

Jurors' Cognitive Depletion and Performance During Jury Deliberation as a Function of Jury Diversity and Defendant Race

Liana Peter-Hagene
Southern Illinois University Carbondale

Racial diversity in juries, meant to ensure representation of minority voices, can also reduce racial bias in verdicts and improve group performance during deliberation (Sommers, 2006). Although jury diversity might increase cognitive depletion because it involves interracial interactions (Richeson & Shelton, 2003), it might also reduce racial disparity in verdicts and deliberation quality by improving the quality of deliberation for Black defendants. In 6-person juries that included 2 White or 2 Black confederates, White jurors viewed trial evidence including a defendant race manipulation (White, Black) and completed measures of cognitive depletion and case facts recall. Deliberation transcripts were coded for the number of total, correct, and new case facts mentioned by jurors. As predicted, jurors in diverse versus all-White juries were more depleted after deliberation, but depletion was not related to deliberation performance. For the Black defendant, jurors on diverse (vs. homogeneous) juries discussed more case facts; for the White defendant, the effect was not significant. Jurors on all-White juries discussed more case facts when they judged a White (vs. Black) defendant, but this difference was not significant for jurors on diverse juries. Thus, jury diversity reduced racial disparity in the quality of deliberation. Before and after deliberation, jurors were less likely to convict the Black (vs. White) defendant regardless of jury composition. These findings reinforce the importance of constructing racially diverse juries, given that they may be better equipped to evaluate trial evidence for both Black and White defendants despite the cognitive demand of deliberating within a diverse group.

Public Significance Statement

Jury racial diversity can help reduce racial disparities in trial outcomes such as verdicts and sentencing. The present study further demonstrates that the mere presence of Black jurors can motivate White jurors to contribute more factual comments during jury deliberation: Although all-White juries performed better when they judged a White (vs. Black) defendant, diverse juries performed equally well regardless of defendant race. This happened despite the fact that interracial interactions tend to require more cognitive and self-control resources than same-race interactions, presumably because people are concerned with either appearing racially biased or with being the target of bias.

Keywords: interracial interactions, jury deliberation, diversity, cognitive depletion

Over the past 40 years, legislative reforms to jury selection have helped increase representation of racial and ethnic minorities on American juries. Despite these efforts, White jurors still constitute a disproportional majority on most U.S. juries, because jury selection relies primarily on voter registration lists where minorities are underrepresented (Sweeney & Dizikes, 2013). In addition, ethnic minorities are often eliminated through peremptory chal-

lenges due to judges' tendency to accept attorneys' ostensibly race-neutral justifications for exclusion (*Foster v. Chatman*, 2016; Sommers & Norton, 2007). Increased diversity in juries ensures broader representation of minority voices, it can reduce racial bias in decisions, and it can improve decision quality (Bowers, Steiner, & Sandys, 2001; Sommers, 2006). A recent archival analysis of real juries demonstrated that the more race-gender subgroups were

The author thanks Bette Bottoms, Courtney Bonam, Dan Cervone, Tomas Stahl, and Samuel Sommers for their feedback as committee members on the author's dissertation. The author also thanks Jessica Salerno, Douglas Knutson, and Chasity Ratliff for helpful feedback on the manuscript and Kelly Burke and Tania Alvarado for coding transcript data. The author also thanks the research assistants who played the roles of confederate jurors: Siobhan Midgley, Sabrina Bellamy, Winifred Obonor, Jasmina Ejupovic, Samantha Dabah, and Kristina Todorovic. This research was supported by an NSF Doctoral Dissertation Improvement Grant, a Psi Chi

Graduate Research Grant, a Society for the Psychological Study of Social Issues Grant-in-Aid, an American Psychology-Law Society Grant-in-Aid, and a UIC Provost Award for Graduate Research. Supplemental materials (data, codebook, confederate scripts) are available through the Open Science Framework, <https://osf.io/5jfe6/>.

Correspondence concerning this article should be addressed to Liana Peter-Hagene, Department of Psychology, Southern Illinois University Carbondale, 1125 Lincoln Drive, Behavioral Sciences Building, Carbondale, IL 62901. E-mail: claudia.peter-hagene@siu.edu

present on a jury, the less likely the jury was to convict (Devine, Krouse, Cavanaugh, & Basora, 2016).

Yet diversity can also place a strain on jurors' cognitive and emotional resources, especially when juries are still predominantly White with small numbers of minority jurors. For example, in an actual corruption trial the only Black juror slammed his hand on the table and accused the others of "wanting to hang the Black man." In another trial, a White juror accused a Black juror of flashing gang signs at her during an argument (Renaud, 2010). Thus, diverse juries might foster higher levels of interpersonal and emotional conflict, as research indicates is indeed the case in other group settings (e.g., Pelled, Eisenhardt, & Xin, 1999; Timmerman, 2000). In turn, interpersonal conflict impairs group performance, arguably because it interferes with information processing capacity (e.g., De Dreu & Weingart, 2003).

Even in the absence of open interpersonal conflict, race dynamics shape jury deliberations. Jurors of color talk less than White jurors (Rose & Diamond, 2008) and seem to exert less social influence during deliberation: Black jurors' dissenting opinions are discounted, for example, especially when expressed with anger (Salerno, Peter-Hagene, & Jay, 2017). All this evidence suggests that diverse juries can face interpersonal challenges unique to interracial interactions, with potential detrimental implications for jurors' cognitive performance.

These challenges are not unique to juries. In organizational research, for example, some found beneficial effects of ethnic (and other types of) diversity (DeGrassi, Morgan, Walker, Wang, & Sabat, 2012; McLeod, Lobel, & Cox, 1996), but others found diversity was detrimental to group performance and task enjoyment (Harvey, 2013; Jehn & Bezrukova, 2004). On the one hand, diversity can provide a valuable range of knowledge and perspectives, which is most helpful in idea-generating tasks (Cox & Blake, 1991; McLeod et al., 1996). For example, the presence of minority (i.e., Black) group members resulted in higher integrative complexity for White college students, because it led them to consider multiple perspectives (Antonio, Chang, Hakuta, Kenny, Levin, & Milem, 2004).

On the other hand, however, in decision-making tasks, where team members must provide one single solution, diversity seems to decrease performance (Harvey, 2013; Triana, Porter, DeGrassi, & Bergman, 2013). A meta-analysis of diversity in work groups from organizational settings revealed that demographic diversity (race/ethnicity, gender, age) had a small but consistently negative effect on performance (Joshi & Roh, 2009). In a longitudinal study of group decision making, racially diverse groups of students performed worse than homogeneous groups on tasks that required providing a single correct solution to a real-life management problem after a 1-hr deliberation (a context not unlike jury deliberation), but the differences diminished and even disappeared in groups that worked together multiple times (Watson, Kumar, & Michaelsen, 1993). It appears that increased personal familiarity reduced the detrimental effect of interacting with dissimilar others on task performance. But why might the difficulty arise in the first place?

Interracial interactions, even when benign (i.e., casual conversations that are not explicitly about race), can result in increased tension, as Whites make efforts to control their behavior to avoid appearing racist, and Blacks become vigilant against prejudiced reactions from others (e.g., Richeson & Trawalter, 2005). There-

fore, such interactions can place a strain on people's cognitive resources, reducing cognitive performance, presumably because people are concerned with either appearing racially biased or with being the target of bias (Apfelbaum, Sommers, & Norton, 2008; Murphy, Richeson, Shelton, Rheinschmidt, & Bergsieker, 2013; Zabel, Olson, Johnson, & Phillips, 2015). In other words, interracial interactions are cognitively depleting.

Yet psychology-and-law research shows that racially diverse juries are not only less racially biased, but also more thoughtful than all-White juries, at least when they judge cases involving a Black defendant (Bowers et al., 2001; Sommers, 2006). There is a crucial difference between the two areas of research: Unlike organizational and social psychology studies, jury research is focused on jurors' judgments about other people (i.e., defendants). In jury trials, jurors might be particularly motivated to reach verdicts that are fair to minority defendants and/or victims, which might overcome cognitive depletion effects. Sommers (2006), for example, argues that White jurors' performance was motivated by their desire to appear unbiased against a Black defendant when they deliberated with Black jurors. To test this explanation, the present study includes experimental conditions in which the defendant is White—that is, conditions in which there is no particular concern for White jurors to appear racially biased against the defendant.

Cognitive Depletion During Interracial Group Deliberations

Self-control (self-regulation) is the limited ability to regulate and control one's behavior, thoughts, and emotions in light of current goals (Muraven, Tice, & Baumeister, 1998). Engaging in self-regulatory processes such as behavior monitoring and emotion regulation can have detrimental effects on aspects of cognitive performance such as memory (Richards & Gross, 2000) and logical decision making (Baumeister, Vohs, & Tice, 2007; Schmeichel, Vohs, & Baumeister, 2003), which is why regulatory depletion has been used as a model for cognitive depletion (Burkley, Anderson, & Curtis, 2011). Although the resource model of depletion has received some recent critiques and amendments (e.g., Inzlicht & Schmeichel, 2012; Tuk, Zhang, & Sweldens, 2015), it has been used successfully to predict that interracial interactions are depleting and result in reduced cognitive performance.

Sources of Depletion During Interracial Interactions

Although there is no research specifically on the depleting effects of group deliberations, there is informative research from the dyadic interactions literature. For example, tasks that require social coordination with even one other person are difficult (Finkel et al., 2006), perhaps because the interactions compete for cognitive resources with the task itself. In a group deliberation context, efforts to argue, persuade, or understand people with multiple perspectives and opinions are likely to tax cognitive resources. The more difficult the communication process, the more depleted group members should become during deliberations. For example, White participants went to great extents to not mention race when they were paired with Black participants in a photo-description task, an ineffective strategy that led to decreased team performance (because race was a central descriptor) and individual depletion

(because it was difficult to inhibit mentioning a salient social category such as race; Apfelbaum et al., 2008). In a mock jury study, Stevenson, Lytle, Baumholser, and McCracken (2017) found that White jurors provided shorter written justifications for their verdict, but took longer to write these justifications, when they were assigned to a diverse (vs. all-White) jury—indicating that White jurors were perhaps self-monitoring more when believing they were interacting with Black jurors.

These efforts to self-monitor and self-regulate during interracial interactions have cognitively depleting effects on Whites (Richeson & Shelton, 2003; Richeson & Trawalter, 2005; Zabel et al., 2015) and on Blacks (Murphy et al., 2013), potentially through mechanisms such as behavior monitoring and anxiety. White participants who interacted with a Black (vs. White) experimenter performed worse on a cognitive-control measure (i.e., Stroop test), indicating cognitive depletion. This effect was even more pronounced for those high in implicit prejudice—perhaps because these participants had to work harder to control prejudiced thoughts and behaviors (Richeson & Shelton, 2003). The mechanism that led to cognitive depletion appeared to be participants' efforts to control their behavior: White participants engaged in more behavioral control when they interacted with Black participants, and behavioral control was related to worse Stroop performance. In addition, concerns about appearing prejudiced are themselves depleting. For example, participants who received false "high prejudice" feedback (vs. control feedback) were more depleted after interracial interactions, while participants who could rely on a script (vs. no script) during interracial interactions were less depleted, ostensibly because they were less worried about saying the "wrong thing" (Richeson & Trawalter, 2005). Therefore, White jurors who deliberate as part of racially diverse (vs. homogeneous) juries might be more depleted after deliberation.

Depleting Effects of Interracial Interactions

Regulatory depletion has robust negative effects on cognitive performance (Muraven et al., 1998; Schmeichel et al., 2003). A meta-analysis found a significant effect of depletion on performance across a variety of cognitive tasks involved in jury and other types of group decision-making, such as memory, consideration of relevant facts, and logical thinking (Hagger, Wood, Stiff, & Chatzisarantis, 2010). Although it is hard to predict how depletion would affect the outcome of a deliberation (i.e., whether it would result in a guilty or not guilty verdict), it could affect the deliberation process. Thus, it is important to study the effects of regulatory depletion in a jury context. Depletion could impair jurors' memory for case facts, decrease the breadth and complexity of issues raised during deliberation, and increase the number of factual errors. Depletion might even impair higher-level cognitive processes such as jurors' ability to distinguish between valid and invalid evidence, to disregard inadmissible evidence, and to integrate multiple factors in the process of evaluating evidence. Therefore, the performance of White jurors (i.e., the ability to accurately recall and discuss a large number of facts from a trial) might suffer when they deliberate as part of racially diverse (vs. homogeneous) juries—especially when jurors are not concerned that their decisions might be biased in favor or against the defendant (i.e., when the defendant is White). These concerns, in contrast, might become particularly salient through the interactive effects of jury racial

composition and defendant race (i.e., when the defendant is Black).

Jury Racial Composition and Defendant Race

Although the depletion account paints a rather grim picture of interracial deliberation, jury research actually illustrates that jury diversity has a *positive* effect on decision quality. Diverse juries are less conviction-prone than all-White juries when the defendant belongs to an ethnic minority group (Bowers et al., 2001; Perez, Hosch, Ponder, & Trejo, 1993). White jurors even render more lenient verdicts for minority defendants at the mere anticipation of interacting with minority jurors (Kerr, Hymes, Anderson, & Weathers, 1995). In addition, diverse (vs. all-White) juries judging a minority defendant discuss more case facts, produce fewer factual inaccuracies, and are more likely to correct inaccurate statements (Sommers, 2006), all indicators of increased effort and motivation to reach a fair verdict. In fact, motivation can counter the pernicious effects of depletion on performance (Inzlicht & Schmeichel, 2012; Masicampo, Martin, & Anderson, 2014). Sommers' seminal study revealed that the presence of Black jurors did not simply motivate White jurors to be lenient toward a Black defendant, but rather motivated them to be "better jurors." Sommers argued that White jurors' desire to avoid prejudice is activated when the presence of Black jurors makes the race of a Black defendant salient to them, in accordance with aversive racism theories (Dovidio & Gaertner, 1998, 2004).

Therefore, when we consider the motivating effects of deliberating as part of a diverse group on individual performance, findings from jury research are actually well aligned with the models of cognitive and regulatory depletion that underlie and explain the depleting effects of interracial interactions. Specifically, people might become depleted when they engage in interracial interactions, but these interactions might also provide the very impetus needed for people to become aware of, and motivated to avoid, racial bias. This would have important implications for the criminal justice system, where racial bias against Black defendants at all steps of the adjudication process continues to be of great concern.

Defendant Race and Jurors' Verdicts

Despite evidence of pervasive racial bias against Black defendants in convictions and death-penalty sentencing revealed by studies and meta-analyses (Baldus, Woodworth, Zuckerman, & Weiner, 1997; Mitchell, Haw, Pfeifer, & Meissner, 2005), some research has failed to find effects of race on jurors' verdicts (Mazzella & Feingold, 1994). Why this disparity? Hunt (2015) distinguishes between experimental research—where White participants might suspect race is a target variable, and therefore might be particularly motivated to hide racial prejudice—and archival research of actual jury decisions. In support, defendant-race effects appear to emerge primarily in archival studies, and less so in mock-jury studies. In addition to this important methodological concern (i.e., mock jurors' suspicion that the research has something to do with race masking racial bias effects that would otherwise manifest in actual trials), social psychological theories can also explain why findings about jurors' racial bias are not consistent.

Social Identity Theory

According to social identity theory, people favor in-group versus out-group members (Tajfel & Turner, 1986). For jurors, this effect is expressed in the similarity-lenience bias (Kerr et al., 1995): White jurors are more lenient toward White versus Black defendants (Lynch & Haney, 2009), whereas Black jurors are more lenient toward Black versus White defendants (Sommers & Ellsworth, 2000). Thus, one would expect White jurors to convict the Black (vs. White) defendant because he is an outgroup (vs. in-group) member.

The pattern of ingroup favoritism is complicated, however, by a long history of racial inequality in the criminal justice system. Sommers and Ellsworth (2000) found that Black jurors were more lenient toward a Black, compared to a White defendant, presumably because they see the verdict as “an opportunity to personally contribute to the elimination of racial inequality” (p. 1376). In contrast, White jurors favored a White (vs. Black) defendant only when race was not highlighted as a salient issue at trial. This suggests that White jurors might also be mindful of the potential for racial bias but are motivated to avoid it only when race is made salient. Thus, one would expect White jurors to convict the Black (vs. White defendant) at higher rates when the jury is all-White, but not when the jury includes Black jurors.

Further, social identity theory also supports the prediction that people do not always favor ingroup (vs. outgroup) members, but instead they are more extreme in both *positive* and *negative* evaluations of ingroup (vs. outgroup) members, because ingroup members' actions are more relevant to people's social identity (black sheep effect; Marques, Yzerbyt, & Leyens, 1988). As a result, people judge unlikeable ingroup members more harshly than unlikeable outgroup members, in an effort to distance themselves from the “black sheep” of their social group and thus maintain a positive self-identity. For jurors, a criminal trial defendant perfectly fits the black-sheep profile—and in fact, jurors judge defendants of their same racial group (vs. a different race) more harshly when the evidence against them is strong (Kerr et al., 1995). Of particular interest for the present study, the black sheep effect appears to be stronger when jurors anticipate deliberating with racial outgroup (vs. ingroup) members—that is, when they might have to account for the indefensible behavior of an ingroup member in front of jurors of other races. Thus, White jurors might actually be more likely to convict a White (vs. Black) defendant when they interact with Black jurors (black-sheep effect), but less likely when they deliberate with other White jurors (similarity-lenience effect).

Aversive Racism Theory

Aversive racism theories stipulate that current cultural values prompt Whites people's concerns with fairness, social justice, and racial equality, while unconscious negative feelings and beliefs toward Black Americans still persist (Dovidio & Gaertner, 1998, 2004). Thus, most people are motivated to be—or at least appear—nonprejudiced and therefore make an effort to inhibit their biases. For this motivational mechanism to be activated, however, people must first be aware that the potential for bias exists. When they are not, racial biases are expressed indirectly. For example, in a jury study, the presentation of inadmissible damning evidence increased judgments of guilt for a Black, but not for a White

defendant (Johnson, Whitestone, Jackson, & Gatto, 1995). In other words, when jurors had a valid rationale for being harsh toward the Black defendant (i.e., damning evidence), their verdicts reflected racial bias. Bias is also most likely when race is not salient and the criteria for judgment are ambiguous (e.g., in jury trials with Black defendants where race is not salient, Sommers & Ellsworth, 2001). In contrast, when race is made salient during the trial, White jurors are motivated to avoid prejudice and disparities in guilt judgments between White and Black defendants are reduced (Sommers & Ellsworth, 2000, 2001).

Sommers (2006) relied on aversive racism theory to explain his findings. The presence of Black jurors (compared to only White jurors) drew White jurors' attention to the defendant's race (i.e., increasing race salience), thereby activating White jurors' motivation to avoid racially prejudiced verdicts and increasing systematic information processing. In turn, this resulted in decreased likelihood of a guilty verdict for a Black defendant. This explanation is in line with another motivation-driven mechanism of bias avoidance: the watchdog hypothesis (e.g., Fleming, Petty, & White, 2005). According to this hypothesis, people act as “watchdogs” and process information more systematically when they evaluate a target who belongs to a stigmatized group (i.e., Black vs. White defendant, Sargent & Bradfield, 2004) in an effort to avoid potential biases.

If diversity generally improves jurors' performance in cases involving a Black defendant, would the same be true if the defendant were White? If race salience is the motivating factor, then its motivating effects should be evident only when both the target of the judgment (i.e., defendant) and some of the jurors are Black—a situation that makes race salient. When that is not the case (e.g., in trials involving White defendants), jury-diversity effects could mirror those demonstrated by Watson et al. (1993) and others who have studied managerial tasks: impaired individual or group performance.

Thus, the present study tests a pattern of results in line with interracial interactions and jury decision-making research. By including experimental conditions in which the defendant is White—that is, conditions in which there is no particular concern for White jurors to appear racially biased—the present study tests whether (a) racially diverse jury deliberations are depleting, (b) whether depletion results in reduced cognitive performance during and after deliberation when it is not overcome by jurors' desire to avoid unfair bias, and (c) jurors' verdicts for ingroup versus outgroup members are sensitive to the jury's racial composition.

Design and Hypotheses

In a mock jury deliberation paradigm, groups of six jurors viewed an evidence presentation from a murder trial (including photographs of either a White or Black defendant) and deliberated for 40 min or until they reached a unanimous verdict. Each group was formed of four White participants and two confederates, the latter who were either both White or both Black, resulting in racially homogeneous or racially diverse juries. Thus, the study had a 2 (defendant race: white, black) \times 2 (jury composition: all White, diverse) design. Because Sommers (2006) suggested that the increased quality and depth of jury deliberations in diverse juries is due not only to Black jurors' informational contributions, but also to White jurors' increased effort, a more controlled par-

adigm ensured that confederates always contributed the exact same informational arguments to deliberation, regardless of race.

Dependent variables were group and individual verdicts, performance on a Stroop task as an indicator of depletion (because the task measures the ability to inhibit a dominant response, which requires self-regulatory resources; e.g., Richeson & Trawalter, 2005); memory for case facts after deliberation; and number of accurate and new facts mentioned during deliberation by each individual juror and also at the group level.

Increasing Depletion Hypothesis

Because interracial interactions are more depleting than same-race interactions (e.g., Richeson & Trawalter, 2005), a main effect of jury composition on postdeliberation levels of depletion was predicted. White jurors would be more depleted and would perform worse on the Stroop task (i.e., fail to inhibit dominant responses) when they deliberated in juries that included Black (vs. White) confederates. Defendant race would not have a significant effect on jurors' cognitive depletion after deliberation.

Increasing Performance Hypothesis

In line with prior research demonstrating that people's cognitive performance can suffer as a result of interracial interactions (e.g., Richeson & Trawalter, 2005), but also that White jurors perform better during deliberation when race is made salient by the presence of Black jurors (Sommers, 2006), I predicted no main effects of defendant race or jury diversity on jurors' performance during and after deliberation. Rather, jurors' performance would be predicted by a significant interaction between jury diversity and defendant race. On diverse juries, White jurors who judge a Black defendant would be particularly motivated to avoid succumbing to racial prejudice, and thus particularly likely to perform well by remembering case facts correctly and contributing more correct and varied comments to the deliberation. Thus, on diverse juries, performance would be better when the defendant is Black (vs. White), due to jurors' motivation to avoid prejudice against the defendant. On all-White juries, however, jurors would perform worse when the defendant is Black rather than White, reflecting less motivation to counter personal biases with careful information processing.

Increasing Leniency Hypothesis

In line with prior research suggesting a general similarity-lenency bias in verdicts and other legal judgments, White jurors should be more lenient toward a White, versus Black defendant. Yet in line with aversive racism and black sheep effects, I hypothesize that this main effect would be qualified by an interaction between defendant race and jury diversity. On all-White juries, jurors would be more lenient toward the White (vs. Black) defendant, demonstrating a similarity-lenency bias. But on diverse juries, the racial disparity in verdicts would be reduced or even reversed. I hypothesized a reduction in White jurors' racial bias when they serve on diverse juries for several reasons. The presence of out-group observers (i.e., Black jurors) would motivate White jurors to (a) distance themselves from the White defendant, a potential murderer (i.e., black sheep effect, Marques et al., 1988),

(b) avoid appearing biased in favor of an in-group member (Kerr et al., 1995; Vorauer, 2003), and (c) avoid appearing racially biased against a Black defendant when race is made especially salient by the presence of Black jurors (Sommers, 2006). Based on prior research indicating the jurors are particularly careful about racial bias at the mere anticipation of deliberating with a diverse group (e.g., Kerr et al., 1995), I hypothesized that these effects would hold true for jurors' individual verdicts both before and after deliberation, although they might be stronger postdeliberation.

Method

Participants

Two hundred and four jury-eligible (i.e., United States citizens older than 18 years), White community members participated in 52 juries (13 juries per experimental cell). Five of the 52 juries included only five members (three participants and two confederates); all the others included six members. Analyses with and without the 5-person juries revealed no differences. Six participants were excluded because they failed the manipulation check for defendant race, leaving a final sample of 198 jurors. Based on medium to large effects ($d_s > .82$) found by Sommers (2006) in a 2×3 design with $N = 200$ and 29 juries, this sample size appeared adequately large, especially because power and precision of coefficient estimates in nested designs depends on the number of clusters which was higher (52) in the present study (Snijders & Bosker, 2012). Following the recommendations of Browne (2006) and Snijders (2005), I weighed the total sample size needed for a non-nested design ($N = 128$ to detect a medium effect with 80% power, Gpower analyses) by the design effect value: design effect = $1 + (n - 1) * ICC$. The highest ICC (.08) resulted in a design effect of 1.23 and recommended sample size of $N = 157$.

Community members were recruited via Craigslist ads, mass emails to faculty and staff, fliers posted in the community, and word of mouth. Interested participants contacted the author by e-mail or phone and subsequently received a demographic questionnaire via e-mail or over the phone, asking for the following information: age, citizenship status, race/ethnicity, English proficiency, and gender.

Participants' age ranged from 18 to 82 years ($M = 38$ years, $SD = 15$ years), and 47% were women. All participants endorsed White as their ethnicity. Almost all mock jurors were native English speakers (95.6%), and all jurors who were not native speakers indicated they had attended an English-speaking school for at least 10 years. The sample was well educated, with 5.9% having a high school degree or GED as the highest degree obtained, 24.3% having some college, 47.4% a college degree, and 22.4% a graduate or professional school degree. All participants were paid \$42 cash at the end of the study.

Materials

Trial stimulus. The trial evidence presentation was based on an actual court case (*R. v. Valevski, 2000*) in which the defendant was accused of murdering his wife after an intense fight witnessed by his parents, during which she threatened to leave him. The victim was found dead from a knife wound to the throat in the

locked bedroom of their home the following day. The defense argued that the victim had committed suicide. Forensic and pathology experts testified for both sides. Several prior studies (Salerno & Peter-Hagene, 2013, 2015) involving this trial stimulus have demonstrated that the evidence is ambiguous enough to result in variability in guilty verdicts (46–30%), which suggests that jurors would have to deliberate to reach a unanimous verdict, allowing for the hypothesized interpersonal processes to occur. A printed list of all the characters involved in the trial was given to participants to keep throughout the study as a memory aid—though no additional information about the content of the testimony was included in this document. The evidence was presented via PowerPoint slides on a computer screen with photographs, printed information, and a voiceover to facilitate understanding. The presentation included written summaries of the opening and closing statements, eyewitness testimony establishing a timeline of events, expert forensic testimony, coroner's report, as well as photographs of the crime scene, weapon, and the defendant's shoe—the same types of information that a real trial would contain. There was a large amount of information, some of it technical, allowing for acceptable variation in responses on a memory test and in verdicts. The presentation lasted 23 min and included 32 slides.

Defendant race (White, Black) was manipulated via arrest photographs, which were actually computer-generated images that maintain the same age, weight, facial symmetry, emotional expression, and image texture while changing only the ethnicity parameter (FaceGen Modeler). Pilot testing revealed no significant differences in perceptions of how dangerous, suspicious, or criminal the defendants looked, all $t_s < .15$, all $p_s < .15$.

Pilot testing indicated that using a White victim in all conditions made jurors suspicious that the study was concerned with race, because the interracial marriage stood out to jurors. In fact, interracial sexual abuse draws fewer guilty verdicts than same-race abuse because jurors think that it is less plausible (Bottoms, Davis, & Epstein, 2004). Therefore, the victim's (i.e., wife's) race was varied along with the race of the defendant (the husband), making the crime always intraracial. The wife's race was manipulated by having her name be either stereotypically White (Emily) or Black (Lakisha; Bertrand & Mullainathan, 2004). No pictures of the wife were included in the trial stimulus.

Illinois pattern jury instructions for first-degree murder. Jurors received the same jury instructions that would be given in the state of Illinois for this crime. Instructions were written and accompanied by a voiceover, and informed jurors of the function

of the jury, the court, and the attorneys, provided guidelines of how to rely on the evidence presented during trial, and stated the importance of the duty of jurors. Then, specific elements of the crime of first-degree murder were provided, as well as the two verdict options (guilty or not guilty).

Individual and group verdicts. Verdicts (guilty, not guilty) were assessed privately before and after deliberation. For postdeliberation private verdicts, jurors were prompted to respond with their personal verdict preference regardless of how their group had voted. After deliberation, a group verdict was also recorded. Because one of the confederates always voted not guilty, the two verdict options for the group were either not guilty or hung jury.

Stroop task (Stroop, 1935). This classic measure of executive control asks participants to report, as quickly as possible, the font color for a stimulus word or string of letters. In control trials, a string of nonsense letters are presented. In congruent trials, the word (i.e., red, green, blue, black) denotes the same color as its font. In incongruent trials, the word is not the color of its font, so participants have to inhibit the dominant response—reading and reporting the word's meaning rather than the font—which requires regulatory resources that should be impaired upon depletion (e.g., Inzlicht, McKay, & Aronson, 2006; Webb & Sheeran, 2003). This measure was used successfully in prior investigations of interracial interactions and depletion (Apfelbaum et al., 2008; Richeson & Shelton, 2003; Richeson & Trawalter, 2005).

The task was administered on individual laptops via E-Prime Software and consisted of 20 practice trials followed by 80 experimental trials. Each word was presented for a maximum of 2,000 ms, preceded by a fixation cross (+), and with an intertrial interval of 1,500 ms. Valid latencies were used to compute Stroop interference scores (mean latency for incongruent trials minus mean latency for control trials), with higher numbers indicating greater cognitive depletion (Table 1; Apfelbaum et al., 2008; Richeson et al., 2005).

Recognition test for case facts. A 35-item true/false memory test measured jurors' recognition of case facts presented in the trial stimulus. Questions were detailed enough to capture variation in careful processing of information and covered all aspects of the case, from forensic evidence (e.g., "Traces of blood were found on the defendant's clothing") to timeline of events (e.g., "The defendant took his parents to his sister's house Sunday morning") and witness testimony ("The neighbor testified the defendant seemed distressed"). Participants received a point for each correct answer for a maximum possible score of 35 (see Table 1).

Table 1
Grand Means, Standard Deviations, and Ranges for Continuous Measures

Main dependent measures	<i>M</i>	<i>SD</i>	Minimum	Maximum
Stroop difference score	91	97	−105	487
Stroop RT incongruent trials	943	209	471	1673
Stroop RT control trials	853	167	506	1410
Recognition test for case facts	28.34	3.27	8.00	35.00
Number of all case facts during deliberation	47.26	29.27	.00	158.00
Number of correct case facts during deliberation	44.77	27.71	.00	151.00
Number of new correct case facts during deliberation	16.61	10.04	.00	60.00

Note. RT = reaction time.

Motivation to control racial prejudice (Dunton & Fazio, 1997). The 17-item questionnaire measures people's motivation to avoid prejudice against Black people on a scale from -3 (*strongly disagree*) to $+3$ (*strongly agree*) with a midpoint, (e.g., "I always express my thoughts and feelings, regardless of how controversial they might be"; "I feel guilty when I have a negative thought or feeling about a Black person"). The scale is reliable with Cronbach's α s ranging from .74 to .81 in the original study, and $\alpha = .80$ in the present study.

Thermometer measure of racial attitudes. A "thermometer" measure of racial attitudes (e.g., Hugenberg & Bodenhausen, 2004; Miller, Smith, & Mackie, 2004) ranging from 1 (*very cold*) to 7 (*very warm*) assesses participants' general feelings of warmth toward Black ($M = 4.76$, $SD = 1.09$) and White Americans ($M = 4.95$, $SD = 1.27$), indicating general positive or negative attitudes toward these groups. A difference score (i.e., [White Americans – Black Americans]; Hugenberg & Bodenhausen, 2004) was used as a relative measure of explicit prejudice, with higher scores indicating higher preference for White versus Black Americans, $M = .16$, $SD = 1.23$.

Manipulation check. Two separate items assessed the success of the defendant and the victim race manipulation by asking participants to indicate whether the defendant and the victim were White or Black.

Demographic information. Participant age, gender, ethnicity, political orientation, and education were assessed. Two questions prompted participants to indicate whether they (a) were native English speakers and (b) attended at least 10 years of school in English.

Procedure

Confederate training. Six undergraduate research assistants (three White, three Black) were selected to play the part of mock jurors and to participate in teams of two for each experimental session. All six confederates were college-aged women, with ages ranging from 21 to 24. Because most of the confederates wore glasses, the ones who did not need glasses wore fake ones. All confederates wore jeans and plain long- or short-sleeved t-shirts in neutral colors and with no defining characteristics for each experimental session. None of the confederates had any stand-out features such as piercings or tattoos, or any speech particularities.

The research team (confederates and other research assistants) identified a list of statements made by most participants in a prior study involving the same trial stimulus. We anticipated these comments would likely be made by actual participants early during deliberation, and therefore the confederates would not have to contribute novel information or shape the deliberation—allowing actual participants to do so themselves. We also included neutral statements that did not clearly side with the defense or the prosecution (e.g., "If she was depressed, that could go both ways, that she killed herself or that she was getting so crazy she was really taking his children.") The list was edited through pilot practice sessions, and the statements were rephrased to sound conversational, until all confederates were comfortable delivering them in the exact same manner. In addition to the list of statements, confederates' scripts included (a) comments to be made before the study started, in the waiting room, meant to signal to participants that confederates were not familiar with the research setting (e.g.,

"This place was so hard to find"), and (b) gestures and body positions (e.g., "look at the character list" "lean forward", "elbow on the table.")

Once the script was finalized through pilot sessions and group discussions, confederates practiced the final script for another semester. Three types of practice sessions were scheduled. First, confederates practiced with the rest of the research team—all of whom were familiar with the trial stimulus and with the general goal to ensure confederates behave exactly the same in all sessions. During these sessions, we practiced several scenarios: juries leaning toward acquittal, juries pressuring confederates to convict, juries where people argued with each other and left little room for confederates to intervene and say their lines, and juries where people talked very little. Second, confederates practiced with the principal investigator's colleagues, who were unfamiliar with the trial evidence but knew confederates were training to achieve uniformity. Third, confederates practiced with student participants in the subject pool, who were unaware they were confederates. None of the pilot participants suspected the research assistants were anything other than fellow subject pool participants.

Standardizing confederates' verbal contributions in this way across conditions controlled for contributions to the discussion, allowing for isolation of the effects of minority jurors' mere presence. That is, confederates maintained a noncommittal stance during deliberation for as long as possible to allow the actual participants to guide discussions. When a jury called for a vote, one confederate always voted not guilty while the other remained undecided and sided with the majority only at the end of the deliberation session. If the majority leaned toward not guilty, this could result in a unanimous verdict (provided there were no holdouts who maintained a guilty vote). If the majority leaned toward guilty, this resulted in a hung jury with at least one confederate maintaining a not guilty vote until the end of the deliberation. The solution of splitting confederates' verdict options in this way stems from the piloting process during which it became apparent that participants were more likely to put confederates "on the spot" and demand long explanations from them when confederates voted to convict.

Experimental sessions. Participants saw study flyers or read recruitment ads online (i.e., Craigslist, reddit, listservs) and, if interested, inquired via e-mail or phone. Interested participants were sent a short screening questionnaire meant to establish eligibility, including ethnicity. The questionnaire included English proficiency questions because non-native English speakers perform worse than monolingual English speakers on the Stroop task (e.g., Rosselli et al., 2002). To avoid proficiency effects, only participants who were native speakers or had attended at least 10 years of school in English were scheduled to participate. A research assistant met scheduled participants in the building lobby and escorted them to the session room. Confederates arrived in lobby, just like all other jurors. One of the confederates was instructed to mention she had a hard time finding the building, to increase her believability as a community member.

The mock-jury sessions took place in a laboratory room provided with a large deliberation table and chairs, a desktop computer, and a mounted camera. Participants were randomly assigned to seats numbered from 1 to 4, with confederates always entering the room last and taking the same two seats in each session (5 and 6). After giving general oral instructions about the session and highlighting the importance of mock trial research, the experi-

menter presented the 23-min trial stimulus. Then, mock jurors heard and read actual Illinois pattern jury instructions for first-degree murder, followed by individual pre-deliberation verdicts.

Mock jurors deliberated for up to 40 min, while the experimenters recorded and watched the deliberation through a live feed from an adjacent room. During deliberation, the confederates delivered scripted lines in a conversational, natural manner. After deliberation, jurors again indicated their individual, private verdict preferences. Then jurors completed the Stroop task and the recognition test for case facts. At this point, before mock jurors saw any of the race-related measures, they indicated what they thought the study was about. After this open-ended question, participants completed measures of prejudice, race manipulation checks, and demographic measures. At the end of the study, participants were debriefed, asked if they had suspected anything about the confederates, and were paid for their participation. No participants indicated suspecting the confederates before or after the study purpose was revealed. Trial materials, instructions, and measures were all reviewed by an attorney to ensure ecological validity.

Deliberation coding. Deliberations were filmed and transcribed by trained research assistants. Using a detailed coding manual adapted after the coding strategy reported by Sommers (2006), two independent coders and the author coded 20% of all deliberations ($N = 11$) to establish reliability. All independent coders were blind to the experimental conditions, which were not evident in the written transcripts. Disagreements in deliberation coding were resolved through discussion with all three coders present. Following discussion, the author made final decisions about coding and we took detailed notes for future reference. After establishing reliability, the remaining 41 sessions were coded separately: One coder coded 13 sessions, and two coders coded 14 sessions each. Overall, independent coders reached acceptable reliability: over 80% agreement with the author and 76% with each other (see Table 2 for all % agreement and mean Kappas). Agreement was above 73% for all codes.

To eliminate redundancy and to avoid giving credit to jurors who repeated the same statements during a single turn, each fact was coded only the first time it was mentioned during a juror's single turn. For example, a juror said, "The door was locked, if you see a locked door, you automatically assume the person is still inside. If you find the door locked, that means they're still in there." This statement was only coded once for the fact that the defendant found the bedroom door locked. If the same juror repeated this statement after other jurors spoke, however, the statement was coded again, as part of a new speaking turn.

Once a statement was coded as factual (i.e., after it received a fact code), coders decided whether the fact was correct or

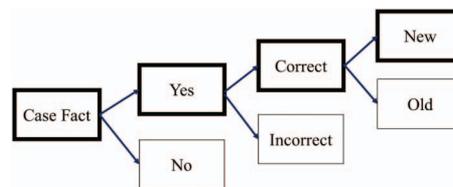


Figure 1. Coding scheme decision tree describing how coding decisions were made at each coding level: Only factual comments were coded correct/incorrect; only correct comments were coded new/old. See the online article for the color version of this figure.

incorrect—the statement received a correct/incorrect code (see Figure 1). Thus, the first coding level was to decide whether a statement was factual in the first place; the second coding level was to decide whether jurors accurately recalled the case facts they were referencing. Only for facts coded as correct, it was then decided whether that fact was being mentioned for the first time during deliberation. Thus, for the third coding level, each fact/correct statement was further coded as new or old. To ensure accuracy, each transcript was accompanied by the detailed list of case facts and coders crossed out each new case fact as they encountered it. This made it easy to decide whether facts were new or old by checking the list. When two jurors spoke at once to contribute a new fact, both contributions were coded as new.

Deliberation coding measures. Measures were developed as a result of deliberations coding to capture several desirable aspects of jury deliberations: reliance on evidence and case facts, accuracy in recalling case details, and breadth of the evidence discussed (e.g., Cowan, Thompson, & Ellsworth, 1984; Horowitz & Bordens, 2002; Sommers, 2006).

Number of case facts mentioned during deliberation. Based on the coding scheme and a list of case facts, the total number of facts mentioned by each juror was calculated to measure jurors' general reliance on evidence and ability to focus on case facts (see Table 1).

Number of correct case facts mentioned during deliberation. The number of times jurors correctly recalled a case fact was tallied to reflect accuracy during deliberation (see Table 1).

Number of new correct case facts mentioned during deliberation. Correct facts were coded as new the first time they were mentioned during that deliberation as a measure of breadth, an indicator of cognitive performance (Sommers, 2006).

Table 2

Frequencies for Each Code as Coded by Each Coder and Agreement Between Coders

Variable	Coder 1	Coder 2	Coder 3	Coders 1 & 2 % agreement	Coders 1 & 3 % agreement	Coders 2 & 3 % agreement	Mean Kappa
Number of case facts	2,012	1840	2074	.87	.87	.81	.76
Number of correct facts	1,898	1765	1940	.83	.83	.76	.75
Number of correct new facts	639	660	733	.82	.81	.73	.73
Number of correct old facts	1,259	1105	1207	.81	.81	.74	.74
Overall				.83	.84	.77	.75

Note. Coder reliability was calculated based on 20% of the data (11 transcripts). Only correct facts were further coded into new or old facts.

Results

Preliminary Analyses

All participants were included in analyses for which they provided data (pairwise deletion). Analyses conducted with and without four participants who did not complete the defendant-race manipulation check revealed no differences, so these participants were not excluded from analyses. Deliberations lasted on average 38 min. Diverse juries deliberated longer ($M = 38.89$, $SD = 3.01$) than all-White juries ($M = 36.50$, $SD = 5.43$), $F(1, 181) = 13.76$, $p < .001$, partial $\eta^2 = .07$, 95% confidence interval [CI] [.02, .15]. This effect, however, was qualified by a Defendant Race \times Jury Composition interaction, $F(1, 181) = 6.96$, $p = .009$, partial $\eta^2 = .04$, 95% CI [.002, .10]. Specifically, the effect of jury composition on deliberation time was not significant for the White defendant, $F(1, 181) = .57$, partial $\eta^2 = .01$, 95% CI [.00, .04], but for the Black defendant, diverse juries deliberated significantly longer than all-White juries, $F(1, 181) = 19.99$, $p < .001$, partial $\eta^2 = .14$, 95% CI [.03, .19].

Racial attitudes, political orientation, gender, and age were considered as potential covariates. Measures of racial prejudice and motivation to control racial prejudice were not related to jurors' depletion levels, performance, and verdicts. Women recognized more case facts than men (see primary hypotheses testing statistics); older (vs. younger) participants were more depleted and remembered fewer case facts (see Table 3). Gender and age were therefore controlled for in all hypothesis testing analyses. With one exception, the pattern of results was identical with or without these controls.

To compute the cognitive depletion measure, Stroop mean latency was calculated for correct trials with response times between 250 ms and 1,813 ms (i.e., within 3 SD s from the mean). Repeated measures analyses of variance (ANOVA) illustrated the classic Stroop effects: Reaction times varied significantly by Stroop condition, $F(2, 372) = 69.25$, $p < .001$, partial $\eta^2 = .27$, 95% CI [.20, .34]. Pairwise comparisons with Bonferroni-adjusted confidence intervals revealed that RTs in incongruent trials ($M = 939$ ms, $SD = 215$ ms) were significantly larger than RTs in congruent ($M = 850$ ms, $SD = 170$ ms), 95% CI for mean difference [113, 65] and control trials ($M = 856$ ms, $SD = 197$ ms), 95% CI for mean difference [60, 105]. There were no significant reaction time (RT) differences between congruent and control trials, 95% CI [-19, 5]. Two participants had unusual difference scores between incongruent and control trials (-890 and -462). The effects reported reflect analyses with these outliers excluded.

Primary Hypotheses Testing

Analyses testing the primary hypotheses (i.e., with depletion, performance measures, and verdicts as dependent variables) were conducted via multilevel models to account for the nested nature of the data, where values for denominator degrees of freedom are obtained by a Satterthwaite approximation (Snijders & Bosker, 2012). Jury composition (all-White, diverse) and defendant race (White, Black) were group-level predictors, whereas verdicts, depletion, memory, and deliberation comments were individual-level outcomes. Control variables (age, gender) were at the individual level. Means and standard deviations (or % guilty verdicts) for each experimental cell are reported in Table 4; significance tests are reported in Table 5.

Increasing depletion hypothesis. I predicted a main effect of jury diversity on depletion: White jurors on diverse juries would be more depleted than White jurors on all-White juries. First, analyses with the difference scores in Stroop latencies between the incongruent and control trials revealed that jurors on diverse juries were significantly more depleted than jurors in all-White juries (see Table 5). This effect was not qualified by defendant race (see Figure 2).

Analyses were also performed on the effects of jury composition and defendant race on RTs on incongruent and on control trials separately. For difficult incongruent trials, the effect of jury composition was significant, with higher RTs in the diverse, versus all-White jury condition, $F(1, 185) = 4.49$, $p = .03$, $b = -63.78$, 95% CI [-123.16, -4.40]. For easy control trials, as theory would predict, the effect of jury composition was not significant, $F(1, 185) = 1.53$, $p = .22$, $b = -30.07$, 95% CI [-78.09, 17.95]. This pattern also supports the hypothesis that the Stroop effect in this study demonstrates cognitive depletion, because jury composition only affected jurors' RTs on the difficult trials that require self-control (i.e., incongruent), but not on easier trials. Thus, results supported the increasing depletion hypothesis: Interracial interactions were cognitively depleting for jurors, even though the Black and White confederates acted in the same way and said the same things.

Increasing performance hypothesis. I predicted no main effects of defendant race or jury diversity. Rather, jurors' performance would be predicted by a significant crossover interaction between jury diversity and defendant race. On diverse juries, White jurors would remember more case facts when the defendant is Black (vs. White), due to jurors' motivation to avoid potential prejudice against the Black defendant. On all-White juries, however, jurors would remember fewer case facts when the defendant

Table 3
Bivariate Correlations Among Main Dependent Measures and Covariates

Variable	1	2	3	4	5
1. Stroop difference score	—				
2. Recognition test for case facts	-.02	—			
3. Number of case facts mentioned during deliberation	.05	.12	—		
4. Number of correct case facts mentioned during deliberation	.04	.12	.99**	—	
5. Number of new case facts mentioned during deliberation	.03	.18*	.86**	.87*	—
6. Juror age	.25**	-.15*	.65	.63	.46

Note. All $ns = 183-197$.

* $p < .05$. ** $p \leq .01$.

Table 4

Means and Standard Deviations for Individual Jurors' Scores on Main Variables as a Function of Jury Composition and Defendant Race

Variable	All-White jury				Diverse jury			
	White defendant		Black defendant		White defendant		Black defendant	
	<i>M</i> (<i>SE</i>)	<i>SD</i>	<i>M</i> (<i>SE</i>)	<i>SD</i>	<i>M</i> (<i>SE</i>)	<i>SD</i>	<i>M</i> (<i>SE</i>)	<i>SD</i>
Stroop difference score	78 ^a (16)	109	69 ^a (10)	73	118 ^b (16)	111	97 ^b (14)	92
Stroop incongruent trials	925 ^a (31)	214	898 ^a (29)	202	1003 ^b (34)	231	948 ^b (26)	171
Stroop control trials	847 ^a (27)	184	828 ^a (24)	171	885 ^a (27)	183	852 ^a (18)	118
Recognition test for case facts	28.13 ^a (.40)	2.76	28.86 ^a (.42)	3.01	27.84 ^a (.59)	4.21	28.53 ^a (.40)	2.77
Number of case facts	51.89 ^a (4.63)	31.78	39.70 ^b (3.34)	23.86	43.96 ^{ab} (3.39)	24.21	54.45 ^c (5.09)	34.87
Number of correct case facts	50.98 ^a (4.39)	30.09	37.14 ^b (3.01)	21.56	40.98 ^{ab} (3.22)	23.03	51.9 ^a (4.88)	33.48
Number of new case facts	17.28 ^{ab} (1.64)	11.25	14.69 ^a (1.11)	7.93	15.75 ^{ab} (1.23)	8.78	18.98 ^b (1.71)	11.72
% individual guilty verdicts								
Predeliberation % guilty verdicts	44% ^a		35% ^b		65% ^a		32% ^b	
Postdeliberation % guilty verdicts	40% ^a		15% ^b		49% ^a		30% ^b	

Note. Superscripts indicate significant cell differences in means or %. Means (or %) with the same superscript (e.g., ^a) are not significantly different from each other; means (or %) with different superscripts are significantly different. Some cell means are not significantly different from two other means that are different from each other; in this case, the means have a dual superscript (e.g., ^{ab}) to indicate similarity/lack of difference with both other means.

is Black rather than White, reflecting less motivation to counter personal biases with careful information processing.

This hypothesis was tested with several measures of jurors' performance: (a) scores on the postdeliberation recognition test (i.e., number of correct responses to true/false questions about the case), (b) the overall number of case facts mentioned during deliberation, (c) the number of correct case facts, and (d) the number of new facts. For all indices of performance during deliberation (i.e., the coded measures, b–d), group-level analyses with two-way ANOVAs were also performed where juries, not jurors, were the level of analysis. These analyses revealed patterns that mirrored juror-level analyses, with significant interactions but simple effects that failed to reach significance, likely due to the low power (i.e., $N = 52$). Means and standard deviations (or % guilty verdicts) for each experimental cell are reported in Table 4; significance tests are reported in Table 5.

Recognition test for case facts. The main effects of defendant race and jury composition and their interaction on the jurors' recognition memory for case facts were not significant, all p s > .12. The pattern of results was the same with and without gender and age as control variables (Tables 4 and 5).

Number of case facts during deliberation. The main effects of defendant race and jury composition on the number of case facts mentioned by each juror during deliberation were not significant; however, as predicted, there was a significant interaction between the two factors (see Table 5). Follow-up analyses of simple effects revealed that, as predicted, on all-White juries, jurors mentioned significantly more case facts when the defendant was White compared to Black, $F(1, 24.47) = 4.65, p = .041, b = 14.11, 95\% \text{ CI } [.62, 27.61]$. In contrast, on diverse juries, the disparity in the number of case facts discussed by jurors who saw a White or Black defendant was reduced and did not reach statistical significance, $F(1, 87) = 2.71, p = .103, b = -10.47, 95\% \text{ CI } [-23.11, 2.17]$ (see Table 4). Thus, jury composition significantly moderated the effect of defendant race on jurors' discussion of evidence, resulting in decreased racial disparity.

To assess whether jury diversity affects the deliberation process differently for Black versus White defendants, I also tested the

simple effects of jury composition separately for the Black and White defendants. Jury diversity increased the number of factual comments when the defendant was Black, $F(1, 92) = 6.03, p = .016, b = -15.06, 95\% \text{ CI } [-27.24, -2.88]$, but did not decrease when the defendant was White, $F(1, 23.29) = 2.93, p = .10, b = 10.13, 95\% \text{ CI } [-2.10, 22.35]$ —illustrating that diversity improves deliberation for Black defendants at no cost to deliberation quality for White defendants.

Number of correct case facts during deliberation. Analyses revealed the predicted significant crossover interaction between defendant race and jury composition on the number of *correct* case facts mentioned by jurors during deliberation (see Table 5). Follow-up analyses revealed that, on all-White juries, jurors mentioned significantly more correct case facts when the defendant was White compared to Black, $F(1, 24.62) = 4.79, p = .038, b = 13.49, 95\% \text{ CI } [.79, 26.18]$. On diverse juries, although the pattern of means was reversed (i.e., more case facts when the defendant was Black compared to White), again the disparity between White and Black defendant was reduced by jury diversity, $F(1, 87) = 2.69, p = .10, b = -10.01, 95\% \text{ CI } [-22.15, 2.11]$ (see Figure 3), because the number of facts did not differ by defendant race.

Testing the simple effects of jury composition separately for the Black and White defendants revealed that, as predicted, jury diversity increased the number of correct factual comments when the defendant was Black, $F(1, 92) = 6.78, p = .011, b = -15.06, 95\% \text{ CI } [-26.54, -3.57]$, and did not decrease it when the defendant was White, $F(1, 23.29) = 2.70, p = .115, b = 9.31, 95\% \text{ CI } [-2.43, 21.04]$.

Number of new case facts during deliberation. The main effects of defendant race and jury composition were not significant, and the interaction term was not significant in the presence of control variables—juror gender and age (see Table 5). The interaction between defendant race and jury composition was significant when control variables were not included in the multilevel model, $F(1, 184) = 3.93, p = .042, b = 5.85, 95\% \text{ CI } [.03, 11.67]$. Given the difference in statistical conclusions based on the inclusion or exclusion of juror gender and age as well as the a priori nature of the hypotheses tested, follow-up analyses were per-

Table 5
Jury Composition and Defendant Race Effects on Individual Jurors' Scores

Variable	<i>F</i>	<i>df</i>	<i>p</i>	<i>b</i>	95% CI	Exp(<i>b</i>)
Stroop difference score						
Defendant race	1.31	48.23	.254	9.60	[-18.88, 77.50]	
Jury composition	5.05	50.16	.030	-29.65	[-57.84, -1.47]	
Defendant × Jury	.01	48.50	.987	-.46	[-57.38, 56.46]	
Juror gender	1.76	179.66	.186	18.80	[-9.15, 46.75]	
Juror age	10.26	181.16	.002	1.53	[.59, 2.48]	
Recognition test						
Defendant race	1.08	52.90	.303	-.50	[-1.47, .47]	
Jury composition	.27	51.58	.609	.25	[-.72, 1.21]	
Defendant × Jury	.01	53.02	.998	.01	[-1.94, 1.94]	
Juror gender	3.62	181.78	.059	.88	[-.03, 1.78]	
Juror age	4.21	182.00	.042	-.03	[-.06, -.001]	
Number of facts						
Defendant race	.04	47.16	.840	.93	[-8.53, 10.41]	
Jury composition	.39	45.78	.538	-2.89	[-12.28, 6.49]	
Defendant × Jury	8.43	45.35	.006	25.00	[7.66, 42.34]	
Juror gender	2.83	178.44	.095	25.00	[-1.25, 15.72]	
Juror age	.19	180.35	.660	.06	[-.22, .35]	
Number of correct facts						
Defendant race	.05	47.21	.821	1.02	[-7.98, 10.00]	
Jury composition	.57	45.83	.57	-3.35	[-12.26, 5.56]	
Defendant × Jury	8.74	45.47	.005	24.12	[7.70, 40.55]	
Juror gender	2.52	178.53	.114	6.47	[-1.56, 14.50]	
Juror age	.22	180.38	.643	.06	[-.21, .34]	
Number of new facts						
Defendant race	.01	181	.96	-.07	[-3.05, 2.91]	
Jury composition	1.16	181	.283	-1.61	[-4.56, 1.34]	
Defendant × Jury	3.66	180	.057	5.72	[-.18, 11.63]	
Juror gender	.86	180	.356	1.39	[-1.58, 4.36]	
Juror age	.55	180	.459	-.04	[-.14, .06]	
Pre-deliberation guilty verdict						
Defendant race	6.86	177	.010	-.83		.44
Jury composition	2.00	177	.159	.44		1.55
Defendant × Jury	3.19	176	.076	1.14		3.12
Juror gender	2.63	176	.107	-.52		.59
Juror age	2.23	176	.137	-.02		.98
Post-deliberation guilty verdict						
Defendant race	7.24	177	.008	-1.15		.32
Jury composition	3.42	177	.066	.78		2.19
Defendant × Jury	.02	176	.921	.09		1.09
Juror gender	.64	176	.424	-.30		.74
Juror age	2.42	176	.121	-.02		.98

Note. Control variable (juror gender and age) statistics are reported from the final model, including the interaction term. Bolded *F* values are statistically significant, $p < .05$.

formed to test the predicted mean differences (control variables included in analyses).

The pattern of means mirrored the one for total and correct facts (see Table 4). The simple slopes for defendant race effects were significantly different from each other, as hypothesized, and trended in opposite directions (i.e., higher number of new facts for the White vs. Black defendant on all-White juries, somewhat higher number of new facts for the Black vs. White defendant on diverse juries), but they were not statistically different from zero on all-White, $F(1, 94) = 2.00, p = .161, b = 2.90, 95\% \text{ CI} [-1.17, 6.96]$, or on diverse juries, $F(1, 87) = 1.45, p = .23, b = -2.70, 95\% \text{ CI} [-7.12, 1.74]$. Testing the simple effects of jury composition separately for the Black and White defendants revealed that, as predicted, jury diversity increased the number of new facts when the defendant was Black, $F(1, 92) = 4.27, p = .042, b = -4.27, 95\% \text{ CI} [-8.37, -.17]$, but did not decrease the

number of new facts when the defendant was White, $F(1, 89) = .53, p = .47, b = 1.56, 95\% \text{ CI} [-2.71, 5.84]$. Thus, all analyses revealed a consistent pattern of crossover interactions between jury racial composition and defendant race: On all-White juries, jurors performed better when the defendant was White (vs. Black), but on diverse juries this disparity was reduced—in line with the increasing performance hypothesis.

Of note, the relationships between postdeliberation depletion, as revealed by Stroop difference scores, and jurors' discussion and recognition of case facts during and after deliberation were not significant (i.e., the measures of performance during deliberation and postdeliberation Stroop difference scores were not significantly correlated, Table 3).

Group-level analyses for deliberation content. Group-level analyses revealed a significant crossover interaction for number of facts mentioned during deliberation (i.e., on all-White juries, more



Figure 2. Increasing depletion hypothesis. Effects of jury composition and defendant race on cognitive depletion measured by the difference score in ms between incongruent trials (i.e., when the word meaning and word font were mismatched) and control trials (i.e., where the nonword stimulus, “XXX” was presented in various font colors). Larger difference scores indicated more depletion. Means are reported for each bar and SEs are represented by error bars.

facts were mentioned when the defendant was White vs. Black; on diverse juries, more facts were mentioned when the defendant was Black vs. White) $F(1, 48) = 7.23, p = .01$. Although significantly different from each other, the simple effects were not significantly different from 0: all-White juries, $F(1, 48) = 3.48, p = .07, d = .73, 95\% \text{ CI} [-.07, 1.52]$; diverse juries, $F(1, 48) = 3.96, p = .06, d = -.78, 95\% \text{ CI} [-1.57, .02]$. The exact same pattern held for number of correct case facts: significant interaction, $F(1, 48) = 7.67, p = .01$, defendant race effects on all-White juries, $F(1, 48) = 3.78, p = .06, d = .89, 95\% \text{ CI} [-.04, 1.55]$, and on diverse juries, $F(1, 48) = 4.03, p = .06, d = -1.03, 95\% \text{ CI} [-1.58, .02]$ were in opposite directions and significantly different from each other, but not significantly different from 0. Finally, identical patterns emerged for number of new case facts: significant interaction, $F(1, 48) = 6.64, p = .01$; opposite but not significant simple effects of defendant race on all-White juries, $F(1, 48) = 3.08, p = .10, d = .68, 95\% \text{ CI} [-.11, 1.46]$, and on diverse juries, $F(1, 48) = 3.64, p = .07, d = -.75, 95\% \text{ CI} [-1.53, .06]$.

Increasing leniency hypothesis. I predicted a main effect of defendant race on White jurors’ verdicts before and after deliberation: Jurors would be more likely to convict a Black versus White defendant. This effect, however, would depend on the racial composition of the jury: On all-White juries, the racial bias effect would be significant; but on diverse juries, racial bias would be reduced (i.e., not significant). I predicted that this pattern would be reflected in pre-deliberation, not just post-deliberation verdicts,

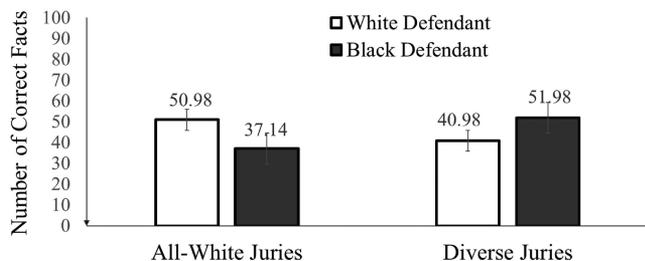


Figure 3. Increasing performance hypothesis. Effects of jury composition and defendant race on the number of correct case facts contributed by each juror during deliberation. Means are reported for each bar and SEs are represented by error bars.

because jurors were already anticipating deliberating in either all-White or racially diverse groups.

Overall, 62% ($N = 32$) of juries reached a not guilty verdict for their public, postdeliberation judgment, while 38% ($N = 20$) hung. Recall that, because of confederates’ script, group-level guilty verdicts were not possible in this study—therefore, the group verdict analyses are descriptive. Juries were more likely to reach a not guilty group verdict when the defendant was Black versus White, $\chi^2 = 19.98, p < .001, \phi = .32, 95\% \text{ CI} [.19, .45]$ and when the jury was all-White versus diverse, $\chi^2 = 23.77, p < .001, \phi = .37, 95\% \text{ CI} [.25, .49]$ (see Figure 4).

Although jury-level verdicts were limited by the methodological choice to restrict confederates to predetermined verdict preferences, individual jurors were able to endorse either a guilty or not guilty verdict in their private verdict decisions before and after deliberating, regardless of the verdict upon which their group agreed (or failed to agree). Overall, before deliberation, 44% of participants gave a guilty verdict, which decreased to 32.5% after deliberation. Multilevel model analyses (which have been used to investigate dichotomous verdicts in jury research, e.g., Ruva & Guenther, 2017) conducted with GENLINUX SPSS syntax revealed a significant effect of defendant race on private, pre-deliberation verdicts, but this effect was in the opposite direction to the one predicted based on prior research: Jurors were significantly more likely to convict the White compared to Black defendant. Although there was not a significant interaction between defendant race and jury composition, Figure 5 suggests that the main effect might be primarily driven by the high conviction rate for the White defendant—in line with the black sheep hypothesis. Postdeliberation conviction rates followed the same pattern as pre-deliberation verdicts: overall higher conviction rates for the White (vs. Black) defendant. The interaction effect was not significant postdeliberation (see Tables 4 and 5, Figure 5).

Discussion

In a highly controlled mock trial study, jury racial composition and defendant race were manipulated to demonstrate that although interracial interactions result in cognitive depletion after jury deliberation, diversity helps jurors perform better during a complex, group deliberation setting. Specifically, jury diversity reduced the disparity in deliberation quality between cases involving White or Black defendants. White jurors on all-White juries discussed a

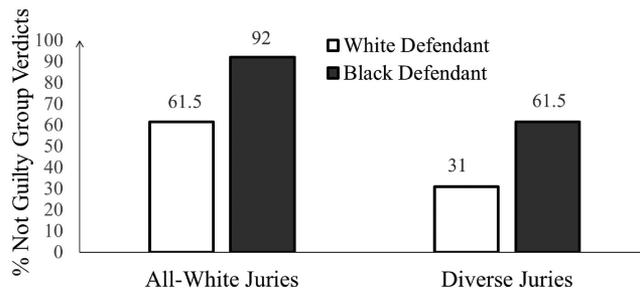


Figure 4. Increasing leniency hypothesis. Effects of jury composition and defendant race on group verdicts. Higher scores indicate higher percentage of juries who reached a unanimous not guilty verdict in each condition, with the alternative decision being hung jury.

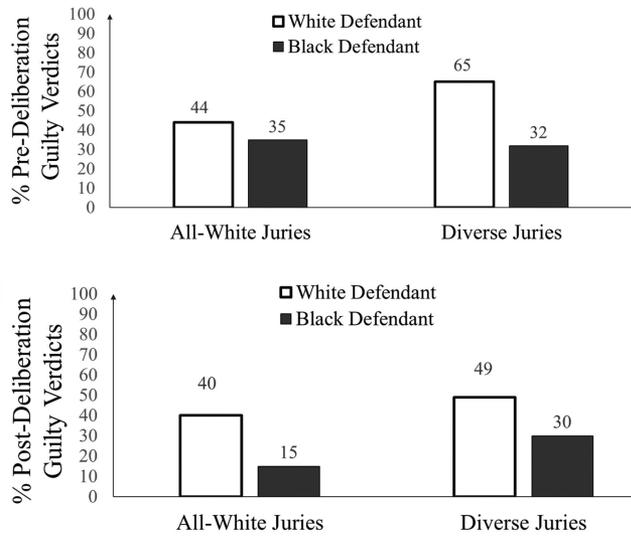


Figure 5. Increasing leniency hypothesis. Effects of jury composition and defendant race on pre-deliberation (above) and postdeliberation (below) % guilty verdicts. Exact % are reported above each bar.

higher number of correct and novel case facts when the defendant was White (vs. Black), indicating that they were more thoughtful when they deliberated to judge a White (vs. Black) defendant. In contrast, when White jurors interacted with Black juror confederates, the racial disparity in performance was eliminated. Contrary to prior research and a priori hypotheses, jurors were overall more lenient toward the Black (vs. White defendant) regardless of jury composition.

Increasing Depletion Hypothesis

Results from this deliberation study extended research on minimal dyadic interactions (e.g., Apfelbaum et al., 2008; Richeson & Trawalter, 2005; Zabel et al., 2015) to complex, realistic group interactions: White jurors who deliberated in racially diverse juries were more depleted than jurors who deliberated with White jurors. This could not be attributed to the *behavior* of Black jurors because the group diversity manipulation ensured that the comments and behavior of White and Black confederate jurors were identical. Thus, results demonstrate that the mere presence of Black group members results in depletion for White jurors.

Deliberation involved remembering case facts, weighing opposing interpretations of forensic evidence, arguing with those who disagreed, and making a final decision—with likely costs to jurors' cognitive and regulatory resources (Vohs et al., 2008). Yet confederates' race still affected postdeliberation depletion, demonstrating that the effect is not limited to laboratory settings where the only possible source of depletion is the interracial interaction.

Of note, the effect of jury diversity on depletion was not moderated by defendant race. On diverse juries, deliberating about a Black defendant with Black jurors could be particularly depleting, because it would (a) prompt people to self-regulate more for fear of saying something racist about the defendant and/or (b) increase anxiety about appearing racist (Plant & Devine, 2003; Vorauer, Hunter, Main, & Roy, 2000). In this study, however,

jurors on diverse juries were not more depleted when they judged a Black versus White defendant. One potential explanation comes from the fact that, overall, jurors were more lenient toward the Black (vs. White) defendant, even before deliberating. Thus, jurors in these conditions were going to predominantly argue for a not-guilty verdict, a verdict that favored the defendant—giving them little reason to feel anxious about appearing racist against the defendant.

An important reason for studying cognitive depletion effects in jury contexts is to investigate its relationship to jurors' ability to perform their duty—that is, to remember, discuss, and weigh the factual information presented during the case in order to reach a fair and reasoned verdict. In this study, postdeliberation depletion was not significantly related to jurors' memory during or after deliberation. A likely explanation for this lack of association between depletion and cognitive performance in general is the brief deliberation time—although jurors were depleted, they were probably able to overcome this by being motivated. In actual deliberations, which can last hours, the effects of depletion on performance might show no matter how motivated the jurors (a limitation of the mock jury paradigm). In future research, it would also be important to consider other mechanisms (e.g., boredom, negative mood, individual differences in need for cognition) that can also influence jurors' performance and mask depletion effects.

Increasing Performance Hypothesis

Performance-related hypotheses were, for the most part, supported. A consistent pattern of crossover interactions revealed that, on all-White juries, jurors' contributions to the deliberation were overall of lower quality (i.e., fewer factual statements, decreased accuracy, and fewer novel contributions) when the defendant was Black (vs. White). The disparity in performance was reduced on diverse juries, where, although performance scores were somewhat higher for the Black (vs. White) defendant, the difference was not statistically significant. These findings are partially in line with the watchdog hypothesis (Fleming et al., 2005)—jurors were careful to process information carefully in an effort to “watch out” for potential biases, but, in line with Sommers' (2006) aversive racism hypothesis, this watchdog effect only occurred when the defendant's minority status was made salient by the presence of Black jurors. Contrary to predictions, jury composition and defendant race had no effect on postdeliberation memory for case facts. All jurors scored very highly on the recognition test, perhaps because jurors had deliberated about the facts for an extended time shortly before taking it.

By including conditions in which the defendant was White, the study provided further support for Sommers' (2006) hypothesis that jury diversity activated White jurors' concern with racial prejudice when they judged a Black defendant. Specifically, jury diversity had the same effect in the Black defendant conditions as it did in Sommers' study (which only included a Black defendant): It improved the deliberation performance of White jurors. In contrast, diversity did not improve White jurors' performance when this potential concern with prejudice was absent—when jurors judged a White defendant. Of importance, the observed differences in deliberation quality were due, in the present study, only to the contributions of White jurors—confederates made the exact same informational contributions regardless of race. This approach allowed me to more directly test the legal system's assumption that

the benefits of diversity lie in the unique contributions of minority jurors and in the diversity of perspectives it provides (*Peters v. Kiff, 1972*). Sommers found this assumption problematic, because it places the burden on minority jurors to educate White jurors by arguing from a “minority perspective,” while at the same time racial minorities struggle with perceived (Bowers et al., 2001) and real (e.g., Salerno et al., 2017) challenges to their opinion in jury deliberation settings. As this study and Sommers’ original study suggest, minority jurors need not carry the entire burden of improving deliberations by providing unique contributions: Their presence seems to serve as a motivator for all jurors to perform their job more thoughtfully.

We strive for an unbiased justice system where corrective responses to long-standing racial disparities (i.e., increasing diversity in juries) do not result in unintended, detrimental effects for White defendants. Results from this study suggest jury diversity would have no such unintended consequences. As one reviewer helpfully noted, individual jurors were more likely to find the White (vs. Black) defendant guilty before deliberation, while confederates were predisposed toward acquittal. As a result, juries might have been further from unanimous agreement when the defendant was White (vs. Black), which should result in higher disagreement and more in-depth discussion of the evidence. Yet diverse juries discussed the evidence just as much when the defendant was Black, even though theoretically there was less disagreement among jurors at the onset of deliberation. Furthermore, when the defendant and victim were Black, diversity (vs. homogeneity) motivated jurors to discuss more case facts, to make fewer factual errors, and to discuss a broader array of evidence. When the defendant and victim were White, jurors made similar contributions regardless of jury composition. Thus, increasing the number of diverse juries would not have detrimental effects on the quality of deliberation for White defendants. This is another strong argument for the practical advantages of increasing jury diversity in general, but especially of ensuring minority defendants are judged by diverse juries.

In addition to these practical implications, the study’s findings contribute to further understanding the theoretical relationship between regulatory depletion and cognitive performance. Although interracial interactions depleted jurors after deliberation, they did not appear detrimental to jurors’ performance during deliberation. This aligns with recent models of self-regulatory depletion: Depletion effects do not necessarily result from insufficient resources to complete a task, but rather from reduced motivation to exert those resources (Inzlicht & Schmeichel, 2012). In this light, depletion is not necessarily a concern, as long as people are motivated to keep working hard. Jurors in this study exerted a significant amount of regulatory and cognitive effort, seemingly more so when they served on diverse than all-White juries. Their performance did not suffer likely because depletion effects did not set in until the second task (i.e., Stroop test) when their motivation was spent on a hard hour’s work of remembering evidence, weighing case facts, and reaching a difficult decision. Yet this explanation is speculative and should be tested empirically in the future.

Implications for Jurors’ Verdicts in Diverse Juries

In line with prior research describing a “leniency effect” (e.g., Diamond, 1997; MacCoun & Kerr, 1988), deliberation reduced the

number of guilty verdicts, likely because jurors are more easily swayed to acquit than to convict, given reasonable doubt standards. Group and individual verdicts favored the Black (vs. White) defendant before and after deliberation. Although these effects were not significantly moderated by jury composition, the pattern of conviction rates suggests that the main effect of defendant race predeliberation might have been primarily driven by the high conviction rate for the White defendant when White jurors anticipated deliberating on diverse juries (see Figure 5). This pattern is more in line with the black sheep effect, and less in line with the similarity-lenency effect. In fact, the largest discrepancy in conviction rates was observed before White jurors deliberated on diverse juries (65% for the White, vs. 32% for the Black defendant); this discrepancy was reduced after deliberation (49% and 30%, respectively). The postdeliberation results could potentially be explained by the fact that the Black confederates were not particularly adamant about convicting the defendant—in fact, one of them consistently favored acquittal—which might have assuaged some of the White jurors’ concerns with distancing themselves from the defendant by voting to convict.

But why would jurors be more lenient toward the Black defendant, especially if they served on all-White juries where race should not be salient (Sommers, 2006)? Racial biases against Black defendants in legal decisions are historically robust in actual trials and in psychological research (e.g., Sommers & Ellsworth, 2001; Sweeney & Haney, 1992). As it has become less socially desirable to express racial prejudice, however, mock jurors’ traditional anti-Black biases can disappear in cases where race is salient (i.e., racially charged cases) because they are afraid of appearing racist by convicting a Black defendant (Sommers & Ellsworth, 2009). But the trial used in this study was not racially salient in an explicit way (i.e., victim and defendant were of the same race, racial issues were not mentioned in the trial stimulus, the confederates did not bring up race).

Yet one could argue that all trials involving Black defendants or victims are, to some degree, racially salient in today’s broader social context. Racial disparities in the criminal justice system are at the forefront of public discourse, providing a background that heightened race salience outside of the narrower context of the trial. Some research suggests that media coverage of racial bias can subsequently reduce it (Pope, Price, & Wolfers, 2018). In the past 5 years in particular, videos and media coverage of Black men killed by police officers sparked a strong social movement and increased awareness of racial bias in the criminal justice system as pervasive and systemic. Even President Obama has spoken out about the issue, calling the criminal justice system “unfair” and openly discussing racial bias (Hudson, 2015).

Increased awareness of racial bias in the legal system might have made race chronically salient for jurors in this study, motivating them to overcorrect and reach more lenient verdicts for a Black defendant, in line with the watchdog hypothesis (Fleming et al., 2005). For example, an American Psychology–Law Conference symposium from 2012 included several studies with counterintuitive results such as leniency toward Black, compared to White, defendants (Salerno & Stevenson, 2012). Although race was not a salient issue in those studies, perhaps concerns with appearing racist are automatically activated whenever some mock jurors are faced with Black defendants in criminal trials, prompting these jurors to be lenient out of egalitarian concerns. Furthermore,

a recent meta-analysis found a similarity-leniency effect of White jurors favoring a White defendant only in comparison to Latino (but not Black) defendants (Devine & Caughlin, 2014), arguably due to people's decreased awareness of racial injustice for Latino defendants.

Thus, awareness of widespread racial bias might motivate jurors to give Black defendants the benefit of the doubt because race is chronically salient as a source of social and legal injustice. In support, measures of racial prejudice and motivation to control prejudice were not related to verdicts or to discussion of case facts during deliberation, indicating that perhaps all jurors (not just those who score high on measures of motivation to avoid prejudice) are particularly careful to avoid racial biases. In fact, research indicates that jurors' preexisting racist attitudes do not predict conviction for Black defendants when race is salient (Cohn, Buccolo, Pride, & Sommers, 2009).

A less optimistic, yet more persuasive complementary explanation is that counterintuitive race effects might be an artifact of experimental research. The findings presented at the 2012 AP-LS Symposium relied on mock-jury studies, where community members or students were aware they were part of a study. According to Hunt (2015), some experimental research, especially in the past decade, fails to find robust racial bias effects—yet archival research of capital cases indicates that White jurors' racial bias against Black defendants is alive and well in actual trials where jurors are less likely to anticipate that their decisions might be scrutinized or investigated in relation to race—in line with aversive racism theory. To continue to study racial bias in ways that are directly informative for the criminal justice system, researchers will likely have to find novel ways to alleviate jurors' concerns with appearing racist—for example, by licensing them to feel egalitarian in a pretrial task (e.g., endorsing a Black political candidate or job applicant, Effron, Cameron, & Monin, 2009).

Contributions, Strengths, and Limitations

Research that increases understanding of the interpersonal and cognitive processes at the heart of jury deliberations can inform courts, policymakers, and the public about the psychological effects of deliberation on jurors and juries, explaining verdicts and deliberation processes and informing court practices. For example, there has been research on depletion mechanisms during dyadic interracial interactions, yet this is the first investigation of the potential depleting effects of these interactions on *group* decision making. Although in this study postdeliberation depletion was unrelated to jurors' performance during deliberation, it is important to address this potential concern in the future. Results can be used to influence policy recommendations about how to maximize the benefits of diversity by reducing depletion—particularly in cases where deliberations last for hours or even days.

The study had several methodological strengths. The community sample was diverse in terms of age and background, and more representative of the jury pool than a student sample—although this sample was overall rather highly educated. Study materials were based on actual trial evidence and jury instructions and reviewed by a legal professional to ensure realism and increase ecological validity. Jurors deliberated to reach a unanimous verdict, under conditions similar to those encountered by real jurors.

In addition to steps taken to increase ecological validity, the study also involved considerable experimental control. Confederates spent several months training in pilot sessions to cover potential situations that could (and did) arise during the actual study. They were therefore able to “stick” to the script lines in every session. This procedure ensured observed effects could be confidently attributed to confederates' race alone. The defendant race manipulation was also successful, with very few manipulation failures. Participants were randomly assigned to conditions, and sessions took place at the same time of day.

The study also had several limitations. First, the trial evidence was ambiguous to ensure variability in verdicts and to keep jurors deliberating at length—therefore one verdict was not more accurate than another. Leniency toward the Black defendant does not necessarily mean jurors were more accurate. Future studies could manipulate evidence strength directly, in addition to defendant race and jury composition. Juries would also have a true option to reach a unanimous guilty verdict, which was not the case in this study due to confederates' script. In fact, had confederates sided with a guilty verdict as part of their script, it is possible that White jurors would have felt licensed to convict the Black defendant at higher rates.

Second, to avoid any racially salient information as part of the trial, the victim's race was manipulated along with the defendant's race—and it is possible that it actually drove jurors' verdicts. Black (vs. White) victims are perceived less favorably (Harrison & Esqueda, 1999; Murray & Stahly, 1987) and elicit less punitive verdicts for defendants (Baldus & Woodworth, 2003; Mazzella & Feingold, 1994). Apart from verdict decisions, the alternative explanation does not change the theoretical and practical implications of this study: Jurors are sensitive to the racial composition of their deliberating group, especially when the decision is relevant to racial minority targets.

Third, jurors were not allowed to deliberate for more than 40 min and only engaged in one deliberating session. This practical consideration reduced the ecological validity of the study especially when it comes to verdicts—but also precluded me from investigating cognitive depletion in more demanding circumstances, such as deliberations that last hours or even days, with increased pressure to reach unanimous verdicts. It would be valuable to know whether racial differences in verdicts are reduced or exacerbated after prolonged jury deliberations, yet such research is difficult to conduct experimentally. In addition, most jury groups did not reach a unanimous decision—which was in part due to the decision to use confederates who were restricted to specific verdict options (i.e., one confederate always voted to acquit). Thus, a limitation to the present study is that it did not involve true deliberation and it restricted the group-level verdict options in ways actual jurors are not.

Fourth, because my purpose was to test the effect of Black jurors' mere presence on White jurors' performance and depletion, confederates were trained to be neutral and contribute little—which differentiates their behavior from that of actual Black jurors in meaningful ways. Actual Black jurors, for example, might contribute more factual comments, argue their points more adamantly, and mention the defendant's race explicitly (as they did in Sommers' study). In addition to extending work on the performance of diverse and all-White juries during deliberation, the study was also specifically intended to isolate the effect of Black

jurors' mere presence on the jury from the informational contributions they would make during deliberation. We already know, from Sommers' (2006) seminal study, that Black jurors make substantially informative contributions in general and can bring up race-related issues in particular. The present study further illustrates that Black jurors' presence can have similar performance effects for White jurors simply by motivating them to "work harder" without actually providing informational support. Although this distinction is of less practical importance, it makes theoretical contributions to our understanding of diverse group performance. It would be interesting to continue to isolate diversity of opinion/perspective from demographic diversity, and to test how they influence group deliberations both separately and jointly.

Given the complexity of the paradigm and difficulty collecting data from community participants, it would not have been feasible to completely cross the race of participants in addition to defendant race and jury composition (i.e., racially homogeneous Black or White juries vs. diverse juries with Black or White majorities). Yet this remains the next important step in studying jury diversity effects on the quality of deliberations and on jurors' verdicts, along with focusing on other racial and ethnic minorities in danger of unequal legal treatment.

Conclusion

The present study increases understanding of the implications of creating diverse juries by testing two competing models of group interactions: one that draws on the detrimental effects of diversity on performance through regulatory depletion, and one that highlights the motivating effect of diversity on jurors' performance when race is salient. Although interracial interactions are depleting, this depletion does not reduce the quality of deliberation or of the verdict. Furthermore, creating diverse juries does not have a negative impact on outcomes for White defendants. Thus, racial diversity in juries not only ensures representation of minority voices, but it also motivates all jurors to perform their duty diligently and thoughtfully regardless of defendant race.

References

- Antonio, A. L., Chang, M. J., Hakuta, K., Kenny, D. A., Levin, S., & Milem, J. F. (2004). Effects of racial diversity on complex thinking in college students. *Psychological Science, 15*, 507–510. <http://dx.doi.org/10.1111/j.0956-7976.2004.00710.x>
- 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. <http://dx.doi.org/10.1037/a0011990>
- Baldus, D. C., & Woodworth, G. (2003). Race discrimination and the legitimacy of capital punishment: Reflections on the interaction of fact and perception. *De Paul Law Review, 53*, 1411–1496.
- Baldus, D. C., Woodworth, G., Zuckerman, D., & Weiner, N. A. (1997). 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.
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*, 351–355. <http://dx.doi.org/10.1111/j.1467-8721.2007.00534.x>
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review, 94*(4), 991–1013.
- Bottoms, B. L., Davis, S. L., & Epstein, M. A. (2004). Effects of victim and defendant race on jurors' decisions in child sexual abuse cases. *Journal of Applied Social Psychology, 34*, 1–33. <http://dx.doi.org/10.1111/j.1559-1816.2004.tb02535.x>
- Bowers, W. J., Steiner, B. D., & Sandys, M. (2001). Death sentencing in black and white: An empirical analysis of the role of jurors' race and jury racial composition. *University of Pennsylvania Journal of Constitutional Law, 3*, 171–274.
- Browne, W. (2006, July). *Sample size calculations in multilevel modelling*. Paper presented at ESRC Research Methods Festival, Oxford, UK. Retrieved from <https://seis.bristol.ac.uk/~frwjb/materials/ssize.ppt>
- Burkley, E., Anderson, D., & Curtis, J. (2011). You wore me down: Self-control strength and social influence. *Social and Personality Psychology Compass, 5*, 487–499. <http://dx.doi.org/10.1111/j.1751-9004.2011.00367.x>
- 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*, 1953–1973. <http://dx.doi.org/10.1111/j.1559-1816.2009.00511.x>
- Cowan, C. L., Thompson, W. C., & Ellsworth, P. C. (1984). The effects of death qualification on jurors' predisposition to convict and on the quality of deliberation. *Law and Human Behavior, 8*, 53–79. <http://dx.doi.org/10.1007/BF01044351>
- Cox, T. H., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *The Executive, 5*, 45–56.
- De Dreu, C. K., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology, 88*, 741–749. <http://dx.doi.org/10.1037/0021-9010.88.4.741>
- DeGrassi, S. W., Morgan, W. B., Walker, S. S., Wang, Y., & Sabat, I. (2012). Ethical decision-making: Group diversity holds the key. *Journal of Leadership, Accountability and Ethics, 9*, 51–65.
- 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*, 109–134. <http://dx.doi.org/10.1037/law0000006>
- Devine, D. J., Krouse, P. C., Cavanaugh, C. M., & Basora, J. C. (2016). Evidentiary, extra-evidentiary, and deliberation process predictors of real jury verdicts. *Law and Human Behavior, 40*, 670–682. <http://dx.doi.org/10.1037/lhb0000209>
- Diamond, S. S. (1997). Illuminations and shadows from jury simulations. *Law and Human Behavior, 21*, 561–571. <http://dx.doi.org/10.1023/A:1024831908377>
- Dovidio, J. F., & Gaertner, S. L. (1998). *Aversive racism*. San Diego, CA: Elsevier Academic Press.
- Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in Experimental Social Psychology, 36*, 1–52. [http://dx.doi.org/10.1016/S0065-2601\(04\)36001-6](http://dx.doi.org/10.1016/S0065-2601(04)36001-6)
- Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 23*, 316–326. <http://dx.doi.org/10.1177/0146167297233009>
- Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring Whites. *Journal of Experimental Social Psychology, 45*, 590–593. <http://dx.doi.org/10.1016/j.jesp.2009.02.001>
- Finkel, E. J., Campbell, W. K., Brunell, A. B., Dalton, A. N., Scarbeck, S. J., & Chartrand, T. L. (2006). High-maintenance interaction: Inefficient social coordination impairs self-regulation. *Journal of Personality and Social Psychology, 91*, 456–475. <http://dx.doi.org/10.1037/0022-3514.91.3.456>
- Fleming, M. A., Petty, R. E., & White, P. H. (2005). Stigmatized targets and evaluation: Prejudice as a determinant of attribute scrutiny and polarization. *Personality and Social Psychology Bulletin, 31*, 496–507. <http://dx.doi.org/10.1177/0146167204271585>

- Foster v. Chatman, 578 U.S. (2016).
- Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, *136*, 495–525. <http://dx.doi.org/10.1037/a0019486>
- Harrison, L. A., & Esqueda, C. W. (1999). Myths and stereotypes of actors involved in domestic violence. *Aggression and Violent Behavior*, *4*, 129–138. [http://dx.doi.org/10.1016/S1359-1789\(97\)00026-8](http://dx.doi.org/10.1016/S1359-1789(97)00026-8)
- Harvey, S. (2013). A different perspective: The multiple effects of deep level diversity on group creativity. *Journal of Experimental Social Psychology*, *49*, 822–832. <http://dx.doi.org/10.1016/j.jesp.2013.04.004>
- Horowitz, I. A., & Bordens, K. S. (2002). The effects of jury size, evidence complexity, and note taking on jury process and performance in a civil trial. *Journal of Applied Psychology*, *87*, 121–130. <http://dx.doi.org/10.1037/0021-9010.87.1.121>
- Hudson, D. (2015, July 15). President Obama: “Our criminal justice system isn’t as smart as it should be.” *White House*. Retrieved from <https://www.whitehouse.gov/blog/2015/07/15/president-obama-our-criminal-justice-system-isnt-smart-it-should-be>
- 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. <http://dx.doi.org/10.1111/j.0956-7976.2004.00680.x>
- Hunt, J. S. (2015). Race, ethnicity, and culture in jury decision making. *Annual Review of Law and Social Science*, *11*, 269–288. <http://dx.doi.org/10.1146/annurev-lawsocsci-120814-121723>
- Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion: How being the target of prejudice affects self-control. *Psychological Science*, *17*, 262–269. <http://dx.doi.org/10.1111/j.1467-9280.2006.01695.x>
- Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, *7*, 450–463. <http://dx.doi.org/10.1177/1745691612454134>
- Jehn, K. A., & Bezrukova, K. (2004). A field study of group diversity, workgroup context, and performance. *Journal of Organizational Behavior*, *25*, 703–729. <http://dx.doi.org/10.1002/job.257>
- Johnson, J. D., Whitestone, E., Jackson, L. A., & Gatto, L. (1995). Justice is still not colorblind: Differential racial effects of exposure to inadmissible evidence. *Personality and Social Psychology Bulletin*, *21*, 893–898. <http://dx.doi.org/10.1177/0146167295219003>
- Joshi, A., & Roh, H. (2009). The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal*, *52*, 599–627. <http://dx.doi.org/10.5465/amj.2009.41331491>
- Kerr, N. L., Hymes, R. W., Anderson, A. B., & Weathers, J. E. (1995). Defendant-juror similarity and mock juror judgments. *Law and Human Behavior*, *19*, 545–567. <http://dx.doi.org/10.1007/BF01499374>
- Lynch, M., & Haney, C. (2009). Capital jury deliberation: Effects on death sentencing, comprehension, and discrimination. *Law and Human Behavior*, *33*, 481–496. <http://dx.doi.org/10.1007/s10979-008-9168-2>
- MacCoun, R. J., & Kerr, N. L. (1988). Asymmetric influence in mock jury deliberation: Jurors’ bias for leniency. *Journal of Personality and Social Psychology*, *54*, 21–33. <http://dx.doi.org/10.1037/0022-3514.54.1.21>
- Marques, J. M., Yzerbyt, V. Y., & Leyens, J. P. (1988). The “black sheep effect”: Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, *18*, 1–16. <http://dx.doi.org/10.1002/ejsp.2420180102>
- Masicampo, E. J., Martin, S. R., & Anderson, R. A. (2014). Understanding and overcoming self-control depletion. *Social and Personality Psychology Compass*, *8*, 638–649. <http://dx.doi.org/10.1111/spc3.12139>
- Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A Meta-Analysis I. *Journal of Applied Social Psychology*, *24*, 1315–1338. <http://dx.doi.org/10.1111/j.1559-1816.1994.tb01552.x>
- McLeod, P. L., Lobel, S. A., & Cox, T. H., Jr. (1996). Ethnic diversity and creativity in small groups. *Small Group Research*, *27*, 248–264. <http://dx.doi.org/10.1177/1046496496272003>
- Miller, D. A., Smith, E. R., & Mackie, D. M. (2004). Effects of intergroup contact and political predispositions on prejudice: Role of intergroup emotions. *Group Processes & Intergroup Relations*, *7*, 221–237. <http://dx.doi.org/10.1177/1368430204046109>
- Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior*, *29*, 621–637. <http://dx.doi.org/10.1007/s10979-005-8122-9>
- Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, *74*, 774–789. <http://dx.doi.org/10.1037/0022-3514.74.3.774>
- Murphy, M. C., Richeson, J. A., Shelton, J. N., Rheinschmidt, M. L., & Bergsieker, H. B. (2013). Cognitive costs of contemporary prejudice. *Group Processes & Intergroup Relations*, *16*, 560–571. <http://dx.doi.org/10.1177/1368430212468170>
- Murray, C. B., & Stahly, G. B. (1987). Some victims are derogated more than others. *The Western Journal of Black Studies*, *11*, 177–180.
- Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the black box: An analysis of work group diversity, conflict and performance. *Administrative Science Quarterly*, *44*, 1–28. <http://dx.doi.org/10.2307/2667029>
- Perez, D. A., Hosch, H. M., Ponder, B., & Trejo, G. C. (1993). Ethnicity of defendants and jurors as influences on jury decisions. *Journal of Applied Social Psychology*, *23*, 1249–1262. <http://dx.doi.org/10.1111/j.1559-1816.1993.tb01031.x>
- Peters v. Kiff, 407 U.S. 493 (1972).
- Plant, E. A., & Devine, P. G. (2003). The antecedents and implications of interracial anxiety. *Personality and Social Psychology Bulletin*, *29*, 790–801. <http://dx.doi.org/10.1177/0146167203029006011>
- Pope, D. G., Price, J., & Wolfers, J. (2018). Awareness reduces racial bias. *Management Science*, *64*(11), 4988–4995.
- Renaud, T. (2010). The biggest bully in the room. *The Jury Expert*, *22*, 23–26.
- Richards, J. M., & Gross, J. J. (2000). Emotion regulation and memory: The cognitive costs of keeping one’s cool. *Journal of Personality and Social Psychology*, *79*, 410–424. <http://dx.doi.org/10.1037/0022-3514.79.3.410>
- Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*, 287–290. <http://dx.doi.org/10.1111/1467-9280.03437>
- Richeson, J. A., & Trawalter, S. (2005). Why do interracial interactions impair executive function? A resource depletion account. *Journal of Personality and Social Psychology*, *88*, 934–947. <http://dx.doi.org/10.1037/0022-3514.88.6.934>
- Rose, M. R., & Diamond, S. S. (2008). Judging bias: Juror confidence and judicial rulings on challenges for cause. *Law & Society Review*, *42*, 513–549. <http://dx.doi.org/10.1111/j.1540-5893.2008.00350.x>
- Rosselli, M., Ardila, A., Santisi, M. N., Arecco, M. D. R., Salvatierra, J., Conde, A., & Lenis, B. (2002). Stroop effect in Spanish–English bilinguals. *Journal of the International Neuropsychological Society*, *8*(6), 819–827.
- Ruva, C. L., & Guenther, C. C. (2017). Keep your bias to yourself: How deliberating with differently biased others affects mock-jurors’ guilt decisions, perceptions of the defendant, memories, and evidence interpretation. *Law and Human Behavior*, *41*, 478–493. <http://dx.doi.org/10.1037/lhb0000256>
- R v. Valevski, 445 NSW Ct. Crim. App. Australia (2000).
- Salerno, J. M., & Peter-Hagene, L. C. (2013). The interactive effect of anger and disgust on moral outrage and judgments. *Psychological Science*, *24*, 2069–2078. <http://dx.doi.org/10.1177/0956797613486988>

- Salerno, J. M., & Peter-Hagene, L. C. (2015). One angry woman: Anger expression increases influence for men, but decreases influence for women, during group deliberation. *Law and Human Behavior, 39*, 581–592. <http://dx.doi.org/10.1037/lhb0000147>
- Salerno, J. M., Peter-Hagene, L. C., & Jay, A. C. (2017). Women and African Americans are less influential when they express anger during group decision making. *Group Processes & Intergroup Relations, 22*, 57–79.
- Salerno, J. M., & Stevenson, M. C. (2012, March). *Counterintuitive race effects in legal decision making: Unexpected leniency toward African American defendants*. Symposium presented at the American Psychology-Law Society, Portland, OR.
- Sargent, M. J., & Bradfield, A. L. (2004). Race and information processing in criminal trials: Does the defendant's race affect how the facts are evaluated? *Personality and Social Psychology Bulletin, 30*, 995–1008. <http://dx.doi.org/10.1177/0146167204265741>
- Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology, 85*, 33–46. <http://dx.doi.org/10.1037/0022-3514.85.1.33>
- Snijders, T. A. (2005). Power and sample size in multilevel linear models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1570–1573). Chichester, U.K.: Wiley.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage.
- 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*, 597–612. <http://dx.doi.org/10.1037/0022-3514.90.4.597>
- Sommers, S. R., & Ellsworth, P. C. (2000). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin, 26*, 1367–1379. <http://dx.doi.org/10.1177/0146167200263005>
- 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*, 201–229. <http://dx.doi.org/10.1037/1076-8971.7.1.201>
- Sommers, S. R., & Ellsworth, P. C. (2009). "Race salience" in juror decision-making: Misconceptions, clarifications, and unanswered questions. *Behavioral Sciences & the Law, 27*, 599–609. <http://dx.doi.org/10.1002/bsl.877>
- Sommers, S. R., & Norton, M. I. (2007). Race-based judgments, race-neutral justifications: Experimental examination of peremptory use and the Batson challenge procedure. *Law and Human Behavior, 31*, 261–273. <http://dx.doi.org/10.1007/s10979-006-9048-6>
- Stevenson, M. C., Lytle, B. L., Baumholser, B. J., & McCracken, E. W. (2017). Racially diverse juries promote self-monitoring efforts during jury deliberation. *Translational Issues in Psychological Science, 3*, 187–201. <http://dx.doi.org/10.1037/tps0000113>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662. <http://dx.doi.org/10.1037/h0054651>
- Sweeney, A., & Dizikes, C. (2013, March 27). The balancing act of jury selection. *Chicago Tribune*. Retrieved from http://articles.chicagotribune.com/2013-03-27/news/ct-met-juries-racial-mix-20130327_1_jury-selection-prospective-jurors-new-jury
- 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.
- Tajfel, H., & Turner, J. (1986). The social identity theory of intergroup behaviour. In S. Worchel & I. W. G. Austin (Eds.), *Psychology of intergroup relations*. Chicago, IL: Nelson Hall.
- Timmerman, T. A. (2000). Racial diversity, age diversity, interdependence, and team performance. *Small Group Research, 31*, 592–606. <http://dx.doi.org/10.1177/104649640003100505>
- Triana, M. C., Porter, C. O. L. H., DeGrassi, S. W., & Bergman, M. (2013). We're all in this together . . . except for you: The effects of workload, performance feedback, and racial distance on helping behavior in teams. *Journal of Organizational Behavior, 34*, 1124–1144. <http://dx.doi.org/10.1002/job.1842>
- Tuk, M. A., Zhang, K., & Sweldens, S. (2015). The propagation of self-control: Self-control in one domain simultaneously improves self-control in other domains. *Journal of Experimental Psychology: General, 144*, 639–654. <http://dx.doi.org/10.1037/xge0000065>
- Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology, 94*, 883–898. <http://dx.doi.org/10.1037/0022-3514.94.5.883>
- Vorauer, J. D. (2003). Dominant group members in intergroup interaction: Safety or vulnerability in numbers? *Personality and Social Psychology Bulletin, 29*, 498–511. <http://dx.doi.org/10.1177/0146167202250917>
- Vorauer, J. D., Hunter, A. J., Main, K. J., & Roy, S. A. (2000). Meta-stereotype activation: Evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction. *Journal of Personality and Social Psychology, 78*, 690–707. <http://dx.doi.org/10.1037/0022-3514.78.4.690>
- Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal, 36*, 590–602.
- Webb, T. L., & Sheeran, P. (2003). Can implementation intentions help to overcome ego-depletion? *Journal of Experimental Social Psychology, 39*, 5–13. [http://dx.doi.org/10.1016/S0022-1031\(02\)00527-9](http://dx.doi.org/10.1016/S0022-1031(02)00527-9)
- Zabel, K. L., Olson, M. A., Johnson, C. S., & Phillips, J. E. (2015). What we talk about matters: content moderates cognitive depletion in interracial interactions. *The Journal of Social Psychology, 155*, 545–552. <http://dx.doi.org/10.1080/00224545.2015.1032197>

Received August 7, 2018

Revision received April 6, 2019

Accepted April 15, 2019 ■