The Implicit Enemy: Countering Unconscious Racial Bias in the Courts

Elizabeth Sykes, Sumeyye Erdem, Adesola Orogade & Shakira Pagan

Faculty Mentor: Dr. Christine Ruva

University of South Florida

Introduction

- **Implicit racial bias** can have an impact on juror decision making and may impact trial outcome (Kang et al., 2012).
- **Juror racial bias** where Black and White jurors show same-race bias by displaying leniency towards defendants of their own race (Mazella & Feingold, 1994; Mitchell et al., 2005).
- **Dual process theory** and **aversive racism theory** provide an explanation for juror bias in sentencing and verdict decisions (Kahneman, 2011; Dovidio & Gaertner, 1994).
- **Counterfactual thinking** ("what if" thinking) has been shown to mediate variables of juror decision-making (Alvarez & Miller, 2016)
- **Unconscious Bias Juror (UBJ) video** is used to combat implicit bias within the courtroom (United States District Court, Western District of Washington, 2017).
    - 11-minute video with a judge and two attorneys as speakers
    - Exemplifies a White male judge who is "smart" or "tough"
- The study examined the effect of the UBJ video, defendant race (Black vs. White), and juror race (Black vs. White) on jurors' decisions in a murder trial (*NJ v Johnson*).

Hypotheses

H1. Jurors who are of the defendant's same race will be less likely to find him guilty, and rate him higher in credibility.

H2. The unconscious bias juror video will have the greatest impact on verdicts and credibility ratings for White jurors.

H3. Credibility and counterfactual endorsement will mediate the effect of race on verdict. Lower credibility ratings, lower endorsement of suicide counterfactuals, and higher endorsement of murder counterfactuals will result in a greater likelihood of a guilty verdict.
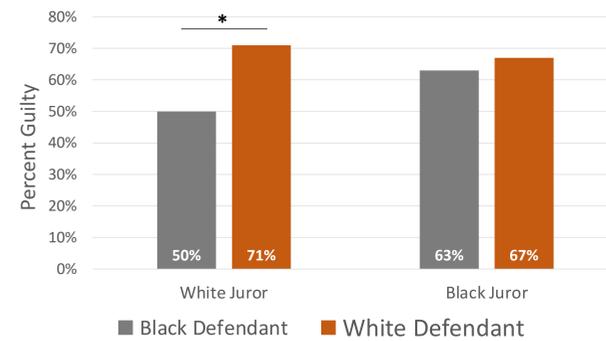
Methods

- **Participants**: (*N*=554) jury-eligible community members recruited through Qualtrics panels, (49%) female, age 18-75 (*M*=46.53, *SD*=17.07), that self-identified as either Black (50%) or White (50%).
- **Design**: 2 (Race of Defendant: Black vs White) x 2 (Race of the Juror: Black vs White) x 2 (UBJ Video: Yes vs. No) between-subjects design. Black and White jurors were randomly assigned to defendant race and UBJ video conditions.
- **Unconscious Bias Juror (UBJ) Video**: The UBJ video (U.S. District Court, Western District of Washington, 2017) described above, is given as part of Voir Dire before a trial. The video defines and provides examples of implicit bias, including implicit racial bias.
- **Trial Stimuli**: The trial transcript was edited from an actual trial and has been used in prior research (Ruva et al., 2007). The defendant was accused of killing his wife and charged with second degree murder. The defendant maintained his wife was trying to commit suicide and the gun went off as he tried to take it away.
- **Measures**: Participants gave a verdict, as well as ratings for credibility (α = .90), and counterfactual endorsement of murder and suicide (α = .71 and .73).
- **Pilot work** was conducted to examine trial ambiguity and racial bias effect, results were presented in a separate study at the 2021 USF Undergraduate Research Conference.
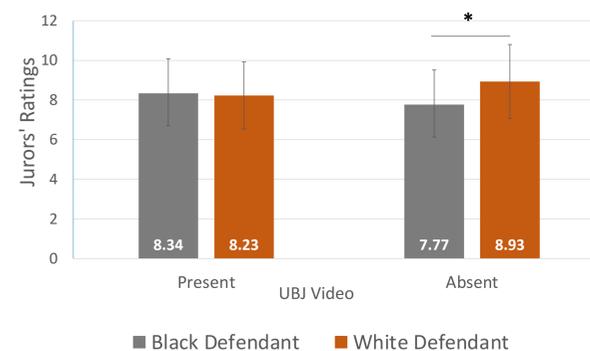
Results

Figure 1. *Interaction of Juror Race X Defendant Race on Murder Verdict*

Note. * signifies a *p* < .05

Verdict. Defendant race had a significant effect on verdict, qualified by a defendant race x juror race interaction, χ^2 (1, *N* = 554) = 3.87, *p* = .05 (Figure 1). Contrary to H1 White jurors were more likely to find the White defendant guilty than the Black defendant. Contrary to H2 the UBJ video did not significantly affect murder verdict.
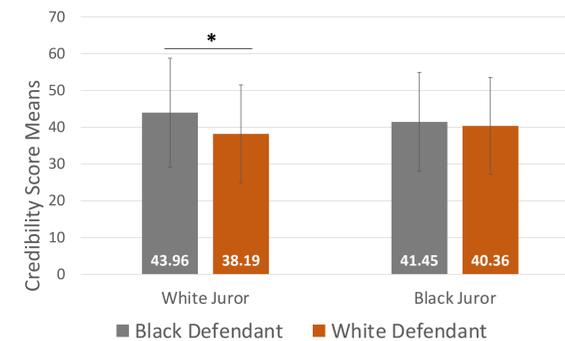
Figure 3. *Jurors' Endorsement of Counterfactuals*

Note. * signifies a *p* < .05. The murder counterfactual scale consisted of two "if-then" statements, for example "If [the victim] had not asked asked for a divorce, she would still be alive today."

Counterfactual Endorsement (exploratory analyses). When the UBJ video was absent, endorsement of murder counterfactuals was higher for the White defendant than the Black defendant, *t*(1, 284) = 2.90, *p* = .004 , *d* = -.33, 95% CI [-.56, -.11] when it was present, defendant race did not affect endorsement of murder counterfactuals *t*(1, 545) = -.25, *p* = .85 (Figure 3).
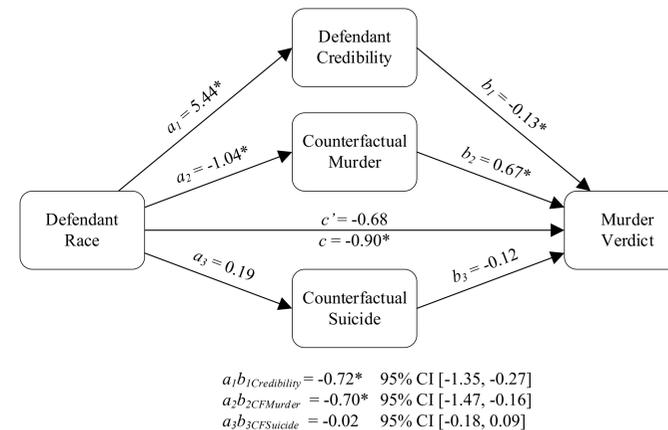
Figure 2. *Interaction of Juror Race X Defendant Race on Defendant Credibility*

Note. * signifies a *p* < .05. The credibility scale consisted of 10 items. Items were rated on a 7-point Likert-type scale.

Defendant Credibility. Defendant race had a significant effect on jurors' ratings of defendant credibility, qualified by a defendant race x juror race interaction, *t*(1, 276) = 3.31, *p* = .001, *d* = .39, 95% CI [-.63,-.15]. The interaction was due to defendant race only affecting White jurors (Figure 2). Contrary to H1 White jurors rated the Black defendant more credible than the White defendant. Contrary to H2 the UBJ video did not affect defendant credibility.

Figure 4. *White Jurors: Mediation Model on Murder Verdict*

$a_1b_{1Credibility} = -0.72^*$ 95% CI [-1.35, -0.27]
$a_2b_{2CFMurder} = -0.70^*$ 95% CI [-1.47, -0.16]
$a_3b_{3CFSuicide} = -0.02$ 95% CI [-0.18, 0.09]

Note. * signifies a *p* < .05. Race and verdict are dummy coded: 0 = White, 1 = Black and 0 = not guilty, 1 = guilty

Mediation Analyses. Only White jurors' verdicts were affected by defendant race and thus we only present mediation analyses for White jurors. Mediation effects were examined using PROCESS Model 4 (Hayes, 2018). Defendant race had a direct effect on White jurors' ratings of defendant credibility (negatively associated with guilty verdicts) and their endorsement of murder counterfactuals (positively associated with guilty verdicts). Supporting H3 these variables mediated the effect of defendant race on murder verdicts (Figure 4).

Discussion

- White jurors displayed an overcorrection of bias in verdict and defendant credibility which may be explained by:
    - Aversive racism theory (Dovidio & Gaertner, 1994)
        - Recent events made race salient and activated jurors' egalitarian views
    - Flexible correction model (Wegener & Petty, 1997)
        - Jurors lacked proper awareness to correct their bias
- The UBJ video did not affect verdict or credibility, however it did affect murder counterfactuals, which mediated the effect of juror race on verdict.
- Prior research on debiasing suggests the UBJ video may be made more effective by:
    - Enabling the participant to take a more active role
    - Combining the video with an Implicit Association Test

Limitations
- The use of shortened trial transcript does not mimic the amount of information jurors receive at trial.
- Deliberation and decision making in diverse groups results in less biased outcomes (Sommers, 2006).

Future Directions
- Further research is needed to draw broader conclusions regarding the effectiveness of the UBJ video in correcting juror bias.
    - Different types of trials (e.g., civil vs. criminal) and charges
- Evaluating improvements to the UBJ video.
- Examining victim race may help gain more information on specific scenarios where bias is most prevalent.

Conclusion

- While the video was not effective at reducing bias, it does have an impact on juror decision making and can be made more effective.
- Further research is needed to evaluate effective methods of combatting racial bias within the courtroom.

Use this QR code to access our references and supplemental results!

UNIVERSITY of SOUTH FLORIDA