The impact of virtual reality on implicit racial bias and mock legal decisions

Natalie Salmanowitz*

Stanford Program in Neuroscience and Society, Stanford Law School, Stanford, CA, 94305, USA
Corresponding author. E-mail: nsalmanowitz@jd19.law.harvard.edu

ABSTRACT

Implicit racial biases are one of the most vexing problems facing current society. These split-second judgments are not only widely prevalent, but also are notoriously difficult to overcome. Perhaps most concerning, implicit racial biases can have consequential impacts on decisions in the courtroom, where scholars have been unable to provide a viable mitigation strategy. This article examines the influence of a short virtual reality paradigm on implicit racial biases and evaluations of legal scenarios. After embodying a black avatar in the virtual world, participants produced significantly lower implicit racial bias scores than those who experienced a sham version of the virtual reality paradigm. Additionally, these participants more conservatively evaluated an ambiguous legal case, rating vague evidence as less indicative of guilt and rendering more Not Guilty verdicts. As the first experiment of its kind, this study demonstrates the potential of virtual reality to address implicit racial bias in the courtroom setting.

KEYWORDS: cognitive bias, criminal law, Implicit Association Test, implicit racial bias, legal decision-making, virtual reality

I. INTRODUCTION

Eric Garner. Freddie Gray. Alton Sterling. In the last three years, these men have become household names, inspiring an onslaught of media coverage, political debate, and community protest.[1] Not only have these events brought to light issues of explicit bias

---

1 See eg Damien Cave & Oliver Rochelle, *The Raw Videos That Have Sparked Outrage over Police Treatment of Blacks*, NEW YORK TIMES (Aug. 19, 2017), http://www.nytimes.com/interactive/2015/07/30/us/police-videos-race.html?_r=0 (accessed Mar. 13, 2018); Bijan Stephen, *Social Media Helps Black Lives Matter*

and systemic racism, but a new buzzword has also risen to center stage: implicit bias.[2] While this term has only recently enjoyed widespread attention, it first became popular among social psychologists in the 1990s, and refers to automatic, unconscious associations that people make on a daily basis.[3]

Implicit biases stem from culturally ingrained stereotypes,[4] and frequently diverge from consciously held beliefs;[5] as a result, a police officer without any overtly racist intentions can still demonstrate strong implicit biases against black individuals.[6] While police encounters offer a prime example of a setting in which implicit biases can lead to consequential actions, the impact of these unconscious associations is far reaching, from discrimination in the workforce, to the quality of medical care, to conviction, and sentencing decisions in the courtroom.[7]

The latter case, and the focus of this article, is particularly concerning given the legal system's ostensible emphasis on impartial trials and presumptions of innocence.[8] Between controlled laboratory studies involving hypothetical cases and analyses using real-world data, researchers have noted disparities in conviction decisions, evaluations of evidence, and sentencing lengths for black and white defendants.[9] Despite mounting research over the past two decades, the impact of implicit bias in the courtroom remains a problem without a solution. This is not to say that scholars have not proposed a wide array of potential strategies;[10] however, existing suggestions are either unsuitable for the courtroom setting or unlikely to cause a significant reduction in bias.

    Fight the Power, Wired (Nov. 2015), https://www.wired.com/2015/10/how-black-lives-matter-uses-social-media-to-fight-the-power/ (accessed Mar. 13, 2018).

2   See eg Shankar Vedantam, *In the Air We Breathe*, NPR (June 5, 2017, 10:07 PM), http://www.npr.org/templates/transcript/transcript.php?storyId=531587708 (accessed Mar. 13, 2018) (noting recent discussions of implicit biases, including those by 2016 Presidential Candidate Hillary Clinton).

3   See David M. Amodio, *The Neuroscience of Prejudice and Stereotyping*, 15 Nat. Rev. Neurosci. 670, 670–71 (2014); Mahzarin R. Banaji et al., *How (Un)ethical Are You?* Harv. Bus. Rev., Dec. 2003, at 56, 58–60.

4   Amodio, *supra* note 3, at 675.

5   Banaji et al., *supra* note 3, at 58; Dale K. Larson, *A Fair and Implicitly Impartial Jury: An Argument for Administering the Implicit Association Test During Voir Dire*, 3 DePaul J. for Soc. Just. 139, 147 (2010); Anna Roberts, *(Re)forming the Jury: Detection and Disinfection of Implicit Juror Bias*, 44 Conn. L. Rev. 827, 834 (2012).

6   See Cheryl Staats et al., *State of the Science: Implicit Bias Review 2016*, 4 Kirwan Inst. for Study Race Ethnicity 25–26 (2016), http://kirwaninstitute.osu.edu/wp-content/uploads/2016/07/implicit-bias-2016.pdf (accessed Sept. 1, 2017); Vedantam, *supra* note 2.

7   See generally Alexander R. Green et al., *Implicit Bias Among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients*, 22 J. Gen. Internal Med. 1231, 1232–37 (2007) (noting implicit bias in the medical setting); Jerry Kang et al., *Implicit Bias in the Courtroom*, 59 UCLA L. Rev. 1124, 1221–26 (2012) (describing implicit bias among judges and jurors); Staats et al., *supra* note 6, at 17–41 (discussing the impact of implicit biases in settings such as the justice system, education, healthcare, housing, and employment).

8   See U.S. Const. amends. V, VI.

9   Kang et al., *supra* note 7 at 1126–68; Justin D. Levinson & Danielle Young, *Different Shades of Bias: Skin Tone, Implicit Racial Bias, and Judgments of Ambiguous Evidence*, 112 W. Va. L. Rev. 307, 331–39 (2010); Kimberly Papillon, *The Court's Brain: Neuroscience and Judicial Decision Making in Criminal Sentencing*, 49 Ct. Rev. 48, 53 (2013); Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?* 84 Notre Dame L. Rev. 1195, 1121–26 (2009); Staats et al., *supra* note 6, at 19–22.

10   See eg Nilanjana Dasgupta & Anthony G. Greenwald, *On the Malleability of Automatic Attitudes: Combating Automatic Prejudice with Images of Admired and Disliked Individuals*, 81 J. Personality & Soc. Psychol. 800, 802–08 (2001); Kang et al., *supra* note 7, at 1174–77; Larson, *supra* note 5, at 162–71; Casey Reynolds, *Implicit Bias and the Problem of Certainty in the Criminal Standard of Proof*, 37 Law & Psychol. Rev. 229, 248 (2013); Roberts, *supra* note 5, at 873–74; Samuel R. Sommers & Phoebe C. Ellsworth, *White Juror Bias: An Investigation of Prejudice Against Black Defendants in the American Courtroom*, 7 Psychol. Pub. Pol'y & L. 201, 216–21 (2001); Staats et al., *supra* note 6, at 43–49.

Given the current state of research, there is a pressing need for novel and creative approaches. This article fills the void, stepping out of the box to introduce and empirically evaluate a revolutionary method: namely, a five-minute virtual reality (VR) paradigm, in which individuals embody an avatar of a different race. Though this study is only a preliminary foray into this potentially expansive field of research, the results are promising. Where existing suggestions fall short, VR offers a mechanism that is interactive and engaging, subtle yet potent, and appropriate for the unique demands of the courtroom environment.

This article proceeds as follows. Part I provides a brief background on implicit bias in the courtroom, covering findings from previous studies and summarizing existing proposals for mitigation. Part I also explains fundamental concepts underlying VR and reviews the current literature on VR for implicit bias reduction. Part II discusses the design, methods, and results of the article's primary experiment (Study 1), which examined the impact of a subtle VR paradigm on both Implicit Association Test (IAT) scores and evaluations of a mock crime scenario. As a follow-up to Study 1, Part III describes two additional studies (Studies 2 and 3) that further investigated the ability of mock legal scenarios to capture the effects of implicit racial biases. While Part IV discusses the studies' limitations, the article ultimately concludes with a call for concerted research efforts in this exciting and innovative field.

II. BACKGROUND

II.A. Implicit Bias in the Courtroom: Evidence of the Problem

In 1998, a team of social psychologists created a test for implicit bias that has since dominated and inspired an entire body of research.[11] The IAT is a computer-based response-time task in which participants sort words or portraits into relevant categories.[12] Specifically, the race version of the IAT examines the strength of positive and negative associations related to black and white individuals.[13] This seminal test has been conducted with millions of Americans, 75% of whom demonstrate a preference toward white faces.[14] While the IAT is not designed to diagnose or predict biased behavior on an individual level, it can successfully predict discriminatory behavior among groups of decision-makers.[15] Over 50 studies have examined the predictive validity of the IAT,[16] and although meta-analyses have reported small to moderate effect sizes,[17] the creators of the IAT have noted that even small effects can have profound societal impacts—particularly when a group (such as a police department) makes decisions

---

[11] See Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J. PERSONALITY & SOC. PSYCHOL. 1464, 1465–78 (1998).

[12] *Id.* at 1466–67.

[13] *Id.* at 1465. A more detailed explanation of this test is provided on pages 14–15.

[14] Banaji et al., *supra* note 3, at 59.

[15] Anthony G. Greenwald et al., *Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects*, 108 J. PERSONALITY & SOC. PSYCHOL. 553, 557 (2015).

[16] *Id.* (noting the number of studies included in meta-analyses that looked specifically at discrimination against stigmatized groups).

[17] Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 97 J. PERSONALITY & SOC. PSYCHOL. 17, 17–32 (2009) (reporting moderate effect sizes); Frederick L. Oswald et al., *Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT-Criterion Studies*, 105 J. PERSONALITY & SOC. PSYCHOL. 171, 171–92 (2013) (reporting small effect sizes and arguing that as a result, the IAT has little predictive validity).

that affect many people, or when slightly discriminatory acts have cumulative effects on a person over time.[18]

In addition to the original IAT, researchers have also created similar tasks illustrating implicit links between blacks and criminal guilt[19] as well as black men and lethal weapons.[20] For instance, in the Weapon Identification Task, participants are briefly shown a black or white male's face before seeing a picture of a weapon or a tool. When asked to identify the object displayed in the picture, participants recognize the weapon more quickly after seeing a black face, and are also more likely to misperceive the tool as a gun.[21] As if these findings were not sufficiently concerning on their own, the courtroom is a breeding ground for factors known to increase individuals' reliance on automatic associations—implicit biases are most influential during stressful situations that are cognitively demanding, emotionally draining, and rife with uncertainty.[22]

To explore the influence of these bias on trial outcomes, researchers have used both mock scenarios and data from actual cases. With respect to mock situations, although studies have revealed mixed findings,[23] black defendants are generally considered guilty more often than their white counterparts, especially when race is introduced in a subtle manner.[24] For example, Samuel Sommers and Phoebe Ellsworth found that in situations where race was made highly salient to mock jurors, participants did not significantly differ in their evaluations of white versus black defendants; however, when race was not overtly brought to participants' attention, black defendants received more guilty verdicts and harsher sentence recommendations than white defendants.[25] In 2009, Jeffrey Rachlinski and colleagues expanded on this study and explored racial bias among black and white judges.[26] During one experiment, the authors subliminally primed judges with either neutral words or words stereotypically related to black individuals; after reading a hypothetical case summary, judges with stronger implicit preferences for whites generally recommended tougher sentences when primed with the race-based words.[27] Yet, when using one of the highly race-salient vignettes from Sommers and Ellsworth's study, Rachlinski and colleagues observed an interesting split between black and white judges.[28] Black judges were less likely to find the defendant guilty when

---

[18] Greenwald et al., *supra* note 15, at 557–60 (responding to Oswald et al.'s (2013) criticism and explaining the observed differences between the two meta-analyses).

[19] Justin D. Levinson et al., *Guilty by Implicit Racial Bias: The Guilty/Not Guilty Implicit Association Test*, 8 Ohio St. J. Crim. L. 187, 189–90 (2010).

[20] B. Keith Payne, *Weapon Bias: Split-Second Decisions and Unintended Stereotyping*, 15 Curr. Direct. Psychol. Sci. 287, 287 (2006); B. Keith Payne et al., *Best Laid Plans: Effects of Goals on Accessibility Bias and Cognitive Control in Race-Based Misperceptions of Weapons*, 8 J. Exp. Soc. Psychol. 384, 388–95 (2002).

[21] Payne, *supra* note 20, at 184, 188–90.

[22] See Daniel Kahneman, Thinking, Fast and Slow 41 (2011); Lucius Caviola & Nadira S. Faber, *How Stress Influences our Morality*, The Inquisitive Mind (Oct. 15, 2014), http://www.in-mind.org/article/how-stress-influences-our-morality (accessed Mar. 13, 2018); Larson, *supra* note 5, at 148–49; Papillon, *supra* note 9, at 52.

[23] See eg Tara L. Mitchell et al., *Racial Bias in Mock Juror Decision-Making: A Meta-Analytic Review of Defendant Treatment*, 29 Law & Hum. Behav. 621, 623–24 (2005).

[24] Sommers & Ellsworth, *supra* note 10, at 220; Rachlinski et al., *supra* note 9, at 1223.

[25] Sommers & Ellsworth, *supra* note 10, at 217–20.

[26] Rachlinski et al., *supra* note 9, at 1204–11.

[27] *Id.* at 1211–15.

[28] *Id.* at 1217–19.

he was black, and were more confident in this decision; comparatively, white judges' verdicts did not significantly differ when the defendant was white or black.[29]

In each of these studies, the authors point to race salience effects as a likely explanation for the white participants' seemingly unbiased responses.[30] According to this theory, when race is strongly emphasized, participants become more attuned to the point of the experiment and respond in ways that are more socially desirable (ie white participants might think that the researchers are expecting them to demonstrate prejudice, and therefore make concerted efforts to appear as unbiased as possible).[31] In contrast, when race is more subtly included in a case summary, participants are less vigilant about monitoring their responses for potential biases.[32] An experiment by Justin Levinson and Danielle Young further illustrates this point.[33] In their study, participants viewed a timed slideshow of photos from a crime scene, including one image that displayed security camera footage of the defendant.[34] Half of the participants saw a dark-skinned perpetrator in the photo, whereas the other half saw a light-skinned version of the same image.[35] Even though many participants could not recall the race of the defendant, participants who viewed the dark-skinned version were more convinced of the defendant's guilt and considered ambiguous evidence more supportive of this verdict than participants who saw the light-skinned defendant.[36] Moreover, participants' IAT scores predicted their verdicts and evidence ratings.[37]

It is worth noting that the studies cited above involve simplified scenarios that are not fully comparable to real-world trials. As a result, these studies do not prove that implicit racial bias is rampant among judges and juries. However, even if actual verdicts and sentence lengths are not directly impacted by implicit biases, these studies still present cause for concern. The fact that a defendant's race can disproportionately influence aspects of the decision-making process runs counter to core notions of impartiality, especially when many people harbor implicitly negative associations toward black individuals upon entering the courtroom.

To gauge the real-world impact of race in the justice system, one can look to disparities in incarceration and sentencing lengths. According to the Bureau of Justice Statistics, black men not only outnumber men of all other races in prison, but they also are up to 10.5 times more likely to be incarcerated than their white peers.[38] Similar disparities exist in sentencing decisions.[39] For example, within white and black subgroups of a Florida prison population, convicts with stronger Afrocentric facial features received longer sentences than inmates with comparable criminal records.[40] Some scholars have expressed the unsurprising nature of these findings: given that decisions to incarcerate

---

29   *Id.*

30   See Sommers & Ellsworth, *supra* note 10, at 220, 222; Rachlinski et al., *supra* note 9, at 1223–24.

31   Sommers & Ellsworth, *supra* note 10, at 216–21.

32   Sommers & Ellsworth, *supra* note 10, at 216–21; Rachlinski et al., *supra* note 9, at 1223–24.

33   See Levinson & Young, *supra* note 9, at 331–39.

34   *Id.* at 332.

35   *Id.*

36   *Id.* at 336–38.

37   *Id.* at 338.

38   Ann Carson, Bureau of Justice Statistics, Dep't of Justice, NCJ 248955, Prisoners in 2014 15 (2015).

39   See eg Irene V. Blair et al., *The Influence of Afrocentric Facial Features in Criminal Sentencing*, 15 Psychol. Sci. 674, 676–78 (2004).

40   *Id.*

and sentence an individual involve calculations of dangerousness, implicit associations between black men and violent crime are likely to hold substantial weight in these analyses.[41]

II.B. Current Suggestions for Mitigation

In light of this growing body of literature, legal scholars and psychologists have proposed several strategies to combat implicit bias in the courtroom. These strategies generally fall into one of two categories: (1) reduction of bias on an individual level and (2) removal of bias on a systemic level. Starting with the first set of strategies, one of the most commonly proposed suggestions involves establishing education or training programs to raise awareness about implicit bias. Pointing to race salience effects, many scholars argue that if judges and jurors were educated about the problem of implicit bias, they could better regulate their impressions during trials and avoid making biased judgments.[42] To help promote self-regulation of implicit biases, advocates have recommended a variety of techniques ranging from the use of checklists when making decisions about a case[43] to discussing implicit bias when delivering jury instructions.[44] In addition, scholars also suggest addressing the underlying stereotypes that give rise to implicit biases by exposing judges and jurors to counterstereotypical exemplars of black individuals;[45] for example, placing portraits of revered black figures in the courthouse could weaken associations between black men and violent crime.[46] Similarly, researchers suggest using perspective-taking exercises to attenuate implicit biases, as stepping into the shoes of one's racial outgroup might diminish the potency of stereotypes and reduce perceived distinctions between members of different races.[47]

Whereas the first set of suggestions attempts to reduce individuals' biases, the second set aims to remove bias from the trial system at large. For instance, some scholars propose altering the composition of judicial panels and jury pools, either by dismissing those with high IAT scores or recruiting a more diverse combination of individuals to counteract the biases' effects.[48] Another strategy involves 'blinding' judges or juries to the race of the defendant;[49] in fact, one scholar has even suggested moving toward

---

[41] See Kang et al., *supra* note 7, at 1150; Papillon, *supra* note 9, at 51, 54.

[42] Kang et al., *supra* note 7, at 1174–77; Reynolds, *supra* note 10, at 248; Roberts, *supra* note 5, at 865–66.

[43] Pamela Casey et al., *Addressing Implicit Bias in the Courts*, 49 CT. REV. 64, 67 (2013).

[44] Mark W. Bennett, *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*, 4 HARV. L & POL'Y REV. 149, 169 (2010).

[45] See eg Patricia G. Devine et al., *Long-Term Reduction in Implicit Race Bias: A Prejudice Habit-Breaking Intervention*, 48 J. EXP. SOC. PSYCHOL. 1267, 1268 (2012); Kang et al., *supra* note 7, at 1171; Staats et al., *supra* note 6, at 47.

[46] See Dasgupta & Greenwald, *supra* note 10, at 802–08; Kang et al., *supra* note 7, at 1169–72.

[47] See Calvin K. Lai et al., *Reducing Implicit Racial Preferences: I. A Comparative Investigation of 17 Interventions*, 143 J. PSYCHOL.GEN. 1765, 1769–70 (2014); Staats et al., *supra* note 6, at 25, 73.

[48] Kang et al., *supra* note 7, at 1180–81; Larson, *supra* note 5, at 167; Rachlinski et al., *supra* note 9, at 1231; Sommers & Ellsworth, *supra* note 10, at 221–22.

[49] See Sunita Sah et al., *Blinding Prosecutors to Defendants' Race: A Policy Proposal to Reduce Unconscious Bias in the Criminal Justice System*, 1 BEHAV. SCI. POL'Y 69, 72–74 (2015) (proposing the use of blinding in prosecutorial decisions and noting its potential application to jurors and judges); see also BLINDING AS A SOLUTION TO BIAS: STRENGTHENING BIOMEDICAL SCIENCE, FORENSIC SCIENCE, AND LAW 25–36, 265–75, 319–31 (Christopher T. Robertson & Aaron S. Kesselheim eds., 2016) (providing additional detail on the concept of blinding, particularly in the legal context).

virtual courtrooms, in which judges and jurors would experience trials through VR, seeing a neutral-skinned avatar instead of the actual defendant.[50]

While these proposals have their benefits, none are likely to offer sufficiently effective solutions. To start, although raising awareness is a useful endeavor, this tactic overlooks fundamental issues with implicit biases and the courtroom environment. Not only are these biases extremely hardwired, automatic, and difficult to self-regulate,[51] but concentrating on a trial subsumes a large portion of a judge or juror's available bandwidth.[52] The remaining cognitive resources are likely too limited to allow hypervigilant self-monitoring, even after being reminded about the perils of implicit biases.[53] And although checklists may promote more careful deliberation by slowing down the decision-making process, they ignore the fact that individuals overestimate their own impartiality[54] and are therefore likely to believe their decisions are grounded in race-neutral factors.[55] Perspective-taking exercises elicit similar cognitive resource constraints,[56] and a short, impersonal exposure to counterstereotypical exemplars cannot be expected to counteract a lifetime of ingrained mental associations.[57] Weakening stereotypes is also unlikely to reduce bias in the courtroom setting. Successful applications of this strategy often involve juxtaposing positive black counterstereotypical exemplars against negative white figures,[58] which further encourages racial stratification rather than limiting it. Likewise, increasing race salience during trials is more undesirable than it is beneficial. By focusing jurors' and judges' attention on the color of a defendant's skin, race becomes a highly emphasized extralegal factor in the case. If part of the problem with implicit biases is that a defendant's race should not play a role in trials one way or the other, increasing race salience seems to undermine efforts to combat racial bias in the courtroom.

Systemic approaches to bias reduction pose their own sets of challenges. First, excluding potential jurors and judges based on IAT scores might severely limit the number of individuals available to hear a case, especially given the prevalence of biases across the country.[59] Second, diversification efforts will frequently be confined to juries, as many cases involve a single judge (who therefore cannot benefit from having a diverse

---

50 ADAM BENFORADO, UNFAIR: THE NEW SCIENCE OF CRIMINAL INJUSTICE 266–71 (2015).

51 Amodio, *supra* note 3, at 679; Banaji et al., *supra* note 3, at 61–62.

52 See KAHNEMAN, *supra* note 21, at 41–42 (noting that cognitively demanding situations can deplete the available resources that individuals can devote to other tasks).

53 See *id.*

54 See eg Kang et al., *supra* note 7, at 1173–74.

55 Banaji et al., *supra* note 3, at 56 (explaining that most people genuinely believe their decisions are not influenced by race); *cf.* Michael I. Norton et al., *Bias in Jury Selection: Justifying Prohibited Peremptory Challenges*, 20 J. BEHAV. DECISION MAKING 467, 467–77 (2007) (finding that mock attorneys used gender as a factor when making peremptory challenges but relied on gender-neutral information to justify their decisions).

56 Sun Joo Ahn et al., *The Effect of Embodied Experiences on Self-Other Merging, Attitude, and Helping Behavior*, 16 MEDIA PSYCHOL. 7, 9 (2013).

57 Jennifer A. Joy-Gaba & Brian A. Nosek, *The Surprisingly Limited Malleability of Implicit Racial Evaluations*, 41 SOC. PSYCHOL. 137, 138–45 (2010).

58 Dasgupta & Greenwald, *supra* note 10, at 802–05; Joy-Gaba & Nosek, *supra* note 57, at 138–40; Lai, *supra* note 47, at 1780.

59 Chris Mooney, *Across America, Whites Are Biased and They Don't Even Know It*, WASH. POST (Dec. 8, 2014), https://www.washingtonpost.com/news/wonk/wp/2014/12/08/across-america-whites-are-biased-and-they-dont-even-know-it/ (accessed Mar. 13, 2018); *Race IAT 2002-2015*, PROJECT IMPLICIT, https://osf.io/52qxl/ (accessed Sept. 1, 2017).

panel of judges to check one another's biases). Regardless, even minorities have been shown to possess implicit biases in favor of the majority group,[60] making it unlikely that biases among jurors would be cancelled out simply by having a diverse combination of backgrounds represented. Third, while blinding may offer a promising solution in the context of prosecutors—whose decisions frequently do not require seeing the defendant in person[61]—it is difficult to envision how blinding could offer a practical solution during the trial itself. The nature of our trial system requires jurors and judges to be able to see the defendant in real time.[62]

Although this analysis paints a fairly bleak picture, VR could offer the effective and appropriate mitigation strategy that is currently lacking. Specifically, judges and jurors could put on a VR headset for five minutes before trial and enter a virtual world in which they embodied an avatar of a racial outgroup. Through this VR experience, judges and jurors could receive the benefits of self-regulation, perspective-taking, and stereotype reduction strategies without suffering the costs noted above. As opposed to jury instructions, checklists, or quick exposures to counterstereotypical exemplars, VR provides an interactive and engaging platform that can induce potent effects without increasing cognitive load.[63] And unlike suggestions that limit the number of individuals who can serve, court systems could employ VR with sitting judges and empaneled jurors as a training exercise. In addition, the flexibility of the VR design process enables participants to embody a member of a different race while minimizing race salience concerns.[64]

II.C. VR and Implicit Racial Bias

The promise of VR lies not just in satisfying the components listed above, but also in its underlying mechanisms. Through physically embodying an avatar, VR can induce an effect called the 'body ownership' illusion.[65] In most paradigms eliciting this effect, tactile stimulation is synchronously applied to a participant's actual body part as well as to some other entity (whether that be a rubber hand, an image of a person on a screen, or an embodied avatar); by watching this other entity be touched while simultaneously feeling their own body's reaction to the stimulation, people can temporarily feel as though this 'other body' belongs to them.[66] For instance, in the 'enfacement' version of this illusion, individuals watch a video in which someone's face is stroked on the

---

[60] Banaji et al., *supra* note 3, at 59; *Frequently Asked Questions*, PROJECT IMPLICIT, https://implicit.harvard.edu/implicit/faqs.html (accessed Sept. 1, 2017).

[61] Sah, *supra* note 49, at 72.

[62] For example, the Confrontation Clause affords defendants the right to cross-examine witnesses before the triers of fact, thereby allowing jurors to scrutinize the witness's and defendant's demeanors during the witness's testimony; the jury cannot properly determine the witness's credibility if it cannot physically see the defendant (and the interaction between the accuser and accused). See U.S. Const. amend. VI; Coy v. Iowa, 487 U.S. 1012, 1019 (1987) (discussing the purposes of the Confrontation Clause and the ability of the jury to assess credibility by examining the witness and defendant's non-verbal behavior during the testimony).

[63] See Ahn et al., *supra* note 56, at 9–10.

[64] For detailed information on how VR would actually be implemented in the courtroom, including practical and ethical considerations, see Natalie Salmanowitz, *Unconventional Methods for a Traditional Setting: The Use of Virtual Reality to Reduce Implicit Racial Bias in the Courtroom*, 15 UNH L. REV. 117, 117–160 (2016).

[65] Lara Maister et al., *Changing Bodies Changes Minds: Owning Another Body Affects Social Cognition*, 19 TRENDS COGN. SCI. 6, 7 (2015).

[66] *Id.* at 7–10.

cheek while the participant's own cheek is stroked concurrently.[67] This effect produces a profound blurring of 'self-other' boundaries, which can in turn reduce implicit racial biases.[68] Specifically, by weakening distinctions between oneself and someone of a different race, the negative associations that are often ascribed to that race can become less potent.[69]

In VR, body ownership illusions can be triggered when the movements of the participant's actual limbs are tracked synchronously with those of the embodied avatar. In a study by Tabitha Peck and colleagues, light-skinned participants entered an immersive virtual environment in which they possessed an avatar's physical body.[70] Using motion-tracking equipment, participants saw their avatar reflected in a mirror and watched as their avatar's arms and legs matched their own body's movements.[71] By analysing IAT scores before and after this VR experience, the authors discovered decreased implicit biases among participants who embodied dark-skinned avatars.[72] A recent study elicited similar results—after embodying a black avatar during a virtual Tai Chi lesson, participants demonstrated reduced IAT scores when they took the test one week after the VR experience.[73] Conversely, the only other study using VR to explore implicit bias found opposite effects: participants who embodied a black avatar demonstrated stronger implicit biases than those who embodied a white avatar.[74] However, this study included multiple features intended to invoke stereotypes, prompting participants to not just embody a black avatar's physical being, but also experience the world through a black individual's perspective.[75] Collectively, these studies suggest that VR can have significant effects on implicit bias scores, but the direction of this impact varies with the design of the VR paradigm. Ideally, a VR exercise could be configured for the courtroom setting, such that the experience reduces judges' and jurors' implicit racial biases on the IAT, while also decreasing the discrepancy in legal judgments that mock crime studies have observed between cases involving black and white defendants.[76]

Given the paucity of literature on VR and implicit racial bias, a multitude of questions remain unanswered. The present experiment sought to both build upon existing studies, and, for the first time, apply this research to the courtroom setting. Accordingly, the goals of the present experiment were twofold: (1) to design an extremely subtle VR paradigm, capable of impacting IAT scores without increasing race salience, and (2) to determine if and how the VR experience influences legal decisions.

[67] *Id.* at 7–8; Maria-Paola Paladino et al., *Synchronous Multisensory Stimulation Blurs Self-Other Boundaries*, 21 PSYCHOL. SCI. 1202, 1203 (2010).

[68] Maister et al., *supra* note 65, at 9–10; Paladino et al., *supra* note 67, at 1205.

[69] See Maister et al., *supra* note 65, at 8–10.

[70] Tabitha C. Peck et al., *Putting Yourself in the Skin of a Black Avatar Reduces Implicit Racial Bias*, 22 CONSCIOUS. & COGN. 779, 780–86 (2013).

[71] *Id.* at 780–82.

[72] *Id.* at 782–86.

[73] Banakou et al., *Virtual Embodiment of White People in a Black Virtual Body Leads to a Sustained Reduction in Their Implicit Racial Bias*, 10 FRONT. HUM. NEUROSCI. 601 (2016).

[74] Victoria Groom et al., *The Influence of Racial Embodiment on Racial Bias in Immersive Virtual Environments*, 4 SOC. INFLUENCE 231, 231–45 (2009).

[75] *Id.* at 8.

[76] Levinson & Young, *supra* note 9, at 344–45; Sommers & Ellsworth, *supra* note 10, at 220–21.

III. STUDY 1: VR FOR THE COURTROOM SETTING

III.A. Method

III.A.1. Participants

The study involved 92 participants (46 male), from both the Stanford University community and the surrounding local region.[77] All participants were 18 years or older (*Md* = 28 years), self-identified as Caucasian,[78] and with normal or corrected-to-normal vision. The sample was predominately liberal and highly educated. (See Table A1 in the Appendix for detailed demographic information.) The study was approved by Stanford University's Institutional Review Board, and each participant provided written informed consent prior to starting the experiment.

Fifty-one percent of participants were recruited through Stanford University Psychology Department's participant pool (comprised of both students and community members). Only participants who met the eligibility criteria (18 years or older, Caucasian, with normal or corrected-to-normal vision, and fluent in English) based on their recorded demographic information were able to view the study's listing on the recruitment database.

In order to increase the size of the participant pool, the remaining 49% of participants were recruited[79] through flyers posted around the local community, which directed respondents to an online eligibility survey. Respondents were asked a series of demographic questions including age, ethnicity, education level, political affiliation, gender, English language fluency, and vision quality. Respondents who satisfied the criteria listed above were notified of their eligibility for the study. Respondents who did not meet one or more of the eligibility specifications were directed to a separate message screen, thanking them for their interest and notifying them of their ineligibility to participate in the experiment. Given that respondents were asked to provide answers to a broad range of demographic questions, it is unlikely that selected respondents were aware of the racial characteristics defining the participant pool, thereby limiting the possibility that participants were primed to think about their race before starting the experiment.

III.A.2. Conditions

Two independent variables were included in this study: VR Type (Embodied vs. Sham) and Defendant Race (Black vs. White).

The first variable, VR Type, refers to the perspective through which participants experienced the virtual paradigm. Those in the Embodied condition adopted the physical appearance of a black avatar, which was matched to the participant in both age and

---

[77] Thirty-nine percent of participants were enrolled as Stanford students at the time of the experiment.

[78] Given the preliminary nature of this research, only Caucasian individuals participated in the study in order to simplify the experiment's design. While biases against outgroup members extend far beyond Caucasian individuals (see *Race IAT 2002-2015*, PROJECT IMPLICIT, https://osf.io/52qxl/ (accessed Sept. 1, 2017)), the extent of biases can diverge from one race to the next and depend on the manifestations of various cultural stereotypes across communities. Since implicit racial biases against black individuals are often strongest among white individuals (*Id.*), and the majority of existing literature on implicit bias in the courtroom focuses on Caucasian participants' biases (see Staats et al., *supra* note 6, at 61, 75–79), a Caucasian-only participant pool presented a logical starting point for this line of investigation.

[79] The differing methods of recruitment did not elicit any statistically significant impact on participant responses.

gender.[80] In contrast, those in the Sham condition experienced the virtual world without any connection to a physical body. Apart from the presence of an avatar in the Embodied condition, all other aspects of the virtual environment were equivalent for the two VR types.

The second variable, Defendant Race, indicates whether the defendant was white or black in two of the response variables (described below): the Mock Crime Scenario and the Follow-up Task. The defendant's race stayed constant across both response measures (ie if participants read about a black defendant in the Mock Crime Scenario, they would also read about a black defendant in the Follow-up Task).

Each participant was randomly assigned to one of four variable combinations (((1) Embodied VR/Black Defendant, (2) Embodied VR/White Defendant, (3) Sham VR/Black Defendant, and (4) Sham VR/White Defendant), with an equal number of participants ($n = 23$) in each group. In order to minimize race salience, participants in the Embodied condition were asked to select their avatar's identity from a pile of cards. As a cover story, participants were told that any combination of race, gender, and age was possible, and that the point of the experiment was simply to experience the world through a different perspective. In reality, all cards in the pile were identical, preselected to match the participant's age and gender. For example, a Caucasian, middle-aged, male participant selected his avatar's identity from a pile that exclusively contained cards with 'black, middle-aged, male' written on the back.[81] (See Figure 1 for an example of the avatars' features). Meanwhile, participants in the Sham condition were told the same story regarding the study's purpose, but without any mention of avatars or identities.

III.A.3. Procedure

After providing their informed consent, all participants completed a five-minute VR paradigm.[82] Before entering the virtual world, participants positioned an HTC Vive headset over their eyes, replacing visual input from the actual surroundings with the scenery from the virtual environment. Participants were also handed two wireless controllers (one for each hand), which tracked their movement and enabled interaction with objects in the virtual setting.

Once ready, participants entered the virtual world: a basic, non-descript room, with a table placed in the center and a mirror stretched across the front wall (see Figure 1). On the table was a pile of wooden blocks to the left and a paintbrush to the right. Adjacent to the table stood an easel with a blank canvas. In both conditions, participants were first asked to wave their 'arms' in front of the mirror. For participants in the Sham condition, only the handheld controllers were visible as they moved their arms in the virtual space. Those in the Embodied condition, however, could see their avatar when looking at the mirror in front of them as well as when looking down at their own arms. As participants moved in the actual world, their avatars moved synchronously in the virtual world. To provoke the body ownership illusion, participants in the Embodied

---

[80]   Four different avatar types were created for this experiment: a female young adult, a male young adult, a female middle-aged adult, and a male middle-aged adult.

[81]   This measure was inspired by the technique used in Elizabeth Behm-Morawitz et al., *The Effects of Virtual Racial Embodiment in a Gaming App on Reducing Prejudice*, 83 COMM. MONOGRAPHS 396, 405 (2016).

[82]   Participants in the Embodied condition selected the identity card after providing informed consent but prior to starting the VR exercise.
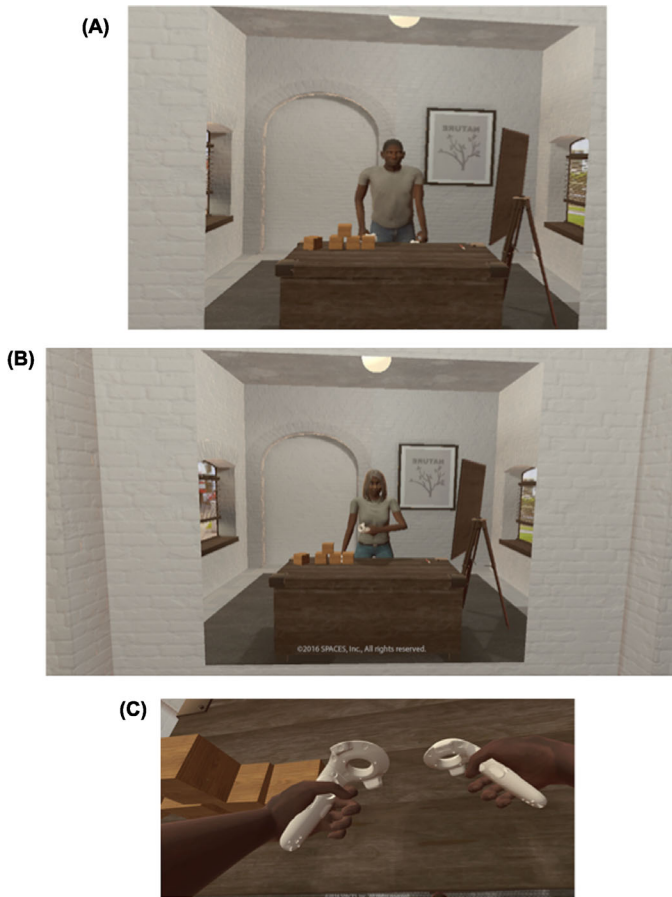
Figure 1. Examples of the visual display within the virtual reality paradigm. Panel A depicts the middle-aged male avatar looking in the virtual mirror, while Panel B portrays the same scene with the young-adult female avatar. Both avatars are holding controllers in their hands. Panel C captures the participants' view when looking down at their arms. All photos were generously provided by SPACES, Inc.

condition were specifically asked to touch their own skin while looking at their avatar's limbs. For instance, participants rubbed their knuckles together, experiencing tactual sensations from their actual body in conjunction with visual cues from their avatar's body. After becoming accustomed to their new physique (whether that be an invisible body or an avatar's body), all participants were asked to play with the blocks and paint on the canvas. These activities were included to subtly increase participants' familiarity with their new physical identities—by completing simple yet engaging tasks, participants could immerse themselves in the experience, seeing a black arm stacking blocks or painting lines without constantly thinking about their avatar's race during the process.

Following the virtual paradigm, participants removed the headset and completed six response measures on a computer: (1) a questionnaire regarding the experience in the virtual world, (2) the Mock Crime Scenario, (3) the race version of the IAT, (4) the gender version of the IAT, (5) the Symbolic Racism Scale (SRS), and (6) the Modern Sexism Scale (MSS).[83] The two gender-related items were included to reduce race salience, maintain the cover story, and assess the scope of the VR paradigm's impact.[84]

Five days after the in-lab portion of the experiment, participants were sent an email with a link to the Follow-up Task. During the informed consent process at the beginning of the experiment, participants agreed to complete an additional survey for extra compensation. They were told that this survey was part of a separate study concerning public opinion on recidivism and sentencing decisions. (To increase the credibility of this cover story, participants were informed that the five-day interval was designed to avoid potential confusion during the data collection processes of the two concurrent studies.)

III.B. Response Measures

III.B.1. VR Questionnaire

Participants filled out a seven-question survey measuring the degree to which they felt immersed in the virtual environment. Questions included both open-ended descriptions and ratings on a five-point scale. Participants in the two VR conditions saw distinct variations of this survey to account for the difference in paradigms. Five out of the seven questions were identical between the two surveys (one opened-ended description and four rated questions about immersion). For those in the Sham condition, two questions addressed the extent of perceived disembodiment (eg 'Did you feel like you had a body in the virtual world?'). For participants in the Embodied condition, the survey included questions assessing the degree of body ownership transfer. Specifically, participants answered the question: 'How strong was the feeling that the avatar's body was your own body?' These participants additionally completed a modified version of the Inclusion of Other in the Self (IOS) Scale.[85] For the IOS Scale, participants selected

---

83   The design of this study was inspired by the method used in Groom and colleagues' between-subject experiment, in which participants completed either a VR exercise or an imagination-only exercise prior to taking the IAT and other explicit attitude measures. Groom et al., *supra* note 74, at 5. While the present study could have elicited more consequential findings by measuring participants' IAT scores before and after the VR experience (see Peck et al., *supra* note 70, at 782), such a design ran the risk of skewing results on the Mock Crime Scenario, as participants might have come away from the first IAT expecting the study to be about race, and subsequently monitored their responses during the experiment in a manner consistent with race salience and social desirability effects.

84   In further efforts to avoid race salience effects and encourage natural, uncontrived responses, the order of response variables was kept consistent across all participants. In general, the order was structured such that the issue of race became incrementally pronounced across tasks, with the exception of the race and gender versions within the IAT and explicit attitude measures. Prior studies have observed diluted effects when participants take multiple IATs within the same experimental session. Bertram Gawronski et al., *Response Interference Tasks as Indirect Measures*, in COGNITIVE METHODS IN SOCIAL PSYCHOLOGY 78, 86 (K. Klauer et al., eds., 2012). Since racial bias was the main focus of this study, all participants took the race IAT before the gender IAT in order to preserve the reflexive nature of the response-time task.

85   Arthur Aron et al., *Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness,* 63 J. PERS. & SOC. PSYCHOL. 596, 597 (1992).

one of seven Venn diagrams, with circles respectively labeled 'Avatar' and 'Self'. The first Venn diagram on the scale presented two separate circles, with each subsequent diagram depicting increasing levels of overlap between the two circles.

III.B.2. Mock Crime Scenario

Following the VR Questionnaire, participants completed the full protocol from Levinson and Young's experiment.[86] After reading a case summary about an armed robbery at a convenience store, participants watched a timed slideshow of five pictures from the hypothetical crime scene. The third picture in the series showed footage from the convenience store's security camera, and depicted the masked perpetrator pointing a gun with his left hand while extending his right arm across the checkout counter. For participants assigned to the 'White Defendant' condition, the suspect's arms were visibly light-skinned, whereas those in the 'Black Defendant' condition saw dark-skinned arms. After the slideshow, participants were asked to evaluate 20 statements of ambiguous evidence on a seven-point scale, ranging from 'Very strongly indicating Not Guilty' to 'Very strongly indicating Guilty', with 'Neutral evidence' as the middle ground. (See Table A2 in the Appendix for a complete list of the evidence items that participants assessed.) Scores were combined from each evidence rating, such that the lowest possible score was 20 and the highest possible score was 140. Upon rating each piece of evidence, participants provided their final verdicts. In Levinson and Young's paradigm, participants were also asked to decide the extent of the defendant's guilt on a 0–100% scale.[87] However, actual jurors are rarely (if ever) asked to rate the *degree* of a defendant's guilt; in order to maintain similarity to Levinson and Young's study, but enhance the measure's ecological validity, the current study asked participants to quantify their level of confidence in their verdict on a 0–100% scale. Inspired by the design of Sommers and Ellsworth's experiment,[88] verdict and confidence ratings were combined to create one 'comprehensive verdict' scale, with confidence scores for Not Guilty verdicts multiplied by –1 (and Guilty verdicts multiplied by +1). For example, a Not Guilty verdict with a 20% confidence score was represented as –0.20. Lastly, as per Levinson and Young's design, participants were questioned about the race of the defendant at the end of the mock scenario. To maintain the cover story in the present experiment, participants were also asked about the defendant's gender.[89]

III.B.3. Implicit Attitudes

Implicit attitudes regarding race and gender were measured with two different versions of the IAT. In the race version, when black and white faces appear on the screen, participants must click either the 'black American' or 'white American' label in the upper corners of the display. Positive and negative words are also presented on the screen, and match with the labels 'good' and 'bad'. Importantly, these four labels are paired such that one corner contains 'black American/bad' and vice versa. The pairings then switch between trials so that 'black American' appears in the same corner as 'good',

---

[86] Levinson & Young, *supra* note 9, at 331–39.

[87] *Id.* at 334.

[88] Sommers & Ellsworth, *supra* note 10, at 217.

[89] Besides adding a gender recall question and modifying the 0–100% verdict scale, all other aspects of the Mock Crime Scenario were identical to Levinson and Young's study.

and the order of these pairings is counterbalanced across participants. The IAT measures the difference in response times during trials with 'stereotype-congruent' pairings (ie 'white American/good; black American/bad') versus 'stereotype-incongruent' pairings (ie 'white American/bad; black American/good'). This calculation produces a 'D-score' that ranges from –2 to +2, and indicates a slight, moderate, or strong preference toward white or black individuals.[90]

Similarly, the Gender/Career IAT examines implicit attitudes about a woman's role in the household.[91] In this test, the four categories are 'Male', 'Female', 'Career', and 'Family'.[92]

III.B.4. Explicit Attitudes

Explicit attitudes were measured using the SRS[93] and MSS.[94] Both questionnaires include eight statements about societal views towards racism or sexism with which participants must agree or disagree. Responses on the SRS vary between three to four point ratings, and total scores range from 8 (no racial bias) to 31 (strong racial bias).[95] Answers on the MSS follow a standard five-point Likert format, with possible scores ranging from 8 (no gender bias) to 40 (strong gender bias).[96]

III.B.5. Follow-up Task

Five days after the VR experience, participants received an emailed link to the Follow-up Task. To maintain the cover story, the first page of the questionnaire reiterated the purpose of this additional survey (ie to gather public opinion about recidivism and sentencing lengths). Participants then read a brief vignette about an armed robbery, in which the defendant and a group of cohorts mugged a female victim at knifepoint (see Appendix B). Following the case summary, participants were asked to rate on a seven-point scale the likelihood that the defendant would reoffend in the future, as well as the severity with which the defendant should be sentenced. To manipulate the factor of race, stereotypically white and black male names were used as subtle priming mechanisms.[97] Half of the follow-up vignettes discussed 'Defendant Jamal H.', while the remaining half concerned 'Defendant Connor H'. As a manipulation check at the end of the survey, participants were asked to recall the race of the defendant. Additionally, to confirm overall comprehension, participants selected from a list of seven items the facts that were presented in the case summary.

This Follow-up Task sought to assess the duration of potential VR effects by evaluating persisting associations between black men and violent crime; theoretically, individuals who more strongly hold this stereotyped belief would view black defendants as

---

90   *Race Attitudes*, PROJECT IMPLICIT, https://implicit.harvard.edu/implicit/demo/background/raceinfo.html (accessed Sept. 1, 2017).

91   Brian A. Nosek et al., *Harvesting Implicit Group Attitudes and Beliefs from a Demonstration Web Site*, 6 GROUP DYNAMICS THEORY RES. & PRAC. 101, 103 (2002).

92   *Id.* at 114.

93   P.J. Henry & David O. Sears, *The Symbolic Racism 2000 Scale*, 23 POL. PSYCHOL. 253, 279 (2002).

94   Janet K. Swim et al., *Sexism and Racism: Old-Fashioned and Modern Prejudices*, 68 J. PERS. & SOC. PSYCHOL. 199, 212 (1995).

95   Henry & Sears, *supra* note 93.

96   Swim et al., *supra* note 94.

97   This design was inspired by Colin Holbrook et al., *Looming Large in Others' Eyes: Racial Stereotypes Illuminate Dual Adaptations for Representing Threat versus Prestige as Actual Size*, 37 EVOL. & HUM. BEHAV. 67, 70 (2016).

more likely to reoffend in the future, more dangerous, and more deserving of punishment than their white counterparts.[98]

III.C. Hypotheses

Given findings from existing literature, four main hypotheses were formed.

H1: For the Mock Crime Scenario, participants in the Sham/Black Defendant condition would (a) be more likely to rate the defendant as Guilty, (b) exhibit greater confidence in their decisions, and (c) evaluate ambiguous evidence as more indicative of guilt than participants in the Sham/White Defendant condition. By contrast, after an immersive experience in a black avatar's body, participants in the Embodied/Black Defendant condition would be less likely to render the defendant Guilty, give lower comprehensive verdict scores, and rate evidence as less indicative of guilt than participants in the Sham/Black Defendant condition.

H2: Participants in the Embodied conditions would demonstrate lower scores on the race IAT than those in the Sham conditions. No group differences would be found for gender IAT scores.

H3: Explicit race and gender scores would not differ across participant groups.[99]

H4: Participants in the Sham/Black Defendant condition would give higher ratings on the recidivism and sentencing measures than those in the Sham/White Defendant condition. If the embodied VR experience had a significant and durable impact on race-based perceptions of criminality, scores from the Embodied/Black Defendant group would be lower than scores from the Sham/Black Defendant cohort.

III.D. Results

III.D.1. VR Questionnaire

Responses on the VR Questionnaire demonstrated strong feelings of presence in both the Embodied and Sham groups. Upon combining scores for the four immersion questions, the mean rating for participants in the Embodied condition was 16.00 (SD = 2.26) out of a possible 20 points. Similarly, those in the Sham condition provided a mean rating of 16.74 (SD = 1.93).

With respect to feelings of disembodiment, participants in the Sham condition appeared to embrace the perspective of a ghost, feeling as though they were in the room despite not having any body. The mean combined response for the two disembodiment questions was 7.24 (SD = 1.92) out of a possible 10 points (lower scores on this measure indicate greater feelings of disembodiment).

Turning to the body ownership illusion, participants in the Embodied condition expressed a substantial degree of unity with their avatar (the mean combined score for the two questions on body ownership transfer was 8.15 (SD = 1.90) out of 12). In particular, the mean response for the IOS Scale was 4.76 (SD = 1.10) out of 7. To appreciate the relative strength of this score, a comparative VR study reported a mean IOS

---

[98]   See Papillon, *supra* note 9, at 49–53.
[99]   Explicit racial biases were not impacted by the VR experience in either of the two existing studies examining VR and implicit bias. Groom et al., *supra* note 74, at 13–14; Peck et al., *supra* note 70, at 783.

rating of 2.98,[100] and an experiment using the enfacement illusion noted a similar score of 2.88.[101]

III.D.2. Mock Crime Scenario

A 2 (VR Type: Embodied vs. Sham) × 2 (Defendant Race: Black vs. White) multivariate analysis of variance (MANOVA)[102] controlling for participant age and gender revealed a significant main effect of VR Type on evaluations of evidence and comprehensive verdicts ($F(2,74) = 6.94$, $P = .002$, $\eta^2 = .16$); participants in the Embodied condition judged evidence as less indicative of guilt and more confidently rated defendants as Not Guilty than participants in the Sham cohort.[103] A chi-squared test also demonstrated a significant impact of VR Type on dichotomous verdicts ($X^2 = 4.25$, $P = .04$). Specifically, 11% of participants in the Embodied condition rendered Guilty verdicts, compared to 30% of participants in the Sham condition.

As opposed to VR Type, the race of the defendant did not produce a significant main effect ($F(2,74) = 0.077$, $P = .93$, $\eta^2 = 0.002$). This finding is markedly inconsistent with hypothesis H1, which was based on Levinson and Young's reported results. In Levinson and Young's experiment, the authors observed a significant impact of the defendant's skin tone on participants' evaluations of ambiguous evidence and the defendant's level of guilt ($F = 3.31$, $P < .043$).[104] Despite employing the same general statistical analysis and controlling for the same participant demographics, the present study did not produce similar results. In fact, evidence ratings in the Sham group—which should have mirrored those in Levinson and Young's study given that identical stimuli and response measures were used—showed no impact of the defendant's race; the median scores were exactly the same between the Black and White Defendant subgroups.[105] Additionally, comprehensive verdict scores did not significantly differ between the two Sham groups,[106] and more participants in the Sham/White Defendant group rendered Guilty verdicts than in any other condition.[107] Nevertheless, a planned contrast between the Embodied/Black Defendant and Sham/Black Defendant groups did produce findings in line with H1, as those in the Sham condition rated evidence as more indicative of guilt ($M = 81.57$, $SD = 3.13$) than their counterparts in the Embodied condition ($M = 78.87$, $SD = 4.28$).[108] Lastly, evidence ratings best predicted

---

[100] Soo Y. Oh et al., *Virtually Old: Embodied Perspective Taking and the Reduction of Ageism Under Threat*, 60 COMPUTERS HUM. BEHAV. 398, 402 (2016).

[101] Paladino et al., *supra* note 67, at 1205.

[102] Since Levinson and Young conducted a MANOVA to analyse data from the mock crime case, this experiment similarly used a MANOVA to mirror their approach.

[103] Overall evidence scores: $M_{Embodied} = 79.02$ (SD = 4.13); $M_{Sham} = 81.89$ (SD = 4.43). Comprehensive verdicts: $M_{Embodied} = -0.36$ (SD = 0.39); $M_{Sham} = -0.15$ (SD = 0.53).

[104] Levinson & Young, *supra* note 9, at 337. The authors additionally reported significant results from univariate ANOVA tests ($F_{Evidence\ Rating} = 4.84$, $p = .032$; $F_{Guilt\ Rating} = 4.40$, $p = .034$). *Id.*

[105] $Md_{Sham/Black} = 82$, $M_{Sham/Black} = 81.57$ (SD = 3.13); $Md_{Sham/White} = 82$, $M_{Sham/White} = 82.22$ (SD = 5.49). By contrast, Levinson and Young reported mean evidence ratings of 86.23 (Black Defendant) and 80.49 (White Defendant). Levinson & Young, *supra* note 9, at 337.

[106] $M_{Sham/Black} = -0.18$ (SD = 0.52); $M_{Sham/White} = -0.12$ (SD = 0.55); $t(44) = -0.40$, $p = .69$.

[107] $n(Guilty)_{Embodied/Black} = 3$; $n(Guilty)_{Embodied/White} = 2$; $n(Guilty)_{Sham/Black} = 6$; $n(Guilty)_{Sham/White} = 8$.

[108] Mean Difference = 2.99, SE = 1.29, $t = 2.33$, $p = .02$, $\eta^2 = 0.07$.

comprehensive verdicts for participants in the Embodied/Black Defendant condition ($\beta = 0.08$, $t(21) = 5.27$, $P < .001$).[109]

As a manipulation check at the end of the Mock Crime Scenario, participants were asked to identify the race and gender of the defendant. In Levinson and Young's study, a considerable portion of participants could not remember the defendant's race, and there was no statistically significant difference among responses between those who could and could not recall this fact.[110] Accordingly, Levinson and Young argued that the defendant's race was probably not an explicit factor in participants' decisions.[111] The present experiment observed similar findings. Approximately half of participants in the Black Defendant condition correctly identified the defendant's race, and this number was not significantly different between participants in the Sham and Embodied groups.[112] In the White Defendant condition, participants' recall was much more limited (26% in the Embodied/White Defendant condition and 4.5% in the Sham/White Defendant condition). Although recall did not significantly impact decisions in the Mock Crime Scenario,[113] this disparity between races might suggest that participants had a difficult time labeling the light-skinned arm with a particular race, or that existing stereotypes linking black men and crime served as a priming mechanism for at least some participants.

III.D.3. Implicit Attitudes

As depicted in Figure 2, implicit racial biases were significantly lower among participants in the Embodied group, confirming hypothesis H2 ($F(1,89) = 7.15$,[114] $P = .009$, $\eta^2 = 0.07$). To put the results in context, IAT scores are generally categorized into four main degrees of bias: none (<0.15), slight (0.15–0.34), moderate (0.35–0.64), and strong (0.65 and above).[115] Moreover, according to a nationwide dataset from 2015, the median IAT score among Caucasians in California was 0.36 (N = 14,174).[116] In the present study, participants in the Sham condition fell squarely within the moderate bias category, with 61% of participants above the state median. Comparatively, participants in the Embodied condition reflected a slight bias on average, with only 50% of participants above the state median. Although the difference in IAT scores between participants in the Sham and Embodied group was not drastic, the effect size of the VR's impact was nonetheless in an intermediate range.[117]

Analysing IAT distributions within the Sham and Embodied conditions provides further insight into the overall effect of the VR paradigm. As Figure 3 shows, the greatest number of participants in the Sham condition presented strong biases, whereas the

---

[109] Embodied/White Defendant: $\beta = 0.04$, $t(21) = 2.54$, $p = .02$; Sham/Black Defendant: $\beta = 0.09$, $t(21) = 2.80$, $p = .01$; Sham/White Defendant: $\beta = 0.05$, $t(21) = 2.58$, $p = .02$.

[110] Levinson & Young, *supra* note 9, at 338.

[111] *Id.*

[112] % (Correct) $_{\text{Embodied/Black}} = 54$; % (Correct) $_{\text{Sham/Black}} = 50$.

[113] $p = .81$.

[114] $M_{\text{Embodied}} = 0.32$ (SD = 0.42); $M_{\text{Sham}} = 0.52$ (SD = 0.41). Gender was included as a covariate in the ANOVA. Male participants demonstrated significantly higher IAT scores on average than female participants ($M_{\text{Male}} = 0.51$ (SD = 0.41); $M_{\text{Female}} = 0.34$ (SD = 0.43)).

[115] *Race Attitudes*, Project Implicit, https://implicit.harvard.edu/implicit/demo/background/raceinfo.html (accessed Sept. 1, 2017).

[116] *Race IAT 2002-2015*, Project Implicit, https://osf.io/52qxl/ (accessed Sept. 1, 2017).

[117] Jacob Cohen, Statistical Power Analysis for the Behavioural Sciences, 278–80 (1969).
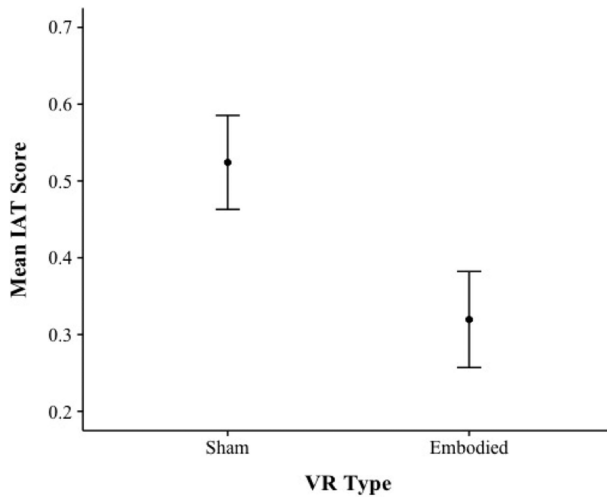
Figure 2. Mean scores on the race IAT among participants in Sham and Embodied groups. The mean score for the Sham condition was 0.52 (SD = 0.41) compared to 0.32 (SD = 0.42) for the Embodied group. Error bars represent standard error of the mean.
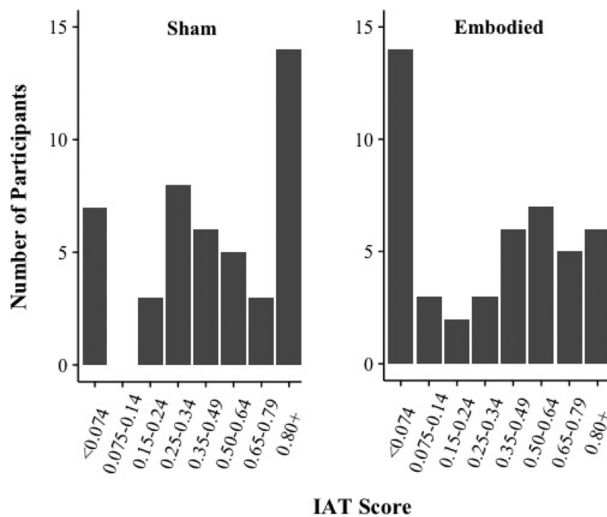

Figure 3. Distribution of race IAT scores among participants in Sham and Embodied groups.

largest subset in the Embodied condition exhibited very little bias. Although there was a great deal of variation across both conditions, it is important to remember that this was a between-subjects, as opposed to a within-subjects, experiment. Participants' IAT scores prior to the VR exercise are thus unknown, meaning that a reduction in bias could still have occurred for those with strong scores in the Embodied condition.

While IAT scores did not predict comprehensive verdict scores,[118] they did marginally predict both evidence scores ($\beta = 2.19$, $t(90) = 2.04$, $P = .04$)[119] and dichotomous verdicts ($\beta = -1.27$, $z(90) = -1.94$, $P = .05$). Lastly, as expected in hypothesis H2, there was no significant difference in implicit gender attitudes between participants in the two VR conditions.[120]

III.D.4. Explicit Attitudes

Mean scores on the SRS reflected low degrees of explicit bias across all participants ($M = 13.14$, $SD = 3.80$). Consistent with hypothesis H3, there was no significant difference in explicit racism or sexism scores between Embodied and Sham groups.[121]

III.D.5. Follow-up Task

Since conviction decisions on the Mock Crime Scenario were not significantly different between the Black and White Defendant groups, the Follow-up Task was rendered moot as a measure of durability for the VR's effect on racial bias. A two-way MANOVA revealed no significant main effects or interactions for VR Type or Defendant's Race.[122] Nevertheless, participants in the Embodied/Black Defendant cohort were the only ones whose recidivism scores predicted sentencing decisions ($\beta = 0.69$, $t(18) = 2.83$, $P = .01$).[123]

With respect to manipulation checks, 99% of participants demonstrated sufficient comprehension of the case summary.[124] For participants who read the vignette about Jamal, 59% identified the defendant as black while 33% selected 'other' from the list of possible races.[125] Similarly, 62% of participants who read about Connor described the defendant as white, with 29% responding 'other'.[126] Based on comments from participants after the study, it is likely that individuals selected 'other' as a neutral response since the vignettes did not explicitly mention race. Given that the majority of participants in each condition did correctly identify the defendant's race, the manipulation appears to have successfully, yet subtly, served its purpose.[127]

---

[118] $\beta = 0.17$, $t(90) = 1.50$, $p = .14$.

[119] The small effect size for this calculation ($r = 0.21$) is consistent with Greenwald et al. (2015)'s reported effect size for the Race IAT ($r = 0.20$). Greenwald et al., *supra* note 15, at 560.

[120] $F(1,90) = 1.16$, $p = .28$, $\eta^2 = 0.01$.

[121] SRS: $F(1,90) = 0.47$, $p = .50$, $\eta^2 = 0.005$; MSS: $F(1,90) = 0.80$, $p = .37$, $\eta^2 = 0.009$.

[122] VR Type: $F(2,83) = 1.21$, $p = .30$, $\eta^2 = 0.03$; Defendant Race: $F(2,83) = 0.49$, $p = .62$, $\eta^2 = 0.01$; Interaction: $F(2,83) = 1.42$, $p = .25$, $\eta^2 = 0.03$.

[123] Embodied/White Defendant: $\beta = 0.04$, $t(22) = 0.18$, $p = .86$; Sham/Black Defendant: $\beta = 0.17$, $t(22) = 0.92$, $p = .37$; Sham/White Defendant: $\beta = 0.59$, $t(18) = 1.96$, $p = .07$.

[124] The one participant who did not satisfy comprehension requirements was excluded from analysis.

[125] Embodied/Black Defendant: 52% 'black', 35% 'other'; Sham/Black Defendant: 65% 'black', 30% 'other'.

[126] Embodied/White Defendant: 68% 'white', 27% 'other'; Sham/White Defendant: 57% 'white', 30% 'other'.

[127] In both the Black and White Defendant groups, a few participants identified the defendant as Hispanic/Latino. These responses ($n = 3$) were removed from the analyses above, since the hypotheses behind the experiment were solely targeted at differences between white and black defendants. Additionally, three participants in the Black Defendant condition incorrectly identified Jamal as white, while two participants in the White Defendant condition incorrectly identified Connor as black. Since the point of the Follow-up Task was to measure the durability of the VR's effects, and these participants clearly visualized the defendant as being either black or white, their responses were re-categorized accordingly in statistical analyses. For instance, if a participant in the Embodied/White condition envisioned Connor as Black, her responses were coded as if she had been in the Embodied/Black condition. The rationale for this decision was that if the VR

III.E. Discussion: The Effect of VR on IAT Scores and Mock Legal Decisions

From its inception, this study had two main objectives: (1) to examine the impact of a subtle yet potent VR paradigm on implicit racial biases and (2) to assess the influence of the VR experience on mock legal decision-making.

With respect to the first goal, the VR paradigm had an observable effect on implicit racial biases between the two participant groups. Not only did participants in the Embodied condition demonstrate lower IAT scores than participants in the Sham condition, but also this result was achieved in a manner subtle enough for the courtroom setting. With a cover story focused on perspective shifting and outgroups, race salience was kept to a minimum in both the design of the overall experiment and the VR content itself. In fact, the race recognition data from the Mock Crime Scenario suggests that participants were not overtly influenced by their avatar's skin color. Although half of the participants in the Black Defendant condition were able to recall the defendant's race, this number did not differ between those in the Embodied and Sham conditions. Had race salience been heightened by the VR experience, participants in the Embodied condition would have been much more likely to notice the race of the defendant after the VR paradigm than those in the Sham group. In addition to the subtle nature of the experiment, the lack of differences between participant groups on the gender IAT, SRS, and MSS speaks to the targeted impact of the VR experience on implicit racial biases. It is also worth noting that the VR paradigm was successfully implemented across four different generations of participants. This is interesting considering that younger generations likely have significantly more exposure to technology, and yet immersion ratings did not differ from one age group to the next. Thus, the VR paradigm used in this study resulted in lower implicit biases for those in the Embodied condition, with minimal race salience, a narrow scope of effects, and a method approachable to all ages.

Besides the notable difference in implicit racial bias scores between the two participant groups, this study also demonstrated the impact of VR on evaluations of mock legal decisions. After embodying a black avatar, participants were more likely to find the defendant Not Guilty, were more confident in this verdict, and were less likely to rate ambiguous evidence as indicative of guilt. Moreover, for participants who evaluated a black defendant after embodying a black avatar, evidence scores were more predictive of comprehensive verdicts than in any other group. Plus, these same participants were the only ones whose recidivism judgments exhibited an observable relationship to their sentencing decisions.

These legal findings are somewhat difficult to interpret. To start, a central hypothesis of this study was that (a) participants in the Sham group would judge the black defendant more harshly than the white defendant, and (b) that embodying a black avatar would shrink the gap between assessments of these two defendants. Contrary to Levinson and Young's reported findings, participants' responses in the current study did not vary based on the defendant's race. As a result, it is difficult to compare the observed differences between the Embodied/Black Defendant group and the Sham/Black Defendant group, as responses in the Sham group were not patently skewed or biased.

---

exercise had a sufficiently potent effect, then participants in the Embodied condition would theoretically find the black defendant less likely to reoffend (or less deserving of severe punishment) than their peers in the Sham/Black group, regardless of whether participants saw the Black or White defendant in the original Mock Crime Scenario.

Specifically, the slight disparity between overall evidence scores is challenging to valuate; even though participants in the Embodied/Black Defendant condition rated evidence as less indicative of Guilt than their counterparts in the Sham condition, both groups' evaluations were relatively neutral. As per Levinson and Young's design, 'each piece of evidence was chosen so that multiple interpretations of that evidence would be possible'.[128] It is therefore unclear whether one group's judgment was objectively better than the other. Nevertheless, the remaining findings do appear normatively positive: according to common legal principles, evidence judgments should predict comprehensive verdict scores, beliefs about a defendant's future recidivism should relate to sentencing decisions, and when enough ambiguous evidence is presented in a trial (as was the case in the Mock Crime Scenario), Guilty verdicts should not be rendered in the presence of reasonable doubt.

Considering the influence of VR Type on legal decisions, the question emerges as to how this effect manifested itself across participant groups. For participants who embodied a black avatar and then evaluated a black defendant, the body ownership illusion (and its associated blurring of 'self-other' boundaries) provides a sensible explanation—seeing a black individual as less foreign from oneself should weaken the negative associations toward blacks that give rise to biased judgments. However, this story does not fully account for the observed effects among participants in the Embodied group who assessed the white defendant. One potential explanation is that adopting a different perspective through the body ownership illusion, regardless of the connection between the embodied experience and the identity of the defendant, was sufficiently potent to provoke more cautious or conservative legal decisions; given recent media coverage documenting racial bias in law enforcement and justice systems, the experience of embodying a black individual might have enhanced the sensitivity or vigilance with which participants approached ambiguous crime scenarios in general. This hypothesized explanation is in line with the growing body of literature on perspective shifting and VR, in which the process of embodying an outgroup member can increase empathetic and prosocial behaviors.[129] In addition, it is worth noting that participants in both VR conditions (Embodied and Sham) were told that the study's purpose was to examine the effects of perspective shifting on legal outcomes. Thus, the impact of the embodied VR on those in the White Defendant condition appears to be rooted in the actual experience of embodying a visibly different human body, as opposed to the mere mention or idea of perspective shifting (otherwise, the data from the Sham groups would have more closely mirrored that of the Embodied/White Defendant group if the idea of perspective shifting alone were sufficient to produce more conservative judgments in the Mock Crime Scenario). That said, matching the race of the avatar and defendant did seem to elicit the most successful effects, as predictive relationships between evidence scores and comprehensive verdicts, as well as recidivism and sentencing decisions, were only present among participants in the Embodied/Black Defendant cohort. In sum, the results suggest a generally positive impact of the VR paradigm on both implicit racial bias and legal decisions.

---

128 Levinson & Young, *supra* note 9, at 334.
129 See Ahn et al., *supra* note 56, at 31; Behm-Morawitz et al., *supra* note 81, at. 397–400.

IV. STUDIES 2 AND 3: ASSESSING LEGAL RESPONSE MEASURES

Though Study 1 achieved its two principle objectives, the lack of disparity between participant responses in the Black and White Defendant groups was fairly surprising in light of existing literature on implicit racial bias. On one level, it is possible that the divergence from Levinson and Young's reported findings was partially caused by Study 1's alteration of the scaled guilt judgment. Whereas Levinson and Young had asked participants to rate the defendant's degree of guilt, Study 1 increased ecological validity by asking participants to render a Guilty or Not Guilty verdict and then rate their degree of confidence in their evaluation of the defendant's guilt. Accordingly, Study 1 was a conceptual extension of Levinson and Young's study rather than a direct replication. But even so, this difference in measures does not appear so drastic as to fully explain the difference in results. For one, a participant who considers the defendant somewhat guilty should theoretically be mildly confident in a Guilty verdict as well.[130] More importantly, however, the evidence measures were the exact same across the two studies, and while Levinson and Young reported a statistically significant difference between black and white defendant groups,[131] Study 1 observed identical medians between the two Sham groups. Plus, Study 1 found no impact of defendant's race on the Follow-Up Task, further contradicting general findings from existing literature.

Given that the participant pool in Study 1 was highly educated and strongly liberal (see Table A1 in the Appendix), it is possible that the lack of disparity between the Sham/Black Defendant and Sham/White Defendant groups was the product of a skewed sample. Alternatively, the Mock Crime Scenario and the Follow-up Task might simply be non-replicable or unsuccessful measures. In order to shed more light on the effect of implicit racial biases in mock legal decisions—while continuing to use the more ecologically valid comprehensive verdict measure—both the Mock Crime Scenario and the Follow-up Task were re-run as online surveys to target a larger and more diverse set of participants.

IV.A. Method

To best enable a comparison to Study 1, the subsequent experiment was divided into two smaller studies, each with its own set of participants. In one subset, Study 2, participants evaluated the Mock Crime Scenario *and* the Follow-up Task, while participants in the other subset, Study 3, assessed only the latter. Whereas Study 2 closely mimicked the design of Study 1,[132] Study 3 provided insight into (a) the success of the Follow-up Task as a standalone measure, and (b) whether completing the Mock Crime Scenario (including its final question about the defendant's race) influences responses on the Follow-up Task.

---

[130] The present study did not empirically evaluate this proposition, as the benefit of continuing to use the comprehensive verdict score (through its enhanced ecological validity) appeared to outweigh the importance of pursuing a pure replication of Levinson and Young's study.

[131] Levinson & Young, *supra* note 9, at 337.

[132] In Study 1, participants evaluated the Follow-up Task five days after the Mock Crime Scenario. With an online sample, this extended interval between response measures can be very difficult to replicate. As a result, participants in Study 2 read the follow-up vignette immediately after completing the Mock Crime Scenario, but were explicitly told that the two cases were unrelated to each other (in keeping with the design of Study 1).

IV.A.1. Participants

Participants (Study 2: $N = 239$; Study 3: $N = 186$) were recruited from Amazon Mechanical Turk.[133] While data were collected from all respondents, only those who demonstrated sufficient quality of responses[134] and self-identified as Caucasian (in the demographic questionnaire at the end of the experiment) were included in the final analysis (Study 2: $n = 167$; Study 3: $n = 152$). In both groups, participants were over the age of 18 (Study 2: $Md_{age} = 33$; Study 3: $Md_{age} = 38.5$), fluent in English, and with normal or corrected-to-normal vision. Political affiliations and educational backgrounds were substantially more diverse than in Study 1. (See Table A1 in the Appendix for additional demographics.)

IV.A.2. Conditions

Defendant Race was the sole independent variable in Studies 2 and 3. Participants were randomly assigned to either the Black Defendant or White Defendant condition; as in Study 1, the defendant's race in Study 2 was kept consistent between the Mock Crime Scenario and the Follow-up Task.

IV.A.3. Procedure

After providing their informed consent, participants in Studies 2 and 3 read and evaluated their respective case(s) before completing a short demographic survey.

IV.B. Hypotheses

Given the lack of expected findings in Study 1, two competing hypotheses were generated.

H1$_A$: If the main issue in Study 1 was the highly liberal and educated nature of the participant pool, having a more diverse sample in Studies 2 and 3 would yield significantly different responses between Black and White Defendant conditions. Specifically, participants in Study 2 who saw the dark-skinned defendant would render more Guilty verdicts, produce higher comprehensive verdict scores, and rate the evidence as more indicative of guilt than participants who saw the light-skinned defendant. Similarly, for the Follow-up Task, participants in Studies 2 and 3 who read about Jamal (and interpreted his race as black) would provide higher scores for the recidivism and sentencing questions than those who read about Connor (and interpreted his race as white).

H1$_B$: If the findings from Levinson and Young's study were nonreplicable, participant responses would not differ across conditions. Similarly, if the Follow-up Task failed to successfully measure racial biases and criminal stereotypes, the defendant's race would not influence ratings.

---

[133] Amazon Mechanical Turk is an online marketplace that enables expedient data collection from a large variety of workers. Amazon Mechanical Turk, *Overview*, AMAZON, https://www.mturk.com/mturk/help?helpPage=overview (accessed Sept. 1, 2017). For the present study, eligibility requirements included United States residency and a 95% approval rating for previous work on the site.

[134] Participants excluded due to quality concerns showed poor comprehension on the manipulation check at the end of the Follow-up Task.

IV.C. Results

IV.C.1. Mock Crime Scenario

A one-way MANOVA detected no significant effect of Defendant Race on evaluations of evidence and comprehensive verdict scores, failing once again to replicate Levinson and Young's findings.[135] It should be noted that in Levinson and Young's study, the participant pool was not restricted to Caucasian individuals[136] (although according to existing literature,[137] implicit racial biases should theoretically be stronger (and more readily visible) among Caucasian populations). Since the original Amazon Mechanical Turk sample in Study 2 included a range of participant ethnicities, a one-way MANOVA was conducted for the entire pool to provide an additional check against Levinson and Young's findings. Even with this larger and more diverse sample, no significant results were observed.[138] Moreover, while more participants in Study 2 correctly identified the white defendant's race than in Study 1, approximately half of participants in both studies correctly identified the black defendant's race;[139] similar to the results in Levinson and Young's experiment[140] and Study 1, participants' ability to recall the defendant's race did not impact responses on the task.[141]

IV.C.2. Follow-up Task

As with Study 1, the use of names to prime racial stereotypes was a successful manipulation; the majority of participants in Studies 2 and 3 identified Jamal as black (Study 2: % (Correct) = 86; Study 3: %(Correct) = 88.5) and Connor as white (Study 2: %(Correct) = 64; Study 3: %(Correct) = 55).[142]

A one-way MANOVA revealed divergent findings for Studies 2 and 3. While no significant results were observed in Study 2,[143] Defendant Race had a significant effect on recidivism and sentencing measures in Study 3 ($F(2143) = 4.34$, $P = .01$, $\eta^2 = 0.06$).[144] Although the data cannot conclusively explain this disparity between results, there was a marginally significant difference in recidivism ratings between the Black Defendant conditions in each study (Study 2: $M = 5.37$, SD = 1.11; Study 3: $M = 5.71$, SD = 1.22; $F(1167) = 3.55$, $P = .06$, $\eta^2 = 0.02$). Accordingly, evaluating

---

[135] $F(2,154) = 2.65$, $p = .07$, $\eta^2 = 0.03$. Age was included as a covariate in the model.

[136] Levinson & Young, *supra* note 9, at 335.

[137] *Race IAT 2002-2015*, PROJECT IMPLICIT, https://osf.io/52qxl/ (accessed Sept. 1, 2017).

[138] $F(2,236) = 1.52$, $p = .22$, $\eta^2 = 0.01$.

[139] In Study 2, 34% of participants in the White Defendant and 57% of participants in the Black Defendant condition correctly identified the defendant's race, compared to 4.5% of participants in Study 1's Sham/White Defendant condition, and 50% of those in the Sham/Black Defendant condition. As with Study 1, this manipulation check demonstrates that a significant number of participants did not explicitly notice the defendant's race, thereby maintaining the subtle aspects of the experiment's priming design. At the same time, the fact that a substantial number of participants *were* able to identify the defendant's race suggests that participants were paying close attention to the slideshow, despite the remote and disinterested nature of online studies.

[140] Levinson & Young, *supra* note 9, at 338.

[141] $F(2,153) = 0.99$, $p = .51$, $\eta^2 = 0.009$.

[142] Consistent with Study 1, participants who identified the defendant as Hispanic/Latino were excluded from statistical analyses, while responses of participants who misidentified Jamal as white or Connor as black were recoded accordingly (ie if a participant who read about Jamal envisioned the situation as involving a white defendant, his or her response was coded as if the participant had read about Connor).

[143] $F(2,160) = 0.45$, $p = .64$, $\eta^2 = 0.006$.

[144] Recidivism: $M_{Black\,Defendant} = 5.71$ (SD = 1.22); $M_{White\,Defendant} = 5.34$ (SD = 1.41). Sentencing: $M_{Black\,Defendant} = 4.97$, (SD = 1.25); $M_{White\,Defendant} = 5.11$ (SD = 1.18).

the Mock Crime Scenario (and being asked to identify the defendant's race) may have contributed to the lower recidivism ratings on the Follow-up Task in Study 2.

IV.D. Discussion: The Shortcomings of Mock Legal Scenarios

Studies 2 and 3 sought to expand on Study 1's findings and investigate the ability of the Mock Crime Scenario and Follow-up Task to capture effects of implicit racial biases. Consistent with Study 1, Study 2 did not find racial disparities using the Mock Crime Scenario, despite Levinson and Young's reported results. Although the Follow-up Task produced significantly different scores between the black and white defendant groups in Study 3, the effect size was relatively small.

While these findings are notable on their own, the more interesting question is why the results diverged from the existing literature. There are likely many possible explanations, but at least three reasons come to the fore. First, the problem could rest in the reliability of the measures. For instance, given that Levinson and Young's experiment involved a relatively small sample size ($N = 66$),[145] it is possible that their study was too underpowered and prone to type I errors (ie false positives). Second, since 2010 (and certainly since Sommers and Ellsworth's study in 2001), a plethora of race-related incidents have taken place across America, particularly involving the shooting of unarmed black men. Not only have these events spurred discussions about race relations in our current society, but people have also become more aware of the presence and impact of implicit racial biases as a result.[146] In light of frequent community protests and an outpouring of responses over social media, people might have become more careful in answering contrived, hypothetical scenarios about race. Over the past few years, implicit racial biases might also have become less pronounced. Data from Project Implicit provides at least some support for this hypothesis.[147] According to their database of IAT responses dating from 2002 to 2015, the median scores among white participants in 2002, 2010, and 2015 were 0.47 ($N = 27{,}068$), 0.43 ($N = 129{,}826$), and 0.37 ($N = 183{,}864$), respectively.[148] In addition to unreliable measures and influential current events, a third possible explanation lies in the complexity of legal decisions. Judgments of guilt, recidivism, and sentencing involve layers of analyses and calculations that are not wholly reflected in the final assessments. Consequently, the dichotomous verdicts and scaled ratings used in mock trial scenarios are too crude to illuminate many aspects of a judge or juror's decision-making process. It is therefore probable that implicit racial biases manifest themselves in a more complex pattern than these tests are capable of capturing. Presumably, these three explanations are not mutually exclusive, and likely all played some role in the observed results.

V. GENERAL DISCUSSION: LIMITATIONS AND FUTURE DIRECTIONS

As a preliminary investigation into VR and implicit bias in the courtroom, the present study illuminated three important points: (1) embodying an outgroup avatar for five

---

[145] Levinson & Young, *supra* note 9, at 331.

[146] See eg Sendhil Mullainathan, *Racial Bias, Even When We Have Good Intentions*, N.Y. Times (Jan. 3, 2015), https://www.nytimes.com/2015/01/04/upshot/the-measuring-sticks-of-racial-bias-.html?mcubz=1&_r=0 (accessed Mar. 13, 2018); Vedantam, *supra* note 2.

[147] *Race IAT 2002-2015*, PROJECT IMPLICIT, https://osf.io/52qxl/ (accessed Sept. 1, 2017).

[148] *Id.*

minutes can have a positive impact on implicit racial biases, (2) this experience can result in more cautious legal judgments in light of ambiguous evidence, and (3) mock legal scenarios may have less success capturing implicit racial biases than the literature might otherwise suggest.

Despite these findings, there are many elements upon which future studies can improve and expand. For one, Study 1 included 92 participants, all self-identified as Caucasian. Although implicit racial biases against black individuals tend to be strongest among white participants,[149] jurors and judges represent a broad array of racial identities. Future studies should increase the number and diversity of participants to assess the VR's impact on individuals from multiple backgrounds. It is also worth noting that the median IAT scores of those in the Sham condition were higher than the median score for Californians,[150] while the median IAT score of participants in the Embodied condition closely approximated California's median. Follow-up studies should examine whether the VR exercise produces similar effect sizes in populations where the control group's scores are significantly higher or lower than the present study. In addition to altering the participant pool, researchers should also experiment with manipulating the VR content itself. For the sake of simplicity, avatars in the present experiment differed from participants in race only. Future studies should incorporate avatars of various races, genders, ages, and socioeconomic backgrounds to examine whether the strength of the body ownership illusion shifts as additional avatar characteristics are added into the mix. Another limitation of the current study was its between-subjects design. In order to gauge the VR's impact on an individual level, researchers should employ a within-subjects procedure, particularly to test pre-VR and post-VR IAT scores. Moreover, future studies could include alternative measures of implicit bias, such as the Weapon Identification Task, which examines associations between black men and lethal weapons.[151] Finally, the shortcomings of the mock legal scenarios speak to the need for paradigms that can more adequately capture the nuances of legal decisions. Without more sensitive procedures, it becomes difficult to confidently grasp the nature of implicit bias in the courtroom, not to mention the subsequent impact of bias mitigation efforts.

VI. CONCLUSION

Since the advent of the IAT, decades of research have exposed the prevalence and persistence of implicit racial biases across the nation. In the case of black defendants in the courtroom, these split-second, unconscious, and negative associations obstruct presumptions of innocence and standards of impartiality. While existing suggestions for bias reduction have significant drawbacks in the courtroom environment, this article presents a novel and promising strategy to address implicit bias. Not only do the results highlight VR's capacity to influence IAT scores, but VR also appears to encourage more cautious evaluations in the face of an ambiguous legal case. The most important finding, however, is the demonstrated potential of this field of research. VR might hold the key to substantial and feasible bias reduction in the courtroom, but without further research, its true promise will never be known. At the very least, this study provides

---

[149]   *Id.*
[150]   *Id.*
[151]   Payne, *supra* note 20, at 287.

sufficient groundwork to warrant continued research in this innovative and uncharted domain.

APPENDIX

A. PARTICIPANT CHARACTERISTICS

Table A1. Demographic information of study participants.

Demographics	Study 1	Study 2	Study 3
Education level			
Less than high school	-	-	1%
High school/GED	2%	11%	11%
College (no degree)	4%	22%	20%
Associate's degree	2%	12%	12%
Bachelor's degree	55%	45%	39%
Advanced degree	36%	10%	17%
Political affiliation			
Democrat	53%	44%	38%
Republican	8%	26%	25%
Registered independent	8%	10%	11%
Libertarian	3%	6%	3%
Other left wing	1%	2%	5%
Unaffiliated	27%	11%	19%
Gender			
Male	46	84	65
Female	46	83	87
Age			
Minimum	18	18	20
Maximum	79	61	72

B. STATEMENTS OF EVIDENCE FROM LEVINSON AND YOUNG'S MOCK CRIME SCENARIO

Table A2: Statements of evidence.

The following information comes from Justin D. Levinson & Danielle Young, *Different Shades of Bias: Skin Tone, Implicit Racial Bias, and Judgments of Ambiguous Evidence*, 112 W. Va. L. Rev. 307, 348–49 (2010).

1. The defendant purchased an untraceable handgun three weeks before the robbery.

2. The store owner identified the defendant's voice in an audio line-up.

3. A week after the robbery, the defendant purchased jewelry for his girlfriend.

4. The defendant's brother is in jail for trafficking narcotics.

5. The defendant recently lost his job.

6. The defendant used to be addicted to drugs.

7. The defendant has been served with a notice of eviction from his apartment.

8. The defendant was videotaped shopping at the same Mini Mart two days before the robbery.

9. The defendant frequently shops at a variety of Mini Mart stores.

10. The defendant used to work at this particular Mini Mart.

11. The defendant is left handed.

12. The defendant was a youth Golden Gloves boxing champ in 2006.

13. The defendant belongs to a local gun club called Safety Shot: The Responsible Firing Range.

14. The defendant had a used movie ticket stub for a show that started 20 minutes before the crime occurred.

15. The defendant wore a plaster cast on his broken right arm around the time of the robbery.

16. The defendant is a member of an antiviolence organization.

17. The defendant's fingerprints were not found at the scene of the crime.

18. The defendant does not have a driver's license or car.

19. The defendant has no prior convictions.

20. The defendant graduated high school with good grades.

*Per Levinson and Young's study, these questions were presented to participants in a randomized order.

C. FOLLOW-UP TASK VIGNETTES

Before answering the recidivism and sentencing questions, participants read the following vignette (with Jamal for the Black Defendant condition and Connor for the White Defendant condition):

Defendant Jamal (Connor) H., 28, was charged in an armed robbery case. On December 20, 2015 at 2:35 am, Jamal (Connor), along with three other men, was loitering in the doorway of a closed storefront. The men surrounded and blocked a middle-aged female who was walking down the street. Jamal and his cohorts, with switchblade knives visibly displayed, demanded the victim surrender her wallet and jewelry. After acquiring her belongings, the men departed the scene in a Toyota Camry driven by the defendant, and headed toward the highway, where they were later intercepted by police.

Jamal (Connor) insists that he was under the influence of alcohol at the time of the crime, and was unduly influenced by his peers. The defendant has a prior misdemeanor conviction for trespassing in 2013.