and suspect race, with one exception. Again consistent with the operation of prevailing racial bias, RB-PJAQ scores were related to harsher punishment for the Hispanic suspect. Thus, although the evidence is far less than what was expected, we do find some indications in the control condition that racial bias was related to case decisions, particularly for those high in EMS.

In the undirected accountability condition for the battery case, we also saw no significant difference by suspect race when looking at outcome means alone. When participants' motivation to control prejudice was taken into account, we saw the predicted pattern regarding IMS scores. Specifically, those high in IMS perceived the crime as more serious for the White suspect than for the Hispanic suspect. EMS scores also moderated several outcomes, with significant race differences emerging among both high and low EMS participants. It is surprising to see the Hispanic suspect sometimes receiving significantly harsher outcomes than the White suspect among those high in EMS, since we would expect the undirected accountability manipulation to exert external pressure on these participants in particular. In terms of the interaction between racial bias and case decisions, different types of racial bias (warmth ratings, explicit group preferences, RB-PJAQ scores, and Good/Bad GNAT scores) related to suspect race for different outcomes. For all of these results the pattern is consistent with the operation of prevailing cultural biases. Thus, the undirected accountability manipulation was not successful in insulating the decision-making process from the influence of racial bias in the battery case.

In the directed accountability condition, we continued to see no significant differences by suspect race when looking at outcome means alone. When participants' motivation to control

prejudice was taken into account, we saw an expected pattern with regard to IMS scores. Specifically, those low in IMS gave significantly higher guilt ratings to the Hispanic suspect compared to the White suspect, consistent with low IMS participants being relatively unconcerned about acting in a prejudiced way. In terms of the relationship between bias measures and case decisions, group preferences and implicit stereotypes were related to suspect race for felony classification decisions and perceived crime seriousness, respectively. In both instances, the pattern was consistent with the direction of prevailing cultural biases. Thus, the directed accountability manipulation was also less than fully effective in insulating case decisions from racial bias in the battery case.

Additional analyses showed that the effects of accountability with regard to suspect race did not differ across the two case files. Measures of cognitive complexity showed some limited effects of the two accountability manipulations. Specifically, while there was no difference by condition on the memory task, those in the two accountability conditions reported greater task engagement and effort than those in the control condition. However, while the accountability manipulations successfully motivated participants to engage more with the task, this additional effort did not seem to translate into less reliance on racial biases. Based on the results from the two case files, it cannot be said that either accountability manipulation was more effective in attenuating racial bias in the decision-making process compared to the control condition. As between the two accountability conditions, the directed accountability manipulation appeared to achieve greater success in reducing the influence of racial bias in case decisions.

## CHAPTER 6: GENERAL DISCUSSION & CONCLUSION

### Overview of Studies 1, 2, & 3

The current project examined the question of whether accountability can serve as an effective intervention to reduce the influence of racial bias in the decision-making process. Over three studies, college students and workers on MTurk made judgments about either a White, Black, or Hispanic suspect across a total of 4 different criminal cases. These studies tested the effectiveness of three different forms of accountability (outcome, process/undirected, directed) in light of participants' explicit and implicit attitudinal and stereotype-based racial biases. The take-away lessons learned from the collected data are summarized here:

*Is accountability effective at reducing the influence of racial bias in the decision process?* Across the three studies we were left with mixed evidence to support the claim that accountability can be an effective bias-reducing intervention. In Study 1, neither accountability condition showed the same anti-Black/pro-White for recommended fine amount that we saw in the control condition. This is consistent with accountability's theorized ability to reduce the influence of bias. However, this was only a single outcome measure and at least one of the accountability conditions (outcome) showed some evidence that participants' bias measures were related to their case decisions in an anti-Black/pro-White direction.

In Study 2, both accountability conditions seemed successful at countering the pro-Black/anti-White race difference found across case files in the control condition in both the drug and robbery cases. But we saw some evidence of an anti-Black/anti-White bias in outcomes for both accountability conditions in the vandalism case. Here, again, some caution is warranted because significant race differences were only found for one measure per accountability condition, which does not amount to a very striking pattern.

In Study 3, the directed accountability condition was the most successful in decoupling the relationship between suspect race and outcomes in the vandalism case. In the vandalism case, both the control and undirected accountability conditions showed some (again limited) evidence of a relationship between bias measures and case decisions, though in a pro-Black/anti-White direction. In the battery case, all three conditions showed evidence of an anti-Black/pro-White bias. Particularly striking is the fact that the undirected condition seemed ineffective at curbing the influence of race among those high in external motivation to control prejudice. Because accountability serves as a kind of external motivational force, one would have expected to see reduced race differences among those most sensitive and responsive to external pressures to be non-prejudiced.

At the very least, the data would caution against any unqualified claims that accountability can effectively reduce the influence of racial bias. Taking all three studies into consideration, it cannot be said that any of the accountability conditions fared particularly better than the control condition. However, as with any study, it is important to take the findings in light of the study's limitations, which are discussed below.

*Does accountability increase cognitive complexity and greater task engagement among decision-makers?* Based on findings from previous research, accountability's effectiveness lies in its ability to motivate decision-makers to engage more with the task and approach the available information in more cognitively complex ways, such as by debating among alternatives and preempting potential criticism. Within the current project, very limited evidence for this

theorized mechanism was found. In the first and second studies, those in the accountability conditions did not report spending more time and effort on the task, being more engaged, or debating more between alternative responses than those in the control condition. Only in the third study we do find that participants in both accountability conditions reported higher levels of task engagement, effort, and importance than those in the control condition. Perhaps this limited evidence can be attributed to the somewhat problematic nature of self-report measures, which may be skewed in ways to address participants' presentational concerns. Yet, performance on two different forms of a memory test in Studies 2 and 3, which were designed to be an indirect measure of task engagement, did not reveal any differences by condition either. Though this was not feasible in the current study, perhaps future research might be able to utilize other objective, yet subtle measures for engagement, such as recording the amount of time it takes participants to complete the task or tracking the extent to which participants seek out additional information before reaching their decision.

*Does accountability's effectiveness remain consistent across multiple decision tasks?*

Within the psychological literature on accountability, the vast majority of studies have participants engage in a single decision task without requiring them to ever actually articulate a justification for their decisions. This is because the motivational efficacy of accountability is thought to stem from the *mere expectation* that one will be held accountable. But this approach to testing accountability's effectiveness may be losing some important information about the psychological relevance of the articulation process. Additionally, it ignores the fact that in many real-world contexts individuals may engage in multiple rounds of decision-making and justification before their decisions are reviewed by someone else. One question raised is whether accountability's effectiveness remains robust over the course of multiple decision tasks. Within the present project, there was no indication in Studies 2 and 3 that differences by suspect race changed across the different case files reviewed. However, it is possible that the differences were too small to be detected by the current analyses given the variation across the reviewed cases. Additionally, as discussed further below, testing the effectiveness of accountability across multiple cases was limited by the fact that we did not observe the expected pattern of racial bias in the control condition.

*Was accountability equally (in)effective at addressing both explicit and implicit forms of racial bias?* The current project proved to be less than ideal in testing accountability's effectiveness in reducing the influence of implicit bias since the implicit bias measures did not moderate the results in the control condition. This is not completely surprising since previous research on implicit bias has shown that its effects often manifest in subtle ways (e.g., non-verbal behaviors, Dovidio, Kawakami, & Gaertner, 2002) and in tasks involving time constraints (Glaser & Knowles, 2007). Yet, some work has successfully shown implicit bias to be related to outcomes on deliberative tasks such as the ones involved here (see, e.g., Green et al., 2007). In Study 3, implicit bias measures related to suspect race for one outcome in each of the accountability conditions in the battery case. For the undirected condition, scores on the Good/Bad GNAT were related to perceived crime seriousness for the Hispanic suspect in a way consistent with prevailing cultural biases. For the directed condition, scores on the race-crime stereotype GNAT were marginally significantly related to perceived crime seriousness for the White suspect, again in a stereotype-consistent way. Although for both conditions the relationship between implicit measures and race is limited to a single outcome, this evidence runs counter to claims that accountability effectively reduces the influence of implicit bias.

One further point is noteworthy about the implicit measures used in Study 3. To my knowledge, no previous study has measured implicit bias towards Hispanics using similar methods. This is perhaps due to the fact that much study on racial bias focuses on the challenges faced by Blacks relative to Whites in the United States. Results from the present study show that Hispanics are also more closely associated with crime and are subjected to more negative attitudes. This finding provides supports the need to expand beyond the Black-White dichotomy which has dominated social psychological research and to more closely examine the ways in which other minority groups are disadvantaged by implicit bias. The need for such research is heightened by the fact that Hispanics represent one of the fastest growing populations in the United States and are also overrepresented with the criminal justice system.

### **Limitations**

Several aspects of the current project limit the conclusions we are able to draw from the results. First, as with all experimental research, there is a concern about the external validity of the study, that is, the extent to which results generalize to more “real-world” contexts. Though fairly typical of studies on jury decision-making (see, e.g., Elek & Hannaford-Agor, 2014; Sommers & Ellsworth, 2000, 2001), the decision tasks used in this project are somewhat artificial compared to how such decisions are made within organizational settings. In particular, the online nature of the study raises some concerns about whether participants were fully attentive to the decision task and whether they took their role as decision-makers seriously. While attempts were made to identify and screen out inattentive or careless participants from the analyses, actual jurors, prosecutors, or other similar decision-makers are likely to be far more invested in the task compared to college students or MTurk workers. The lack of consequences to any real person stemming from their decision arguably changes how people approach the task. Similarly, the lack of consequences to the decision-maker in terms of receiving feedback or criticism from a legitimate third party perhaps makes the test of accountability a conservative one. Presumably, accountability will have even stronger effects than observed in the present project when one is faced with a realistic prospect of being reviewed by a supervisor. As argued in Chapter 1, there are several phases to accountability and the current set of studies can only speak to the efficacy of certain aspects of the phenomenon. Nonetheless, the fact that we see the accountability manipulations producing different results than the control condition suggests that there is some important psychological value attached to the expectation and actual justification of one’s decisions.

Despite the threats to external validity, the experimental nature of the project provides certain unique advantages to studying the psychological phenomenon of interest. Most importantly, the study designs allowed for precise manipulation of the factors of interest while keeping all other aspects of the decision task identical across conditions. This permits us to make causal inferences about the specific effects attributable to accountability itself. Though variations in organizational setting and the specific implementation of accountability certainly matter, there are basic, fundamental psychological processes captured by these studies that arguably can inform policy choices across multiple contexts.

A second limitation is that because the present project largely failed to generate the traditional pattern of bias in the control condition, we were unable to fully examine the effectiveness of accountability in reducing bias. Consistent with prior research (Bodenhausen, 1988; Bodenhausen et al., 1994; Sommers & Ellsworth, 2000, 2001; Cohn et al., 2009), we

expected to find participants giving harsher outcomes to the minority suspect relative to the White suspect in the control condition. But we actually found only limited evidence for this bias across the three studies. For most outcomes in the first and third studies, there are no significant differences by suspect race. In the second study, we see the racial disparity running in the opposite direction, with harsher outcomes being given to the White suspect. While an emergent anti-White cultural shift cannot be ruled out as a possibility, this is not a likely explanation given that participants still exhibited a relative pro-White bias on both implicit and explicit measures across all the studies. Rather, the observed patterns in the control condition are better explained by an (over)compensation effect, whereby participants are cautious about doling out harsh outcomes to minority suspects out of concern of acting or being perceived as prejudiced. Indeed, several vignette-based studies conducted around the same time as this project also failed to replicate the baseline bias effect (Elek & Hannaford-Agor, 2014). Elek & Hannaford-Agor (2014) have attributed such failures to spontaneous self-correction since “Americans may have become increasingly aware of the cultural attention to race bias over the past decade and are now more sensitive to the possibility of revealing such bias, particularly in research settings” (p. 16). They note that several high-profile media stories pertaining to race and the American criminal justice system in 2013 may have, perhaps temporarily, increased the salience of race issues and primed participants as they approach these kinds of experimental studies. This self-correction explanation is supported by research that successfully reversed the expected bias pattern by intentionally making race salient (Sommers & Ellsworth, 2001; Cohn et al., 2009). Although we were limited in testing accountability’s effectiveness at reducing traditional forms of racial bias, we did see some evidence that accountability countered the tendency of participants to over-correct for the bias, which resulted to harsher outcomes for the White suspect (Study 2).

A third limitation of this project is the questionable effectiveness of some of the manipulations used. In particular, responses to the manipulation check question in Study 1 suggest that participants in the outcome accountability condition did not expect to have to justify their decisions any more than those in the control condition. It is unclear why this is the case since the manipulations used in the current study were close adaptations from what has been successfully used in previous research (e.g., Davis, Mero, & Goodman, 2007). Additionally, the process and outcome accountability manipulation did not differ from one another in terms of strength, only wording. Regardless of the reason, though, conclusions about the effectiveness of outcome accountability need to be taken with some caution.

Responses to the race manipulation check question for the first two studies also indicated that the manipulation for suspect race was less than fully effective, particularly for the Black suspect. Again, it is unclear why this was the case since previous studies have successfully manipulated race by using stereotypical names only. While the suspect names were presented with several other pieces of background information, which may have detracted attention from the name manipulation, race was also manipulated a second way through a checked demographic box for suspect race. It is also striking that participants correctly identified the race for the White suspect with greater frequency than for the Black suspect (Study 2: 82% vs. 51%). Responses to the race manipulation check question from the non-descript race condition in Study 1 suggests that this pattern is not simply due to people “defaulting” to think of a White suspect. A more likely possibility is that, consistent with people being concerned about being politically correct and avoiding race, participants were more reticent to label the Black suspect as Black. However, because the basic pattern of outcomes remains the same regardless of whether participants



incorrectly identifying suspect race are excluded, the effectiveness of the race manipulation may not be as concerning.

### **Conclusion**

Despite lay intuitions and scholarly claims that accountability is an effective way to reduce the influence of racial bias in the decision-making process, the results from the current project suggest the need for caution. Although there were limitations in fully testing accountability's effectiveness in addressing prevailing cultural biases, none of the accountability conditions consistently fared better at decoupling the link between suspect race and case decisions than the control condition. Moreover, we still found some (limited) evidence that implicit bias related to case outcomes among those held accountable. Although accountability structures may fulfill many different functions, organizations may do well to consider additional means by which to reduce bias rather than relying solely on accountability to serve this purpose. Future research should continue to test different forms of accountability, especially ones that are most easily adoptable within different organizational settings. Finally, in the present project, the salience of race issues may have been at play and triggered an over-compensation effect amongst those in the control condition. Additional research should directly explore how making race salient affects the decision-making process of those held accountable. Doing so is particularly relevant to contexts where an accountability system is set up in direct response to race-based problems (e.g., police districts with a history of racial profiling).

## REFERENCES

- Allport, G. W. (1954). *The nature of prejudice*. Garden City, NY: Doubleday.
- Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology, 95*, 918-932.
- Baldus, D. C., Woodworth, G., Zuckerman, D., Wiener, N. A., & Broffitt, B. (1998). Racial discrimination and the death penalty in the post-*Furman* era: An empirical and legal overview, with recent findings from Philadelphia. *Cornell Law Review, 83*, 1638-1770.
- Baron, J. N., & Pfeffer, J. (1994). The social psychology of organizations and inequality. *Social Psychology Quarterly, 57*, 190-209.
- Bartlett, K. T. (2009). Making good on good intentions: The critical role of motivation in reducing implicit workplace discrimination. *Virginia Law Review, 95*, 1893-1972.
- Batson v. Kentucky, 476 U.S. 79 (1986).
- Baumgartner, M. P. (1992). The myth of discretion. In K. Hawkins (Ed.), *The uses of discretion* (pp. 129-162). New York: Oxford University Press.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review, 94*, 991-1013.
- Bielby, W. T. (2003). Expert report of William T. Bielby in *Betty Dukes, et al. v. Wal-Mart Stores, Inc.* Retrieved from [http://www2.law.columbia.edu/fagan/courses/law\\_socialscience/documents/Spring\\_2006/Class%2017-Gender%20Discrimination/BD\\_v\\_Walmart.pdf](http://www2.law.columbia.edu/fagan/courses/law_socialscience/documents/Spring_2006/Class%2017-Gender%20Discrimination/BD_v_Walmart.pdf)
- Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues, 57*, 707-724.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-261.
- Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142-1163.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality & Social Psychology, 81*, 828-841.
- Bobo, L. D. (2011). Somewhere between Jim Crow & post-racialism: Reflections on the racial divide in America today. *Dædalus, 140*, 11-36.
- Bodenhausen, G. V. (1988). Stereotypic biases in social decision making and memory: Testing process models of stereotype use. *Journal of Personality and Social Psychology, 55*, 726-737.
- Bodenhausen, G. V., Kramer, G. P., & Susser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology, 66*, 621-632.
- Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgment: The differential impact of anger and sadness. *European Journal of Social Psychology, 24*, 45-62.
- Bonilla-Silva, E. (2010). *Racism without racists: Color-blind racism & racial inequality in contemporary America*. (3<sup>rd</sup> Edition). Lanham: Rowman & Littlefield Publishers, Inc.

- Brandeis, L. D. (1914). *Other people's money: And how the bankers use it*. New York: Frederick A. Stokes Company.
- Bridges, G. S., & Steen, S. (1998). Racial disparities in official assessments of juvenile offenders: Attributional stereotypes as mediating mechanisms of juvenile offenders. *American Sociological Review*, *63*, 554–571.
- Brown, A. (2014). The U.S. Hispanic population has increased sixfold since 1970. Pew Research Center. Retrieved from <http://www.pewresearch.org/fact-tank/2014/02/26/the-u-s-hispanic-population-has-increased-sixfold-since-1970/>
- Brtek, M. D., & Motowidlo, S. J. (2002). Effects of procedure and outcome accountability on interview validity. *Journal of Applied Psychology*, *87*, 185–191.
- CALCRIM No. 960 (2014). Judicial Council of California Criminal Jury Instructions.
- CALCRIM No. 2304 (2014). Judicial Council of California Criminal Jury Instructions.
- CALCRIM No. 2900 (2014). Judicial Council of California Criminal Jury Instructions.
- California Penal Code § 290.006 (West, 2014).
- Castilla, E. J. (2012). Accounting for the gap: A firm study manipulating organizational accountability in pay promotions. Paper presented at MORS colloquium, Berkeley, CA, March 22, 2012.
- Center for Policing Equity (November, 2013). Nation's first police profiling database awarded grant by NSF. Retrieved from [http://cpe.psych.ucla.edu/images/uploads/database\\_release\\_final\\_\(1\).pdf](http://cpe.psych.ucla.edu/images/uploads/database_release_final_(1).pdf).
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Cohn, E., Bucolo, D., Pride, M., & Sommers, S. (2009). Reducing white juror bias: The role of race salience and racial attitudes. *Journal of Applied Social Psychology*, *39*, 1953–1973.
- 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*, 1314–1329.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.
- Davis, W. D., Mero, N., & Goodman, J. M. (2007). The interactive effects of goal orientation and accountability on task performance. *Human Performance*, *20*, 1–21.
- Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.
- Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology*, *48*, 1267–1278.
- Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*, 835–848.
- Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup bias. In S. T. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, 5th ed., pp. 1084–1121). New York: Wiley.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–68.
- Draper v. United States, 358 U.S. 307 (1959).

- Dubnick, M. (2005). Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review*, 28, 376-417.
- Dubnick, M. J., & Justice, J. B. (2004). Accounting for accountability. Paper presented at the 2004 Annual Meeting of the American Political Science Association. Retrieved from <http://pubpages.unh.edu/~mjc46/papers/2004/dubjusacctg2004.pdf>
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, 17, 383-386.
- Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology*, 87, 876-893.
- Edelman, L. B. (1992). Legal ambiguity and symbolic structures: Organizational mediation of civil rights law. *American Journal of Sociology*, 97, 1531-1576.
- Edelman, L. B., Erlanger, H. S., & Lande, J. (1993). Internal dispute resolution: The transformation of civil rights in the workplace. *Law & Society Review*, 27, 497-534.
- Edelman, L. B., Krieger, L. H., Eliason, S. R., Albiston, C. R., & Mellema, V. (2011). When organizations rule: Judicial deference to institutionalized employment structures. *American Journal of Sociology*, 117, 888-954.
- Edelman, L. B., Uggen, C., & Erlanger, H. (1999). The endogeneity of legal regulation: Grievance procedures as rational myth. *American Journal of Sociology*, 105, 406-54.
- Elek, J. K., & Hannaford-Agor, P. (2014). Can explicit instructions reduce expressions of implicit bias? New questions following a test of a specialized jury instruction. National Center for State Courts. Retrieved from <http://www.ncsc-jurystudies.org/What-We-Do/~media/Microsites/Files/CJS/What%20We%20Do/Can%20Explicit%20Instructions%20Reduce%20Expressions%20of%20Implicit%20Bias.ashx>
- Federick, B., & Stemen, D. (2012). The anatomy of discretion: An analysis of prosecutorial decision making—technical report. Vera Institute of Justice. Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/grants/240334.pdf>
- Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, 73, 31-44.
- Feldman, M. (1992). Social limits to discretion: An organizational perspective. In K. Hawkins (Ed.), *The uses of discretion* (pp. 163-184). New York: Oxford University Press.
- Felstiner, W. L. F., Abel, R. L., & Sarat, A. (1980). The emergence and transformation of disputes: Naming, blaming, claiming . . . *Law & Society Review*, 15, 631-654.
- Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, 43, 522-527.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.
- Fisher v. University of Texas at Austin, 133 S. Ct. 2411 (2013).
- Fiske, S. T. (2004). Intent and ordinary bias: Unintended thought and social motivation create casual prejudice. *Social Justice Research*, 17, 117-127.
- Fleming, M. A., & Petty, R. E. (2000). Identity and persuasion: An elaboration likelihood approach. In D. J. Terry & M. A. Hogg (Eds.), *Attitudes, behavior, and social context: The role of norms and group membership* (pp. 171-199). Mahwah, NJ: Lawrence Erlbaum.
- Ford, T. E., Gambino, F., Lee, H., Mayo, E., & Ferguson, M. A. (2004). The role of accountability in suppressing managers' preinterview bias against African-American sales job applicants. *Journal of Personal Selling and Sales Management*, 24, 113-124.

- Ford, J. K., & Weldon, E. (1981). Forewarning and accountability: Effects on memory-based interpersonal judgments. *Personality and Social Psychology Bulletin*, 7, 264-268.
- Fox, D. R. (1993). Psychological jurisprudence and radical social change. *American Psychologist*, 48, 234-241.
- Fyock, J., & Stangor, C. (1994). The role of memory biases in stereotype maintenance. *British Journal of Social Psychology*, 33, 331-343.
- Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia: Psychology Press.
- Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality & Social Psychology*, 78, 708-724.
- Gelsthorpe, L., & Padfield, N. (2003). Introduction. In Authors (Eds.), *Exercising discretion: decision making in the criminal justice system and beyond* (pp. 1-28). Oregon: Willan Publishing
- Gilliam, F. D., Jr., & Iyengar, S. (2000). Prime suspects: The influence of local television news on the viewing public. *American Journal of Political Science*, 44, 560-573.
- Gilliam, F. D., Jr., Iyengar, S., Simon, A., & Wright, O. (1996). Crime in black and white: The violent, scary world of local news. *Harvard International Journal of Press/Politics*, 1, 6-23.
- Glaser, J., & Knowles, E. D. (2007). Implicit motivation to control prejudice. *Journal of Experimental Social Psychology*, 44, 164-172.
- 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, 292-306.
- 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, 91-107.
- Gordon, R. A., Rozelle, R. M., & Baxter, J. C. (1988). The effect of applicant age, job level, and accountability on the evaluation of job applicants. *Organizational Behavior and Human Decision Processes*, 41, 20-33.
- Graham, S., & Lowery, B. S. (2004). Priming unconscious stereotypes about adolescent offenders. *Law and Human Behavior*, 28, 483-504.
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22, 1231-1238.
- Green, T. K., & Kalev, A. (2008). Discrimination-reducing measures at the relational level. *Hastings Law Journal*, 59, 1435-1462.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Greenwald, A. G., Poehlman, A. T., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97, 17-41.
- Hagan, J., & Dinovitzer, R. (1999). Collateral consequences of imprisonment for children, communities, and prisoners. *Crime and Justice*, 26, 121-162.

- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814-834.
- Hartney, C. & Vuong, L. (2009). Created equal: Racial and ethnic disparities in the US criminal justice system. National Council on Crime and Delinquency. Retrieved from [http://nccd-crc.issuelab.org/research/listing/created\\_equal\\_racial\\_and\\_ethnic\\_disparities\\_in\\_the\\_us\\_criminal\\_justice\\_system](http://nccd-crc.issuelab.org/research/listing/created_equal_racial_and_ethnic_disparities_in_the_us_criminal_justice_system)
- Hattrup, K., & Ford, J. K. (1995). The roles of information characteristics and accountability in moderating stereotype-driven processes during social decision making. *Organizational Behavior and Human Decision Processes*, *63*, 73-86.
- Hawkins, K. (1992). The use of legal discretion: Perspectives from law and social science. In Author (Ed.), *The uses of discretion* (pp. 11-46). New York: Oxford University Press.
- Heilman, M. E., & Haynes, M. C. (2008). Subjectivity in the appraisal process: A facilitator of gender bias in work settings. In E. Borgida & S.T. Fiske (Eds.), *Beyond common sense: Psychological science in the courtroom* (pp. 127-156). Oxford: Blackwell Publishing.
- Hodson, G., Dovidio, J. F., & Gaertner, S. L. (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin*, *28*, 460-471.
- Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *14*, 640-643.
- Hugenberg, K. & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science*, *15*, 342-345.
- In re Bye, 12 Cal.3d 96 (1974).
- In re Podesto, 15 Cal.3d 921 (1976).
- In re Spears, 157 Cal.App.3d 1203 (1984).
- In re Sturm, 11 Cal.3d 258 (1974).
- Jacoby, J. E., Mellon, L. R., Turner, S. H., & Ratledge, E. C. (1982). *The standard case set. A tool for criminal justice decision makers*. Washington, DC: National Institute of Justice.
- Jones-Brown, D., Gill, J., & Trone, J. (2010). Stop, question & frisk policing practices in New York City: A primer. Center on Race, Crime and Justice. Retrieved from [http://www.jjay.cuny.edu/web\\_images/PRIMER\\_electronic\\_version.pdf](http://www.jjay.cuny.edu/web_images/PRIMER_electronic_version.pdf)
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R. (2013). Presumed fair: Ironic effects of organizational structures. *Journal of Personality and Social Psychology*, *104*, 504-519.
- Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, *71*, 589-617.
- Kang, J. (2008). Trojan horses of race. In G. S. Parks, S. Jones, & W. J. Cardi (Eds.), *Critical race realism: Intersections of psychology, race, and law* (pp. 125-154). New York: The New Press.
- Kawakami, K., Dovidio, J. F., Moll, L., & Hermsen, S. (1999). *The role of practice in reducing automatic stereotype activation*. Unpublished manuscript, University of Nijmegen, Nijmegen, The Netherlands.
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotype associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*, 871-888.

- Kawakami, K., Dovidio, J. F., & van Kamp, S. (2005). Kicking the habit: Effects of nonstereotypic association training and correction processes on hiring decisions. *Journal of Experimental Social Psychology, 41*, 68-75.
- Kawakami, K., Dovidio, J. F., & van Kamp, S. (2007). The impact of counterstereotypic training and related correction processes on the application of stereotypes. *Group Processes & Intergroup Relations, 10*, 139-156.
- Krieger, L. H. (1995). The content of our categories: A cognitive bias approach to discrimination and equal employment opportunity. *Stanford Law Review, 47*, 1161-1248.
- Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology, 19*, 448-468.
- Lambert, A. J., Payne, B. K., Jacoby, L. L., Shaffer, L. M., Chasteen, A. L., Khan, S. R. (2003). Stereotypes as dominant responses: On the "social facilitation" of prejudice in anticipated public contexts. *Journal of Personality and Social Psychology, 84*, 277-295.
- Lambert, A. J., Cronen, S., Chasteen, A. L., Lickel, B. (1996). Private vs public expressions of racial prejudice. *Journal of Experimental Social Psychology, 32*, 437-459.
- Langer, E. J., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology, 36*, 635-642.
- Lazarus, R. S. (1982). Thoughts on the relation between emotion and cognition. *American Psychologist, 37*, 1019-1024.
- Lazarus, R. S. (1984). On the primacy of cognition. In K. R. Scherer and P Ekman (Eds.), *Approaches to emotion* (pp. 247-257). Hillsdale, N.J.: Erlbaum.
- Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PLoS ONE, 4*, e4215. doi:10.1371/journal.pone.0004215.
- 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*, 2010-2038.
- Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages: How motivational interventions can reduce (but also increase) prejudice. *Psychological Science, 22*, 1472-1477.
- Lenton, A. P., Blair, I. V., & Hastie, R. (2001). Illusions of gender: Stereotypes evoke false memories. *Journal of Experimental Social Psychology, 37*, 3-14.
- Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin, 24*, 563-574.
- Lerner, J. S. & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*, 255-275.
- MacCoun, R. J. (2005). Voice, control, and belonging: The double-edged sword of procedural fairness. *Annual Review of Law and Social Science, 1*, 171-201.
- MacCoun, R. J. (2006). Psychological constraints on transparency in legal and government decision making. *Swiss Political Science Review, 12*, 112-121.
- Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*, 93-120.
- Madva, A., & Brownstein, M. (2013). The blurry boundary between stereotyping and evaluation in implicit cognition. Unpublished manuscript.

- Malinowska-Sempruch, K. (February 8, 2014). Liberals and republicans signal huge shift in attitude to US drug laws. *The Guardian*. Retrieved from <http://www.theguardian.com/commentisfree/2014/feb/09/liberals-tea-party-republicans-unison-drug-law-reform-america>
- McCleskey v. Kemp, 753 F.2d 877 (11th Cir. 1985).
- McCleskey v. Kemp, 481 U.S. 279 (1987).
- Mero, N. P. & Motowidlo, S. J. (1995). Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology*, 80, 517-524.
- Mero, N. P., Motowidlo, S. J., & Anna, A. L. (2003). Effects of accountability on rating behavior and rater accuracy. *Journal of Applied Social Psychology*, 33, 2493-2514.
- Mitchell, G. (2009). Second thoughts. *McGeorge Law Review*, 40, 687.
- Mitchell, O. (2005). A meta-analysis of race and sentencing research: Explaining the inconsistencies. *Journal of Quantitative Criminology*, 21, 439.
- Mustard, D. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the US federal courts. *Journal of Law and Economics*, 44, 285-314.
- Neuberg, S. L. (1989). The goal of forming accurate impressions during social interactions: Attenuating the impact of negative expectancies. *Journal of Personality and Social Psychology*, 56, 374-386.
- Neuberg, S. L. (1994). Expectancy-confirmation processes in stereotype-tinged social encounters: The moderating role of social goals. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario symposium volume 7* (pp. 103-130). New Jersey: Lawrence Erlbaum Associates, Inc.
- Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology*, 53, 431-444.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Norton, M. I., Sommer, S. R., Vandello, J. A., & Darley, J. M. (2006). Mixed motive and racial bias: The impact of legitimate and illegitimate criteria on decision making. *Psychology, Public Policy, and Law*, 12, 36-55.
- Norton, M. I., Vandello, J. A., Biga, A., & Darley, J. M. (2008). Colorblindness and diversity: Conflicting goals in decisions influenced by race. *Social Cognition*, 26, 102-111.
- Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology*, 87, 817-831.
- Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, 19, 625-664.
- Orbe, M. P., & Harris, T. M. (2008, 2<sup>nd</sup> ed.). *Interracial communication: Theory to practice*. Thousand Oaks, CA: Sage Publications.
- Page, A. (2005). Batson's blind-spot: Unconscious stereotyping and the preemptory challenge. *Boston University Law Review*, 85, 155.
- Pager, D., & Quillian, L. (2005). Walking the talk? What employers say versus what they do. *American Sociological Review*, 70, 355-380.
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181-192.



- Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology, 38*, 384-396.
- People v. Martin, 42 Cal.3d 437 (1986).
- Peruche, B. M., & Plant, A. E. (2006). The correlates of law enforcement officers' automatic and controlled race-based responses to criminal suspects. *Basic and Applied Social Psychology, 28*, 193-199.
- Petersen, T., & Saporta, I. (2004). The opportunity structure for discrimination. *American Journal of Sociology, 109*, 852-901.
- Pettigrew, T. F., & Tropp, L. R. (2000). Does intergroup contact reduce prejudice? Recent meta-analytic findings. In Oskamp, S. (Ed.), *Reducing prejudice and discrimination: The Claremont Symposium on Applied Social Psychology* (pp. 93-114). Mahwah, NJ: Erlbaum.
- Pettit, B., & Western, B. (2004). Mass imprisonment and the life course: Race and class inequality in U.S. incarceration. *American Sociological Review, 69*, 151-169.
- Petty, R. E., & Brinol, P. (2010). Attitude structure and change: Implications for implicit measures. In B. Gawronski, & K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 335-352). New York: Guilford Press.
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer-Verlag.
- Petty, R. E., & Cacioppo, J. T. (1990). Involvement and persuasion: Tradition versus integration. *Psychological Bulletin, 107*, 367-374.
- Pew Center on the States. (2008). One in 100: Behind bars in America. Washington, DC: Author. Retrieved from <http://www.pewcenteronthestates.org/uploadedFiles/One%20in%20100.pdf>
- Pew Research Center for the People & the Press (April 2, 2014). America's new drug policy landscape: Two-thirds favor treatment, not jail, for use of heroin, cocaine. Retrieved from <http://www.people-press.org/2014/04/02/americas-new-drug-policy-landscape/>
- Pippen v. Iowa, No. LACL107038 (Iowa Dist. Ct. Polk County Apr. 17, 2012).
- Plant, A. E., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*, 811-832.
- Plant, A. E., & Peruche, B. M. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science, 16*, 180-183.
- Plaut, V. C. (2010). Diversity science: Why and how difference makes a difference. *Psychological Inquiry, 21*, 77-99.
- Quillian, L., & Pager, D. (2001). Black neighbors, higher crime? The role of racial stereotypes in evaluations of neighborhood crime. *American Journal of Sociology, 107*, 717-67.
- Quillian, L., & Pager, D. (2010). Estimating risk: Stereotype amplification and the perceived risk of criminal victimization. *Social Psychology Quarterly, 73*, 79-104.
- Rattan, A., Levine, C. S., Dweck, C. S., & Eberhardt, J. L. (2012). Race and the fragility of the legal distinction between juveniles and adults. *PLoS ONE, 7*, 1-5.
- Reskin, B. F. (2000). The proximate cause of employment discrimination. *Contemporary Sociology, 29*, 319-328.
- Ricci v. DeStefano, 557 U.S. 557 (2009).
- Richman, D. (1998.) *Terry and the fourth amendment: The process of Terry-lawmaking*. *St. John's Law Review, 72*, 1043-1051.

- Roets, A., Van Hiel, A., & Kruglanski, A. W. (2013). When motivation backfires: Optimal levels of motivation as a function of cognitive capacity in information relevance perception and social judgment. *Motivation & Emotions*, *37*, 261-273.
- Rooth, D. O. (2007). Implicit discrimination in hiring: Real world evidence. *IZA Discussion Paper No. 2764*. Retrieved from [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=984432](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=984432)
- Rudman, L. A., & Ashmore, R. D. (2007). Discrimination and the implicit association test. *Group Processes & Intergroup Relations*, *10*, 359-372.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*, 743-762.
- Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, *39*, 590-598.
- Salonga, R. (2013). San Jose: Breakthrough 'curb sitting' policy to roll out following delay. *San Jose Mercury News*. Retrieved from [http://www.mercurynews.com/ci\\_23606015/san-jose-breakthrough-curb-sitting-policy-roll-out#](http://www.mercurynews.com/ci_23606015/san-jose-breakthrough-curb-sitting-policy-roll-out#)
- Schadewald, M. S., & Limberg, S. T. (1992). Effect of information order and accountability on causal judgments in a legal context. *Psychological Reports*, *71*, 619-625.
- Schuetz v. Coalition to Defend Affirmative Action, 133 S. Ct. 1633 (2013).
- Schuman, H., Steeh, C., Bobo, L., & Krysan, M. (2001). *Racial attitudes in America: Trends and interpretations*. (Revised ed.) Cambridge, MA: Harvard Univ. Press.
- Schlenker, B. R., & Weigold, M. F. (1989). Self-identification and accountability. In R. A. Giacalone & P. Rosenfeld (Eds.), *Impression management in the organization* (pp. 21-43). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality & Social Psychology*, *80*, 645-654.
- Securities and Exchange Commission v. Chenery Corp., 318 U.S. 80 (1943).
- Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, *49*, 11-36.
- Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes*, *65*, 1-17.
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158-174.
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, *51*, 416-446.
- Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology*, *77*, 885-904.
- Sommers, S., & Ellsworth, P. (2000). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin*, *26*, 1367-1379.
- Sommers, S., & Ellsworth, P. (2001). White juror bias: An investigation of prejudice against Black defendants in the American courtroom. *Psychology, Public Policy, and Law*, *7*, 201-229.
- Spencer, S. J., Fein, S., Wolfe, C. T., Fong, C., & Dunn, M. A. (1998). Automatic activation of stereotypes: The role of self-image threat. *Personality and Social Psychology Bulletin*, *24*, 1139-1152.

- Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*, 42-61.
- Stangor, C., Sechrist, G. B., & Jost, J. T. (2001). Changing racial beliefs by providing consensus information. *Personality and Social Psychology Bulletin*, *27*, 484-494.
- Taylor, K. A. (1995). Testing credit and blame attributions as explanation for choices under ambiguity. *Organizational Behavior and Human Decision Processes*, *64*, 128-137.
- Tedeschi, J. T., Schlenker, B. R., & Bonoma, T. V. (1971). Cognitive dissonance: Private ratiocination or public spectacle? *American Psychologist*, *26*, 685-695.
- Terry v. Ohio, 392 U.S. 1 (1968).
- Tetlock, P. E. (1983). Accountability and the complexity of thought. *Journal of Personality and Social Psychology*, *45*, 74-83.
- Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. *Advances in Experimental Social Psychology*, *25*, 331-376.
- Tetlock, P. E., & Boettger, R. (1994). Accountability amplifies the status quo effect when change creates victims. *Journal of Behavioral Decision Making*, *7*, 1-23.
- Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology*, *52*, 700-709.
- Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality & Social Psychology*, *57*, 632-640.
- Thompson, E. P., Roman, R. J., Moskowitz, G. B., Chaiken, S., & Bargh, J. A. (1994). Accuracy motivation attenuates covert priming: The systematic reprocessing of social information. *Journal of Personality and Social Psychology*, *66*, 474-489.
- Topanga Assn. for a Scenic Community v. County of Los Angeles, 11 Cal.3d 506 (1974).
- Tyler, T. R. (1990). *Why people obey the law*. New Haven, CT: Yale University Press.
- Tyler, T. R. (2006). *Why people obey the law: Procedural justice, legitimacy, and compliance*. Princeton, NJ: Princeton University Press.
- Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. New York, NY: Russell-Sage Foundation.
- Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, *16*, 474-480.
- United States v. Capriola, 537 F.2d 319 (9th Cir. 1976).
- United States v. Kent, 383 U.S. 541 (1966).
- Wal-Mart v. Dukes, 131 S. Ct. 2541 (2011).
- Wallaert, M., Ward, A., & Mann, T. (2010). Explicit control of implicit responses: Simple directives can alter IAT performance. *Social Psychology*, *41*, 152-157.
- Walsh, W., Banaji, M. R., & Greenwald, A. G. (1995). A failure to eliminate race bias in judgments of criminals. Paper presented at the Annual Convention of the American Psychological Society, June-July, New York.
- Webster, D. M., Richter, L., & Kruglanski, A. W. (1996). On leaping to conclusions when feeling tired: Mental fatigue effects on impression primacy. *Journal of Experimental Social Psychology*, *32*, 181-195.
- Wells, G. L., Petter, R. E., Harkins, S. G., Kagehiro, D., & Harvey, J. (1977). Anticipated discussion of interpretation eliminates actor-observer differences in the attribution of causality. *Sociometry*, *40*, 247-253.

- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.
- Zajonc, R. B. (1984). On the primacy of affect. In K. R. Scherer and P. Edman (Eds.), *Approaches to emotion* (pp. 259-270). Hillsdale, N.J.: Erlbaum.

## APPENDIX A: CASE FILE NARRATIVES

### **Drug Case Narrative from Study 1 and Study 2, Black Condition**

Responded to dispatch call at 10:57 pm to investigate loud noises reported from neighbor as coming from the suspect's apartment. DeShawn Williams answered the door and identified himself as the apartment's legal occupant. While speaking with the suspect, Officer Howe and I detected an odor of marijuana coming from the unit. Williams said he was having a party with a few friends, and consented to our entering the unit. Several persons were present in the living room, and a protective sweep of the premises was conducted for officer safety. During a search of a small office space towards the back of the apartment I observed in plain view on a desk a plastic ziplock bag containing a white substance, which was later confirmed to be 4.93 grams of cocaine. No one was present in the room at the time of the search. DeShawn Williams appeared nervous throughout the search of the unit. When questioned about the cocaine, he stated that the bag was not his and that several others—including the guests at the party—had been in his apartment with access to the office space. Williams was taken into custody, read his Miranda rights, and he declined to answer any questions.

### **Vandalism Case Narrative from Study 2, Black Condition**

At approximately 1:50AM on April 16, 2012, the police were dispatched to Lincoln High School after a witness called 911 reporting acts of vandalism being committed at the school. At the scene, the investigating officer was told by the reporting party that four males had emerged from a red vehicle, climbed a short fence onto the property and proceeded to push over three light poles along the driveway of the high school and break one of the windows of the main building. The reporting party provided the license plate number which he had copied down. The license number was reported to central communications and shortly thereafter the vehicle in question was stopped by the arresting officer. Darnel Rogers, who was the only person in the vehicle, was taken back to the scene. The reporting party positively identified him as one of the four persons he had seen pushing over the light poles. Rogers stated he had no knowledge about what happened at the school and that he had just gotten back the car from a friend who had borrowed it for the day. It was subsequently determined that the damage done at the crime scene amount to \$825.

### **Robbery Case Narrative from Study 2, Black Condition**

On November 26, 2012, at approximately 7:30 PM, the victim accompanied by her son, 19 years of age, was travelling in her automobile on a city street and noticed that the fire hydrant was spraying across the entire street ahead of her. As she slowed her vehicle someone ran up to the driver's side of the car and threw a bucket of water into her car. Two other suspects ran up to the vehicle, one on each side. The suspect on the passenger's side then reached inside the car window and grabbed the victim's purse which was on the floor of the car. At the same time he also struck the victim's son in the face. The victim then struggled with the suspect who had taken her purse. As she grabbed the purse and started to pull, the purse ripped and he then took it out of her hand. In the purse was a wallet containing \$125, jewelry worth \$200, and credit cards. Though the son's face showed some redness, no other injury was sustained.

The victim gave a description of the suspect who had taken the purse as a Black male, approximately 6 feet tall, and wearing dark pants and shirt. After canvassing the area where the crime had occurred, Tyrrell Hill, who matched the available description, was taken back to the scene to be identified by the victim and her son. The son stated that things happened so fast he wasn't able to get a look at the suspect's face and was unable to positively identify the attacker. The victim said she was almost certain Tyrrell Hill was the person who grabbed her purse.

### **Vandalism Case Narrative from Study 3**

At approximately 1:50AM on 4/16/12, the police were dispatched to Lincoln High School after a witness called 911 reporting acts of vandalism being committed at the school. At the scene, the investigating officer was told by the reporting party that four males had emerged from a red sedan and climbed a short fence onto the property. The witness then heard sounds of banging and glass breaking. The reporting party provided the license plate number which he had copied down.

An inspection of the school grounds revealed that the vandals had pushed over three light fixtures along the driveway of the high school, sprayed graffiti around the property, and broke several windows of the main building. A light in one of the classrooms was on and the room had been trashed, with two of the computer monitors smashed.

The license number was reported to police dispatch. At approximately 3am, the vehicle matching the license plate was stopped about two miles from the school by the arresting officer. The suspect, who was the only person in the vehicle, was taken back to the scene. The reporting party identified the suspect, saying that he had the physical build and was wearing a green t-shirt and a hat just like one of the persons he had seen emerging from the car. Suspect stated he had no knowledge about what happened at the school and that he had just gotten back the car from a friend who had borrowed it for the day. A search of the car uncovered a baseball bat in the truck. Suspect was arrested for vandalism.

### **Battery Case Narrative from Study 3**

On 12/15/12, at approximately 10:30 PM, I received a dispatch call about a disturbance at Dunphy's Bar. When I arrived on scene, the bar manager informed me that the suspect had been drinking at the bar since approximately 8:30pm. Shortly after his arrival, the suspect's ex-girlfriend came into the bar with a few of her friends. One of the friends, later identified as Catherine Burbank, was involved in an altercation with the suspect and the manager called 911 after the verbal argument became physical.

I first spoke with Ms. Burbank. She stated that she was just talking with her friends when the suspect yelled at her to "Shut up, bitch. You know better than to talk that way about a man." She replied that she could say anything she wanted. The suspect then lunged at her and pushed her down hard by hitting her in the chest. She fell backwards onto the ground and she twisted her ankle. She also had a bruise on her right elbow and hip. While speaking, she was very agitated and was not able to put pressure on her left ankle.

I then spoke with the suspect, who said that he was just having a good time when one of his ex-girlfriend's friends started mouthing off loudly by making fun of his physique and his sexual performance. He said that he told her to be quiet, but she got in his face. He denied having pushed her during this interaction. He said that she tripped on a chair and fell down on her own because she was drunk. He was sorry that she had hurt her ankle, but said he didn't play any role in her injury.

Finally, I spoke with the suspect's ex-girlfriend who said she was in the bathroom when the suspect pushed her friend down. She said the suspect had always had a hot temper and that's why she broke up with him. She said that her friend had only had a drink or two and wasn't drunk enough to fall down on her own. Suspect was arrested for Battery.

## APPENDIX B: LEGAL INSTRUCTIONS

### Drug Case:

#### Possession of Controlled Substances:

Provisions of the California Health and Safety Code state that every person who **possesses** any controlled substance shall be punished accordingly under state law.

“Controlled substances” includes methamphetamines (including crystal meth), ecstasy, heroin, cocaine, “GHB,” “PCP,” specific anabolic steroids, and ketamine (“special K”).

In order to be in violation of this law, it must be proven that:

1. The defendant possessed a controlled substance
2. The defendant knew of its presence
3. The defendant knew of the substance’s nature or character as a controlled substance

A person is said to “possess” drugs when either they have **direct and immediate physical control over it** (e.g., on their body, in a pocket, in a purse or briefcase) **OR** when the **drugs were discovered in a location over which the person exercises control**.

### Vandalism Case:

#### Vandalism:

Provisions of the California Penal Code state that every person who willfully damages or destroys the property of others shall be punished accordingly under state law.

In order to be in violation of this law, it must be proven that:

1. The defendant maliciously damaged or destroyed real or personal property;
- AND
2. The defendant did not own the property

#### Criminal Trespass:

Provisions of the California Penal Code state that every person who willingly trespasses upon the property of others shall be punished accordingly under state law.

In order to be in violation of this law, it must be proven that:

1. The defendant willfully entered land or a building belonging to someone else;



2. When the defendant entered, he intended to damage someone else's property;  
AND
3. The defendant actually did damage someone else's property.

**Robbery Case:**

**Robbery:**

Provisions of the California Penal Code state that every person who takes property from the person of another shall be punished accordingly under state law.

In order to be in violation of this law, it must be proven that:

1. The defendant took property that was not (his/her) own;
2. The property was taken from another person's possession and immediate presence;
3. The property was taken against that person's will;
4. The defendant used force or fear to take the property or to prevent the person from resisting;  
AND
5. When the defendant used force or fear to take the property, he intended to deprive the owner of it permanently.

**APPENDIX C: MEANS AND STANDARD DEVIATION TABLES  
FOR OUTCOME MEASURES. STUDIES 1-3**

*Table 12: Study 1 Drug Case Sample Means & Standard Deviations*

	Control		Process Acct.		Outcome Acct.	
	White	Black	White	Black	White	Black
Seriousness	5.10 (1.06)	4.77 (1.20)	4.64 (1.21)	4.77 (1.11)	4.79 (1.27)	4.76 (1.14)
Suspect Guilt	5.29 (1.07)	4.98 (1.34)	4.89 (1.25)	4.93 (1.02)	4.85 (1.30)	4.76 (1.20)
Fine Amount	633.19 (287.87)	749.75 (376.87)	733.46 (375.15)	700.26 (400.89)	621.03 (287.60)	608.71 (445.20)
Comm. Serv. Hours	180.43 (182.00)	134.96 (150.68)	120.11 (109.99)	202.60 (286.12)	189.83 (272.99)	112.25 (115.36)
Days Incarcerated	126.91 (155.67)	134.54 (154.69)	112.44 (130.54)	119.40 (135.44)	132.82 (151.32)	93.09 (128.51)

*Table 13: Study 2 Drug Case Sample Means & Standard Deviations*

	Control		Process Acct.		Directed Acct.	
	White	Black	White	Black	White	Black
Seriousness	4.62 (1.36)	4.57 (1.62)	4.34 (1.48)	4.40 (1.43)	4.63 (1.35)	4.92 (1.44)
Suspect Guilt	5.31 (1.19)	4.76 (1.61)	4.84 (1.56)	5.23 (1.45)	5.00 (1.25)	5.12 (1.39)
Fine Amount	711.01 (292.26)	540.74 (323.17)	667.16 (615.67)	609.08 (272.86)	670.02 (540.84)	675.52 (297.69)
Comm. Serv. Hours	179.35 (236.44)	84.89 (84.83)	97.06 (79.56)	125.38 (158.08)	129.24 (130.29)	187.30 (246.17)
Days Incarcerated	144.21 (184.14)	106.39 (117.53)	115.20 (132.75)	103.67 (109.99)	119.13 (167.00)	145.37 (176.93)
Punishment Composite	0.17 (0.82)	-0.25 (0.83)	-0.06 (0.86)	-0.03 (0.73)	-0.07 (0.94)	0.14 (0.82)

*Table 14: Study 2 Vandalism Case Sample Means & Standard Deviations*

	Control		Process Acct.		Directed Acct.	
	White	Black	White	Black	White	Black
Seriousness	4.88 (1.16)	4.80 (1.35)	4.56 (1.48)	5.09 (1.10)	4.83 (1.18)	5.00 (1.21)
Suspect Guilt	5.79 (1.30)	5.44 (1.50)	5.53 (1.33)	5.72 (1.30)	5.30 (1.48)	5.38 (1.39)
Fine Amount	894.47 (251.08)	762.14 (365.42)	877.80 (214.77)	901.33 (219.32)	860.78 (252.19)	852.00 (199.91)
Comm. Serv. Hours	168.24 (231.31)	105.40 (106.64)	161.83 (198.51)	177.68 (245.16)	165.02 (185.99)	222.69 (250.63)
Days Incarcerated	93.96 (137.84)	92.20 (139.14)	115.02 (150.77)	90.36 (101.18)	99.93 (162.28)	86.06 (109.02)
Punishment Composite	-0.01 (0.71)	-0.29 (0.89)	0.08 (0.57)	0.06 (0.66)	0.01 (0.83)	0.15 (0.62)

*Table 15: Study 2 Robbery Case Sample Means & Standard Deviations*

	Control		Process Acct.		Directed Acct.	
	White	Black	White	Black	White	Black
Seriousness	6.16 (0.94)	6.22 (0.79)	6.04 (0.92)	6.30 (0.86)	6.26 (0.79)	6.47 (0.70)
Suspect Guilt	5.49 (1.27)	4.90 (1.54)	5.15 (1.63)	4.66 (1.59)	4.83 (1.73)	4.92 (1.67)
Fine Amount	928.70 (219.70)	925.45 (285.80)	970.17 (252.05)	1274.33 (1256.27)	962.79 (279.75)	953.14 (288.64)
Comm. Serv. Hours	298.52 (347.09)	244.06 (258.41)	313.24 (245.33)	459.75 (726.85)	473.61 (817.69)	503.32 (508.79)
Days Incarcerated	364.80 (312.72)	377.64 (299.00)	477.60 (427.08)	491.09 (324.90)	502.88 (357.99)	544.40 (522.83)
Punishment Composite	-0.17 (0.64)	-0.19 (0.75)	0.03 (0.64)	0.11 (1.03)	0.13 (0.77)	0.18 (0.63)

*Table 16: Study 3 Vandalism Case Sample Means & Standard Deviations*

	Control		Process Acct.		Directed Acct.	
	White	Hispanic	White	Hispanic	White	Hispanic
Seriousness	4.32 (1.24)	4.39 (1.20)	4.42 (1.20)	4.41 (1.24)	4.47 (1.15)	4.35 (1.14)
Suspect Guilt	4.67 (1.47)	4.77 (1.40)	4.98 (1.13)	4.77 (1.52)	4.65 (1.40)	4.85 (1.32)
Aggressiveness	3.88 (1.48)	4.02 (1.49)	3.83 (1.34)	4.00 (1.38)	3.90 (1.29)	4.00 (1.52)
Poor Moral Character	4.36 (1.41)	4.69 (1.60)	4.49 (1.40)	4.51 (1.42)	4.44 (1.43)	4.44 (1.50)
Punishment Severity	3.86 (1.17)	3.95 (1.35)	4.00 (1.15)	3.91 (1.19)	3.88 (1.05)	3.76 (1.24)
Fine Amount	5.86 (2.40)	5.82 (2.44)	6.12 (2.37)	5.80 (2.58)	6.01 (2.46)	6.02 (2.50)
Days Incarcerated	3.09 (2.18)	3.42 (2.31)	3.07 (2.17)	3.53 (2.31)	2.74 (1.72)	3.53 (2.50)
Character Composite	-0.06 (0.94)	0.10 (0.97)	-0.04 (0.87)	0.03 (0.90)	-0.03 (0.84)	0.01 (0.95)
Punishment Composite	-0.05 (0.74)	0.03 (0.82)	0.03 (0.75)	0.03 (0.84)	-0.07 (0.67)	0.01 (0.85)

*Table 17: Study 3 Battery Case Sample Means & Standard Deviations*

	Control		Process Acct.		Directed Acct.	
	White	Hispanic	White	Hispanic	White	Hispanic
Seriousness	5.00 (1.28)	5.28 (1.31)	4.80 (1.38)	4.86 (1.21)	4.83 (1.35)	4.73 (1.31)
Suspect Guilt	4.91 (1.42)	5.09 (1.38)	4.62 (1.67)	4.76 (1.52)	4.47 (1.55)	4.73 (1.56)
Aggressiveness	5.09 (1.30)	5.46 (1.33)	4.93 (1.57)	5.09 (1.39)	4.68 (1.42)	5.02 (1.56)
Poor Moral Character	4.88 (1.43)	5.16 (1.45)	4.48 (1.58)	4.63 (1.46)	4.52 (1.45)	4.59 (1.52)
Punishment Severity	4.55 (1.40)	4.49 (1.41)	4.08 (1.52)	4.29 (1.46)	3.96 (1.49)	4.02 (1.47)
Fine Amount	5.15 (2.55)	5.08 (2.88)	4.90 (3.08)	4.64 (2.84)	4.49 (2.99)	4.66 (2.66)
Days Incarcerated	3.73 (2.55)	4.20 (2.75)	3.51 (2.59)	3.67 (2.70)	3.38 (2.75)	3.65 (2.57)
Character Composite	-0.11 (0.93)	0.13 (0.93)	-0.05 (1.01)	0.06 (0.87)	-0.06 (0.93)	0.07 (0.99)
Punishment Composite	-0.02 (0.82)	0.02 (0.89)	-0.01 (0.91)	0.02 (0.86)	-0.04 (0.91)	0.04 (0.86)