# SIMILARITY LENIENCY IN MENS REA DETERMINATIONS AND THE MEDIATING ROLE OF CAUSAL ATTRIBUTIONS

**Alexander C.V. Jay**
**Charles B. Stone**
**Mark R. Fondacaro**
**John Jay College of Criminal Justice**
**The Graduate Center, City University of New York**

**Joonwon Yoon**
**Klaudia Zuraw**
**John Jay College of Criminal Justice**

Jurors are tasked with determining if the defendant's alleged transgression is the result of the defendant's guilty mind or other mitigating situational factors. The extant research, though, suggests that jurors tend to err in making these mens rea judgments. Jurors may have particular difficulty in judging mens rea when the defendant is a different race than the juror (i.e., harsher treatment of cross-raced defendants and less harsh treatment of same raced defendants, what are known as similarity leniency effects; Mitchell et al., 2005). The present research examined jurors' mens rea judgments of cross-raced defendants, whether there is any relation between similarity leniency effects and implicit racial biases, and whether causal attributions (i.e., emphasizing dispositional vs. situational causal attributions) mediate the relation between juror race and cross-race judgments of a defendant's mens rea (i.e., similarity leniency effects).

The results indicate that causal attributions impact Black jurors' same-raced judgments, and that implicit racial bias is not related to similarity leniency. That is, for Black jurors tasked with determining a defendant's culpability, Black defendants will elicit situational causal attributions, and White defendants will elicit dispositional causal attributions, and these disparate causal attributions can influence mens rea determinations regardless of the jurors' implicit racial bias. The present results have important implications for researchers and legal system practitioners interested in ensuring that the jury system provides defendants with a fair and unbiased trial.

The Sixth Amendment of the United States Constitution guarantees every criminal defendant the right to a trial by an impartial jury of his or her peers. Once empaneled, jurors are tasked with determining whether the defendant is criminally responsible for an event, and this task has two primary components. One component is *actus reus*, which re-

Authors' note. Correspondence should be sent to Alexander C. V. Jay, Department of Psychology, John Jay College of Criminal Justice, The City University of New York, 524 West 59th St, New York, NY 10019 E-mail: acjay6@gmail.com. The authors declare no conflicts of interest. The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

quires jurors to consider if the defendant committed the criminal act voluntarily (Gordon & Fondacaro, 2018). A second component is *mens rea*, which requires jurors to determine the defendant's culpability, that is, whether the defendant had a guilty mind at the time of the alleged event. Few studies have directly examined jurors' mens rea determinations, but those that have suggest that people can make errors in rendering mens rea judgments (e.g., inferring a higher category or level of culpability than is legally appropriate, such as a *purposeful* mens rea rather than *reckless* when *reckless* is appropriate; Beattey & Fondacaro, 2018).

In addition to more general errors in mens rea judgments, jurors might also make biased mens rea judgments as a result of racial bias (e.g., Mitchell et al., 2005; Sommers, 2006) and biased causal attributions (i.e., emphasizing dispositional over situational causal attributions for some people but not others; Sommers & Ellsworth, 2000), as both can affect jurors' legal judgments. That is, both racial bias and biased causal attributions may lead to unfavorable and even discriminatory legal outcomes for defendants, particularly when the defendant is a different race than the juror (i.e., harsher treatment of cross-raced defendants and less harsh treatment of same raced defendants, what are known as similarity leniency effects; Mitchell et al., 2005). However, the precise nature of the relation between racial bias and biased attributions, on the one hand, and judgments of mens rea, on the other, have not been examined empirically. The present study fills this gap in the literature.

The primary aim of our study is to examine whether similarity leniency is evident in jurors' judgments of mens rea, and whether causal attributions (i.e., dispositional vs. situational) mediate this relationship, such that jurors are more lenient on same- (vs. other-) raced defendants, and their causal attributions explain this in-group leniency/out-group punitiveness. Within the present study, mens rea is assessed using a scale of culpability developed by Graham and Lowery (2004) to provide more sensitive (vs. a dichotomous yes/no) juror judgments of a defendant whose mental state was previously determined by legal experts to evince a reckless mens rea (i.e., Shen et al., 2011). Additionally, the present study will assess whether jurors' implicit racial biases predict racially biased mens rea judgments, as implicit racial biases have been identified as one possible explanation for the robust similarity leniency effect (e.g., Hunt, 2015). However, research suggests that Black research participants' show no clear implicit preference for their in-group (e.g., Black people vs. White people, as measured by the Race Implicit Association Test; Greenwald & Krieger, 2006). Thus, it remains unclear whether jurors' implicit racial biases predict racially biased judgments that are consistent with the direction of their implicit bias, regardless of a jurors' in-group (e.g., whether a Black juror implicitly biased in favor of White people demonstrates racially biased judgments against Black people), or whether jurors would make biased judgments that strictly follow in-group/out-group status (i.e., White and Black jurors demonstrate in-group/out-group bias in the form of similarity leniency regardless of the direction of their implicit racial bias).

In what follows, we will first provide a review of the literature on jurors' mens rea judgments. Second, we will discuss the relevant research examining racial bias as similarity leniency in juror decision-making, including a discussion of implicit racial bias and how

jurors' implicit racial biases might predict similarity leniency effects (e.g., Hunt, 2015), as well as how they might not (e.g., Greenwald & Krieger, 2006). Third, we describe the literature assessing how causal attributions affect jurors' decision-making, and how they might mediate the effect of a juror's race on their cross-raced legal judgments. Last, we outline the present study.

### *Jurors' Mens Rea Judgments*

To render a guilty verdict at trial, jurors must determine whether the defendant committed an illegal act with a culpable mental state at or above the level of culpability specified in a criminal statute (with increasing levels of culpability ranging from negligence, to recklessness, to knowledge, to purposeful). Although some research suggests that jurors make mens rea determinations accurately and reliably (e.g., Shen et al., 2011), other research (e.g., Beattey & Fondacaro, 2018) and legal analyses (e.g., Fondacaro & O'Toole, 2015) suggest that jurors have difficulty distinguishing between the different mens rea categories noted above (Severance & Loftus, 1982), particularly the two middle categories of knowing and reckless (Severance et al., 1992). For example, Severance et al. (1992) found that mock-jurors could reliably distinguish between the lowest (i.e., negligent) and highest (i.e., purposeful) mens rea categories, but erred in distinguishing any other mens rea categories (i.e., purposeful vs. knowing, knowing vs. reckless, and reckless vs. negligent).

Consequently, jurors' mens rea determinations can be unreliable and might undermine the likelihood of a fair and just outcome for criminal defendants (Fondacaro & O'Toole, 2015). For example, Shen and colleagues (2011) found that jurors had difficulty distinguishing between reckless and knowing mental states, and a closer examination of their results indicates that there was low reliability between the mock jurors' judgments. That is, the consistency between mock-jurors' judgments was at or below 50% for negligent, reckless, and knowing mental states and only 78% for a purposeful mental state (Shen et al., 2011). Additionally, mock jurors exhibited inaccurate judgements, inferring higher (i.e., more intentional/culpable) mens rea categories, such as knowing or a purposeful mental state, in 47% of the cases when reckless was the legally correct choice (Shen et al.). Similarly, Beattey and Fondacaro (2018) found that people often infer higher (i.e., more purposeful) mens rea categories than is warranted given the fact pattern of the case. For example, when participants were presented with fact patterns that were previously rated by legal professionals as establishing a reckless mental state, nearly 75% of participants indicated that the prosecutor had met his burden of establishing a *purposeful* mental state beyond a reasonable doubt.

Overall, the available research on jurors' mens rea determinations suggests that, at best, jurors have difficulty distinguishing between mens rea categories, and at worst, jurors err consistently and in a way that is detrimental to defendants. Yet, these errors in mens rea judgments might be amplified further by the racial characteristics of the case, and this might be true for both racial minority (e.g., Black) and majority (i.e., White) defendants (Devine & Caughlin, 2014; Mitchell et al., 2005).

### *Racial Bias and Similarity-Leniency*

Despite some inconsistent results (e.g., see Mazzella & Feingold, 1994), research at the intersection of psychology and law demonstrates how the racial characteristics of a criminal case such as the race of the defendant and/or juror affect case outcomes (e.g., Devine & Caughlin, 2014; Hunt, 2015; 2017; Mitchell et al., 2005). This literature suggests that White jurors can make racially biased judgments, leading to discriminatory legal outcomes for minority defendants (e.g., Sweeney & Haney, 1992). However, it is not simply the case that White jurors are uniquely biased against Black defendants, but rather that this pattern of racially biased judgment is one part of a broader similarity-leniency effect pattern (Devine & Caughlin, 2014; Kerr et al., 1996; Mitchell et al., 2005).

Similarity leniency is found when jurors of one race (e.g., White or Black) are more lenient when judging a defendant of the same race (e.g., White or Black, respectively), and harsher or more punitive when judging a defendant of a different race (Devine & Caughlin, 2014; Kerr et al., 1996; Mitchell et al., 2005). As such, jurors' racial bias threatens the rights of all defendants with cross-raced jurors. In fact, two recent meta-analyses demonstrated that similarity leniency effects are stronger for Black mock jurors judging Black defendants than they are for White mock jurors judging White defendants. This robust similarity-leniency effect is thus likely to extend to both majority (i.e., White) and minority (e.g., Black) jurors' mens rea judgments, leading to harsher judgments toward defendants of a different race, but this hypothesis has yet to be tested empirically. Moreover, the mechanism driving this phenomenon remains unclear (Devine & Caughlin, 2014).

Some have speculated that jurors' implicit racial biases might be one factor underlying the similarity-leniency effect (Hunt, 2015). However, for practical reasons and applied concerns (i.e., the considerable disproportionate rate of racial minorities interfacing with the criminal justice system; Kang et al., 2012), the implicit racial bias literature has focused primarily on how implicit racial biases affect jurors' judgments and decision-making processes with Black defendants (e.g., Mazzella & Feingold, 1994; Sweeney & Haney, 1992), without adequate consideration of the interactive effect of the mock jurors' race and defendant race (Devine & Caughlin, 2014). After discussing the literature on implicit racial bias in juror decision-making with Black defendants, we will consider how implicit racial bias might, and how it might not, predict similarity leniency effects in Black and White jurors' legal judgments of Black and White defendants.

### *Implicit Racial Bias and Black Defendants*

People who are explicitly non-prejudiced still possess implicit racial biases that can affect their judgments and behaviors toward racial minorities (e.g., Dovidio & Gaertner, 2004; Dovidio et al., 1997; Gaertner & Dovidio, 1986; Oliver, 2003). In the context of a criminal trial, implicitly held stereotypes and prejudicial beliefs can affect jurors' legal judgments and behaviors without their awareness, ultimately leading to discriminatory outcomes for Black defendants (Hunt, 2015; 2017; Kang et al., 2012). For instance, tasked with judging a Black (vs. White) criminal defendant, negative implicit associations (e.g., Black people are implicitly associated with the concept of guilt; Levinson et al., 2010) can increase jurors' perceptions of evidence strength against a Black defendant (Levinson et

al., 2010) and bias their memory and evidence interpretations toward inculpating (vs. exculpating) evidence (Gaertner & Dovidio, 1986; Hodson et al., 2005; Johnson et al., 1995; Levinson, 2007; Levinson & Young, 2010).

Implicit racial biases are also believed to be responsible for more guilty verdicts for Black (vs. White) defendants (e.g., Levinson & Young, 2010), as well as a possible explanation for White jurors' increased punitiveness toward Black (vs. White) defendants in capital (i.e., death penalty) trials (e.g., Lynch & Haney, 2000, 2009). Implicit racial biases might thus predict discriminatory mens rea judgments for minority defendants. However, there is an acknowledged lack of research on the effect of implicit racial biases and juror decisions (Devine & Caughlin, 2014; Hunt, 2015; 2017). As a result, little is known about how implicit racial biases operate in cross-raced contexts. Given the field's focus on the effects of jurors' implicit biases on Black defendants without consideration of how they operate in cross-raced judgment scenarios, it is understandable that to the best of our knowledge there are no studies that actually measure Black and White jurors' implicit racial biases and assess whether they predict their legal judgments of same and/or other-raced defendants. Thus, even though implicit racial bias is one possible cause of the robust similarity-leniency effect (Hunt, 2015), the lack of research testing this relationship raises questions as to whether it exists, and there is reason to believe it might not.

***Similarity Leniency***

As stated above, similarity leniency is seen in both White and Black jurors, but the effects are consistently stronger for Black jurors (i.e., they are more lenient toward Black and harsher toward White defendants; Devine & Caughlin, 2014; Mitchell et al., 2005). If implicit racial bias produces similarity leniency effects, one might expect Black people to show an implicit bias for their in-group that is at least as strong as, if not stronger than, the implicit bias White people have for other White people. But there is evidence to the contrary, showing that White and Black people share many of the same implicit biases (including those that favor White over Black people; Dovidio et al., 2000; Greenwald et al., 1998; Greenwald et al., 2009; Krueger, 1996). Moreover, whereas most White people show an implicit preference for White people (approximately 72% favoring White vs 7% favoring Black and 21% favoring neither), Black people show no discernable implicit preference for their own group (34% favoring Black vs 32% favoring White and 34% favoring neither; Greenwald & Krieger, 2006). This discrepancy in White and Black people's implicit in-group preferences is difficult to reconcile with the results of the two meta-analyses demonstrating similarity leniency effects in jurors' cross raced judgments (i.e., Devine & Caughlin, 2014; Mitchell et al., 2005) if implicit racial bias does indeed produce similarity leniency.

A single study has examined the effect of Black people's implicit racial biases on their decision-making or behavior toward Black and White targets, albeit not in a jury decision-making context. Ashburn-Nardo et al. (2003) assessed whether Black participants' pro-White or pro-Black implicit racial bias would predict whether they wanted to work with a White or Black partner on an upcoming task that was supposedly difficult. The researchers found that Black participants with pro-White implicit bias (i.e., those participants show-

ing an implicit preference for White rather than Black people) indicated that they preferred to work with a White rather than a Black partner on an upcoming challenging task. This finding is inconsistent with the robust similarity-leniency effect found in meta-analyses of cross-raced judgments in mock jury research (e.g., Devine & Caughlin, 2014). Although unlikely given experimental controls such as random assignment, it is at least possible that jurors' implicit biases predict their case judgments, and that the minimal extant literature on cross-raced juror judgments demonstrating a robust similarity leniency effect had, by chance, samples of White and Black mock jurors with implicit biases that favored their in-groups. This cannot be ruled out entirely, however, due to the lack of measurement of implicit racial biases in the extant literature.

Given that the stand-alone study assessing the effect of Black participants' implicit biases on their cross-raced judgments (i.e., Ashburn et al., 2003) was not within the jury decision-making arena, and considering the data showing that Black people's implicit racial biases show no discernible in-group preference, more research is needed to clarify the relationship between Black and White jurors' implicit racial biases and their judgments with same or cross-raced defendants, including their mens rea assessments and any similarity leniency effects. Thus, the present research directly examines whether Black and White jurors' level of implicit racial bias predicts similarity leniency in their mens rea judgments. In addition to addressing these important *why* questions of racial bias in juror decision-making, the present research also examines the *how* questions, and examines whether causal attributions mediate the relationship between jurors' racial bias and their mens rea judgments for same and other raced defendants (Pettigrew, 1979; Sommers & Ellsworth, 2000).

***Causal Attributions as a Mediator***

To the extent that racial biases are associated with mens rea assessments, it raises the question of what psychological processes, if any, mediate this relationship. Given that race is known to affect whether people make biased causal attributions to explain others' behavior (e.g., Pettigrew, 1979), and that biased causal attributions predict more punitive legal judgments (e.g., Sommers & Ellsworth, 2000), one potential candidate is biased causal attribution. A person's causal attribution for why another person behaved as they did may vary depending on whether the person being judged is an in-group or out-group member. Pettigrew's (1979) Ultimate Attribution Error (UAE) is the tendency to ascribe positive behaviors to dispositional characteristics of in-group members and to situational influences on out-group members. The reverse is true for negative behaviors, which are attributed to situational influences on in-group members and to dispositional characteristics of out-group members. In the context of a criminal trial and juror decision-making, White and Black jurors might make different attributions (i.e., dispositional or situational) for the cause of a crime depending on whether the defendant is an in-group member (i.e., White or Black, respectively).

Importantly, whether a juror attributes a crime to a defendant's disposition, or the situation, is directly relevant to their legal judgments and the case outcome, as the extant literature suggests that emphasizing negative dispositional over situational causal attributions for a crime may lead to unfavorable verdicts and sentencing outcomes for defendants. For

example, mock-jurors who make negative dispositional attributions or fail to consider situational factors have increased perceptions of guilt (e.g., Sommers & Ellsworth, 2000), favor more punitive (vs. lenient) punishment recommendations (e.g., Hawkins, 1981; Kubota et al., 2014), render more guilty (vs. not guilty) verdicts (Pope & Meyers, 1999), and are more likely to vote for a death (vs. life) sentence (Lynch & Haney, 2000; Stevenson et al., 2010). Thus, it is possible that an association between dispositional or situational causal attributions and jurors' mens rea judgments exists, and that these causal attributions in fact mediate the relationship between racial characteristics of a case (i.e., juror and defendant race) and judgments of culpability. Moreover, this mediating role of causal attributions would suggest that racial bias toward out-group members might be explained, in part, by the tendency to attribute the criminal behavior of an out-group member to dispositional characteristics rather than situational influences.

Likewise, the similarity-leniency effect might be explained in part by the tendency to attribute the criminal behavior of an in-group member to situational influences rather than dispositional characteristics. If so, it would suggest that racial bias and biased causal attributions may be intricately linked in jurors' decision-making process when the defendant is of another race (vs. of the same race) in that biased attributions for an other-raced defendant's behaviors might explain racially discriminatory case outcomes. Indeed, the extant research suggests that this does occur (e.g., Sommers & Ellsworth, 2000).

***Racial Bias and Biased Causal Attributions in Juror Decision-making***

Research suggests that racially biased juror judgments are often accompanied by biased causal attributions, whereby inculpating dispositional factors are emphasized for out-group members or minorities. For example, Duncan (1976) found that White participants are more likely to rate an act of aggression (i.e., a shove) as more violent, and more representative of the perpetrator's disposition when the perpetrator is Black, but not when he is White. In addition to dispositional factors being emphasized when the perpetrator is an out-group member, racially biased decision-making processes can also occur when situational factors are emphasized for in-group members. Levinson (2007) examined whether implicit racial bias might affect jurors' memory for evidence and found that (Japanese, White, Hawaiian, and 'mixed') participants remembered information about a crime event in a racially biased way. That is, participants remembered more aggressive acts (e.g., the main character threw the first punch) when the main character was Black. Conversely, when the character was White, participants remembered more mitigating situational information (e.g., it was an accident). Other research demonstrates that negative dispositional causal attributions often occur in mock jurors' biased cross-raced judgments (e.g., Sommers & Ellsworth, 2000).

Sommers and Ellsworth (2000) examined the interplay of participant race and defendant race on mock jurors' causal attributions and verdict decisions across two studies. They found that jurors attributed the causes of criminal behavior to more dispositional characteristics for defendants of other races (i.e., when the participant was White, and defendant was Black and vice versa in Study 2) but more situational influences for the defendants of their own race (Study 1 and 2). Additionally, the negative dispositional attributions made

for cross-raced defendants (e.g., aggressive, violent) had a strong and positive association with mock jurors' judgments of guilt. Thus, racial bias and biased causal attributions appear linked in producing discriminatory judgments in jurors' decision-making.

Similarly, Graham and Lowery (2004) examined the effects of subliminally priming probation officers and police officers with categorically relevant Black or neutral words on their attributions of hypothetical juvenile offenders and ratings of culpability, likelihood of recidivism, and punishment recommendations. Despite no indication of whether the hypothetical offenders were of a particular race, those officers primed with Black (vs. neutral) words made more negative trait/dispositional causal attributions about the offenders, and more negative trait/dispositional causal attributions predicted higher culpability ratings. Thus, at least for officer's legal decisions, biased causal attributions appear to mediate the relationship between racial bias and their legal judgments. However, it remains unclear whether causal attributions do in fact mediate this relationship in instances of mens rea determinations on the part of jurors. The present study attempts to address this gap in the literature.

## STUDY OVERVIEW AND HYPOTHESES

The present study examines whether Black and White jurors demonstrate racial bias in their mens rea judgments (as measured with the proxy of perceived culpability; Graham & Lowery, 2004) through the causal attributions they make for a crime committed by a Black or White defendant. This proposed moderated mediation model, whereby the effect of a defendant's race on jurors' culpability judgments is explained by jurors' causal attributions, and conditional on whether the juror is of the same or different race, is the primary focus of the present study. However, before testing the proposed moderated mediation model, the present study also assessed whether jurors' IAT scores relate to their case judgments. Although implicit racial biases are one possible factor that underly racially biased juror judgments (i.e., similarity leniency; Hunt, 2015), it is unclear how jurors' implicit racial biases will relate to their judgments of perceived culpability. On the one hand, meta-analyses show that White and Black mock jurors demonstrate similarity leniency effects, but on the other hand, Black mock-jurors consistently demonstrate a stronger similarity leniency effect than White mock-jurors (Devine & Caughlin, 2014; Mitchell et al., 2005), but no obvious implicit preference for their in-group (i.e., compared to White people, who do show such a preference; Greenwald & Krieger, 2006). Thus, in addition to testing the proposed moderated mediation model, this research also addresses a gap in the extant literature by directly measuring one type of jurors' implicit racial bias (i.e., White/Black, good/bad associations), and specifically examining whether a jurors' implicit racial bias relates to their mens rea judgments (and verdicts and punishment preferences[1]) for White and Black defendants who committed a crime rated by legal experts as exhibiting a mens rea of recklessness. To this end, the study tests several hypotheses including a pair of compet-

1 Jurors' verdicts and punishment preferences were included to have a more comprehensive assessment of the effect of their implicit racial bias or the defendant's race on their judgments, although jurors' culpability judgments are the primary focus of this research as a proxy for mens rea judgments. Thus, only jurors' culpability judgments will be assessed in the proposed moderated mediation model.

ing hypotheses to reflect the differing predictions of similarity leniency effects and implicit racial biases, as well as the proposed moderated mediation model:

1.a ***Similarity-Leniency Hypothesis.*** Black and White jurors' judgments will reflect similarity-leniency, such that White jurors will be more lenient toward White (vs. Black) defendants, and Black jurors more lenient toward Black (vs. White defendants), in terms of their judgment of the defendant's culpability, punishment preferences, and verdicts.

1.b. ***Implicit Racial Bias Hypothesis****.* White and Black jurors, regardless of their own race, will be more lenient in their judgment of a defendant's culpability, punishment preferences, and verdicts, the more implicitly biased they are toward that defendant's race (i.e., the more implicitly pro-White the juror is, the more lenient they will be toward the White defendant/the harsher they will be toward the Black defendant and vice versa).

2. ***Attribution Hypothesis.*** Dispositional causal attributions are predicted to increase jurors' perceptions of culpability, and situational causal attributions are predicted to decrease jurors' perceptions of culpability.

3. ***Similarity Leniency Moderated Mediation Hypothesis.*** Causal attributions are predicted to mediate the effect of defendant race on jurors' culpability determinations, such that the race of the defendant will increase jurors' dispositional causal attributions, which will increase judgments of the defendant's culpability, when the defendant is a different race than the juror.

## METHODS

### *Participants and Design*

Black ($n$ = 141) and White ($n$ = 277) jury eligible community members ($N$ = 418)[2] were recruited through Amazon's Mechanical Turk (M-Turk[3]) and randomly assigned to one of the four conditions comprising our 2 (Defendant Race: Black, White) x 2 (Crime Type: Rock Kick, Cycle Crash) x 2 (Participant Race: Black, White) quasi-experimental between-subjects design, with participant race as a quasi-experimental variable. Power analyses for an *a priori* sample for our proposed mediation model indicates that our sample size is sufficiently large ($n \geq 77$) to detect a small effect ($f^2 = .15$) with a statistical power level of 0.80, and alpha = .05 (Cohen, 1988; Cohen et al., 2003). We oversampled to ensure sufficient power for the moderated mediation model, and to account for sample attrition on M-Turk from failed attention checks. The majority of the sample (53%) was between the ages of 26 and 40, and there were slightly more women (53%) than men (47%). Three ($n$ =

2 The difference in sample sizes between Black and White participants was a consequence of how the data were collected via MTurk and a tool for social scientists called Turk Prime, which did not create balanced samples. However, independent samples t-tests confirmed that the Black and White participants had equal variances on all critical variables (i.e., IAT Score, Culpability Ratings, Causal Attribution ratings, and Deserved Punishment Ratings), all Levene's test $F$'s $\leq$ 0.88, all $p$'s $\geq$ .349. Therefore, the unequal sample sizes, although not ideal, are not a threat to the statistical analyses.

3 M-Turk provides samples that are more demographically diverse than other traditional samples (e.g., college students or online community member samples) and is a legitimate source of quality data (Buhrmester et al., 2011; Paolacci et al., 2010).

3) participants failed manipulation checks and were, therefore, excluded from the analyses. Participants were compensated $1.50 for approximately 20 minutes of their time.

***Materials***

*Implicit Association Test*

The IAT (Greenwald et al., 1998) measures the strength of implicit cognitive associations by examining people's reaction times to specific pairings of attitude objects such as people or things and positively or negatively valanced words. The present study used the race IAT, which consists of 12 Black and White male and female faces and six positively valenced words (joy, love, wonderful, pleasure, laughter, and happy) and six negatively valenced words (terrible, horrible, evil, awful, agony, and hurt). If people respond more quickly to certain pairings (e.g., White and positive words or Black and negative words presented simultaneously on the screen), they are said to have an implicit preference for that specific association, or in other words, an implicit bias. That is, people will respond faster to pairings that are consistent with their implicit associations (i.e., stronger associations), and slower to pairings that are inconsistent with their implicit associations (i.e., weaker associations).

*Crime Scenarios*

Participants were assigned to one of four brief crime scenarios, each involving one (Black or White) criminal defendant who allegedly committed one of two recklessness mens rea crimes (Rock Kicking or Cycle Crash; Shen et al., 2011; see Appendix A). A reckless mens rea was selected because culpability for reckless crimes is more ambiguous compared to those that require purpose or evidence of intent, creating a decision-making context where implicit biases are likely to exert their influence (Gaertner & Dovidio, 1986). Defendant race was manipulated by including photographs (from the shoulder up, in white t-shirts and a white background) of the defendants in the vignettes. The crime type was manipulated in the content of the scenarios which were written in the style of short vignettes, and described the context of the crime, the defendant's actions, and the outcome. Shen et al. (2011) pretested the vignettes with nine criminal law professors who confirmed that the language within the vignettes correctly communicated a reckless mental state. We included two crimes in order to boost the reliability of how we assessed a reckless crime. In the Rock Kick conditions, John allegedly kicked a stone off of an overpass which resulted in the rock hitting a passing car and causing the driver serious injury. In the Cycle Crash conditions, John was a cyclist who was racing against his top rival in a big championship race through hilly terrain. He allegedly passed his rival cyclist too tightly, causing the rival cyclist to fall down and suffer serious injuries. After the crime scenario, participants received the Model Penal Code § 2.02 (i.e., a reckless mens rea requirement for guilt; The American Law Institute, 1962).

*Perceived Culpability*

Participants rated the defendant's perceived culpability on four, 7-point items with scales ranging from 1 (*Very Unlikely*) to 7 (*Very Likely*) (Graham & Lowery, 2004). For example, participants were asked how likely it was that the defendant was aware that his

actions were a criminal act for which he could be prosecuted. The four items were averaged to create a composite score of the defendant's perceived culpability (Cronbach's $a$ = .86).

*Verdict and Deserved Punishment*

Participants indicated if they believed the defendant was either a) Guilty or b) Not Guilty. Participants also rated on a 10-point scale from 1 (*No Punishment*) to 10 (*Extreme Punishment*), what severity of punishment they believed the defendant deserved (Shen et al., 2011).

*Crime Attribution*

Participants reported which of the two possible causes, a) situational pressure or b) the defendant's personal character, was more responsible for the defendant's alleged crime on a 7-point scale ranging from 1 (*Definitely Situational Pressure*) to 7 (*Definitely his Personal Character*).

*Demographics*

Participants were asked to provide basic demographic details (e.g., age, gender, race, etc.).

***Procedure***

After agreeing to participate online, participants were randomly assigned to one of the four conditions, and began the study by completing the Implicit Association Test (IAT; Greenwald et al., 1998). Immediately after completing the IAT, participants read the crime scenarios, followed by a description of the Model Penal Code § 2.02 (i.e., a reckless mens rea requirement for guilt; The American Law Institute, 1962). Participants then completed, always in the following order, the defendant's culpability measure, followed by the dichotomous verdict measure, a situational versus dispositional causal attribution measure, and finally, the demographics questionnaire. Upon completion of the survey, participants were debriefed and thanked for their time.

## RESULTS

The moderated mediation analyses are the primary interest of this paper, but first we describe the sample's IAT results followed by several correlational analyses that were conducted as an initial assessment of the relationship amongst participants' IAT scores, causal attributions, culpability ratings, verdicts (dichotomous, 0 = Not Guilty, 1 = Guilty), and punishment preferences as a function of participant and defendant race (see Table 1 and Table 2 for descriptive statistics of the key variables in each of the crime vignette scenarios).

Table 1. Descriptive Statistics by Juror Race and Defendant Race in the Rock Kick Condition

| | | Participant Race | | | | | |
|---|---|---|---|---|---|---|---|
| | | Black | | | White | | |
| Defendant Race | Variable | *M* | *SD* | *N* | *M* | *SD* | *N* |
| Black | Casual Attributions | 3.94 | 1.50 | 32 | 5.22 | 1.52 | 66 |
| | Culpability | 4.87 | 1.5 | 32 | 5.10 | 1.19 | 66 |
| | Punishment Preferences | 3.84 | 1.62 | 32 | 3.78 | 1.41 | 66 |
| White | Casual Attributions | 5.33 | 1.59 | 40 | 5.40 | 1.74 | 73 |
| | Culpability | 4.79 | 1.37 | 40 | 5.23 | 1.29 | 73 |
| | Punishment Preferences | 3.75 | 1.41 | 40 | 4.15 | 1.47 | 73 |

Note. Causal Attribution ratings ranged on a scale of 1 (Situational Pressure) to 7 (Defendant's Personal Character). Culpability was the average of four items related to their perceptions of the likelihood of the defendant's culpability, all of which ranged on a scale from 1 (Very Unlikely) to 7 (Very Likely). Punishment Preference ratings ranged from 1 (No Punishment) to 10 (Extreme Punishment).

Table 2. Descriptive Statistics by Juror Race and Defendant Race in the Cycle Crash Condition

| | | Participant Race | | | | | |
|---|---|---|---|---|---|---|---|
| | | Black | | | White | | |
| Defendant Race | Variable | *M* | *SD* | *N* | *M* | *SD* | *N* |
| Black | Casual Attributions | 3.15 | 1.68 | 33 | 3.39 | 1.60 | 67 |
| | Culpability | 4.30 | 1.68 | 33 | 4.18 | 1.76 | 67 |
| | Punishment Preferences | 3.06 | 1.48 | 33 | 2.72 | 1.51 | 67 |
| White | Casual Attributions | 4.67 | 1.47 | 36 | 3.85 | 1.65 | 71 |
| | Culpability | 5.19 | 1.33 | 36 | 4.33 | 1.59 | 71 |
| | Punishment Preferences | 3.92 | 1.59 | 36 | 3.00 | 1.65 | 71 |

Note. Causal Attribution ratings ranged on a scale of 1 (Situational Pressure) to 7 (Defendant's Personal Character). Culpability was the average of four items related to their perceptions of the likelihood of the defendant's culpability, all of which ranged on a scale from 1 (Very Unlikely) to 7 (Very Likely). Punishment Preference ratings ranged from 1 (No Punishment) to 10 (Extreme Punishment).

Next, a series of multiple regression analyses assess main and interaction effects between IAT Sores (as a continuous variable), Defendant Race (0 = White, 1 = Black), Participant Race (0 = White, 1 = Black), and Crime (i.e., Rock Kick or Cycle Crash) on participants' perceptions of culpability and punishment preferences (Hypothesis 1a and 1b). Logistic regressions will then assess main and interaction effects of IAT Sores (as a continuous variable), Defendant Race (0 = White, 1 = Black), Participant Race (0 = White, 1 = Black), and Crime (i.e., Rock Kick or Cycle Crash) on participants' verdict preferences (Hypothesis 1a and 1b). Finally, a moderated mediation analysis to determine whether the defendant's race predicts culpability ratings through jurors' causal attributions as a function of the jurors' race (i.e., conditional indirect effects; Hypothesis 2 and 3).

***IAT results***

Greenwald et al.'s (1998) race IAT scoring conventions maintain that high numbers (i.e., > 0) indicate a pro-White implicit racial bias, or preference, whereas low numbers (i.e., < 0), indicate a pro-Black implicit racial bias or preference. The higher the number (up to a maximum of 2) or the lower the number (down to a minimum of -2) reflect the strength of the bias. The overall sample (i.e., both Black and White participants) showed a small pro-White implicit bias ($M_D$ = 0.29, $SD_D$ = 0.40). When broken down by participants race, White participants showed a significantly stronger pro-White implicit bias ($M_D$ = 0.36, $SD_D$ = 0.38) compared to Black participants' pro-White implicit bias ($M_D$ = 0.16, $SD_D$ = 0.41), $t(416) = 4.93$, $p < .001$, $d = .51$. When assessing the absolute breakdown, according to IAT conventions, 82% of White ($n$ = 227) and 65% of Black ($n$ = 91) jurors showed a pro-White implicit bias, compared to 18% of White ($n$ = 50) and 36% of Black ($n$ = 50) jurors who showed a pro-Black implicit bias. These IAT results suggest that our sample is more pro-White implicitly biased (for both White and Black jurors) compared to data from the general population (Greenwald & Krieger, 2006).

***Correlations***

Bivariate correlations were used to assess the overall relation among jurors' IAT scores, their causal attributions, culpability ratings, verdicts, and punishment preferences (see Table 3 for the overall correlations across conditions), broken down by participant race and defendant race.

Table 3. Correlations Between Key Variables Across Conditions

| Variable | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. IAT Score | -- | -.09* | -.07 | -.02 | -.12* |
| 2. Attributions | -- | -- | .55*** | .41*** | .51*** |
| 3. Culpability | -- | -- | -- | .65*** | .71*** |
| 4. Verdict | -- | -- | -- | -- | .56*** |
| 5. Punishment | -- | -- | -- | -- | -- |

Note. All correlations based on the full sample ($N$ – 418, $df$ – 416).
*$p < .05$. **$p < .01$. ***$p < .001$.

*Correlations between IAT Scores, Causal Attributions, and Culpability Ratings*

**White Jurors.** ***White Defendant.*** There was a small to moderate, negative correlation between White participants' IAT scores and their causal attributions of White defendants, whereby more pro-White implicit bias was associated with more situational (i.e., less dispositional) causal attributions of the White defendant (see Table 4 for all relevant correlations).

Table 4. Correlations Between Key Variables for White Jurors Judging White Defendants

| Variable | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. IAT Score | -- | -.24** | -.14 | -.12 | -.17* |
| 2. Attributions | -- | -- | .61*** | .54*** | .57*** |
| 3. Culpability | -- | -- | -- | .70*** | .72*** |
| 4. Verdict | -- | -- | -- | -- | .61*** |
| 5. Punishment | -- | -- | -- | -- | -- |

Note. All correlations based on a partial sample ($n$ – 144, $df$ – 142).
*$p$ < .05. **$p$ < .01. ***$p$ < .001.

Conversely, there was no significant relationship between participants' IAT scores and their culpability ratings or verdicts. White participants' IAT scores had a small to moderate, negative correlation with their punishment preferences, indicating that more pro-White implicit bias was associated with preferring less severe punishment for the White defendant. White participants causal attributions had a strong positive association with their culpability ratings, verdicts, and punishment preferences, such that more dispositional causal attributions were associated with higher culpability ratings, guilty verdicts, and preferring more severe punishment for the White defendant. Similarly, White participants' culpability ratings had a strong, positive correlation with their verdicts and punishment preferences, such that higher culpability ratings were associated with more guilty verdicts and preferring a more severe punishment for the White defendant. Finally, White participants' verdicts had a strong, positive correlation with their punishment preferences, such that guilty verdicts were associated with preferring a more severe punishment for the White defendant.

***Black Defendant****.* There were no significant correlations between White participants' IAT scores and their causal attributions, culpability ratings, verdicts, or punishment preferences of Black defendants (see Table 5 for all relevant correlations). Conversely, there were strong, positive correlations between White participants' causal attributions and their culpability ratings, verdicts, and punishment preferences, such that more dispositional causal attributions were associated with higher culpability ratings, guilty verdicts, and preferring more severe punishment for the Black defendant. Similarly, White participants' culpability ratings had a strong, positive correlation with their verdicts and punishment preferences, such that higher culpability ratings were associated with more guilty verdicts and preferring more severe punishment for the Black defendant. Finally, White

participants' verdicts had a strong, positive correlation with their punishment preferences, such that guilty verdicts were associated with preferring more severe punishment of the Black defendant.

Table 5. Correlations Between Key Variables for White Jurors Judging Black Defendants

| Variable | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. IAT Score | -- | -.12 | -.04 | -.06 | -.14 |
| 2. Attributions | -- | -- | .54*** | .40*** | .51*** |
| 3. Culpability | -- | -- | -- | .70*** | .76*** |
| 4. Verdict | -- | -- | -- | -- | .60*** |
| 5. Punishment | -- | -- | -- | -- | -- |

Note. All correlations based on a partial sample ($N$ – 133, $df$ – 131).
*$p$ < .05. **$p$ < .01. ***$p$ < .001.

*Correlations between IAT Scores, Causal Attributions, and Culpability Ratings*

**Black Jurors. *White Defendant.*** There were no significant correlations between Black participants' IAT scores and their causal attributions, culpability ratings, verdicts, or punishment preferences of White defendants (see Table 6 for all relevant correlations). Conversely, there were moderate to strong, positive correlations between Black participants' causal attributions and their culpability ratings, verdicts, and punishment preferences, such that more dispositional causal attributions were associated with higher culpability ratings, guilty verdicts, and preferring more severe punishment for the White defendant. Similarly, Black participants' culpability ratings had a strong, positive correlation with their verdicts and punishment preferences, such that higher culpability ratings were associated with more guilty verdicts and preferring more severe punishment for the White defendant. Finally, Black participants' verdicts had a moderate to strong, positive correlation with their punishment preferences, such that guilty verdicts were associated with preferring more severe punishment of the White defendant.

Table 6. Correlations Between Key Variables for Black Jurors Judging White Defendants

| Variable | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. IAT Score | -- | .09 | -.12 | -.02 | -.01 |
| 2. Attributions | -- | -- | .48*** | .28* | .38*** |
| 3. Culpability | -- | -- | -- | .56*** | .53*** |
| 4. Verdict | -- | -- | -- | -- | .41*** |
| 5. Punishment | -- | -- | -- | -- | -- |

Note. All correlations based on a partial sample ($n$ – 76, $df$ – 74).
*$p$ < .05. **$p$ < .01. ***$p$ < .001.

***Black Defendant***. There were no significant correlations between Black participants' IAT scores and their causal attributions, culpability ratings, verdicts, or punishment preferences of Black defendants (see Table 7 for all relevant correlations). Conversely, there were moderate to strong, positive correlations between Black participants' causal attributions and their culpability ratings, verdicts, and punishment preferences, such that more dispositional causal attributions were associated with higher culpability ratings, guilty verdicts, and preferring more severe punishment for the Black defendant. Similarly, Black participants' culpability ratings had a strong, positive correlation with their verdicts and punishment preferences, such that higher culpability ratings were associated with more guilty verdicts and preferring more severe punishment for the Black defendant. Finally, Black participants' verdicts had a moderate to strong, positive correlation with their punishment preferences, such that guilty verdicts were associated with preferring more severe punishment of the Black defendant.

Table 7. Correlations Between Key Variables for Black Jurors Judging Black Defendants

| Variable | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. IAT Score | -- | .02 | .10 | .21 | .01 |
| 2. Attributions | -- | -- | .49*** | .32** | .50*** |
| 3. Culpability | -- | -- | -- | .59*** | .75*** |
| 4. Verdict | -- | -- | -- | -- | .57*** |
| 5. Punishment | -- | -- | -- | -- | -- |

Note. All correlations based on a partial sample ($n$ – 65, $df$ – 63).
*$p$ < .05. **$p$ < .01. ***$p$ < .001.

### *Regression Analyses*

*Multiple Regression of Culpability Ratings and Punishment Preferences*

A multiple regression analysis was conducted to assess whether participants' IAT Score (as a continuous variable), Participant Race (0 = White, 1 = Black), Defendant Race (0 = White, 1 = Black), and Crime (0 = Rock Kick, 1 = Cycle Crash) predicted participants' culpability ratings or punishment preferences, as well as any interaction effects. In each multiple regression analysis, Step 1 included all main effects, and Step 2 included all main and interaction effects.

**Culpability Ratings.** A multiple regression assessed main (Step 1) and interaction effects (Step 2) of participants' IAT Score (as a continuous variable), Participant Race (0 = White, 1 = Black), Defendant Race (0 = White, 1 = Black), and Crime (0 = Rock Kick, 1 = Cycle Crash) predicting participants' culpability ratings. In Step 1, results showed that the overall model explained 5.2% of the variance, and was a significant predictor of culpability ratings, $F(4, 413) = 5.62$, $p < .001$. Only Crime was a significant predictor of culpability ratings, such that the Cycle Crash (vs. Rock Kick) crime was associated with a significant decrease in culpability ratings, $B = -0.62$, $t(416) = -4.25$, $p < .001$, 95% CI = [-0.90, -0.33].

Counter to Hypothesis 1a and 1b, Participant Race, Defendant Race, and IAT scores were not significant predictors of culpability ratings.

In Step 2, results showed that the overall model explained 9.1% of the variance, and was a significant predictor of culpability ratings, $F(15, 402) = 2.68$, $p = .001$. Crime was a significant predictor of culpability ratings, such that the Cycle Crash (vs. Rock Kick) crime was associated with a significant decrease in culpability ratings, $B = -0.98$, $t(416) = -3.03$, $p = .003$, 95% CI = [-1.62, -0.35], and Participant Race was a marginally significant predictor of culpability ratings, such that Black (vs. White) participants gave marginally significantly lower ratings of culpability, $B = -0.58$, $t(416) = -1.75$, $p = .081$, 95% CI = [-1.23, 0.07]. These effects, however, were qualified by a significant Crime x Participant Race interaction, $t(415) = 2.85$, $p = .005$, 95% CI = [0.43, 2.33], as well as a significant Crime x Participant Race x Defendant Race interaction, $t(414) = -1.99$, $p = .048$, 95% CI = [-0.02, -0.09].

To probe the three-way interaction, the data were split by Crime and Participant Race, and the multiple regression analysis was re-run with IAT Score and Defendant Race in the model. Counter to Hypothesis 1a and 1b, this analysis revealed no significant models for the White or Black participants in the Rock Kick crime condition, or the White participants in the Cycle Crash crime condition. Conversely, in support of Hypothesis 1b, the model for Black participants in the Cycle Crash condition predicted 8.4% of the variance in culpability ratings, and the overall model was marginally significant, $F(2, 66) = 3.00$, $p = .056$. Simple slope analysis indicated that for Black participants in the Cycle Crash crime condition, IAT Score was not a significant predictor contrary to Hypothesis 1a, but Defendant Race significantly predicted culpability ratings, such that the Black (vs. White) defendant predicted significantly lower culpability ratings, $B = -0.90$, $t(67) = -2.45$, $p = .017$, 95% CI = [-1.63, -0.17], in full support of Hypothesis 1b.

**Punishment Preferences.** A multiple regression assessed main (Step 1) and interaction effects (Step 2) of participants' IAT scores (as a continuous variable), Participant Race (0 = White, 1 = Black), Defendant Race (0 = White, 1 = Black), and Crime (0 = Rock Kick, 1 = Cycle Crash) predicting participants' punishment preferences. In Step 1, results showed that the overall model explained 9.5% of the variance, and was a significant predictor of punishment preferences, $F(4, 413) = 10.83$, $p < .001$. Crime was a significant predictor of punishment preferences, such that the Cycle Crash (vs. Rock Kick) crime predicted a significant decrease in punishment preferences, $B = -0.83$, $t(416) = -5.55$, $p < .001$, 95% CI = [-1.12, -0.53]. Defendant Race was also a significant predictor of punishment preferences, such that the Black (vs. White) defendant predicted significantly less severe punishment, $B = -0.31$, $t(416) = -2.07$, $p = .039$, 95% CI = [-0.60, -0.02]. Similarly, participants' IAT scores were a significant predictor of punishment preferences, such that higher (i.e., pro-White) implicit bias was associated with significantly lower punishment preferences, $B = -0.41$, $t(416) = -2.14$, $p = .033$, 95% CI = [-0.78, -0.53]. Participant Race was not significant predictor of punishment preference.

In Step 2, results showed that the overall model explained 12.6% of the variance, and predicted punishment preferences, $F(15, 402) = 3.85$, $p < .001$. Only Crime remained a

significant predictor of punishment preferences, such that the Cycle Crash (vs. Rock Kick) crime predicted a significant decrease in punishment preferences, $B$ = -0.98, $t(416)$ = -3.11, $p$ = .002, 95% CI = [-1.69, -0.38]. Contrary to Hypothesis 1a and 1b, IAT Scores, Defendant Race, and Participant Race were not significant predictors of punishment preferences. The effect of Crime, however, was qualified by a significant Crime x Participant Race interaction, $t(415)$ = 2.43, $p$ = .015, 95% CI = [0.23, 2.18].

To probe the two-way interaction, the data were split by Crime type and the multiple regression analysis was re-run with IAT Score, Defendant Race, and Participant Race in the model. This analysis revealed the model was not significant for the Rock Kick crime condition. Conversely, the model for the Cycle Crash condition predicted 7.2% of the variance in punishment preferences, and the overall model was significant, $F(3, 203)$ = 5.26, $p$ = .002. Simple slope analysis indicated that in the Cycle Crash crime condition, Defendant Race significantly predicted punishment preferences, such that the Black (vs. White) defendant predicted significantly lower punishment preferences, $B$ = -0.13, $t(205)$ = -2.07, $p$ = .040, 95% CI = [-0.88, -0.02]. Similarly, Participant Race significantly predicted punishment preferences, such that Black (vs. White) Participants predicted significantly higher punishment preferences, $B$ = 0.53, $t(205)$ = 2.21, $p$ = .028, 95% CI = [0.56, 1.00]. Conversely, IAT Score was not a significant predictor of punishment preferences in the Cycle Crash crime condition. Because there were no significant interaction effects between IAT Scores or Defendant Race with Participant Race, this analysis revealed no support for Hypothesis 1a or 1b.

*Logistic Regression of Verdict Preferences*

As with the multiple regression analyses, the logistic regression analysis of Verdicts (0 = Not Guilty, 1 = Guilty) assessed the main effect of the individual predictors (i.e., IAT Score, Participant Race, Defendant Race, and Crime) with all main effects in Step 1 and main and interaction effects in Step 2. This analysis revealed that in Step 1, the logistic regression model was significant, $\chi^2(4, N = 418) = 41.55$, $p < .001$, and predicted a significant proportion of verdicts (Nagelkerke $R^2$ = 0.15). Only Crime was a significant predictor of verdicts, such that compared to participants in the Rock Kick condition, participants in the Cycle Crash condition were 81% less likely to return a Guilty Verdict, $B$ = -1.68, *S.E.* = 0.287, *Wald* = 34.26, $p < .001$, *OR* = 0.19, 95% CI = [0.11, 0.33]. Conversely, contrary to Hypothesis 1a and 1b, IAT Score, Participant Race, and Defendant Race were not significant predictors of verdicts.

In Step 2, the addition of the interaction terms did not significantly improve the model, $\chi^2(11, N = 418) = 16.00$, $p$ = .141, although the overall model was significant, $\chi^2(15, N = 418) = 57.56$, $p < .001$, with acceptable fit, and predicted a significant proportion of the verdicts (Nagelkerke $R^2$ = 0.20). In this model, Crime was the only significant predictor of verdicts, such that compared to participants in the Rock Kick condition, participants in the Cycle Crash condition were 92% less likely to return a Guilty Verdict, $B$ = -2.57, *S.E.* = 0.90, *Wald* = 8.21, $p$ = .004, *OR* = 0.08, 95% CI = [0.01, 0.44]. Conversely, contrary to

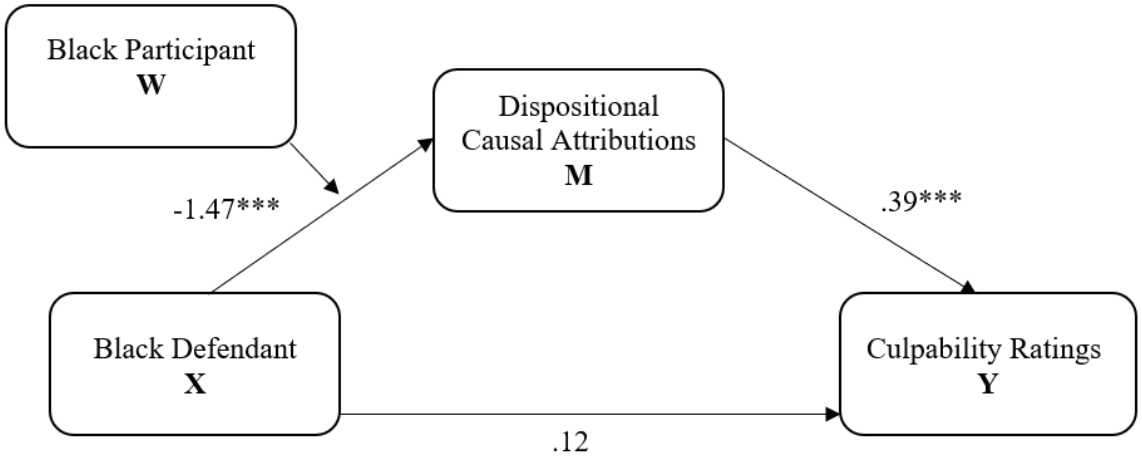
Hypothesis 1a and 1b, IAT Score, Participant Race, and Defendant Race, and the interaction effect, were not significant predictors of verdicts.

***Moderated Mediation***

We conducted a moderated mediation analysis using model 7 of Haye's (2013; 2017) PROCESS macro in which we assessed the conditional indirect effect of defendant race on participants' culpability judgments through the causal attributions they made about the criminal event as a function of the participants' race. In other words, to test Hypothesis 2 and 3, the proposed moderated mediation model examined whether White and Black participants made disparate culpability judgments of White and Black Defendants as a result of the causal attributions they made for the defendants' crime. This analysis thus assessed moderated mediation of similarity leniency, whether White jurors would be more lenient on White defendants (and harsher on Black defendants), and Black jurors would be more lenient on Black defendants (and harsher on White defendants), as a function of making more situational (vs. dispositional) causal attributions.

This analysis revealed significant conditional indirect effects of Defendant Race on participants' culpability judgments through causal attributions as a function of Participant Race, and White and Black participants exhibited different conditional indirect effect patterns depending on the race of the defendant, index = .46, *SE* = .19, 95% CI [.08, .84]. More specifically, there was a significant interaction between Defendant Race and Participant Race in predicting causal attributions, $B = -1.14$, $SE = .37$, $t(417) = -3.08$, $p = .002$, 95% CI = [-1.87, -.41]. Black (vs. White) participants made significantly more situational causal attributions when the defendant was Black (vs. White), $B = -1.47$, $SE = .30$, $t(417) = -4.88$, $p < .001$, 95% CI = [-2.07, -.88]. Furthermore, dispositional causal attributions predicted significantly higher culpability ratings, $B = 0.39$, $SE = .05$, $t(210) = 8.64$, $p < .001$, 95% CI = [.30, .48], and the indirect effect was significant, indirect effect = -.54, *SE* = .16, 95% CI = [-.88, -.22] (see Figure 1).

Figure 1. Moderated Mediation Model for Black Participants and Black Defendants




Note. The conditional indirect effect of the defendant's race (X) on Black jurors' culpability ratings (Y), mediated by their casual attributions (M).
***$p < .001$

Conversely, there was no significant indirect effect for White (vs. Black) participants, indirect effect = -.33, *SE* = .22, $p = .124$, 95% CI = [-.75, .09]. Thus, in support of Hypothesis 2 and 3, the moderated mediation analysis showed that Black participants were more lenient in their culpability judgments of Black (vs. White) defendants because they made more situational causal attributions for their crime.

## DISCUSSION

The present study examined whether a) Black and White jurors' culpability judgments (as a proxy for mens rea), verdict judgments, and/or punishment preferences would reflect similarity leniency effects for Black and White defendants, respectively, b) this effect would be related to their level of implicit racial bias, and c) causal attributions for the crime would mediate the effect of defendant race on jurors' judgments of defendant explained any differences in culpability. Results indicate partial support for our hypotheses, and each will be discussed in turn, followed by a consideration of broader issues addressed or highlighted by of our findings and future directions.

### *Hypothesis 1a: Similarity Leniency*

Some evidence of a similarity leniency effect emerged from the regressions with the significant interaction between crime, participant race, and defendant race, on mock jurors' perceptions of culpability, but not their verdicts or punishment preferences. Additionally, we found further support for the similarity leniency effect in the moderated mediation model of Black jurors' culpability judgments. More specifically, Black jurors were more lenient on Black (vs. White) defendants as a function of making more situational (vs. dispositional)

causal attributions for the Black (vs. White) defendant's behavior. Thus, our moderated mediation results contribute to the refinement of prior research by demonstrating Black mock jurors tend to be more lenient toward ingroup members, and less lenient toward outgroup members, in this case, White defendants (i.e., similarity leniency; Cohn et al., 2009; Devine & Caughlin, 2014; Mitchell et al., 2005, Sommers & Ellsworth, 2000).

***Hypothesis 1b.: Implicit Racial Bias***

Although suggested as one possible factor to underly the similarity leniency effect (e.g., Hunt, 2015), we found no evidence that jurors' implicit racial bias, as operationalized here and measured by the Race IAT, was related to whether participants made disparate culpability judgments, verdicts, or punishment recommendations for Black and White defendants. For White jurors, as will be discussed further below, the results suggest race was a salient factor (e.g., Sommers & Ellsworth, 2000; 2001) in the experimental juror decision-making context, suppressing the possibility of racial bias from White participants and an true assessment of whether similarity-leniency varies with their level of implicit racial bias. For Black jurors, however, the results of the regressions suggest that their level of implicit racial bias operates independently of biased culpability judgments, as there was no interaction of IAT scores with either participant race or defendant race in any of the crime scenarios.

***Hypothesis 2: Causal Attributions***

As predicted and demonstrated in the correlations and moderated mediation analysis, both White and Black jurors' causal attributions were directly related to their perceptions of culpability for both White and Black defendants, such that more dispositional (vs. situational) causal attributions predicted more perceived culpability. As such, the present results are consistent with the literature demonstrating that dispositional causal attributions are related to more negative outcomes for defendants than situational causal attributions for both White and Black jurors and White and Black defendants (e.g., Sommers & Ellsworth, 2000). Moreover, these results support literature suggesting people make causal attributions that are consistent with the Ultimate Attribution Error (i.e., Pettigrew, 1979), as Black jurors made dispositional attributions for out-group members' (i.e., the White defendant's) transgressions but situational attributions for in-group members' (i.e., the Black defendant's) transgressions.

***Hypothesis 3: Similarity-Leniency Moderated Mediation***

The moderated mediation hypothesis received partial support as causal attributions mediated the effect of defendant race on Black (but not White) jurors' culpability determinations, such that when the defendant was Black (vs. White), they made more situational causal attributions, which led to less perceived culpability. Conversely, when faced with a White defendant, Black jurors made more dispositional causal attributions for the crime, which in turn led them to perceive the defendant as more culpable. To the best of our knowledge, this is the first demonstration of causal attributions mediating the effect of defendant race on jurors' cross-raced judgments of culpability.

***Broader Issues and Future Directions***

By examining both Black and White participants, measuring their implicit bias, and examining how these variables relate to similarity leniency effects in downstream judgments, the present research speaks to outstanding issues in the literature. For example, Sommers and Ellsworth (2000) found that participant race was a relevant factor in how legal decision-makers rendered judgments, but Graham and Lowery (2004) did not. The present study thus supports the literature suggesting that the decision-makers' (in this instance the mock jurors') race is an important variable in considering how the racial characteristics of a case affect case outcomes (e.g., Devine & Caughlin, 2014). Similarly, the present research and results address and highlight several other theoretical (and practical) issues in the relevant bodies of literature, such as: a) The suppressive role of race salience in racial bias and decision-making research in the context of a racially charged society; b) issues with the IAT, including specific practical administration considerations and general questions of the IAT's validity; c) the conflict between similarity leniency effects and what we presume about implicit racial bias; and d) the generalizability of the results to other categories of mens rea. Each will be discussed in turn, including recommendations for future directions.

*White Jurors' Null Effects and Race Salience in a Racially Charged Society*

No effects emerged for White jurors, who rendered equivalent culpability ratings, verdicts, and punishment recommendations for White and Black defendants, regardless of their implicit racial bias. The present research thus appears inconsistent with the research demonstrating similarity-leniency in White jurors' crossed-race legal judgments (e.g., Mitchell et al., 2005) as well as the literature demonstrating discrimination against minority defendants (e.g., Levinson & Young, 2010). As mentioned above, however, these results *are* consistent with the literature suggesting that when racial issues are salient, White mock jurors will take strides to correct for their biases, and the discriminatory effects will disappear (e.g., Cohn et al., 2009; Maeder & Ewanation, 2018; Peter-Hagene, 2019; Sommers & Ellsworth, 2000, 2001).

Assessing participants' responses to our manipulation check question supports the possibility of race being salient and White jurors being on guard for bias, as almost half of our White (45%; $n = 125$) participants mentioned something about race. Although race salience effects are not expected by the mere presence of a Black person (i.e., race salience effects are triggered by the salience of the potential for appearing prejudiced or discriminatory in behavior toward a Black person; Sommers & Ellsworth, 2000), racial tensions in the U.S. are increasingly high. Most of the general public (i.e., 65%) believes it is now more commonplace to behave in a racist manner or express racist views, 45% think it has become more socially acceptable (e.g., Horowitz et al., 2019), and perceptions of race relations are on a downward trajectory (Gallup, 2021).

Future research should thus assess whether the increased perceptions of racist behaviors and the acceptability of those behaviors in the country affect peoples' proclivity toward suppressing bias when race is a possible factor in the decision-making context. Recent changes in the U.S. social and political climate (i.e., Horowitz et al., 2019) require an updat-

ing of what is known about how race salience affects jurors' case judgments with minority defendants. Future research also should consider experimentally controlling for racial salience (e.g., stereotypical names vs. pictures) to assess whether biased causal attributions mediate the effect of White jurors' racial bias on their mens rea judgments, and whether that racial bias differs by their level of implicit racial bias. Future research should also carefully consider what might cause race salience effects within the study's methodology.

*IAT Practical and Theoretical Issues and Considerations*

It is debatable whether one should administer the IAT at the beginning of the study or at the end. But the positioning of the IAT in the study methodology is one plausible cause of the race salience effects found in the present research. On the one hand, if taken at the beginning of the study, one risks tipping off the purpose of the study based on the content of the IAT (as we might have seen in the present results). On the other hand, IAT results are known to vary given the context in which the test is taken (e.g., see Azar, 2008 for a discussion on IAT critiques), suggesting that administering the IAT at the end of the study might produce different results than administering it at the beginning. Ideally, the IAT would be administered in an ostensibly unrelated fashion, but given resource and methodological constraints, this might not always be possible and future researchers should carefully consider whether administering the IAT at the beginning or end of the study is best. The possibility of the IAT revealing the study's purpose to participants also raises an interesting question about the use of the IAT on platforms such as M-Turk. Many participants on M-Turk appear to have considerable experience completing IATs which may tip them off to the purpose of the studies that follow.

There is also a noteworthy tension in the literature over the validity of the IAT. For instance, it has been suggested that millisecond reaction times to stimulus pairings might not actually reflect real biases, but rather knowledge of common stereotypes and incidental associations of varying strengths (e.g., Fazio & Olsen, 2003). Some scholars have also criticized the IAT because IAT scores are hard to interpret and based on arbitrary zero points (e.g., Blanton & Jaccard, 2006; Fiedler et al., 2006). Clearly, more research is needed measuring people's implicit biases and assessing their impact on actual behaviors and judgments. Moreover, progress in this line of research is likely to be advanced by increasing refinements in the conceptualization and measurement of implicit bias.

Although conjecture, one alternative explanation for the null effects surrounding jurors' implicit racial biases is that the IAT is not accessing the right implicit associations. There remains a debate in the field as to what the IAT (and other implicit bias measures) actually measure. That is to say, it might be that implicit racial bias *is* what drive's racial bias and similarity leniency effects, but that we presently lack the tools to measure it. In any case, the present results suggest that, at least for Black jurors, implicit racial bias as measured here (i.e., Black/White, good/bad) are less predictive of their judgments than whether they are the same or different race of the defendant, and thus more or less likely to treat the defendant more leniently, respectively. Future research should examine why Black participants' IAT scores fail to predict racially biased behavior.

*Similarity Leniency Versus Implicit Racial Bias*

The present study also brings attention to a conflict in the jury decision-making literature on the effects of a case's racial characteristics that does not receive much consideration or debate. That is, in the context of the robust literature demonstrating the similarity-leniency effect when jurors make cross-raced judgments (e.g., Devine & Caughlin, 2014; Mitchell et al., 2005), it is unclear what happens when a decision-maker such as a juror is implicitly biased against their own in-group, and instead (implicitly) prefers the out-group. The present results suggest that, at least for Black jurors, similarity-leniency is more likely than any effect based on someone's level of implicit racial bias (at least as measured by the Race IAT employed here). This, however, is a tentative conclusion pending further assessment of White jurors when racial salience is controlled experimentally, and more research comparing implicit racial bias and similarity leniency effects.

*Mens Rea Category Generalizability*

The study also relied on crimes that were exclusively from the reckless category of mens rea, and it is presently unclear if the results will generalize to higher (or lower) levels of mens rea. Given that research is burgeoning on direct assessments of mens rea judgments, consideration of how juror and defendant race interact to produce discriminatory mens rea judgments (through their causal attributions) at different levels of mens rea is a promising future avenue for research. It is also worth noting that the two crimes, previously determined by legal experts to evince recklessness mens rea (Shen et al., 2011), were perceived significantly different in terms of the offender's culpability, and participants' verdicts and punishment preferences. More specifically, the Rock Kick crime offenders were seen as more culpable, deserving of more severe punishment, and more likely to receive a guilty verdict, compared to Cycle Crash crime offenders. This disparate treatment of offenders of different crimes that are presumed to reflect the same seriousness (Shen et al., 2011) suggests that legal experts and lay persons might diverge in their perspectives on what crimes and conduct represent a certain level of a guilty mind. Future research should consider what might produce these differences in perspective on crimes that are presumably equivalent in their level of mens rea.

## CONCLUSION

Research shows that people often err in a fundamental task asked of all criminal juries, making mens rea determinations (Beattey & Fondacaro, 2018). The present research demonstrates that extralegal variables, such as the race of the juror and the defendant might play an important role in influencing mens rea judgments through the attributions they make (at least in the case of Black jurors). The present research also addresses a noted gap in the limited implicit racial bias and jury decision-making literature (Hunt, 2017): the empirical assessment of the mechanisms through which racial biases influence jurors' decision-making (Kang et al., 2012; Levinson & Young, 2010; Sommers, 2007). The results show that, at least for Black jurors, causal attributions help explain how the racial dynamics of a case might exert their influence at trial and affect case judgments. And last, the present research adds to the limited literature examining Black jurors' decision-making and rep-

resents one of the first attempts to compare the effects of Black and White jurors' implicit racial biases on their case judgments with actual measurements of their implicit racial bias.

## REFERENCES


Ashburn-Nardo, L., Knowles, M.L., & Monteith, M.J. (2003). Black Americans' implicit racial associations and their Implications for intergroup judgment. *Social Cognition*, *21*(1), 61-87. doi: 10.1521/soco.21.1.61.21192

Azar, B. (2008). IAT: Fad or fabulous? *Monitor on Psychology*. Retrieved from: https://www.apa.org/monitor/2008/07-08/psychometric

Beattey, R.A., & Fondacaro, M.R. (2018). The misjudgment of criminal responsibility. *Behavioral Sciences & the Law*, *36*(4), 457-469. doi: 10.1002/bsl.2354

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*(1), 27-41. doi: 10.1037/0003-066X.61.1.27

Buhrmester, M., Kwang, T., & Gosling, S.D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science, 6*(1), 3-5. doi: 10.1177/1745691610393980

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cohn, E.S., Bucolo, D., Pride, M. & Sommers, S.R. (2009). Reducing white juror bias: The role of race salience and racial attitudes. *Journal of Applied Social Psychology*, *39*(8), 1953-1973. doi: 10.1111/j.1559-1816.2009.00511.x

Devine, D.J., & Caughlin, D.E. (2014). Do they matter? A meta-analytic investigation of individual characteristics and guilt judgments. *Psychology, Public Policy, and Law*, *20*(2), 109-134. doi: 10.1037/law0000006

Dovidio, J.F., & Gaertner, S.L. (2004). Aversive racism. In M. P. Zanna (Ed.) *Advances in experimental social psychology*, *36*, 1-52. San Diego, CA: Elsevier Academic Press.

Dovidio, J.F., Kawakami, K., & Gaertner, S.L. (2000). *Reducing contemporary prejudice: Combating explicit and implicit bias at the individual and intergroup level*. In S. Oskamp (Ed.), Reducing prejudice and discrimination: The Claremont Symposium on applied social psychology (pp. 137-163). Mahwah, NJ: Lawrence Erlbaum Associates.

Dovidio, J.F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*, 510-540. doi: 10.1006/jesp.1997.1331

Duncan, S.L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology, 34*(4), 590–598. doi: 10.1037/0022-3514.34.4.590

Fazio, R.H., & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, *54*, 297-327. doi: 10.1146/annurev.psych.54.101601.145225

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I', the "A", and the "T": A logical and psychometric critique of the implicit association test (IAT). *European Review of Social Psychology*, 17, 74-147. doi: 10.1080/10463280600681248

Fondacaro, M.R., & O'Toole, M.J. (2015). American punitiveness and mass incarceration. *New Criminal Law Review, 18* (4), 477-509.

Gaertner, S.L., & Dovidio, J.F. (1986). *The aversive form of racism*. (pp. 61-89). Academic Press, San Diego, CA.

Gallup. (2021). *Race relations*. Retrieved May 08, 2021, from: https://news.gallup.com/poll/1687/race-relations.aspx

Gordon, N., & Fondacaro, M.R. (2018). Rethinking the voluntary act requirement: Implications from neuroscience and behavioral science research. Behavioral Sciences & the Law, 36, 426-436. doi: 10.1002/bsl.2352

Graham, S., & Lowery, B.S. (2004). Priming unconscious racial stereotypes about adolescent offenders. *Law and Human Behavior*, *28*(5), 483-504. doi: 10.1023/B:LAHU.0000046430.65485.1f

Greenwald, A.G., & Krieger, L.H. (2006). Implicit bias: Scientific foundations. California Law Review, 94(4), 945-967.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464-1480.

Greenwald, A.G., Poehlman, T.A., 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*(1), 17-41. doi: 10.1037/a0015575.supp

Hawkins, D.F. (1981) Causal attribution and punishment for crime. *Deviant Behavior: An Interdisciplinary Journal, 2*, 207-230. doi: 10.1080/01639625.1981.9967554

Hayes, A.F. (2013). *Introduction to mediation, moderation and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Hayes, A.F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Hodson, G., Hooper, H., Dovidio, J.F., & Gaertner, S.L. (2005). Aversive racism in Britain: The use of inadmissible evidence in legal decisions. *European Journal of Social Psychology, 35*(4)*,* 437–448. doi:10.1002/ejsp.261

Horowitz, J.M., Brown, A., & Cox, K. (2019). Race in America 2019. *PEW Social Trends*. Retrieved from https://www.pewsocialtrends.org/2019/04/09/race-in-america-2019/

Hunt, J.S. (2015). Race, ethnicity, and culture in jury decision making. *Annual Review of Law and Social Science*, *11*, 269-288. doi: 10.1146/annurev-lawsocsci-120814-121723

Hunt, J.S. (2017). Studying the effects of race, ethnicity, and culture on jury behavior In M. Bull Kovera (Ed.), *Psychology of Juries* (pp. 118-153). NY: American Psychological Association.

Johnson, J.D., Whitestone, E., Jackson, L.A., & Gatto, L. (1995). Justice is not colorblind: Differential racial effects of exposure to inadmissible evidence. *Personality and Social Psychology Bulletin*, *21*(9), 893-898. doi: 10.1177/0146167295219003

Kang, J., Bennett, M., Carbado, D., Casey, P., Dasgupta, N., et al. (2012). Implicit bias in the courtroom. *UCLA Law Review*, *59*, 1124-1186.

Kerr, N.L., MacCoun, R.J., & Kramer, G.P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review*, *103*(4), 687-719. doi: 10.1037/0033-295X.103.4.687

Krueger, J. (1996). Personal beliefs and cultural stereotypes about racial characteristics. *Journal of Personality and Social Psychology*, *71*(3), 536-548. doi: 10.1037/0022-3514.71.3.536

Kubota, J.T., Mojdehbakhsh, R., Raio, C., Brosch, T., Uleman, J.S., & Phelps, E.A. (2014). Stressing the person: Legal and everyday person attributions under stress. *Biological Psychology, 103*, 117-124. doi: 10.1016/j.biopsycho.2014.07.020

Levinson, J.D. (2007). Racial equality: Implicit bias, decisionmaking, and misremembering. *Duke Law Journal*, *57*(2), 345-424.

Levinson, J.D., Cai, H., & Young, D. (2010). Guilty by implicit racial bias: The guilty/ not guilty implicit association test. *Ohio State Journal of Criminal Law*, *8*(187), 187–208.

Levinson, J.D., & Young, D. (2010). Different shades of bias: Skin tone, implicit racial bias, and judgments of ambiguous evidence. *West Virginia Law Review*, 112, 307–350.

Lynch, M., & Haney, C. (2000). Discrimination and instructional comprehension: Guided discretion, racial bias, and the death penalty. *Law and Human Behavior*, *24*(3), 337-358. doi:10.1023/A:1005588221761

Lynch, M., & Haney, C. (2009). Capital jury deliberation: Effects on death sentencing, comprehension, and discrimination. *Law and Human Behavior*, *33*(6), 481-496. doi: 10.1007/s10979-008-9168-2

Maeder, E.M. & Ewanation, L. (2018). What makes race salient? Juror decision-making in same-race versus cross-race identification scenarios and the influence of expert testimony. *Criminal Justice and Behavior, 45(8)*, 1234-1251. doi: 10.1177/0093854818776998

Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendant and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology, 24*(15), 1315–1344. doi: 10.1111/j.1559-1816.1994.tb01552.x

Mitchell, T.L., Haw, R.M., Pfeifer, J.E., & Meissner, C.A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior*, *29*(6), 621–637. doi:10.1007/s10979-005-8122-9

Oliver, M.B. (2003). African American men as "criminal and dangerous": Implications of media portrayals of crime on the "criminalization" of African American men. *Journal of African American Studies, 7*(2), 3–18. doi: 10.1007/s12111-003-1006-5

Paolacci, G., Chandler, J., & Ipeirotis, P.G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411-419.

Peter-Hagene, L. (2019). Jurors' cognitive depletion and performance during jury deliberation as a function of jury diversity and defendant race. *Law and Human Behavior, 43*(3), 232-249. doi: 10.1037/lhb0000332

Pettigrew, T.F. (1979). The Ultimate Attribution Error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin, 5*(4), 461-476. doi: 10.1177/014616727900500407

Pope, J., & Meyer, R. (1999). An attributional analysis of jurors' judgments in a criminal case: A preliminary investigation. *Social Behavior and Personality, 27*(6), 563-574. doi: 10.2224/sbp.1999.27.6.563

Severance, L.J., & Loftus, E.F. (1982). Improving the ability of jurors to comprehend and apply criminal jury instructions. *Law & Society Review, 17*(1), 153–197. doi:10.2307/3053535

Severance, L.J., Goodman, J., & Loftus, E.F. (1992). Inferring the criminal mind: toward a bridge between legal doctrine and psychological understanding. *Journal of Criminal Justice*, *20*(2), 107-120.

Shen, F.X., Hoffman, M.B., Jones, O.D., Greene, J.D., & Marois, R. (2011). Sorting guilty minds. *New York University Law Review*, *86*, 1306-1360.

Sommers, S.R. (2006). On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Journal of personality and Social Psychology*, *90*(4), 597-612. doi: 10.1037/0022-3514.90.4.597

Sommers, S.R. (2007). Race and the decision making of juries. *Legal and Criminological Psychology, 12,* 171–187. doi:10.1348/135532507X189687

Sommers, S.R., & Ellsworth, P.C. (2000). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin*, *26*(11), 1367–1379. doi:10.1177/046167200263005

Sommers, S.R., & Ellsworth, P.C. (2001). White juror bias: An investigation of prejudice against black defendants in the American courtroom. *Psychology, Public Policy, and Law*, *7*(1), 201–229. doi:10.1037/1076-8971.7.1.201

Stevenson, M.C., Bottoms, B.L., & Diamond, S.S.S. (2010). Jurors' discussions of defendants history of child abuse and alcohol abuse in capital sentencing deliberations. *Psychology, Public Policy, and Law*, *16*(1), 1–38. doi:10.1037/a0018404

Sweeney, L.T., & Haney, C. (1992). The influence of race on sentencing: A meta-analytic review of experimental studies. *Behavioral Sciences & the Law*, *10*(2), 179–195. doi:10.1002/bsl.2370100204

The American Law Institute. (1962). *The Model Penal Code.* Philadelphia, PA: The American Law Institute.

U.S. Const. amend. VI

Date Received: 10/2021
Date Accepted: 10/2021

Suggested citation: Jay, A.C.V., Stone, C.B., Fondacaro, M.R., Yoon, J., & Zuraw, K. (2021). Similarity leniency in mens rea determinations and the mediating role of casual attributions. [Electronic Version]. *Applied Psychology in Criminal Justice, 16*(2), 129-155.