DOI: 10.1111/lapo.12181

ORIGINAL ARTICLE

Law & Policy University of Denver WILEY

# The subtle effects of implicit bias instructions

**Mona Lynch**[1] | **Taylor Kidd**[1] | **Emily Shaw**[2]

[1]University of California, Irvine, Criminology, Law & Society, 2340 Social Ecology II, Irvine, California, USA

[2]University of California, Irvine, Psychological Science, 4208 Social & Behavioral Sciences Gateway, Irvine, California, USA

**Correspondence**
Mona Lynch, University of California, Irvine, Criminology, Law & Society, 2340 Social Ecology II, Irvine, CA 92697-7080, USA.
Email: lynchm@uci.edu

**Funding information**
National Institute of Justice, Office of Justice Programs, U.S. Department of Justice, Grant/Award Number: 2017-IJ-CX-0044

**Abstract**
Judges are increasingly using "implicit bias" instructions in jury trials in an effort to reduce the influence of jurors' biases on judgment. In this article, we report on findings from a large-scale mock jury study that tests the impact of implicit bias instructions on judgment in a case where defendant race was varied (Black or White). Using an experimental design, we collected and analyzed quantitative and qualitative data at the individual and group levels obtained from 120 small groups who viewed a simulated federal drug conspiracy trial and then deliberated to determine a verdict. We find that while participants were sensitized to the importance of being unbiased, implicit bias instructions had no measurable impact on verdict outcomes relative to the standard instructions. Our analysis of the deliberations, however, reveals that those who heard the implicit bias instructions were more likely to discuss the issue of bias, potentially with both ameliorative and harmful effects on the defendant. Most significantly, we identified multiple instances where, in an effort to avoid bias, participants who heard the implicit bias instructions interfered with their own or other participants' appropriate assessments of witness credibility.

## 1 | INTRODUCTION

We have experienced a dramatic shift in recent years in how racial and other biases are conceptualized and understood by scholars and the general public alike. Most notably, implicit biases—"attitudes or stereotypes that affect our understanding, decision making, and behavior, without our even realizing it" (Kang et al., 2012, p. 1126)—are now widely understood to be a driving force behind social inequality and discrimination. A diverse array of organizations and institutions have become attuned to the problem of implicit bias and its impact on members of underrepresented racial and ethnic groups, including in education (Gullo et al., 2018; Jackson et al., 2014), private industry (Bertrand & Mullainathan, 2004), public agencies (Beniwal, 2016;

Foley & Williamson, 2019), health systems (FitzGerald & Hurst, 2017; Zestcott et al., 2016), and legal institutions (Kang et al., 2012; Negowetti, 2014). These entities have instituted various programs and interventions to mitigate the potential impact of implicit bias (Burns et al., 2017; Casey et al., 2013; Lai et al., 2014).

In the legal context, one such intervention is the delivery of "implicit bias instructions" (Su, 2020, p. 90) by judges in jury trials. Because judicial instructions that explain the relevant substantive law, as well as the procedural rules for decision making, are a universal component of jury trials, implicit bias instructions have been recommended by scholars, judges, and attorneys as a feasible and straightforward method of educating jurors and promoting unbiased deliberations (Elek & Hannaford-Agor, 2014; Su, 2020). Indeed, some argue that pre-deliberation interventions such as these kinds of instructions offer a promising way of discouraging jurors from relying on inherent biases, without placing undue burden on the court system (West, 2011).

Implicit bias instructions vary in their specific wording, but they essentially expand upon the standard cautionary instruction (Tanford, 1990) that directs jurors to be free from bias when arriving at a decision. For example, in 2017, the federal courts in the Western District of Washington adopted pattern jury instructions that define and describe forms of "unconscious" bias and that direct jurors to arrive at a verdict in a manner that relies solely on the facts of the case and is free of prejudice and unconscious bias (Doyle, 2017). State courts have also adopted versions of implicit bias instructions, including in Texas (Wise, 2020), Massachusetts (Olson, 2019), Illinois (Illinois Civil Jury Instructions, 2019), and California (Judicial Council of California, 2020). While these instructions typically specify a range of potential biases that jurors are to guard against—including on the basis of race, national origin, age, religion, gender identity, or sexual orientation—a primary motivation of this intervention has been racial bias, especially against Black people (see, e.g., Doyle, 2017).

Implicit bias instructions are delivered to seated juries prior to their deliberations as part of the general instructions to the jury (sometimes at the start of a trial, sometimes at the conclusion). They ostensibly serve several purposes. First, they are meant to be educative. The typical implicit bias instruction defines implicit bias for the jury members in order to ensure that they are aware of the concept (Bennett, 2010). Second, they aim to transform what is implicit or unconscious cognition into something recognizable to the jurors in the hope that greater awareness of implicit bias will improve jurors' ability to recognize such bias in themselves (Bennett, 2010). Finally, they aim to steer jurors away from relying on implicit biases when considering evidence and making judgments, just as they are directed not to rely on explicit biases or prejudices (Kang et al., 2012).

Yet, while implicit bias instructions are predicated upon and justified by social science research on cognitive bias, very little empirical work has examined how such instructions shape the judgment process. So even though changes to jury instructions that promote bias awareness sound commonsensical and harmless, the social-psychological phenomena underpinning this intervention are more complex and have the potential to produce unanticipated effects in the group decision-making process. Moreover, in regard to jurors' anti-Black racial bias, the underlying causal forces appear to be more multifaceted than simply individual cognition (Lynch & Haney, 2011), thereby limiting the potential ameliorative impact of instructions aimed at changing cognitive processes.

In this article, we report on findings from a large-scale mock jury study that tests the impact of implicit bias instructions on case outcomes and explores how such instructions shape deliberations. Using a $2 \times 2 \times 2$ factorial design, where the race of the defendant (Black or White), the race of the key witness (Black or White), and the type of instructions (standard or implicit bias) were manipulated, we test whether our participants exhibited anti-Black bias in their judgments and whether the implicit bias instructions impacted any demonstrated bias. We collected quantitative and qualitative data at the individual and group levels obtained from 120 small

group "juries" (each containing four to seven jurors) who viewed a simulated federal drug conspiracy trial and then deliberated to determine a verdict.

In the next section, we review current scholarship on implicit biases, including what empirical research indicates about bias in jury decision making, how jury instructions mediate bias, and the nascent research on implicit bias instructions. Following that review, we describe the present study's design and procedures, and then we detail our findings from both our quantitative analyses and our qualitative analysis of the deliberations. While we obtained no measurable effect of the implicit bias instructions on verdict outcomes, our quantitative and qualitative findings indicate that these instructions did increase recognition of the general problem of bias in judgment. Jurors in the implicit bias instructions condition were more likely to discuss bias, and they did so across three general themes: avoiding bias as an ethic of decision making; recognizing jurors' own, and each other's, biases; and bias in the assessment of witnesses. Most significantly, we identified multiple instances where, in an effort to avoid bias, participants who heard the implicit bias instructions interfered with their own or other participants' appropriate assessments of witness credibility. We conclude by considering the implications of these findings for the goal of mitigating jurors' racial bias against defendants, including whether the versions of implicit bias instructions being used could be tailored to better historicize and contextualize how racism produces inequality with respect to defendants of color.

## 1.1 | The psychology of bias and jury judgment

While there are both complementary and competing theories explaining how racial inequality is produced in the legal system, ranging from purely psychological explanations to social-structural ones (Lynch, 2016), the literature on racially discriminatory jury decision making has primarily been framed by psychological theories, including but not limited to implicit bias theories (e.g., Elek & Hannaford-Agor, 2014; Hunt, 2015; Levinson et al., 2014; Lynch & Haney, 2011; Sommers & Ellsworth, 2001). The general basis for understanding why jurors, as individuals, and juries, as small groups, may exhibit racial bias comes from the field of social cognition. In that regard, psychological research has established that the human cognition process relies on categories, or schemas, to allow us make sense of our surroundings in an efficient manner (Fiske, 2000; Wilder, 1986). Although the categorization process is fundamental to cognition, categories may be created that contain inaccurate beliefs or information about members of a group (i.e., stereotypes) (Fiske, 2000; Levinson et al., 2014), which are classified as either explicit (conscious) or implicit (unconscious) (Gawronski, 2019; Kang et al., 2012).

Contemporary research suggests that while explicit racial stereotyping and prejudice have not disappeared, implicit biases, operating outside of conscious control in human social judgment, are more prevalent in modern society (Dovidio et al., 2000; Greenwald et al., 2009; Nosek et al., 2007). Implicit racial biases are commonly assessed using the implicit association test (IAT), which measures an individual's reaction time in a sorting task with two categories in order to uncover unconscious stereotypes and biases that a person may hold (Greenwald et al., 2003, 2009; Kang et al., 2012). While a body of work critiques the IAT as a measurement instrument, suggesting that it does not necessarily measure implicit bias (see Gawronski, 2019), proponents of the IAT tool argue that it is able to uncover biases that better correlate with discriminatory behavior, relative to measures of explicit bias (Greenwald et al., 2009; Kang et al., 2012).

The debates over the assessment of implicit bias by the IAT sometimes conflate the measurement issue of what is said to be "implicit" with the actual cognitive constructs of bias and racism. Recent work suggests a more fluid relationship between what have been characterized as "implicit" and "explicit" biases (Gawronski, 2019) and investigates their relative power and directionality in predicting behavior under different conditions (Axt et al., 2016). For instance,

"subtle racism" (forms of which include modern racism and aversive racism) represents a category of bias that may be consciously held but is neither as motivational nor as affective in nature as traditional forms of explicit racial prejudice (Fish & Syed, 2020; Sue et al., 2007). Moreover, such bias may be either activated or inhibited, depending on situational and structural conditions. For instance, the theory of aversive racism suggests that people may harbor implicit or subtle biases but will not act in a biased manner unless they can mask any biased intent by, for instance, pointing to a nondiscriminatory reason for their actions (Dovidio, 2001). Similarly, the racial salience of a given judgment context may inhibit biased action by putting people on notice about the potential for bias (Sommers & Ellsworth, 2001).

These forms of racial bias—both implicit and subtle—appear to interact with situational factors such that their strength and influence on decision making is at least partly the product of the decision-making context (P. G. Devine, 1989; Kawakami et al., 1998; Payne et al., 2017). To that end, jurors do not appear to be immune to the influence of racial bias in their consideration of evidence and in determining verdicts (Mitchell et al., 2005; Sommers & Ellsworth, 2001; West, 2011; for reviews, see Bell & Lynch, 2016; Hunt, 2015). Racial biases can influence jurors' perceptions of case evidence, shaping the way they construct a narrative of a trial (Levinson et al., 2014; West, 2011). Jurors may also rely on race-consistent stereotypes about defendants (Jones & Kaplan, 2003; Smalarz et al., 2018) when making determinations about guilt.

Moreover, while most implicit bias research has focused on how bias shapes judgment at the individual level, there is reason to expect that group decision making, such as the kind that juries make, may differ in the way in which bias is activated and how it influences judgment (Bell & Lynch, 2016; Lynch & Haney, 2015; Sommers, 2006). Thus, racial bias has the potential to become amplified through the group-level process of deliberation (MacCoun, 1990). Recent empirical research suggests that in an emotionally charged capital penalty phase decision-making context, racial bias may move from being a subtle, even implicit individual-level influence into an explicit group-level force in jury decision making through such cues as individual invocations of stereotypes during deliberations (Lynch & Haney, 2015). Empirical research on guilt determinations by juries, however, suggests that deliberations can sometimes attenuate bias (Miller et al., 2011).

The jury deliberation process is also shaped by the demographic make-up of the group members, including the foreperson. White people, men, and more highly educated jurors are disproportionately likely to become forepersons (D. J. Devine, 2012). Forepersons, in turn, can be influential in shaping deliberations and outcomes (D. J. Devine et al., 2007; Diamond & Casper, 1992; York & Cornwell, 2006). In addition, those in lower-status social groups sometimes subjectively experience marginalization on juries (Hickerson & Gastil, 2008; Winter & Clair, 2018). For example, Winter and Clair (2018) found that Black and Hispanic jurors with lower levels of education felt less positive about their experiences and felt that they had fewer opportunities to express their thoughts and influence deliberations relative to White jurors of all educational backgrounds. Nevertheless, diverse jury groups are found to engage in higher-quality, less biased deliberations (Peter-Hagene, 2019; Sommers, 2006).

## 1.2 | The role of instructions in jury decision making

Jury instructions, and jurors' ability to understand and apply them, have been shown to attenuate the effect of racial bias on judgment in some instances. In their ideal form, jury instructions should decrease biased outcomes because they moderate the contextual conditions that make bias especially influential. For instance, psychological research suggests that in ambiguous decision-making contexts (Rector et al., 1993) or when the decision-making task is cognitively taxing (Bodenhausen & Lichtenstein, 1987), people are more likely to rely on stereotypes,

heuristics, and biases in their judgments. Therefore, to the extent that they are comprehensible, jury instructions should improve those conditions by providing clarity about decision-making rules and by helping decision makers make sense of complex and/or conflicting information offered as evidence.

The use of jury instructions that outline key legal concepts, principles, and decision-making procedures can reduce racially biased verdicts by individual mock jurors, relative to those who receive no instructions (Pfeifer & Ogloff, 1991). Comprehension of jury instructions has also been associated with reduced discrimination against Black defendants by individual mock jurors in a capital case setting (Lynch & Haney, 2000), but the ameliorative effect of individual comprehension appears to be somewhat diluted by the deliberation process (Lynch & Haney, 2009). On the other hand, some forms of jury instructions, such as those that ask jurors to disregard prejudicial or irrelevant evidence or testimony to which they have been exposed (Steblay et al., 2006) or those asking jurors to limit how they use certain types of evidence (Lieberman & Arndt, 2000), are generally less effective in guiding decision making. Indeed, these kinds of instructions can produce a backfire effect, wherein the evidence that is supposed to be limited or disregarded because it is prejudicial becomes more influential in decision making (Lieberman & Arndt, 2000).

Whether and how implicit bias instructions are used by jurors remains understudied. On the one hand, by making the issue of race and racial bias salient, they may prompt a "watchdog effect" whereby jurors consciously guard against making decisions that could be interpreted as biased (Sargent & Bradfield, 2004). The mere fact of making racial bias a salient issue for jurors may put them on notice and reduce their reliance on racial cues in decision making (Ingriselli, 2014; Sommers & Ellsworth, 2001). On the other hand, implicit bias instructions could elicit a backfire effect, either one analogous to that associated with instructions that ask jurors to disregard certain forms of evidence (Lieberman & Arndt, 2000) or by triggering resentment among jurors who may feel they are being accused of harboring bias (Elek & Hannaford-Agor, 2014).

Elek and Hannaford-Agor (2014) were the first to empirically test for the effects of implicit bias jury instructions on juror expressions of racial biases. Using an experimental design, they tested whether the race of the defendant, the race of the victim, and the type of instructions (standard or implicit bias) predicted individual laypersons' guilt judgments and assessments of case elements in an assault case. They found no overall effects of defendant or victim race on their outcome measures across the jury instruction conditions. Nor did they find either an ameliorative or backfire effect for the implicit bias instructions on the participants' assessments of the case. In sum, Elek and Hannaford-Agor (2014, p. 15) found "no conclusive evidence regarding whether or not the specialized implicit jury instruction used . . . is effective as a bias-reduction intervention."

In an examination of how different kinds of instructions impact judgment, Ingriselli (2014) used an "egalitarian" instruction, which explicitly educated participants about implicit bias, as one of four instructions conditions in an experiment that again used individual participants who considered a criminal case. She found that those participants who were assessed to be "aversive" racists—and who were therefore motivated to appear nonracist—were less likely to find the Black defendant guilty when they heard the "egalitarian" instruction compared with when they heard an instruction that focused on procedural justice (Ingriselli, 2014).

The two studies described above, which are the only publicly available studies examining the role of implicit bias instructions on jury decision making (Ingriselli, 2014; Elek & Hannaford-Agor, 2014), used online participants who were provided written case summaries and who made individual judgments without any group deliberative process. Moreover, given the equivocal findings of those studies, neither an ameliorative effect nor a potential backfire effect can be ruled out. Because bias can be activated and enhanced through the deliberation process (Lynch & Haney, 2011, 2015), it is important to address how implicit bias jury

instructions influence decision-making processes in a group setting in order to more closely approximate the conditions in which jury verdicts are rendered. It is possible that implicit bias instructions may prompt a backfire effect that manifests during the deliberation process. Or, conversely, implicit bias instructions may promote deliberations that are more consciously evidence-based (Elek & Hannaford-Agor, 2014).

## 2 | THE STUDY

### 2.1 | Overview of design

The present study was designed with two goals in mind. First, it assesses whether verdict outcomes were influenced by the race of the defendant and informant witness by testing the alternative hypotheses that (1) the implicit bias instructions will reduce anti-Black bias against the defendant in determining guilt, or (2) the implicit bias instructions will, consistent with a backfire effect, increase anti-Black bias against the defendant in determining guilt.

Second, the study examines *how* the use of implicit bias instructions impacts the deliberation decision-making processes, including whether such instructions raise awareness of and concern with the problem of bias in decision making. Therefore, we assess the impact of implicit bias instructions relative to standard instructions on verdict outcome measures and on group decision-making processes.

The study employed a $2 \times 2 \times 2$ factorial design, in which the race of the defendant (Black or White), the race of the informant witness (Black or White), and jury instructions (implicit bias or standard bias instruction) was varied, creating eight experimental conditions. To better approximate an actual jury trial, the case was presented as a 70-min voice-recorded and visual trial presentation. The voice recording was completed using actors trained to play all of the speaking roles (prosecutor, defense attorney, FBI agent, informant, judge). That audio was then overlaid on a digital capture of a quick-moving slide show of 366 photographs representing the trial. The versions of that slide show vary only in terms of the race of the defendant and informant, as well as the instructions given, to capture the eight experimental conditions; otherwise, they are identical. We did not manipulate the race of other trial participants, nor did we test for any potential interactions between judge race and the instructions. All trial participants shown in the slide show, besides the defendant and the informant, were presented across all conditions as White.

The trial presentation was loosely based on an actual federal narcotics conspiracy trial transcript. The defendant in the simulated case, Harold Williams, is charged with conspiracy to distribute cocaine in the Central District of California. The government alleges that Mr. Williams entered into an agreement to sell 500 g of cocaine to an associate, Sheldon Smith, for $10,000. Unbeknownst to the defendant, the associate himself has been arrested and is cooperating as an undercover informant for the FBI. The trial presentation features opening statements from the prosecutor and defense attorney, followed by the testimony of two prosecution witnesses. The first witness is an FBI agent who testifies about her experience working in narcotics and managing informants, then about the specifics of the case, including about text messages between the defendant and the informant arranging the drug sale, and how the defendant was ultimately arrested with marked cash. The second witness is the informant, Sheldon Smith, who testifies about his prior history selling drugs for the defendant and about his plan to acquire drugs from the defendant as part of this sting operation. Both witnesses are subjected to cross-examination by the defense after their direct examination testimony. The defense attorney challenges the FBI agent's training in handling informants and her procedure in this case, and attacks the informant's credibility on the grounds that he has a history of lying to law enforcement and that he expects a reduced sentence for agreeing to cooperate and testify.

After witness testimony concludes, the judge reads the relevant jury instructions. These instructions are drawn from the actual case transcript and generally reflect the pattern jury instructions used in federal criminal trials. Specifically, the instructions explain the duties of jurors, the elements of the specific offense charged that must be proved, the burden of proof and the presumption of innocence, considerations of witness credibility, and the defendant's right not to testify. The instructional manipulation occurs in the first section of the instructions, which explains the duties of the juror. The standard instruction we used was the Federal Criminal Jury Instruction 3.1, which includes shorter, more generic language instructing jurors to avoid bias and prejudice (Model Criminal Jury Instructions, *Ninth Circuit*, 2010); the implicit bias version was adapted from instructions developed and used by judges in the Western District of Washington. The trial concludes with closing arguments by the prosecution and defense.

We opted to use the Western District of Washington's implicit bias instructions because they are the most often used in the federal system. We maintained the standard cautionary language about bias in the standard instructions condition, even though it potentially made the manipulation less robust, for ecological validity reasons. That is, a general caution against using bias is universally used in actual federal criminal trials, and to omit it would render our standard instructions condition too unrealistic.

The specific implicit bias instruction we included is: "You must decide the case solely on the evidence and the law before you and must not be influenced by any personal likes or dislikes, opinions, prejudices, sympathy, or biases, including unconscious bias. Unconscious biases are stereotypes, attitudes, or preferences that people may consciously reject but may be expressed without conscious awareness, control, or intention. Like conscious bias, unconscious bias, too, can affect how we evaluate information and make decisions. It is important that you discharge your duties without discrimination, meaning that bias regarding the race, color, religious beliefs, national origin, sexual orientation, gender identity, or gender of the defendant, witnesses, and the lawyers should play no part in the exercise of your judgment."[1] The standard bias language used in federal court, and included in our standard instructions condition, says simply, "You must decide the case solely on the evidence and the law and must not be influenced by any personal likes or dislikes, opinions, prejudices, or sympathy."

To ensure that our instructional manipulation was detected, we pretested whether the two sets of instructions were perceived differently with regard to bias. Using an online MTurk (Irvine et al., 2018) sample of 352 jury-eligible citizens, we randomly assigned one of the two sets of full jury instructions to each participant. Participants read the instructions, then were asked to describe the duties and responsibilities of a juror in their own words. In the implicit bias condition, 77% of the participants explicitly mentioned that jurors have a duty to be unbiased and/or unprejudiced, compared with only 22% of those in the standard instructions condition ($\text{Exp}(B) = 6.78$, 95% CI [4.27, 10.88], Wald = 62.91, $p < .001$). We then asked participants to rank-order seven specific duties and responsibilities of jurors, which included "To consider the case without bias or prejudice" as one of the items. We created a dummy variable that coded whether or not participants ranked this as the most important duty. Compared with those in the standard instructions condition, those in the implicit bias condition were significantly more likely to rank being unbiased as the most important duty ($\text{Exp}(B) = 1.69$, 95% CI [1.04, 2.74], Wald = 4.50, $p = .034$). Based on this pretest of our instruction manipulation, we were confident that the two sets of instructions were perceived differently with respect to bias.

## 2.2 | Participants and jury group composition

The study was conducted in a simulated "jury room" suite in a centrally located office building within the Central District of California to facilitate the recruitment of participants who were eligible to serve on a federal jury within the district. We recruited participants through a

multipronged outreach strategy that advertised the study opportunity and noted the $100 payment offered to participants. First, we were given permission to place business cards that advertised the study in the jury assembly room at a local county courthouse, allowing us to directly recruit individuals who showed up to jury service in the area. The advertising cards were also placed in local businesses and recreational areas. In addition, display and classified advertisements were placed in local community newspapers and on online outlets such as Craigslist. Potential participants were screened for jury eligibility by phone[2]; those deemed eligible were then scheduled to participate in an upcoming session. Sessions were available at multiple times of the day and days of the week, and conditions were randomized to ensure that there were no time or day-of-the-week confounds with the experimental conditions.

A total of 639 eligible participants were successfully recruited to participate in this study. Participants were assigned to one of 123 jury groups consisting of four to seven individuals, which were scheduled over an eight-month period of data collection. Following the protocol of previously published research utilizing a large number of small jury groups (Dahl et al., 2007; Lynch & Haney, 2009), we scheduled seven persons per jury group, with the goal of ultimately obtaining six-person groups. While federal criminal juries are composed of 12 individuals, the smaller juries allowed for a sufficient number of jury units to meaningfully analyze, while not sacrificing a group decision-making process. This approach has been successfully used in multiple prior experimental jury decision-making studies (e.g., Lynch & Haney, 2009; MacCoun & Kerr, 1988; Peter-Hagene, 2019; Sommers, 2006; Shaw et al., 2021). Data from three groups, totaling 16 participants, were removed from the analyses: two were removed due to technical difficulties during data collection, rendering the data unreliable, and one was removed due to the dismissal of a participant for disruptive behavior, resulting in too few participants remaining to comprise a jury unit of at least four people. The analyses are thus based on data from 623 participants and 120 jury groups.

The majority of participants identified as women (62%), with 38% identifying as men and five participants (<1%) identifying as nonbinary or other. While women were overrepresented in our participant pool, this is consistent with actual jury pools, where women tend to be slightly overrepresented relative to their share of the general population (D. J. Devine et al., 2016; Rose et al., 2018). Participants ranged in age from 18 to 88 years old, with a mean age of 42 years. More than half of our participants self-identified as White (56%), while 44% identified as non-White. Specifically, 13% identified as Asian, 13% as Latinx, 6% as Black/African American, and roughly 12% as a different race or ethnicity. Roughly 32% of participants self-identified as Democrats and 23% as Republicans, with the remainder indicating a different political affiliation or none altogether. Regarding political ideology, 23% of participants identified as conservative, 41% as moderate, and 27% as liberal. Approximately 17% of participants had served on a jury prior to study participation.

The jury groups on the whole were relatively diverse, with just three groups of the 120 including only White participants. Sixty of the 120 groups comprised 50% or more non-White participants. Table 1 shows the relative share of jurors per group who identified as White in each instructional condition, broken out by defendant condition.

In the standard instruction condition, 49% of the groups were majority White, while 51% of the implicit instruction groups were majority White. These were well distributed across

**TABLE 1** Jury group (% White participants) composition by instructional condition and defendant race

| | Black defendant groups ($n = 59$) | | White defendant groups ($n = 61$) | |
|---|---|---|---|---|
| % White jurors | Standard instruction | Implicit instruction | Standard instruction | Implicit instruction |
| 0%–50% | 16 (27%) | 14 (24%) | 15 (25%) | 15 (25%) |
| 51%–100% | 13 (22%) | 16 (27%) | 17 (28%) | 14 (23%) |

defendant race conditions, as indicated in Table 1. There were no jurors who identified as Black in 69% (standard instructions) and 73% (implicit bias instruction) of the jury groups, respectively. Otherwise, the groups typically had one Black juror (one group in each condition had more than one Black juror).

## 2.3 | Experimental procedure

When participants arrived at our study site, they were seated in a controlled room with a large video screen and seven chairs around a table. Each participant was provided a badge indicating their assigned juror number, which corresponded with their seat at the table. They were then given a study information sheet that explained the study and their rights as participants. The researcher described the study procedure and obtained verbal consent from the participants indicating they understood what the study entailed and agreed to participate. Next, the trial video was presented to the group, which was described as an actual trial that had taken place in the Central District of California. After viewing the trial, participants each submitted a "straw" vote form privately on paper, indicating their personal verdict preference (i.e., guilty or not guilty) and their confidence in that verdict on a five-point scale. This "straw" vote was confidential and nonbinding, and jurors were told they could amend their verdict preference at any time.

After a short break, participants conducted deliberations as a group. Each group was instructed to choose a foreperson and to deliberate to reach a unanimous verdict—a dichotomous choice of guilty or not guilty—on the single count of conspiracy to sell cocaine. They were provided with copies of the judicial instructions, identical to the ones read by the judge during the trial video, as is standard federal trial practice. The groups were left in the closed "jury room" to deliberate, and a video camera was used to record the deliberations. Juries were given a maximum of 90 min to deliberate. Upon reaching a verdict, the foreperson completed a verdict form for the group that recorded the group verdict and jurors' individual levels of confidence in the decision (measured on a five-point scale). Failure to reach a unanimous verdict within the time limit resulted in a mistrial, and the nature of the split was recorded by the foreperson on a separate "mistrial voting form," which included individual verdict choices and confidence ratings. Some groups declared mistrials prior to the 90-min time limit when they were certain that they absolutely could not reach a unanimous decision.

Approximately 19% of deliberations ended in mistrials. The mean deliberation time was 29 min and 14 s. Deliberations reaching unanimous verdicts were significantly shorter (24 min, 43 s) than deliberations ending in mistrials, which lasted just under 48 min on average ($t$ [116] = 7.75, Cohen's $d = 1.82$, $p < .001$). Although the experiment was a simulated trial, thus reducing its external validity, deliberations were often highly intense; participants were engaged and expressive, and many defended their positions actively and passionately, indicating the study's high level of "experimental realism" (Lynch & Haney, 2015). Very few participants commented about being recorded or speculated about whether the trial they watched was fictitious; even when such comments were made, other group members spoke up to direct the group back to the task of reaching a verdict.

Following deliberations and documentation of the verdict, participants individually completed an electronic survey on laptops provided by the researchers. The survey assessed individual perceptions of the witness testimony, the attorneys, the defendant, and the judge; comprehension of jury instructions and perceived influence of those instructions; memory of case facts; attitudes about a variety of issues; and demographic information. We included measures that assessed racial bias among the attitude measures, drawn from Williams and Eberhardt's (2008) Race Conceptions Scale and from Pettigrew and Meerten's (1995) Subtle Racism Scale. After participants completed the individual measures, they were debriefed and paid $100 for their participation. The study took approximately 3.5–4 hours for participants to complete.

## 2.4 | Analytic strategy

To analyze our quantitative data, we first ran a series of binary logistic regressions to see how the instructional condition (implicit bias vs. standard) and race conditions impacted verdicts. We report odds ratios (Exp(*B*)) for our logistic regressions for ease of interpretation and to convey effect sizes. We then used binary logistic regression and t-tests to assess differences by instruction condition in the quantitative measures derived from our manifest content analysis of the deliberations. To prepare for the analysis of the deliberations, all video recordings were transcribed by a professional transcription service. Two jury groups were not recorded due to experimenter error, so the deliberations analysis is based on 118 deliberations. The transcripts were systematically coded by a team of graduate and undergraduate research assistants in order to measure how manipulated aspects of the case (i.e., instructions and race) shaped deliberations.

We used a mixed strategy employing both concept-driven and data-driven qualitative content analysis (Schreier, 2012) and more traditional manifest content analysis (Krippendorff, 2012) in order to develop an extensive coding frame for analyzing the deliberations. The frame included 27 discrete thematic categories, each with two or more subcategories to code discussions about the evidence and witnesses, the defendant, the attorneys, the jury instructions, bias and the avoidance of bias, among other topics. While several measures were quantified in the coding, our primary approach was systematic qualitative coding, using the qualitative data analysis software Dedoose to code discursive units from the transcripts.

For the present analysis, we focus on instructions-related and bias-related content, including whether references were made to the instructions in the deliberations, the number of discrete instances in which the instructions were referenced or discussed, and the number of explicit discussions or comments that addressed the issue of bias or avoiding bias. We draw on the deliberation discourses themselves to illustrate the ways in which these topics played a role in deliberations. Three key themes emerged from the qualitative analysis as to how the groups talked about bias, their need to avoid it, and how it impacted their decision-making process. The first theme was a general principle or ethic regarding avoiding bias during deliberations. The second was about jurors' own biases, which took two primary forms: jurors recognizing and acknowledging their own biases, and jurors accusing others in the group of harboring bias. The third thematic category was concerned with bias in how the jurors assessed witnesses and their testimony.

# 3 | RESULTS

## 3.1 | Quantitative findings

### 3.1.1 | Impact of instructions on judgment: Analyses of verdict outcomes

Recall that we expected that the implicit bias instructions would have an impact on juror verdicts, alternatively hypothesizing that the instructions would either reduce anti-Black bias or increase anti-Black bias in a backfire effect. Generally, neither hypothesis was supported. As we detail here, our quantitative findings indicated no differences in verdict outcomes between those who received the implicit bias instructions compared with those who heard the standard instructions, either as a main effect or as a function of the race conditions.

We first conducted a binary logistic regression to test whether the instructions condition or race conditions predicted pre-deliberation verdict preference. There was no significant impact of instructions, defendant race, or informant race in predicting pre-deliberation straw poll verdicts, nor were there any significant interactions between instructions and the race

manipulations (all *p*s > .40). (See Figure 1 for individual condition pre-deliberation conviction preferences).

As illustrated in Figure 2, post-deliberation, fewer participants supported conviction after deliberation across the instructional conditions, with the largest decreases in the Black defendant–Black informant condition in both instructional conditions. In order to test whether the experimental manipulations impacted participants' post-deliberation verdicts, we conducted a general linear mixed binary logistic model to control for the influence of jury group and tested the effect of defendant race, informant race, instructions condition, and the interaction of instructions with both defendant and informant race. A substantial share of variance was explained by jury group assignment ($Z = 5.37$, $p < .001$), and the overall fixed effects model was not significant ($p = .485$).

We also tested to see whether any of our independent variables predicted convictions at the group level. We again conducted a binary logistic regression to test whether the instructions

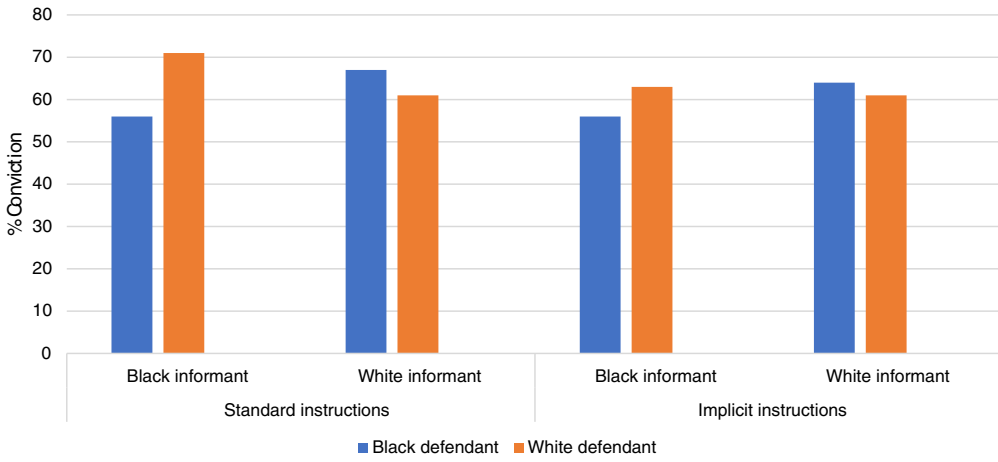


**FIGURE 1** Pre-deliberation support for convictions across all eight conditions ($N = 622$)

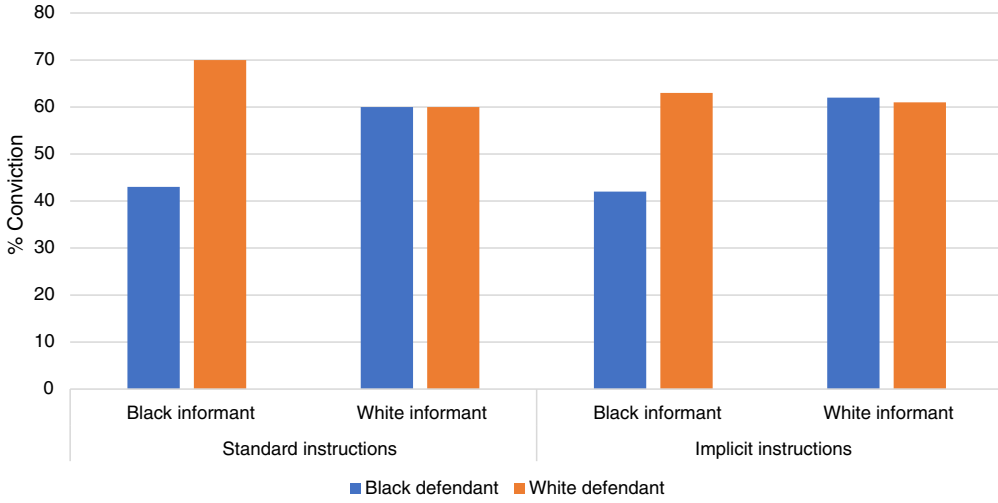


**FIGURE 2** Post-deliberation support for convictions across all eight conditions ($N = 623$)

condition or race conditions predicted a conviction verdict (as opposed to either an acquittal or mistrial). Again, there was no significant impact of instructions, defendant race, or informant race in predicting group conviction, nor any significant interactions between instructions and the race manipulations (all $ps > .15$). Based on these findings from both the individual and group-level analyses, it does not appear that the implicit bias instructions had a significant impact in either mitigating or exacerbating bias against the Black defendant that was manifested in the verdict outcome.

Given prior research indicating racial differences among jurors in conviction-proneness, we also examined whether participant race was associated with verdict preferences. We tested the possibility that White jurors would be more likely to support a guilty verdict against a Black defendant compared with a White defendant, using a binary logistic regression to test for interactions between dichotomized juror race (White/non-White) and defendant race in predicting individual straw poll verdicts. The overall model was not significant ($p = .106$). Among White jurors, 68% supported a guilty verdict prior to deliberations for the White defendant compared to 65% for the Black defendant. White participants were more conviction-prone overall, however, relative to non-White participants. At the pre-deliberation stage, White participants were significantly more likely to support conviction (66% of White participants) compared to non-White participants (57% of non-White participants) (Exp($B$) = 1.43, 95% CI [1.03, 1.99], Wald = 4.57, $p = .033$), but that preference did not interact with defendant or informant race. At the group level, post-deliberation, the proportion of White jurors in a jury group was not associated with conviction ($p = .715$), nor did it interact with defendant race ($p = .552$).[3]

### 3.1.2 | Racial bias, bias awareness, and instructional conditions

Given our interest in examining whether implicit bias instructions increase awareness of bias or decrease expressions of bias, we conducted several analyses designed to measure this. First, we tested whether the implicit bias instructions reduced biased cognitions among our participants. To do so, we created three composite scores drawn from our racism measures. The first (four-item) composite represents racial resentment, the second (two-item) composite represents racial empathy, and the third (two-item) composite represents belief in race as a biological/physiological phenomenon.[4] $T$-tests revealed that White participants expressed more racial resentment ($t$ [621] = 3.49, Cohen's $d = -0.29$, $p = .001$), less racial empathy ($t$ [621] = 4.53, Cohen's $d = 0.37$, $p = .000$), and a higher degree of belief that race is a biological category ($t$ [621] = 2.19, Cohen's $d = -0.18$, $p = .029$) than did non-White participants.

We then tested whether there were any differences in these three composite measures of racial bias between our two instructional groups, since one of the premises of using implicit bias instructions is that they should reduce bias by making jurors aware of their own implicit biases so that they can avoid relying on them. So, while our randomization procedure would be expected to evenly distribute individual differences in the bias measures across the instruction conditions, we hypothesized that biased cognitions may be reduced among those who were exposed to the instructions. Across three composite measures of bias, representing racial resentment ($t$ [621] = 1.10, Cohen's $d = 0.08$, $p = .271$), racial empathy ($t$ [621] = 0.96, Cohen's $d = 0.08$, $p = .336$), and biological beliefs about race ($t$ [621] = 1.05, Cohen's $d = 0.07$, $p = .293$), there were no differences between the instruction conditions, indicating that the instructions had no measurable effect on reducing biased cognitions.

We did find evidence that jurors were attuned to the messaging of the implicit bias instruction, however. As we did in our pretest of the instructions, we asked our participants to rank-order the importance of specific duties and responsibilities of jurors, including "to consider the case without bias or prejudice." We then created a dummy variable to code whether or not participants ranked this item as the most important duty. Jurors in the implicit instruction

condition were significantly more likely to assert that avoiding bias or prejudice was the most important duty of a juror than were jurors in the standard instruction condition (Exp ($B$) = 1.44, 95% CI [1.01, 2.06], Wald = 4.11, $p$ = .043).

Given the differential ratings of the importance of being unbiased, we next analyzed the quantified deliberations data to assess whether the instructions, in general, were more frequently referenced or discussed by participants in the implicit bias condition. We coded all direct references to the instructions or any of their specific elements, any time a group member read from the instructions, and specific references to the judge's words (he only spoke when reading the instructions in our simulated trial). We then examined whether groups were more likely to discuss instructions in the implicit bias condition compared to the standard condition.

The groups in the implicit bias condition were marginally more likely to discuss aspects of the jury instructions than those groups in the standard instruction condition. Specifically, 73% of those in the standard instruction condition discussed the jury instructions at least once during deliberations, while 86% of the groups in the implicit bias condition discussed the instructions at least once (Exp($B$) = 0.44, 95% CI [0.17, 1.13], Wald = 2.98, $p$ = .087). Among the groups that did discuss the instructions, the instructions were referenced at roughly the same frequency across the two instruction conditions. Implicit bias juries had a mean of 3.74 distinct mentions of instructions per jury deliberation, while standard instructions juries had a mean of 3.18 mentions ($t$ [116] = 1.02, Cohen's $d$ = 0.21, $p$ = .312).

We then coded explicit references to bias, prejudice, stereotyping, discrimination, or racism in the deliberations to test whether the implicit bias groups were more likely to discuss such topics. Those in the implicit bias condition were significantly more likely to explicitly discuss bias-related issues in their deliberations. Bias was raised as a discussion point twice as often in the implicit bias condition, coming up at least once during deliberations in 45% (26/58) of the juries in the implicit bias condition, compared with just 23% (14/60) of the groups in the standard instruction condition (Exp($B$) = 0.375, 95% CI [0.17, .826], Wald = 5.92, $p$ = .007). The number of distinct mentions/discussions per group ranged from 1 to 13, with a mean of 1.93 mentions within the standard instructions groups and 2.22 in the implicit bias groups (*n.s.*).

## 3.2 | Qualitative findings

Given our findings that participants in the implicit instructions condition were more attuned to the importance of avoiding bias and that those groups were especially likely to discuss bias in their deliberations, we next examined *how* our jury groups talked about bias in their deliberations by conducting a qualitative content analysis of the subset of 40 deliberations that had explicit discussions of "bias."

The three key themes that emerged from this analysis as to how the groups talked about bias and its effect on their decision-making process are (1) avoiding bias in the deliberations as an ethic of decision making, (2) self-examination of jurors' own biases, and (3) concern about bias in assessing witnesses' credibility.

### 3.2.1 | Avoiding bias as an ethic of decision making

The first core theme that emerged from this analysis was a general value or ethic that avoiding bias should be a goal during deliberations. Nine implicit bias instruction juries and one standard instruction jury engaged in this kind of discourse. Some jurors in the implicit bias instruction condition declaratively stated the importance of remaining impartial and unbiased, often with direct reference to the instructions, as a way to set the tone for the deliberations, and as a statement of values. For instance, in Jury 9, an implicit bias instruction group, Juror 9-7 directly

quoted from the instructions to impress upon the others the importance of remaining unbiased: "'You must not be influenced by any personal likes or dislikes, opinions, prejudices, sympathy, or biases, including unconscious.' I mean wow. Wow." In the implicit bias instruction condition, the general value statement was often greeted with agreement by others, suggesting that it was useful in setting the tone for unbiased deliberations, as illustrated by Juries 87 and 111:

> **87-3**: I appreciate the fact that . . . according to the judge's instructions, we're not going on prejudices about the way people look. You know?
> **87-4**: Yeah.
> **87-6**: Yeah.

> **111-7**: If you remember from the instructions of the judge, we're looking at how we're not going to be influenced by biases or what we think, et cetera.
> **111-3**: Exactly.

Conversely, the one mention of bias as a value in the standard instruction jury did not generate affirmation from others. Thus, in Jury 123, Juror 5, stated, "I like that in this type of setting, you should not base it on character or prejudice," then linked that sentiment to evaluating evidence fairly. The next speaker did not acknowledge that sentiment, and instead stated why she supported a guilty verdict.

In some groups, the implicit bias instruction language was also used to direct how the jurors fulfilled their duty. In several juries in the implicit bias condition, this manifested in the group forgoing an initial verdict poll of individual members as a way to avoid bias. At the start of deliberations for Jury 13, for instance, Juror 13-5 immediately asked to poll everyone on their verdict preferences. Two of the other jurors jumped in to protest on the grounds that it would bias the decision-making process.

> **13-5**: How did you guys all vote? Let's take a quick vote where we're at right now.
> **13-3**: Well, that would probably be [crosstalk] biased.
> **13-2**: Yeah, we don't want to start with any bias.
> **13-5**: Okay.

Likewise, early in Jury 107's deliberation, the five-person group realized that they agreed on a verdict but collectively decided that they could not conclude their deliberation without "figuring out some more stuff." They then referred to the instructions to address their required duties:

> **107-6**: Well, let's look at, can you see anything in the jury instructions that we're missing? You know, was it supposed to be that cut and dry—
> **107-1**: Basically, we can't, we can't make any, like, based on, you know, our personal likes, dislikes, stereotypes, those kinds of things.
> **107-5**: Yeah.
> **107-4**: Okay, we're not using that he's, that they're both [defendant and informant]—
> **107-1**: Scumbags.
> **107-4**: Scumbags.
> **107-5**: Yeah.
> **107-1**: Can't say that. That's one of the things you can't do.
> **107-4**: Well, we're not saying that.
> **107-1**: Yeah.
> **107-4**: We might feel that.
> **107-5**: Yeah. I know, exactly.
> **107-4**: We might know that, but we're not supposed to—

**107-1**: But we can't use that, yeah. We can't use that.
**107-5**: All feelings aside.

While jurors interpreted avoidance of bias in very expansive ways, they generally did *not* explicitly link it to race. In fact, Jury 56 (implicit bias instruction condition), which had five distinct discussions of bias in their deliberations, was the only group in which a juror connected the implicit bias instructions to avoiding racial bias. Juror 56-1 raised a concern about using "common sense" in assessing the credibility of testimony, since that could lead to an "emotion-based" judgment. Juror 56-7 took that opportunity to remind the others about the bias instructions, which Juror 56-1 then illustrated with an example that specifically referenced race.

**56-7**: That's why he [the judge] was talking about unbiased consciousness—
**56-1**: Yeah.
**56-7**: —and prejudices and—
**56-1**: Well, yeah, it would be like, say, because the defendant was Black, that would be, like, you going, oh, like . . . oh, he's Black, so he's guilty probably.

### 3.2.2 | Jurors' own (and each other's) biases

Especially in the implicit bias instruction conditions, some jurors were openly self-reflective about their own biases, a stance that seemed aimed at recognizing and controlling the influence of bias. This came up in 10 of the 40 juries in which bias was discussed during deliberations, only three of which were in the standard instruction conditions. For instance, in Jury 72 (implicit bias instruction condition), Juror 72-6 shared his struggle with judging the defendant for his decision to not testify and by his appearance. His fellow jurors encouraged him in his effort to put those thoughts aside:

**72-6**: And that's why I think guilty to begin with. The guy not saying nothing, the guy's appearance. I know some of those things you're not supposed to—
**72-5**: It's right here [referencing the instructions].
**72-6**: Yeah.
**72-5**: Bias, yeah.
**72-6**: Yeah, I'm biased. Forgive me.
**72-5**: So you wouldn't have made it on the jury anyway.

Jurors also pointed out others' bias, sometimes productively, and sometimes as a rhetorical weapon during conflict. There were 20 distinct instances of jurors calling out others for bias across 12 jury deliberations. Eight of the 12 were in the implicit bias instruction condition. In a more productive version of this, in the previously-highlighted Jury 56 (implicit bias instruction condition), Juror 56-1 used the overarching message of the implicit bias instruction to gently point out another juror's bias and to redirect her. Juror 56-3 asserted that "We should all use our common sense here," to which Juror 56-1 replied, "Well, there's common sense, and then there's going by the jury instructions. No offense. I'm not trying to sound rude, but that's what they said in here [referencing copy of instructions]." Juror 56-3 then agreed that she would try to avoid bringing in too much of her own personal experience. A short group discussion of "unconscious bias" then ensued, concluding with Juror 56-7 reminding everyone that they must make their decision "off the evidence and what they presented." Juror 56-3 herself then picked up the instructions and read aloud the elements of the crime that needed to be proved.

However, the directive about bias as laid out in the implicit bias instructions was also at times contested in response to an accusation of bias. For instance, in Jury 93, a group that

heard implicit bias instructions, Juror 93-1 tried to steer Juror 93-4 away from relying on personal bias in assessing evidence. Juror 93-4 resisted by calling into question the whole premise that "unconscious bias" can be controlled, continually interrupting Juror 93-1 in the process:

**93-1**: You're not supposed to use your personal bias.
**93-4**: No. No, the difference is—
**93-1**: [overlapping] You are supposed to—
**93-4**: [overlapping] It's my unconscious bias.
**93-1**: You're not supposed to—
**93-4**: [overlapping] And how am I supposed to know if it's unconscious?
**93-1**: You are—
**93-4**: [overlapping] It's unconscious, so how would I know?
**93-1**: You were specifically directed not to use conscious or unconscious bias. You're specifically—
**93-4**: [overlapping] It is physically, it is mentally and physically impossible for people to not take in their biases or whatever. You can't do it.
**93-1**: And yet, you have been—
**93-4**: [overlapping] They can *tell* you to do it—
**93-1**: Specifically directed—
**93-4**: [overlapping] Whatever.

In addition, some jurors in both conditions weaponized the language of the instructions to label fellow jurors as biased as part of a persuasion strategy. For instance, in Jury 70, a group that heard the standard instructions, Juror 70-6 tried to redirect Juror 70-7, who supported a guilty verdict, to focus on assessing the credibility of the informant, which Juror 70-7 resisted by speculating that the informant may not be getting anything for his testimony: "I want to ask you just to think to yourself, whether you care what we think or not . . . if it's possible that *you* have a bias existing in your mind about the fact that an informant informs for a reason?"

While Jury 70's deliberations eventually moved from a personal accusation of bias to a discussion centered around the instructions and evidence, the deliberation in Jury 84, another standard instruction group, depicts how the discussion sometimes devolved into more personal animosity when accusations of bias were made, losing focus on the evidence and the law. There were five distinct accusations of bias that occurred during this deliberation. Nearly a half-hour into this 68-min deliberation, a heated discussion ensued, in which two jurors together argued that the case lacked direct evidence of the crime, while a third juror, Juror 84-5, argued that the evidence was sufficient for guilt. Juror 84-5 then accused the other two of being biased against law enforcement. While the discussion came back to the evidence at some points, it continued to deteriorate to the point that Juror 84-5 accused one of the other jurors of being a drug dealer. This group ultimately ended in a mistrial, with Juror 84-5 holding out for guilt:

**84-4**: They don't have concrete evidence. They don't have any audio or visual.
**84-5**: You're assuming they're crooked!
**84-4**: No, I'm not assuming.
**84-5**: But that's from the movies that all cops are dirty. And the FBI is dirty.
**84-4**: No.
. . .
**84-5**: Have you been around? Are you a drug dealer? Have you been around drugs?
**84-3**: No, no, no. I've been around it, but not a drug dealer. So, I know more about it.
**84-5**: But that's your bias going into assuming—
**84-7**: That's his experience.

Juror 84-7 then accused Juror 84-5 of reverse racism against the defendant, a White man, in the presentation this group viewed: "Is it because he's White that you think they're not planting it on him? Would it be different if Harold was Black?" Juror 84-5 denied seeing race at all: "Who said he's White? No, color's got nothing to do with it."

The implicit bias instruction groups were not immune to devolving into these kinds of unproductive or hostile deliberations, either. For instance, two jurors in Jury 50, in the implicit bias instruction condition, became fixated on accusing each other of harboring biases. Jurors 50-4 and 50-7 were opposed to each other on the verdict, and the discussion devolved into a more personal argument that Juror 50-1 tried to referee:

> **50-7**: I'm not biased. I'm not biased because he's a drug dealer.
> **50-4**: I'm not biased either!
> **50-7**: I'm saying not guilty because they [law enforcement] set stuff up all the time.
> **50-4**: [overlapping] I'm not biased.
> . . .
> **50-4**: I don't know her [Juror 50-7's] experiences with the law and what side [of the law]. I know what my experience is with the law, and what I've experienced in my lifetime. I'm a foster parent. I deal with the law every day.
> **50-1**: That's bias and they said not to use that.
> **50-4**: I know, but I'm not biased. I'm just saying, I deal with them [law enforcement].

Finally, in Jury 41, another group that heard the implicit bias instructions, bias accusations continued even after deliberations were done. After a contentious deliberation that ended in a mistrial, several jurors got into an argument over filling out the mistrial polling form, during which one juror called another an "idiot." The jurors then took turns suggesting that each other would be tossed out by the judge for bias:

> **41-6**: The judge would throw you way, way, way, way, way—not only out of the courtroom but . . .
> **41-3**: Arguably, he would throw you out as well because you were biased from the very beginning.
> **41-6**: No, he would throw *him* out [referring to Juror 41-4].

### 3.2.3 | Improper bias or appropriate scrutiny of witnesses?

Finally, the issue of bias intersected with the assessment of witnesses in two ways. First, some jurors, especially in the standard instruction condition, were concerned with whether the witnesses were biased against the defendant. This made sense, given that the instructions on assessing the credibility of witnesses directed jurors to consider whether any witness might have something to gain or lose by testifying, or may be biased for one side or the other (a standard criminal trial instruction).[5] Therefore, jurors often focused on whether the informant or FBI agent might have a bias in this case. Fifteen of the 40 juries had at least one mention of this kind of witness bias, including eight of the 14 in the standard instruction condition.

Sometimes the assessment started with the actual reading of the relevant instruction, as in Jury 3, in the standard instruction condition, and Jury 49, in the implicit bias instruction condition:

> **3-7**: Is the witness biased?
> **3-6**: These people [informants], they'd eat their mother. I mean, c'mon. They'll kill each other for anything.

**3-7**: [reading instruction] "If you believe that the witness has something to gain, ask yourself whether this would make him less inclined to be truthful. Bear these things in mind as you try to decide if a witness is telling the truth or not."
**3-4**: I think he has to tell the truth, though. They never told him what he would get, so there was no inclination to lie.

**3-7**: They told him they'd get him out of jail. They told him if he snitched, he'd serve less time.

**49-5**: [reading instruction] "Does the witness have some bias toward one side or the other that might cause him to shade the truth?" Well, yeah. He doesn't want to go to jail.

**49-1**: That's true.

While the majority of the discussions about witness bias centered on the informant, several also raised questions as to the FBI agent's potential bias. For instance, in Jury 68, which was in the implicit bias instruction condition, Jurors 68-4 and 68-5 called into question the agent's testimony about the informant's trustworthiness.

**68-4**: She said she felt comfortable with him not lying to her because she didn't think he was capable of—
**68-5**: Because his demeanor.
**68-3**: Because he's unsophisticated.
**68-4**: I get it, but—
**68-5**: That's making a personal—like you said, interjecting her own personal discrimination, saying, "Well, just because he's a coke head, I don't think he's capable of being this intelligent." Well . . . that's biased.
**68-4**: Yeah. But clearly he was intelligent enough to both be selling cocaine and being an informant.
**68-5**: Yeah, that's what I'm saying.

In the implicit bias instruction condition, some deliberations revealed a more troubling trend in how bias was considered in assessing credibility, elucidating a tension between the implicit bias instruction and the instruction to consider and assess witness credibility. While jurors in all conditions were instructed to consider the credibility of the witnesses, some juries in the implicit bias condition seemed to overcorrect for bias by discouraging critical scrutiny of the witnesses. Specifically, in seven of the 26 implicit bias instruction groups that discussed bias (but in none of the standard instructions groups), credibility assessments appeared to be *stunted* by inappropriate concern about being biased against the witnesses. That is, some jurors' concern with checking and containing their own and others' bias appeared to impede their ability to critically assess the government witnesses' bias, ultimately disadvantaging the defendant.

In Jury 101 (implicit bias instruction condition), for instance, Juror 101-4 opened the deliberation with a statement about the inappropriateness of being biased against the witnesses. Before giving her "thesis" about why the defendant should be found guilty, she accused the defense of bias for raising questions about the credibility of the witnesses:

**101-4**: The law states to dismiss any bias. So with that being said, I think the defendant's points of pointing out Sheldon Smith's [the informant's] past, as with questioning the agent's education, is irrelevant because in doing so, the defendant

> is attempting to create a bias, rather than look at focal evidence such as testimonies. The defendant keeps trying to name an emotional plea, specifically towards the end at their conclusion by . . . emphasizing Smith's past, so in doing so, he is subconsciously creating a bias in the minds of many people.

Later in the deliberations, when another juror tried to push back on this in regard to assessing the informant, Juror 101-4 raised the inappropriate question of whether the defendant not testifying indicated something about his credibility.

> **101-7**: The whole purpose of the informant is to test his credibility. The judge said that's our primary job.
> **101-3**: Exactly.
> **101-4**: So if we're speaking with a lot of credibility, if the defendant knew that his butt is on the line, do you think there's a reason in particular why he chose not to speak in trial?
> **101-5**: Because if he hits the stand, they could ask him about his past.
> **101-3**: Yeah.
> **101-7**: The judge specifically said—
> **101-4**: I know that's what the judge said, but it's not—
> **101-7**: He said you were not—not to—that should not have any bearing on this because he—by law, does not have to [testify].

In Jury 102's (implicit bias instruction condition) four-person deliberation, jurors discussed evidence in relation to the charge of conspiracy, as outlined in the jury instructions, but then put aside the question of witness credibility. This group consciously made an effort to identify facts presented during trial but, in doing so, downplayed the testimony that emerged about the informant directly pertaining to the credibility assessment, including testimony concerning previous incidents of him lying to law enforcement authorities. The group discussed the credibility of both the FBI agent and the informant at length from the start of the deliberations, raising some skepticism about both witnesses in the process. Three of the jurors, 102-1, 102-3, and 102-5, all agreed that the informant had no credibility. Ultimately, though, that assessment was put aside, and the group voted unanimously for guilt, after Juror 102-3 suggested that the judge had told them not to consider anything but the "law" concerning the elements of the crime.

Jurors in the implicit bias instruction condition also at times policed themselves on their own biased expressions regarding the witnesses, again sometimes at the cost of scrutinizing credibility. For instance, Juror 23-1 told the others he did not trust the informant due to the testimony about him illegally dealing drugs on the side while being an informant (a valid credibility assessment), but then appeared to dismiss this concern as instance of his own bias against the informant: "We know the guy selling the drugs isn't an angel, but that's not what we're deciding." Ultimately, this juror's support for acquittal was not based on the informant credibility issue, but on facts related to how the FBI handled the surveillance.

The tension between avoiding bias and critically assessing witness credibility was full-blown in Jury 88's (implicit bias instruction condition) five-person deliberation, where Juror 88-7, who supported acquittal, went into detail as to why he did not find the informant credible, based in part on the testimony about the informant previously using his brother's driver's license as ID to avoid arrest when stopped by police. Juror 88-7 asserted, "I wouldn't trust him or anything that that guy said." Other jurors expressed agreement with this juror as he raised these issues. Then Juror 88-4, who supported a guilty verdict, interjected by accusing Juror 88-7 of relying on emotion, not facts, which was deemed to be biased:

**88-4**: Let's address your concerns . . . One of the things I realized that you said in terms of verbiage—you said, "feel." I think one of the biggest things we have to consider is just taking feeling out of the equation.
**88-7**: Taking what now?
**88-4**: Taking feeling out of the equation.
**88-7**: Yeah?
**88-4**: And we're looking at it through a lens as a judge would.
**88-7**: Yeah.
**88-4**: Unbiased. No thinking just because someone has that background. We don't know much of Williams' [the defendant's] background as well.

The disagreement between these two jurors continued throughout the deliberations. Nearly a half-hour after the above exchange, Juror 88-4 specifically referenced the implicit bias instruction, which he had written down, to challenge Juror 88-7's continued insistence on evaluating the informant's credibility. By this point, two other jurors worked to counter 88-4's misconception, and Juror 88-7 specifically cited the credibility instruction.

**88-4**: What I was questioning, too . . . when it comes to when the judge said making a decision based on no bias, stereotypes conscious and unconscious. You know, just whether it's unbiased or not relying on stereotypes. I mean, for me, the way I see it is I feel biased if I'm judging his [the informant's] character and making a decision on whether or not this testimony is credible . . .
**88-5**: You *can* base your decision on that.
**88-1**: Because they did catch him in a lie.
**88-7**: And I *am* trying to eliminate the prejudice, but when he [the judge] says you have to take into account—I wrote it down, not verbatim—but the credibility of the witness, and does that witness have something to gain . . .

Jury 88 never resolved this conflict and ended in a mistrial, with Juror 88-7 and Juror 88-1 supporting acquittal, and the others supporting a guilty verdict.

On the other hand, not all juries in the implicit bias instruction condition had such difficulty or conflict over distinguishing inappropriate bias from appropriate scrutiny of the witnesses. For example, in Jury 4, two of the four jurors worked to distinguish their negative assessment of the informant from any judgment of the defendant's character, while tamping down the temptation to draw inferences from the fact that the defendant did not testify.

**4-5**: He probably is guilty, but they didn't have enough evidence to send him to jail with.
**4-1**: I could go with along with that. He didn't get on the stand. So, I guess we don't know from him. That's the way it goes.
**4-5**: Sometimes it's better for them not to get on the stand and talk. Let the lawyers do the talking. He might incriminate himself.
**4-1**: We don't have the bad opinion of him, like we do of Smith [the informant].
**4-5**: Williams [the defendant] might be nice. It's all corrupt, but just because we're not involved in a corrupt living situation like that, we can't just stereotype that guy [the defendant].

Similarly, Jury 49 began its deliberations by having each juror take a turn at reading the instructions out loud until it was fully read, after which their first topic of discussion was the credibility of the witnesses. They stuck to the actual testimony, without engaging in speculation. Juror 49-5 affirmed his fellow jurors' comments regarding the informant: "Also, like you guys said, [the informant] blatantly lied on the stand. [mimicking informant] 'Oh, I thought you were

talking about felonies.' He never once mentioned that at all. And, basically, we're trying to just decide if there is reasonable doubt." Juror 49-1 verbally reinforced this analysis with "Good answer." After a few more confirmatory exchanges, Juror 49-5 referred to his notes from the trial: "Even his responses . . . I wrote down, his first response was, 'I don't remember.' You gave him your brother's ID, you don't remember *that*?" Others again agreed with him. Then Juror 49-5, in order to clarify the law that they were to follow, read directly from the judge's instruction on witness credibility, including the section asking jurors to consider "Does the witness have some bias toward one side or the other that might cause him to shade the truth?" In the end, this jury, which was originally split (three for not guilty, two for guilty) voted unanimously to acquit the defendant.

## 4 | DISCUSSION AND CONCLUSION

The primary focus of the current study was to understand the effects of implicit bias jury instructions on verdict outcomes and on group deliberation processes. Because so little empirical research has been done addressing the impact of specialized instructions on reducing jurors' expressions of implicit bias, and as more courts are beginning to implement tailored instructions in an effort to mitigate bias in the jury trial context, the present research comes at a critical moment. In line with Elek and Hannaford-Agor's (2014) findings, we did not find that the implicit bias jury instruction exerted influence on verdict outcomes when compared with the standard instruction condition. We also did not find a main effect of defendant race, although the Black defendant was nonsignificantly less likely to be convicted than the White defendant across instructions conditions. Nor did the implicit bias instructions interact with defendant race, contrary to our expectations. Nonetheless, jurors who received the implicit bias instructions did indicate that they were more attuned to the prohibition against bias, in that they were more likely to express that their most important role as a juror was to avoid bias or prejudice, and they were more likely to discuss the issue of bias during their deliberations.

Despite this increased awareness, however, we found that our measures of individual racial bias, which were administered after deliberations were complete and near the end of the battery of individual measures, showed no differences between those who had heard the standard instructions and those who had heard the implicit bias instructions. To the extent that such biases are stable attributes, and to the extent that the variations were evenly distributed through our random assignment procedure in the experiment, this is not a surprise. But to the extent that the implicit bias instructions are predicated upon the notion that biases are malleable, and can be moderated through the act of instruction, this bodes poorly for the remedial goal of these instructions.

The deliberations data revealed a more complex story about how the implicit bias instructions were interpreted and used in the decision-making process. The groups in the implicit bias instruction conditions were more likely to explicitly discuss issues of bias, and those discussions had several key features. In some of the small groups, participants who received implicit bias instructions engaged in conversations that aimed for a fair, unbiased process when discussing the facts of the case, suggesting that these instructions can promote meaningful discussions focused on the legal requirements to determine a verdict, resulting in decisions that have been thoroughly reflected upon. Some of these deliberations also revealed how groups of strangers tasked with coming to a verdict decision can engage in non-defensive dialogue about the problem of bias. In that regard, the instructions sometimes catalyzed an educative process that also could aid in keeping a check on group members' bias.

On the other hand, the concept of bias was elastic enough in these deliberations to sometimes impede other aspects of the jurors' duties in a manner that could potentially harm defendants, especially in the implicit bias instruction conditions. In more than a quarter of the implicit bias groups that discussed the problem of bias, mock jurors—seemingly in good faith—

interfered with their own and their peers' lawful and appropriate assessments of witness credibility, which was an absolutely essential task in the case we presented. Because informant evidence is notoriously unreliable, especially when it is obtained through incentives (Natapoff, 2009; Swanner & Beike, 2010), a defendant's fate can rest upon having a jury that is willing and able to scrutinize the informant-related evidence and testimony with respect to the question of credibility. In a number of instances, the implicit bias instructions were invoked to shut down that critical inquiry. Research suggests that it is already difficult for jurors to give weight to the effect of incentives that informants receive, impeding a fulsome credibility assessment (Neuschatz et al., 2008, 2012; Wetmore et al., 2014). If the implicit bias instructions further impede that assessment, as these findings suggest can happen, the defendant is even more disadvantaged in cases that rely upon informants.

Finally, given that jury deliberations are a venue for persuasive engagement, as jurors seek to change others' minds and come to a unanimous verdict (Lynch & Haney, 2015), it is not surprising that we found some participants using elements of the instructions as rhetorical weapons to counter those with whom they disagreed. In groups that heard the implicit bias instructions, a charge of "bias" could be a particularly powerful and nasty accusation, given the heightened perceived value of being unbiased among those in that condition. The group that devolved into a squabble over filling out the mistrial polling form, resulting in serial accusations of bias *after* deliberations were completed, provides a telling example of the power of that accusation as a weapon.

Although this study was able to provide new insights into how implicit bias instructions influence and shape group-level processes during deliberations, the findings do not offer conclusive evidence as to whether the advantages of using implicit bias instructions outweigh the potential costs. Additional research is needed to fully understand whether implicit bias instructions can reduce biased decision making in specific kinds of cases and with different jury configurations. More research is also needed to further explore some of the potential complications that come with implicit bias instructions. While we did not specifically uncover any backfire effects, we did find that the elastic interpretation of bias could be detrimental to the defendant, whose due process and equal protection rights are at stake in criminal trials.

While the findings speak to our theoretical understanding of bias, including in small group contexts such as juries, there is also more work to be done on how individuals' attributes and experiences contribute to the group-level process. Prior research indicates that jurors subjectively experience the deliberation process differently as a function of "status characteristics" such as gender and racial identity (Winter & Clair, 2018) and that decision-making processes and outcomes vary considerably as a function of jury diversity (Shaw et al., 2021; Sommers, 2006). While we preliminarily examined how juror race impacted verdict outcomes in this study, further research on the racial and gender dynamics of deliberation and decision making is warranted, including whether these instructions may improve the subjective experience for jurors from marginalized groups.

Ultimately, our findings raise as many questions as they answer about the value and utility of implicit bias instructions in reducing the specific problem of racial bias against criminal defendants in jury decision making. One of our unanticipated findings from the deliberations analysis was that while nearly half of the implicit bias instructions groups explicitly discussed issues around bias, only one of them, a group whose case had a Black defendant, identified racial bias against the Black defendant as something they were to avoid. Otherwise, the groups tended to treat bias as a generic term, without relating it to the problem of racism. So while implicit bias instructions have aimed to address racial bias against defendants of color (Bennett, 2010; Doyle, 2017; Su, 2020), our findings indicate that the lesson about bias that is conveyed by the instructions was so generalized that most groups did not explicitly identify racial bias as a problem to be avoided. Rather, bias was often articulated as any negative assertion about any of the legal actors in the case, which had its own potentially negative consequence for the defendant.

In retrospect, this finding is not surprising given the generic, one-size-fits-all language of the implicit bias instructions, both as to the protected categories and the potential targets of bias, including witnesses and lawyers. If the solution to jury bias in criminal cases is to be specialized instructions, there is considerable room for increasing the robustness of those instructions in cases where racial bias, especially against the defendant, is a risk. Specifically, given the multiple ways in which implicit, explicit, and institutionalized racism have created longstanding patterns of inequality in the criminal system (Haney López, 2000), including in jury verdicts, the instructions could be tailored to provide a more historicized, contextualized lesson about the many ways defendants are disadvantaged by different forms of racism. To do so may make jurors more realistically "race-conscious" (Do, 2000) in their decision making.

## ACKNOWLEDGMENTS

We would like to thank the reviewers and the *Law & Policy* editorial team for their helpful comments on earlier drafts of this article, and Lynch Lab manager Blake Song and the Lynch Lab research assistants for their contributions to this project. This research was supported by Grant No. 2017-IJ-CX-0044, awarded to the first author by the National Institute of Justice, Office of Justice Programs, US Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice.

## ORCID

*Mona Lynch* https://orcid.org/0000-0002-5594-9016

## ENDNOTES

[1] These instructions are offered at different points in the trial process, and some juror orientations include a video about implicit bias. For internal validity reasons we presented the instructions in both the standard and implicit bias conditions at the end of evidence, prior to closing arguments when instructions are normally presented.

[2] Federal jury eligibility requirements include: US citizenship, 18 years or older, ability to speak and understand English, and no felony charges pending or prior felony convictions, unless civil rights have been restored.

[3] We recognize that statistical power is an issue for our group-level analyses. Due to the time-intensive, high-cost nature of this kind of study, we had to sacrifice statistical power at the group level to be able to manipulate all three variables of interest.

[4] The racial resentment items were: "Everyone has an equal chance to succeed in this country, if they are just willing to try hard enough"; "Some 'disadvantaged' groups in this country push too hard to get what they want"; "Many groups are able to overcome prejudice and work their way up in this country. Blacks should do the same without special treatment"; and "Members of some groups try to use their disadvantaged status in society as excuses for their criminal behavior." The racial empathy items were: "I feel sympathy for minority groups who are less well off than Whites in this country" and "I feel admiration for members of minority groups who succeed in this country despite the obstacles they face." The race as a biological phenomenon items were "Racial groups are primarily determined by biology" and "It's easy to tell what race people are by looking at them."

[5] On this subject, the instructions in both conditions read, in part: "One of the prime jobs of the jury is to determine the credibility—that is, the believability—of the witnesses. You must decide whether these witnesses were telling the truth or not. A person can tell the truth in whole, in part, or not at all. So I encourage you to run through a mental checklist as you evaluate what you heard today. For example, ask yourself: Did the witness have a good opportunity to see what he or she testified about today? Does the witness have an accurate recollection of the events he or she is recounting? Does the witness have some interest in the outcome, something to gain or lose by what the jury decides? Does the witness have some bias toward one side or the other that might cause him or her to shade the truth? If you believe that the witness has something to gain, ask yourself whether this would make him or her more or less inclined to be truthful."

## REFERENCES

Axt, Jordan R., Charles R. Ebersole, and Brian A. Nosek. 2016. "An Unintentional, Robust, and Replicable Pro-Black Bias in Social Judgment." *Social Cognition* 34(1): 1–39.

Bell, Jeannine, and Mona Lynch. 2016. "Cross-Sectional Challenges: Gender, Race, and Six-Person Juries." *Seton Hall Law Review* 46: 419–69.

Beniwal, Rakesh. 2016. "Implicit Bias in Child Welfare: Overcoming Intent." *Connecticut Law Review* 49: 1021–67.

Bennett, Mark W. 2010. "Unraveling the Gordian Knot of Implicit Bias in Jury Selection: The Problems of Judge-Dominated Voir Dire, the Failed Promise of Batson, and Proposed Solutions." *Harvard Law & Policy Review* 4: 149–71.

Bertrand, Marianne, and Sendhil Mullainathan. 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.

Bodenhausen, Galen V., and Meryl Lichtenstein. 1987. "Social Stereotypes and Information-Processing Strategies: The Impact of Task Complexity." *Journal of Personality & Social Psychology* 52(5): 871–80.

Burns, Mason D., Margo J. Monteith, and Laura R. Parker. 2017. "Training Away Bias: The Differential Effects of Counterstereotype Training and Self-Regulation on Stereotype Activation and Application." *Journal of Experimental Social Psychology* 73: 97–110.

Casey, Pamela M., Roger K. Warren, Fred L. Cheesman, and Jennifer K. Elek. 2013. "Addressing Implicit Bias in the Courts." *Court Review* 49: 64–70.

Dahl, Janne, Enemo, Ida, Drevland, Guri C. B., Wessel Ellen, Eilertsen Dag Erik, Magnussen Svein. 2007. "Displayed Emotions and Witness Credibility: A Comparison of Judgements by Individuals and Mock Juries." *Applied Cognitive Psychology* 21(9): 1145–1155.

Devine, Dennis J. 2012. *Jury Decision Making: The State of the Science*. New York: NYU Press.

Devine, Dennis J., Jennifer Buddenbaum, Stephanie Houp, Dennis P. Stolle, and Nathan Studebaker. 2007. "Deliberation Quality: A Preliminary Examination in Criminal Juries." *Journal of Empirical Legal Studies* 4(2): 273–303.

Devine, Dennis J., Paige C. Krouse, Caitlin M. Cavanaugh, and Jaime Colon Basora. 2016. "Evidentiary, Extra-evidentiary, and Deliberation Process Predictors of Real Jury Verdicts." *Law & Human Behavior* 40(6): 670–82.

Devine, Patricia G. 1989. "Stereotypes and Prejudice: Their Automatic and Controlled Components." *Journal of Personality & Social Psychology* 56(1): 5–18.

Diamond, Shari Seidman, and Jonathan D. Casper. 1992. "Blindfolding the Jury to Verdict Consequences: Damages, Experts, and the Civil Jury." *Law & Society Review* 26: 513–63.

Do, Long X. 2000. "Jury Nullification and Race-Conscious Reasonable Doubt: Overlapping Reifications of Commonsense Justice and the Potential Voir Dire Mistake." *UCLA Law Review* 47(6): 1843–84.

Dovidio, John F. 2001. "On the Nature of Contemporary Prejudice: The Third Wave." *Journal of Social Issues* 57(4): 829–49.

Dovidio, John F., Kerry Kawakami, and Samuel L. Gaertner. 2000. "Reducing Contemporary Prejudice: Combating Explicit and Implicit Bias at the Individual and Intergroup Level." In *Reducing Prejudice and Discrimination*, edited by Stuart Oskamp, 137–63. Hillsdale, NJ: Erlbaum.

Doyle, T. 2017. "US District Court Produces Video, Drafts Jury Instructions on Implicit Bias." In *Bar Bulletin*, April 2017. Seattle, WA: King County Bar Association.

Elek, Jennifer, and Paula Hannaford-Agor. 2014. "Can Explicit Instructions Reduce Expressions of Implicit Bias? New Questions Following a Test of a Specialized Jury Instruction." https://doi.org/10.2139/ssrn.2430438.

Fish, Jillian, and Moin Syed. 2020. "Racism, Discrimination, and Prejudice." In *The Encyclopedia of Child and Adolescent Development*, edited by Steven Hupp and Jeremy D. Jewell, 1–12. San Francisco, CA: Wiley-Blackwell.

Fiske, Susan T. 2000. "Stereotyping, Prejudice, and Discrimination at the Seam Between the Centuries: Evolution, Culture, Mind, and Brain." *European Journal of Social Psychology* 30(3): 299–322.

FitzGerald, Chloë, and Samia Hurst. 2017. "Implicit Bias in Healthcare Professionals: A Systematic Review." *BMC Medical Ethics* 18(1): 1–18.

Foley, Meraiah, and Sue Williamson. 2019. "Managerial Perspectives on Implicit Bias, Affirmative Action, and Merit." *Public Administration Review* 79(1): 35–45.

Gawronski, Bertram. 2019. "Six Lessons for a Cogent Science of Implicit Bias and Its Criticism." *Perspectives on Psychological Science* 14(4): 574–95.

Greenwald, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji. 2009. "Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity." *Journal of Personality & Social Psychology* 97(1): 17–41.

Greenwald, Anthony G., Brian A. Nosek, and Mahzarin R. Banaji. 2003. "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm." *Journal of Personality & Social Psychology* 85(2): 197–216.

Gullo, Gina Laura, Kelly Capatosto, and Cheryl Staats. 2018. *Implicit Bias in Schools: A Practitioner's Guide*. New York, NY: Routledge.

Haney López, Ian F. 2000. "Institutional Racism: Judicial Conduct and a New Theory of Racial Discrimination." *Yale Law Journal* 109(8): 1717–884.

Hickerson, Andrea, and John Gastil. 2008. "Assessing the Difference Critique of Deliberation: Gender, Emotion, and the Jury Experience." *Communication Theory* 18(2): 281–303.

Hunt, Jennifer S. 2015. "Race, Ethnicity, and Culture in Jury Decision Making." *Annual Review of Law & Social Science* 11: 269–88.

Illinois Civil Jury Instructions. 2019. 1.08. https://courts.illinois.gov/CircuitCourt/CivilJuryInstructions/1.08.pdf.
Ingriselli, Elizabeth. 2014. "Mitigating Jurors' Racial Biases: The Effects of Content and Timing of Jury Instructions." *Yale Law Journal* 124: 1690–745.
Irvine, Krin, David A. Hoffman, and Tess Wilkinson-Ryan. 2018. "Law and Psychology Grows Up, Goes Online, and Replicates." *Journal of Empirical Legal Studies* 15(2): 320–55.
Jackson, Sarah M., Amy L. Hillard, and Tamera R. Schneider. 2014. "Using Implicit Bias Training to Improve Attitudes Toward Women in STEM." *Social Psychology of Education* 17(3): 419–38.
Jones, Christopher S., and Martin F. Kaplan. 2003. "The Effects of Racially Stereotypical Crimes on Juror Decision-Making and Information-Processing Strategies." *Basic & Applied Social Psychology* 25(1): 1–13.
Judicial Council of California. 2020. CACI 113. Bias: 29. https://www.courts.ca.gov/partners/documents/Judicial_Council_of_California_Civil_Jury_Instructions.pdf.
Kang, Jerry, Mark Bennett, Devon Carbado, Pam Casey, Nilanjana Dasgupta, David Faigman, Rachel Godsil, Anthony G. Greenwald, Justin Levinson, and Jennifer Mnookin. 2012. "Implicit Bias in the Courtroom." *UCLA Law Review* 59: 1124–86.
Kawakami, Kerry, Kenneth L. Dion, and John F. Dovidio. 1998. "Racial Prejudice and Stereotype Activation." *Personality & Social Psychology Bulletin* 24(4): 407–16.
Krippendorff, Klaus. 2012. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage Publications.
Lai, Calvin K., Maddalena Marini, Steven A. Lehr, Carlo Cerruti, Jiyun-Elizabeth L. Shin, Jennifer A. Joy-Gaba, Arnold K. Ho, et al. 2014. "Reducing Implicit Racial Preferences: I. A Comparative Investigation of 17 Interventions." *Journal of Experimental Psychology: General* 143(4): 1765–85.
Levinson, Justin D., Robert J. Smith, and Danielle M. Young. 2014. "Devaluing Death: An Empirical Study of Implicit Racial Bias on Jury-Eligible Citizens in Six Death Penalty States." *New York University Law Review* 89: 513–81.
Lieberman, Joel D., and Jamie Arndt. 2000. "Understanding the Limits of Limiting Instructions: Social Psychological Explanations for the Failures of Instructions to Disregard Pretrial Publicity and Other Inadmissible Evidence." *Psychology, Public Policy, & Law* 6(3): 677–711.
Lynch, Mona. 2013. "Institutionalizing Bias: The Death Penalty, Federal Drug Prosecutions, and Mechanisms of Disparate Punishment." *American Journal of Criminal Law* 41(1): 91–131.
Lynch, Mona, and Craig Haney. 2000. "Discrimination and Instructional Comprehension: Guided Discretion, Racial Bias, and the Death Penalty." *Law & Human Behavior* 24(3): 337–58.
Lynch, Mona, and Craig Haney. 2009. "Capital Jury Deliberation: Effects on Death Sentencing, Comprehension, and Discrimination." *Law & Human Behavior* 33(6): 481–96.
Lynch, Mona, and Craig Haney. 2011. "Mapping the Racial Bias of the White Male Capital Juror: Jury Composition and the 'Empathic Divide.'" *Law & Society Review* 45(1): 69–102.
Lynch, Mona, and Craig Haney. 2015. "Emotion, Authority, and Death: (Raced) Negotiations in Mock Capital Jury Deliberations." *Law & Social Inquiry* 40(2): 377–405.
MacCoun, Robert, J., Kerr, Norbert, L. 1988. "Asymmetric Influence in Mock Jury Deliberation: Jurors' Bias for Leniency." *Journal of Personality and Social Psychology* 54(1): 21–33.
MacCoun, Robert J. 1990. "The Emergence of Extralegal Bias during Jury Deliberation." *Criminal Justice & Behavior* 17(3): 303–14.
Miller, Monica K., Jonathan Maskaly, Morgan Green, and Clayton D. Peoples. 2011. "The Effects of Deliberations and Religious Identity on Mock Jurors' Verdicts." *Group Processes & Intergroup Relations* 14(4): 517–32.
Mitchell, Tara L., Ryann M. Haw, Jeffrey E. Pfeifer, and Christian A. Meissner. 2005. "Racial Bias in Mock Juror Decision-Making: A Meta-analytic Review of Defendant Treatment." *Law & Human Behavior* 29(6): 621–37.
Model Criminal Jury Instructions. 2010. *Ninth Circuit*. https://www.rid.uscourts.gov/sites/rid/files/documents/juryinstructions/otherPJI/9th%20Circuit%20Model%20Criminal%20Jury%20Instructions.pdf.
Natapoff, Alexandra. 2009. *Snitching: Criminal Informants and the Erosion of American Justice*. New York: NYU Press.
Negowetti, Nicole E. 2014. "Implicit Bias and the Legal Profession's Diversity Crisis: A Call for Self-Reflection." *Nevada Law Journal* 15: 930–58.
Neuschatz, Jeffrey S., Deah S. Lawson, Jessica K. Swanner, Christian A. Meissner, and Joseph S. Neuschatz. 2008. "The Effects of Accomplice Witnesses and Jailhouse Informants on Jury Decision Making." *Law & Human Behavior* 32(2): 137–49.
Neuschatz, Jeffrey S., Miranda L. Wilkinson, Charles A. Goodsell, Stacy A. Wetmore, Deah S. Quinlivan, and Nicholaos J. Jones. 2012. "Secondary Confessions, Expert Testimony, and Unreliable Testimony." *Journal of Police & Criminal Psychology* 27: 179–92.
Nosek, Brian A., Frederick L. Smyth, Jeffrey J. Hansen, Thierry Devos, Nicole M. Lindner, Kate A. Ranganath, Colin Tucker Smith, et al. 2007. "Pervasiveness and Correlates of Implicit Attitudes and Stereotypes." *European Review of Social Psychology* 18(1): 36–88.
Olson, Kris. 2019. "New Jury Instructions Take Aim at Implicit Bias." *Massachusetts Lawyers Weekly*. https://masslawyersweekly.com/2019/06/20/new-jury-instructions-take-aim-at-implicit-bias/.

Payne, B. Keith, Heidi A. Vuletich, and Kristjen B. Lundberg. 2017. "The Bias of Crowds: How Implicit Bias Bridges Personal and Systemic Prejudice." *Psychological Inquiry* 28(4): 233–48.
Pettigrew, Thomas F., Meertens, Roel, W. 1995. "Subtle and Blatant Prejudice in Western Europe." *European Journal of Social Psychology* 25(1): 57–75.
Peter-Hagene, Liana C. 2019. "Jurors' Cognitive Depletion and Performance During Jury Deliberation as a Function of Jury Diversity and Defendant Race." *Law & Human Behavior* 43(3): 232–49.
Pfeifer, Jeffrey E., and James R. P. Ogloff. 1991. "Ambiguity and Guilt Determinations: A Modern Racism Perspective." *Journal of Applied Social Psychology* 21(21): 1713–25.
Rector, Neil A., R. Michael Bagby, and Robert Nicholson. 1993. "The Effect of Prejudice and Judicial Ambiguity on Defendant Guilt Ratings." *The Journal of Social Psychology* 133(5): 651–9.
Rose, Mary R., Raul S. Casarez, and Carmen M. Gutierrez. 2018. "Jury Pool Underrepresentation in the Modern Era: Evidence from Federal Courts." *Journal of Empirical Legal Studies* 15(2): 378–405.
Sargent, Michael J., and Amy L. Bradfield. 2004. "Race and Information Processing in Criminal Trials: Does the Defendant's Race Affect How the Facts Are Evaluated?" *Personality & Social Psychology Bulletin* 30(8): 995–1008.
Schreier, Margrit. 2012. *Qualitative Content Analysis in Practice*. Thousand Oaks, CA: Sage Publications.
Shaw, Emily, Mona Lynch, Sofia Laguna, and Steven Frenda. 2021. "Race, Witness Credibility, and Jury Deliberation in a Simulated Drug Trafficking Trial." *Law & Human Behavior* 45(3): 215–28.
Smalarz, Laura, Stephanie Madon, and Anna Turosak. 2018. "Defendant Stereotypicality Moderates the Effect of Confession Evidence on Judgments of Guilt." *Law & Human Behavior* 42(4): 355–68.
Sommers, Samuel R. 2006. "On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations." *Journal of Personality & Social Psychology* 90(4): 597–612.
Sommers, Samuel R., and Phoebe C. Ellsworth. 2001. "White Juror Bias: An Investigation of Prejudice Against Black Defendants in the American Courtroom." *Psychology, Public Policy, & Law* 7(1): 201–29.
Steblay, Nancy, Harmon M. Hosch, Scott E. Culhane, and Adam McWethy. 2006. "The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis." *Law & Human Behavior* 30(4): 469–92.
Su, Anona. 2020. "A Proposal to Properly Address Implicit Bias in the Jury." *Hastings Women's Law Journal* 31(1): 79–100.
Sue, Derald Wing, Christina M. Capodilupo, Gina C. Torino, Jennifer M. Bucceri, Aisha Holder, Kevin L. Nadal, and Marta Esquilin. 2007. "Racial Microaggressions in Everyday Life: Implications for Clinical Practice." *American Psychologist* 62(4): 271–86.
Swanner, Jessica K., and Denise R. Beike. 2010. "Incentives Increase the Rate of False but Not True Secondary Confessions from Informants with an Allegiance to a Suspect." *Law & Human Behavior* 34(5): 418–28.
Tanford, J. Alexander. 1990. "The Law and Psychology of Jury Instructions." *Nebraska Law Review* 69: 71–111.
West, Jessica L. 2011. "12 Racist Men: Post-verdict Evidence of Juror Bias." *Harvard Journal on Racial & Ethnic Justice* 27(1): 165–204.
Wetmore, Stacy Ann, Jeffrey S. Neuschatz, and Scott D. Gronlund. 2014. "On the Power of Secondary Confession Evidence." *Psychology, Crime & Law* 20(4): 339–57.
Wilder, David A. 1986. "Social Categorization: Implications for Creation and Reduction of Intergroup Bias." *Advances in Experimental Social Psychology* 19: 291–355.
Williams, Melissa J., and Jennifer L. Eberhardt. 2008. "Biological Conceptions of Race and the Motivation to Cross Racial Boundaries." *Journal of Personality & Social Psychology* 94(6): 1033–47.
Winter, Alix S., and Matthew Clair. 2018. "Jurors' Subjective Experiences of Deliberations in Criminal Cases." *Law & Social Inquiry* 43(4): 1458–90.
Wise, Kathy 2020. "Groundbreaking Implicit Bias Project Takes Shape in Dallas County Civil Courts." D Magazine, January 16, 2020. https://www.dmagazine.com/frontburner/2020/01/dallas-county-implicit-bias-civil-court-pilot/.
York, Erin, and Benjamin Cornwell. 2006. "Status on Trial: Social Characteristics and Influence in the Jury Room." *Social Forces* 85(1): 455–77.
Zestcott, Colin A., Irene V. Blair, and Jeff Stone. 2016. "Examining the Presence, Consequences, and Reduction of Implicit Bias in Health Care: A Narrative Review." *Group Processes & Intergroup Relations* 19(4): 528–42.

## AUTHOR BIOGRAPHIES

**Mona Lynch** is a professor in the Department of Criminology, Law & Society at the University of California, Irvine. Her research focuses on jury decision making, plea bargaining, criminal sentencing, and punishment processes, with a focus on institutionalized forms of bias within the criminal legal system.

**Taylor Kidd** is a PhD candidate in the Department of Criminology, Law & Society at the University of California, Irvine. Her research interests include implicit bias, jury decision making, and youth diversion programs.

**Emily Shaw** earned her PhD from the Department of Psychological Science at the University of California, Irvine, in 2021. Her research focuses on jury decision making and perception of witness testimony. She now works in litigation consulting.

**How to cite this article:** Lynch, Mona, Taylor Kidd, and Emily Shaw. 2022. "The Subtle Effects of Implicit Bias Instructions." *Law & Policy* 1–27. https://doi.org/10.1111/lapo.12181