Evidence of Racial Bias Using Immersive Virtual Reality: Analysis of Head and Hand Motions During Shooting Decisions

Tabitha C. Peck *Senior Member, IEEE*, Jessica J. Good, and Katharina Seitz

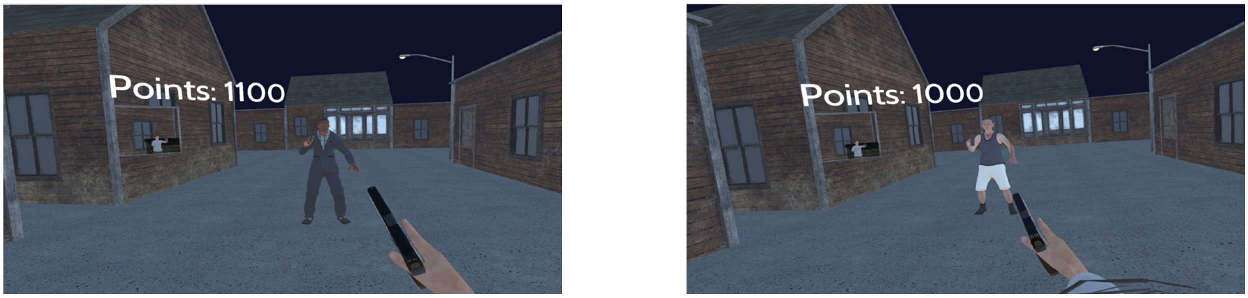

Fig. 1: Left: A participant shooting at a black avatar dressed in high socioeconomic status clothing. Right: A participant shooting at a white avatar dressed in low socioeconomic status clothing. In both images the target avatar is carrying a cell phone. The participant holds a virtual gun directed at the avatar. The participant's self-avatar is reflected in the mirror to the left of the target avatar.

Abstract—Shooter bias is the tendency to more quickly shoot at unarmed Black suspects compared to unarmed White suspects. The primary goal of this research was to investigate the efficacy of shooter bias simulation studies in a more realistic immersive virtual scenario instead of the traditional methodologies using desktop computers. In this paper we present results from a user study (N=99) investigating shooter and racial bias in an immersive virtual environment. Our results highlight how racial bias was observed differently in an immersive virtual environment compared to previous desktop-based simulation studies. Latency to shoot, the standard shooter bias measure, was not found to be significantly different between race or socioeconomic status in our more realistic scenarios where participants chose to raise a weapon and pull a trigger. However, more nuanced head and hand motion analysis was able to predict participants' racial shooting accuracy and implicit racism scores. Discussion of how these nuanced measures can be used for detecting behavior changes for body-swap illusions, and implications of this work related to racial justice and police brutality are discussed.

Index Terms—Shooter bias, virtual reality, body-swap illusions, user studies, racism, bias, implicit bias, tracker data

1 Introduction

On May 25th, 2020 George Floyd was callously murdered by a White police officer kneeling on his neck for 8 minutes and 46 seconds [4]. Witness accounts show that police drew their weapon on Floyd, a Black man arrested for allegedly passing a counterfeit $20 bill (USA), despite Floyd being unarmed and compliant with police orders. Disturbing video footage of Floyd's murder led to widespread outrage and protests around the United States and beyond. George Floyd was sadly the latest victim in a history of racist violence that stretches back hundreds of years. Though much of the focus has been on racist violence within the United States, leaders from around the world (including Australia, the EU, Iran, Zimbabwe, Kenya, and South Africa) have both condemned the murder of George Floyd and drawn parallels to instances of racist violence within their own countries [25].

Scholars of all types have identified and studied racism at its individual, structural, and systemic levels. Social psychologists have examined police data to identify biased policing practices as well as general population-level bias that contributes to disparate use of force directed at Black Americans. In one well-replicated area of research, psychologists have studied "shooter bias" – the tendency to mistakenly shoot unarmed Black targets more often than unarmed White targets, and more quickly shoot at armed Black suspects compared to armed White suspects [10]. Shooter bias has been demonstrated across multiple populations, such as civilians [10, 26, 23] and police officers [11, 13, 51], and with different experimental materials, such as static images [10, 11, 51, 26, 23] and short videos [24, 13]. Currently police officers are using virtual reality (VR) simulators to address racial bias [1], however to date no published research has studied the feasibility of immersive virtual environments (VEs) for studying shooter bias.

In the present research, we argue that VR offers important advantages to the study of shooter bias. VR enables participants to be fully immersed in a realistic environment, experiencing cognitions, emotions, and behaviors in a more realistic manner [9]. Additionally, VR enables ecological validity and realism as well as the collection of more nuanced movement data. This work provides a critical foundation for understating shooter bias within VR as well as provides a baseline control condition for future research in understanding if and how virtual self-avatars could be used to change negative shooter-bias behavior.

We argue that utilizing VR in the study of shooter bias not only provides a more accurate understanding of behavior, but also offers a critically important avenue for testing training and interventions designed to reduce racial bias in shooting decisions. Although several studies have sought to reduce or eliminate shooter bias [33, 47, 54], the phenomenon persists. The U.S. public has reached a boiling point, with increased civil unrest due to continued disparate use of force toward Black suspects. Politicians, funders, and police departments themselves are turning to academic researchers for help in understanding and reducing racially biased policing. It is imperative that we utilize the best technology and methodology in providing answers to their questions, as lives are literally at stake.

• *Tabitha C. Peck. Jessica J. Good, and Katharina Seitz are with Davidson College. E-mail: tapeck/jegood/kaseitz@davidson.edu.*

Manuscript received 9 Sept. 2020; revised 15 Dec. 2020; accepted 8 Jan. 2021. Date of publication 22 Mar. 2021; date of current version 7 Apr. 2021. Digital Object Identifier no. 10.1109/TVCG.2021.3067767

2 Background

2.1 Socioecological Evidence of Shooter Bias

Sociological data provide evidence supporting public claims of police brutality against Black Americans. Several sociological studies using police data have demonstrated the disproportionate amount of lethal force used on Black suspects, finding that police officers do not make use-of-force decisions independent of race [19, 28, 38, 49, 53]. For example, researchers used ecological-level data to test whether differential criminal activity or racially biased policing explained disparate use of deadly force based on race [53]. The researchers found police were more likely to use deadly force on Black suspects compared to White suspects, even when controlling for levels of criminal activity. Thus, in the case of police activity, the evidence clearly supports the notion that race (intentionally or unintentionally) influences the use of force.

It is important to note however, that sociological data does not account for confounding variables, such as socioeconomic status (SES). A sociological review of shooter data [19] notes that police may actually rely on socioeconomic status information as well as race, while making shooting decisions. In the United States, Black individuals are disproportionately represented in the lower SES strata (US Census Bureau, 2019). Just as people perceive race within a fraction of a second upon meeting someone [30], people can also accurately perceive others' socioeconomic status by rapidly evaluating status signals including clothing, body language, linguistic style, and leisure activities [17, 29]. People also associate criminality with both low SES and Black people [34]. When asked to describe criminals, with the exception of depicting lawyers, participants were more likely to depict a Black male working in a low-income job, such as mechanic, fast-food worker, and janitor [34]. Since Black, male, low-income individuals are perceived as more criminal, it is likely that they are also perceived as more threatening. Sociological research raises the question of whether race, SES, or both influence shooter bias. It is difficult to determine causality using sociological data, however. Instead, experimental data must be used to determined causality.

2.2 Desktop Experimental Tests of Shooter Bias

In order to study shooter bias in controlled, experimental settings, psychologists developed a desktop computer simulation in which participants had to quickly decide if the pictured White or Black target was holding a threatening or non-threatening object and then make a "shoot" or "don't shoot" decision by pressing corresponding keys on the keyboard [10]. Results from the initial study on shooter bias [10] demonstrated that participants made the decision to shoot armed Black targets more quickly than armed White targets, and decided to not shoot unarmed White targets more quickly than unarmed Black targets. Those findings have been extensively replicated [11, 12, 13, 23, 37], finding that shooter bias exists across a variety of groups and with varied methodologies. Although police officers are often the prototypical perpetrator of shooter bias, evidence of shooter bias has been found in non-police participants [10, 22, 26]. This is of note because armed civilians can also utilize deadly force (for example, in the case of Trayvon Martin). Shooter bias has also been detected among both White and non-White participant samples [10, 26].

Previous work suggests that police and civilians show biased shooting patterns between Black and White suspects due to perceiving Black individuals as more threatening. One study examining primes and object recognition found that participants primed with Black faces mistook objects for weapons more readily than those primed with White faces [40]. In another weapon identification study, participants struggled to distinguish weapons from harmless objects when held by Black targets and responded more quickly and accurately to Black targets holding weapons than White targets holding weapons [23]. These studies all suggest that Black individuals are perceived as more threatening and it is more difficult to differentiate harmless objects from weapons when a Black individual is holding them.

Paralleling the sociological research, a few desktop shooter bias studies have also investigated the impact of SES. In this research, SES has been manipulated by varying neighborhood context and the targets' clothing choices [12, 13, 26, 37]. For example, in one study participants viewed pictures of targets in front of dangerous or neutral neighborhood backgrounds and in another participants viewed targets in threatening (baggy gray sweatshirt, a gray headband, and a black baseball cap, worn to the side) or non-threatening clothing (a button-down shirt with a tie) [26]. Shooter bias effects increased in the dangerous neighborhood setting or with threatening clothing, suggesting that race is not the only contributing factor to quicker shooting of Black targets.

A major criticism of experimental shooter bias methodology stems from its lack of ecological validity. Sitting in front of a computer in a lab lacks the same stress that a real-life shooting incident would incur. Rather than shooting a gun, participants press keys designated as "shoot" and "don't shoot". Importantly, researchers have sought to improve this methodology by increasing the realness of the situation. For example, studies have improved the methodology by removing the "don't shoot" button, instead requiring participants to inhibit their shooting response on unarmed avatars while also holding a gun-like controller rather than unrealistically clicking a key to indicate not shooting [13, 26]. Other research has used animated video clips of targets to increase the complexity and stress of the task [24, 13]. While an improvement over the original methodology, it is clear that non-immersive methodologies cannot approximate the real threat perception and stress level of an actual encounter with an armed person.

2.3 Shooter Bias in Virtual Reality

Using VR to study psychological phenomena has many advantages, one of which is the increased realism of the participant's experience [9] and realistic responses to the environment [55]. Studying shooter bias in an immersive VE has the potential to allow participants to genuinely feel as if they are in a threatening environment and must make a real choice to shoot armed targets and not shoot unarmed targets. Participant shooting responses, compared to desktop simulation key presses, display 'truer' shooting behaviors due to place illusion—a sensation of being in a real place—and plausibility illusion—that the scenario is depicted as actually occurring [55]. In essence, VR removes the typical internal versus external validity trade-off by allowing controlled, precise stimuli and measurement but in a realistically experienced environment.

Another advantage is that participants can experience virtual embodiment illusions: the illusion of ownership over one's virtual, embodied form [20, 27, 56]. The first evidence of embodiment was found using the rubber hand illusion, in which a rubber hand was placed near participants' actual, hidden arm [8]. The rubber hand and the participant's actual hand were stroked and tapped synchronously, resulting in the cognitive merging of the visual feedback from the rubber hand and the tactile feedback from their actual arm in a proprioceptive illusion [8]. Understanding the relationship between users, their self-avatars, and their subjective sense of embodiment is an active research area within VR [16]. Further, the existence of self-avatars has been shown to affect cognition and perception within VR [36, 45, 57]. Numerous studies have found that people quickly feel ownership over their self-avatar and behave in stereotypical ways related to their self-avatar's appearance [63]. For example, embodying a visual representation of Einstein improves cognitive performance [7], and women were buffered from stereotype threat when embodying masculine avatars while men experienced stereotype threat when embodying feminine avatars [41, 43]. Most relevant to the present research, white participants embodying Black avatars demonstrated a reduction in implicit racial bias [44, 6, 52].

Additionally, VR hardware enables easy data collection of head and hand tracker data. Commodity VR systems (e.g. HTC Vive and Facebook Oculus) include 6-degrees of freedom (DOF) tracking of the head and hands via the head-mounted display (HMD) and hand-held trackers. Head and hand trackers provide movement data during shooting and non-shooting actions. These additional data can lead to better understanding of shooting decisions by providing a more nuanced look at participants' physical actions throughout the experiment. Whereas dichotomous shoot/not shoot behaviors have typically been used in shooter bias studies, research shows that racial bias is often indicated in more complex ways. For example, subtle facial movements have been linked to racially biased selection of job applicants [61]. VR offers an important solution to ecological validity and experimental concerns, allowing researchers to create more realistic simulations and to easily collect data that may highlight previously unknown implicit bias measures.

Finally, an advantage of using VR to study shooter bias is that the technology is already being used in industry as a training tool. Military and first-response industries like policing have already begun to use VR for training purposes [32, 62, 1]. Thus, if we can better understand biased shooting behavior through VR, we can apply those lessons learned to improving training protocols.

Although the purpose of the present study is a basic science investigation of the use of VR in testing shooter bias, the applied goal of shooter bias research is ultimately to alleviate it. Given that the most effective interventions to reduce implicit bias are generally affective or experience-based [50, 15] it follows that shooter bias might also be mitigated through experiential training. Research from the computer science literature has demonstrated that VR can be successfully used to diminish certain biases [6, 44]. For example, White participants embodied in Black, rather than White, avatars demonstrated a reduction in racial implicit bias [6]. Studying shooter bias in VR therefore, opens up the possibility of applied training and police reform in order to reduce the disproportionate use of force against Black individuals.

2.4 Current Study

In the present study, we had three main research goals. First, we aimed to investigate shooter bias within a fully immersive VE, making several key methodological improvements that we argue increase the generalizability of the findings. Through VR, participants were placed in a dark, abandoned city alley with realistic sound effects. They held a virtual gun in their right hand and had to raise their arm and pull the trigger to shoot, or re-holster their weapon at their hip to not shoot. The targets were animated avatars that appeared in random positions, requiring the participant to be more vigilant than with static images of targets. In addition to traditional measures of shooting accuracy and latency, we also measured position and orientation of both participants' shooting hand and head. With these methodological improvements, we argue that VR can provide a more nuanced understanding of shooter bias. To our knowledge, no published literature has investigated shooter bias within an immersive VE. Given the methodological changes from prior desktop-based work, it was unclear whether or not we would replicate common shooter bias patterns. Thus, part of our goal was to compare our testing of shooter bias within an immersive VE to the results typically found in studies presented on a desktop monitor.

Our second research goal was to assess implicit racial bias and embodiment in relation to shooter bias. Although implicit racial bias measures have not typically been included in shooter bias research, they were included in the present research in order to explore whether implicit racial bias was related to shooting accuracy or latency as a function of both race and SES. Additionally, we included a self-report measure of embodiment to determine if embodying a gender and race-matched avatar was related to participants' shooting decisions or accuracy. This embodiment measure provides a base-line for future research that may involve gender, age, race etc. swapped self-avatars when further investigating embodiment effects for mitigating shooter bias.

Our final goal was to explore how motion tracking could advance the understanding of shooting decisions. We collected 6-DOF hand and head tracker data at approximately 120 Hz. We did not make specific predictions about participants' motion paths since, to the best of our knowledge, shooter bias studies in VR have not been published and motion paths within shooter bias research have not been investigated. Previous work by Dovidio et al. [14] found that White participants' implicit bias scores were related to observers' ratings of their racial bias and friendliness toward Black confederates (as indicated by nonverbal behaviors). This suggests that nonverbal behaviors, or motion over various parts of the body that are measurable with head and hand trackers, may be able to predict bias. Additionally, Kilteni et al. [27] found that self-motion was affected by the appearance of a self-avatar. For this exploratory research, we followed a similar analysis approach to Kilteni et al. and calculated the variance explained by position and orientation data as an operationalization of motion complexity. In other words, when greater variance in position and orientation could be explained, participants were engaging in simpler movements [27]. In the case of shooter bias, greater complexity of hand movement could be interpreted as hesitation, indecision, or even inhibition of an automatic movement. Greater complexity of head motion could be interpreted as a more holistic examination of the scene rather than a singular focus on one point in space. We therefore explored movement complexity as a function of target race and SES to determine if these nonverbal behaviors could be used to predict bias similar to the nonverbal behaviors identified by Dovidio et al. [14].

3 METHOD

3.1 Experimental Design

We used a 2 (target race: Black or White) x 2 (target SES: high or low) x 2 (object: gun or phone) within-subjects design to evaluate whether both target race and SES had an impact on shooting decisions. In an immersive VE, participants first completed the shooter bias task in which they made shoot/don't shoot decisions when confronted with 160 target avatars. Next they completed a racial implicit association test also in VR. Finally, participants completed self-report measures on a laptop computer. The study design and all materials were approved by the (removed) Institutional Review Board prior to beginning data collection. Participants were informed before the experiment that they would be holding and firing a virtual gun at avatars. Avatars did not experience visual harm and would appear and disappear with each trial.

3.2 Participants

We planned to collect a sample of 100 participants in order to adequately power our analyses, given that several were exploratory. We reviewed prior literature, noting that initial shooter bias studies testing for a race $\times$ object interaction on latency and accuracy used sample sizes ranging from N = 42 to N = 52 [10]. Studies that most closely matched our design (race $\times$ ses $\times$ object) used samples ranging from N = 56 to N = 80 [26]. The reason that these relatively small samples generate sufficient statistical power is the large number of measurements generated by the within-subjects, multi-trial design.

Since shooter bias has been shown to exist in non-law enforcement populations (Correll et al., 2002), the sample of this study was comprised of undergraduate students and community members from Davidson, North Carolina (N = 102). During debriefing, one person declined for their data to be used in analyses. Two female participants were removed from the data for electing to not shoot at armed participants. Analysis is reported on the final sample of 99 participants. All participants had no prior military or law enforcement training, had normal or corrected to normal vision, had full use of both hands, were in their usual state of good health, had no prior history of motion sickness or epilepsy, and were naive to the purpose of the experiment. The majority of participants (67%) had no prior firearm experience. On a scale from 1 (none) to 7 (very much) only 7% of participants self-rated their firearms experience above a 4. Participants ranged in age from 18 to 76 years (M = 23.08, SD = 9.91), though the majority (88%) of participants were between the ages of 18 and 22 years. The sample was split approximately evenly between men and women (55.5% women, 42.5% men, and 2.0% non-binary). There were 80 participants who identified as White, 9 Asian, 8 Hispanic, 3 Black, and 1 multi-racial. The majority of the participants described themselves as non-gamers (rarely or never playing computer games: 56%), while a quarter of the sample identified as casual gamers (playing for enjoyment or relaxation: 24%) and the rest as core gamers (deeper involvement or tactical challenges: 17%) or hard core gamers (competitive, high action games: 3%). Participants were recruited using emails, flyers, and word of mouth. The study lasted approximately 45 minutes, and all participants were compensated with a $5 Amazon gift card or 0.5 research credits at the completion of the study.

3.3 Experimental Manipulations

Target Race: All target avatars were male. The race of the target avatars was manipulated as either Black or White by adjusting skin tone and facial features. Pilot testing with an independent sample (N = 101) yielded 10 Black and 10 White avatars matched on subjective ratings (1-7 Likert scale) of attractiveness and aggressiveness such that the Black and White avatars were rated, on average, within .1 of each other on the subjective scales. All target avatars were presented in forward facing view, with their full bodies visible (see Figure 1).

Target SES: The targets' SES was indicated using clothing. Consistent with previous methodology [26, 37], high SES targets wore business attire (suits), whereas low SES targets wore street clothing (shorts and tank-tops). The perception of high and low SES attire was further validated within the same pilot test mentioned above by choosing the ten outfits perceived with the highest economic and social position and the ten outfits with the lowest. Each outfit was worn by both Black and White avatars.

Item: The target either held a gun or cell phone in every trial. The gun and phone were matched on size and color to maximize similarity in the two objects. The gun/phone was always attached to the target's right hand (see Figure 1).

3.4 Measures

Race Implicit Association Task (IAT): Participants completed a race IAT while in virtual reality but after avatar trials so as not to prime participants about racial bias. Using the methodology proposed by Greenwald et al. [21], participants were asked to categorize pictures of the White or Black avatars' faces from the shooter bias simulation with positive and negative attribute words, by pulling either the left or right trigger to indicate their decision. The words and faces appeared suspended in front of the participant. Participants were told to work as quickly and accurately as possible.

Implicit racial bias scores (d-scores) were calculated based on both participant's accuracy and speed throughout the combined practice and test trials, using the Greenwald et al. [22] revised scoring algorithm. The final score was calculated by subtracting response latency of the Black-positive and White-negative trials from the White-positive and Black-negative trials. Thus, higher positive scores indicate greater implicit racial bias against Blacks and more negative scores indicate implicit racial bias against Whites. As is suggested by Nosek et al. [39], the order of the pairings was counterbalanced.

Embodiment: Immediately after the experiment, subjective embodiment was measured using the standardized embodiment questionnaire [20]. Following the authors' updated recommendation, final scores were calculated using a revised scale [42]. Embodiment was defined as the average of four sub-measure scores: appearance, response, ownership, and multi-sensory and were calculated as follows:

- Appearance = (Q15 + Q16 + Q8 + Q17 + Q18 + Q20 + Q21 + Q13) / 8
- Response = (Q17 + Q20 + Q24 + Q22 + Q21 + Q12) / 6
- Ownership = (Q18 + Q1 + Q19 + Q14 + Q6 + Q10) / 6
- Multi-Sensory = (Q8 + Q14 + Q6 + Q10 + Q12 + Q13) / 6

The final embodiment score was calculated as the average of the sub-measure scores. Participants responded to statements using a Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). A higher score indicates a stronger feeling of embodiment, whereas a lower score indicates a weaker feeling of embodiment.

Accuracy: A correct choice was determined as correctly shooting at an armed avatar or not shooting at an unarmed avatar. Each participant's percentage accuracy for each condition (race $\times$ ses $\times$ item) was calculated yielding eight accuracy measures per participant. Outlier trials, determined to be more than 1.5 times the inter-quartile range (IQR) from the median were replaced with the median value $\pm 1.5 \times IQR$ (6% of the data).

Latency: Latency to shoot was calculated as the time to shoot from the start of the avatar animation to the time the participant pulled the trigger. Analysis was performed only on correct shooting decisions (armed trial only). Outliers were removed and replaced in the same way as in the accuracy data (less than 1% of the data).

Motion Paths: During the experiment participants' right-hands and heads were tracked via the hand-held tracker and the HMD, respectively. From each device the (x, y, z) position data and $(yaw,\ pitch,\ roll)$ orientation data were logged. For each trial the raw position and orientation data were normalized by subtracting the starting value, subsequently making the tracker data a measure of change in position or orientation from the beginning of each trial. The motion path data for the hand and head were analyzed separately to determine if there were nuanced motion differences that could be detected in either.

Principal component analysis (PCA) was performed on the change in the six tracker-logged dimensions for each trial. The eigenvalues, ordered from highest to lowest, of the covariance matrix were calculated and analysis was performed on *p95*, the sum of the variance accounted for by the first principal components that on average accounted for at least 95% of the variance in the data. This measure is similar to the one proposed by Kilteni et al. [27] who analyzed many more principal components and instead used the total number of principal components to account for at least 95% of the variance. For hand-tracked data the first three principal components on average accounted for at least 95% of the variance in the data, and for head-tracked data it was the first four principal components.

Bias Scores: Additionally, head and hand variance bias scores were calculated following a similar procedure to Vanman et al. [61]. Mimicking the calculation for IAT d-scores, for each participant the mean p95 score for Black avatar trials was subtracted from the mean p95 score for White avatar trials. A positive racial variance bias score indicates more variance in motion explained for White avatar trials while a negative racial variance bias score indicates more variance in motion explained for Black avatar trials. Similar head and hand bias scores were calculated for high and low SES avatars. An additional accuracy bias score was also computed. A positive accuracy bias implies that participants made more correct decisions on trials with White avatars compared to trials with Black avatars while a negative accuracy score implies that participants made more correct decisions on trials with Black avatars compared to trials with White avatars. We investigated if head and hand bias scores were able to predict accuracy bias, d-scores of the race IAT, or embodiment.

3.5 Equipment

Participants were fitted with an HTC Vive Pro head-mounted display (HMD) with approximately 110° field-of-view. The experiment was run in a 4.2m x 5m tracking space and rendered on an Alienware computer running on Windows 10 Pro edition with an Intel i7-7820X processor (3.6 GHz), 32GB RAM, and NVIDIA GeForce GTX1080 Ti GPU. The VE was created using Unity version 2018.6.1f1 using Steam VR libraries.

3.6 Experimental Procedure

Participants were recruited from the Davidson College campus and surrounding community for a study ostensibly on video game interactions in VR. Upon arriving at the Davidson Research in Immersive Environments (DRIVE) Lab, participants were greeted by a White female experimenter who oriented them to the study procedures. All participants were tested individually. First, participants read and signed an informed consent form, and then completed a checklist to ensure they met the inclusion criteria for safe VR research. These included that participants must be 18 or older; cannot be knowingly pregnant; must have normal or corrected to normal vision; cannot have a history of seizures, epilepsy, or strong susceptibility to motion sickness; must be proficient in spoken and written English due to instructions being given in English; have normal or corrected to normal hearing; must be in their usual state of good health at the time of the experiment; must be able to use both hands to manipulate a controller and buttons; must be naïve to the experiment; and must be aware that they will be firing a virtual gun at avatars during the experiment. Participants were also reminded that they could opt out of the experiment at any time without penalty.

The experimenter then explained how the simulation would work and the steps that the participant needed to follow. Participants were then asked to repeat those instructions back to the experimenter to ensure comprehension. Next, while participants were in a seated position the experimenter attached five HTC Vive trackers to participants: at both elbows, both feet, and the right hip. The experimenter then led participants to the center of the room and assisted them with donning an HTC Vive Pro HMD.

To begin the experiment, participants were asked to stand in a T-pose for self-avatar calibration and assignment of the participant's avatar. Each participant was assigned one of four self-avatars depending on their gender and race: a light-skinned male, a light-skinned female, a dark-skinned male, or a dark-skinned female. The self-avatar was co-located with and moved synchronously with the participant in the VE. After avatar assignment and calibration, the participant entered the VE, an alleyway that was much larger than the physical room and created to emulate an industrial district. The participant was positioned at the end of an $8.22m \times 24m$ alleyway surrounded by buildings on three sides in an attempt to induce a feeling of being trapped in order to increase the urgency to make the shoot/don't shoot decision (see Figure 1). To provide participants with visuomotor synchrony a reflective window was positioned directly to their right and a mirror was placed on a building to their front-left. The buildings looked old and unused. It was night time and the only illumination was provided by streetlamps. Through headphones, participants heard footsteps, voices, and dogs barking. To increase realism, a virtual gun was placed in the participant's right hand. Position and orientation data were collected from both the right controller and the HMD at over 120 frames-per-second.

Participants had two minutes to explore the virtual environment. After the two minute acclimation period was over, the experimenter informed the participant that the trials of the experiment were beginning and that participant would need to holster the virtual gun on their right hip to initiate the appearance of the first target. During the trial the participant drew their weapon at a target avatar and then chose to either shoot, by pulling the trigger, on not shoot, by inhibiting the trigger pull. Once a trial was completed, the target avatar disappeared. For the next target to appear, the participant needed to re-holster their gun. Target avatars appeared in a random position in a 5.5m $\times$ 3.25m rectangle beginning 1m in front of the participant. Upon appearing, the avatar remained static for a random time interval ranging from 0 to 1.5 seconds before an animation started where the avatar then reached behind his hip with his right hand and revealed either a cell phone or a gun. After the animation finished, participants had an additional one second to make the shoot/don't shoot decision before the trial timed out and the target disappeared. The next target would only appear when the participants re-holstered their gun, by bringing it towards their right hip. Participants completed a total of 160 trials (40 White $\times$ high SES, 40 White $\times$ low SES, 40 Black $\times$ high SES, 40 Black $\times$ low SES) in 4 blocks of 40. Each block of 40 had equal numbers of White/high, White/low, Black/high, and Black/low avatars presented in a random order.

In order to encourage participants to make their shooting decisions quickly and accurately, as well as to reduce any potential ethical concern on the part of the participants due to the nature of the task, we utilized a point system to "gamify" the procedure. Participants started with 1000 points. A scoreboard was on the left side of the seen, floating above where the avatars would appear. For correctly shooting an armed target, 100 points were added to participant's scores. If an unarmed avatar was incorrectly shot, 100 points were deducted. There was no point penalty for timing out on trials. In the initial introduction to the study, participants were motivated to shoot by being given a fictitious high score (8000 points) supposedly from the prior week of data collection.

Immediately following the experimental trials, and while still wearing the HMD and trackers, participants completed the race IAT. The virtual setting was changed to a dark room, with black walls and black ceilings. They were no longer able to see their embodied self-avatar. Instructions for the IAT were presented on the wall approximately 8 ft in front of them. The IAT used the same faces as the male avatar faces from the shooter bias trials. Results were input via the hand-held tracker buttons.

At the conclusion of the IAT, the experimenter assisted participants with removing the HMD, led them back to the desk chair, and removed the trackers. Participants were then directed to complete a Post-Experiment Survey on the desktop computer. The survey was hosted via Qualtrics, and included the Embodiment questionnaire as well as demographic questions. Additional measures included the Modern Racism Scale [35], the Discrimination Scale, and the External and Internal Motivation to Respond without Prejudice scale [46] were also included but are not relevant to the primary goals of the present research and are therefore not analyzed in detail. Finally, all participants were verbally debriefed by the experimenter, compensated, thanked, and escorted from the lab.

4 Results

4.1 Implicit Association Test

Overall participants demonstrated a pro-Black bias, d=-0.19. However, a significant order effect for the IAT was found, $t(97.99) = -2.49$, $p = .01$, $d = .50$. Participants who completed the Black-negative, White-positive condition first demonstrated significantly higher pro-Black bias ($M = -0.27$, $SE = .007$) compared to participants who completed the Black-positive, White-negative ordering first ($M = -0.10$, $SE = .007$).

4.2 Embodiment

Each Embodiment sub-measure, appearance, response, ownership, and multi-sensory, had high reliability with Chronbach's α ranging from .73 to .82. Embodiment scores varied across the scale with average embodiment scores falling exactly in the middle of the scale ($M = 3.48$, $SD = 1.02$).

4.3 Accuracy

After removing outliers the accuracy data violated the assumption of homogeniety of variance as calculated using Levene's Test, $F(7,800) = 27.45$, $p < .001$. Therefore, accuracy data were analyzed using Wilcox's robust statistical tests since they are robust against violations of parametric assumptions. Participants had high accuracy and ceiling effects were seen in the data. See Table 1 for descriptive statistics.

Table 1: Descriptive statistics of the mean accuracy and standard error (se) for each of the eight race $\times$ SES $\times$ item conditions.

race	SES	item	n	mean	se
Black	high	gun	99	0.96	0.01
Black	high	phone	99	1.00	0.00
Black	low	gun	99	0.95	0.01
Black	low	phone	99	0.99	0.00
White	high	gun	99	0.96	0.01
White	high	phone	99	0.98	0.00
White	low	gun	99	0.96	0.00
White	low	phone	99	0.99	0.00

The accuracy data analysis was performed with a robust three-way ANOVA in R using the WRS2 package comparing (race: White, Black) $\times$ (SES: high, low) $\times$ (item: gun, phone) with corresponding post-hoc tests and a conservative trim value of 10% instead of the standard 20% due to the updated outlier values. A significant main effect of item was found, $Q = 92.13$, $p < .001$. Participants were more accurate at not shooting at avatars holding a phone ($M = 98.81\%$, $SE = .21\%$) compared to shooting at avatars holding a gun ($M = 96.03\%$, $SE = .26\%$). However this was qualified by a higher-order race $\times$ item interaction $Q = 6.46$, $p = .01$.

Race $\times$ Item: Two post-hoc comparisons, considering differences in race for each item, were performed using linear contrasts with Bonferroni corrections applied and significance reported at $p < .025$. A significant accuracy difference was found between Black and White

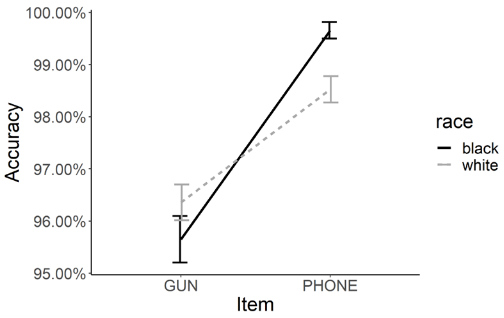

Fig. 2: Average accuracy, with standard error bars, of participant decisions to shoot at avatars holding a gun or not shoot at avatars holding a phone. The item the avatar was holding is along the x-axis. White avatars are dashed-grey lines and Black avatars are solid-black lines.

avatars holding a phone ($\hat{\Psi} = .01$, CI = $(.005, .02)$, $p < .001$) where participants were less accurate with White avatars holding a phone ($M = 98.46\%$, $SE = .31\%$) compared to Black avatars holding a phone ($M = 99.65\%$, $SE = .20\%$). In other words, participants were less likely to shoot unarmed Black avatars than unarmed White avatars. No difference was found between Black and White avatars holding a gun ($\hat{\Psi} = -0.007$, CI = $(-0.02, .004)$, $p = .21$). See Figure 2.

Race $\times$ SES: Additionally, a race $\times$ SES interaction was found, $Q = 4.68$, $p = .03$. Four post-hoc comparisons were performed using linear contrasts. Due to multiple tests Bonferroni corrections were applied and significance is reported at $p < .0125$. A significant accuracy difference was found between high and low SES Black avatars ($\hat{\Psi} = .02$, CI = $(.006, .02)$, $p = .001$) where participants were less accurate with low SES Black avatars compared to high SES Black avatars. No difference was found between high and low SES White avatars ($\hat{\Psi} = -0.002$, CI = $(-0.01, .006)$, $p = .68$). Additionally, a significant accuracy difference was found between Black and White high SES avatars ($\hat{\Psi} = .01$, CI = $(.003, .02)$, $p = .01$) where participants were less accurate with high SES White avatars compared to high SES Black avatars. Finally, no difference was found between Black and White low SES avatars ($\hat{\Psi} = -0.006$, CI = $(-0.01, .003)$, $p = .20$). In other words, participants made the fewest mistakes when presented with high SES Black avatars compared to all other conditions. Additionally SES was not a factor in determining shooting decisions for White avatars, but it was for Black avatars. See Figure 3.

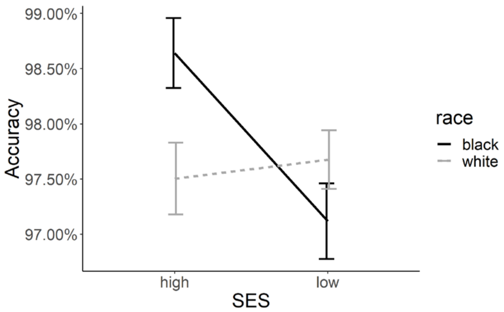

Fig. 3: Average accuracy, with standard error bars, of participant shooting decisions for high and low SES avatars along the x-axis. White avatars are dashed-grey lines and Black avatars are solid-black lines.

4.4 Latency

The assumption of homogeneity of variance was not violated, $F(3, 7436) = 1.124$, $p = .29$. Analysis was performed using a 2 (race: White, Black) $\times$ 2 (SES:high, low) ANOVA with race and SES as within-participant measures. No significant main effects or interactions were found.

4.5 Motion Paths and Bias Scores

Both the p95 hand and head motion data violated the assumption of homogeneity of variance as calculated using Levene's Test, Hand $F(15, 15635) = 25.08$, $p < .001$, Head $F(15, 15635) = 3.35$, $p < .001$. Therefore, the p95 data were analyzed with robust three-way repeated measures ANOVAs comparing (race: White, Black) $\times$ (SES: high, low) $\times$ (decision: correct, incorrect) from the *WRS*2 package in R with corresponding post-hoc tests using a standard trim value of 20%. A robust four-way ANOVA was not used because it does not exist. Data collected for trials were grouped and analyzed by item, either gun or phone, due to motion paths likely being different between the two groups.

Within each type of item and type of motion (head or hand), we first tested for differences in p95 scores as a function of target race, SES, and decision. Then we tested whether motion bias predicts accuracy bias and implicit racial bias (d-scores). For analyses of implicit racial bias, IAT test order was added as a predictor due to the significant order effect found in d-scores. For parsimony, only significant results are reported.

4.5.1 Hand Data

The p95 data was determined to be the variance accounted for in the first three principal components(3PCs): percentage variance in 3PCs ($M = 97.30\%$, $SE = .02\%$), and 2PCs, ($M = 90.09\%$, $SE = .05\%$)).

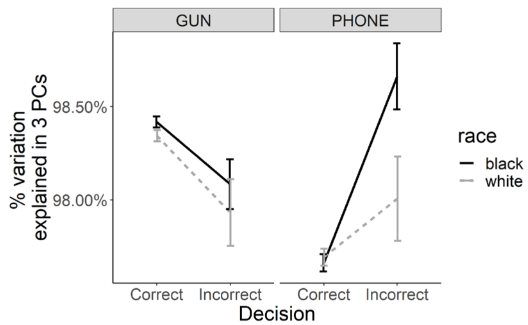

Fig. 4: The percentage of the variation explained by the first three principal components of the participants' hand tracker data with mean and standard error bars.

Gun Data Motion Paths: See Figure 4, left. When the avatar was holding a gun a significant main effect of decision was found, $Q = 10.21$, $p = .002$. Less variance was accounted for in p95 when participants did not shoot at avatars carrying guns ($M = 98.02\%$, $SE = .11\%$, N=386) compared to shooting at avatars carrying guns ($M = 98.38\%$, $SE = .02\%$, N=7,440). Significantly less variance in participant arm movement can be explained (i.e. more complicated hand movements) in p95 when making a decision to not shoot compared to making a decision to shoot. No other significant main effects or interactions were found.

Gun Data Bias Scores: A trending result was seen with the racial hand variance bias score predicting racial accuracy bias, $p = .05$. Visual inspection of the plot identified an outlier that was further confirmed using the Bonferroni outlier test. After removing this participant the hand variance bias score was able to predict participant accuracy bias using the following formula: $\textit{accuracy bias} = 0.01 + 2.19 \times \textit{hand bias}$, $F(1, 96) = 4.34$, $r = .21$, $p = .04$. Participants with more variance explained in their hand motions on White avatar trials compared to Black avatar trials (a positive bias score) tended to show greater shooting decision accuracy on White avatars compared to Black avatars. In other words, participants with greater hand movement complexity for Black avatar trials were less accurate at making shooting decisions for Black avatars. See Figure 5.

Phone Data Motion Paths: See Figure 4, right. When an avatar was holding a phone, significant main effects of race, $Q = 4.35$, $p = .04$ and decision $Q = 20.60$, $p = .001$ were found. These are qualified by the

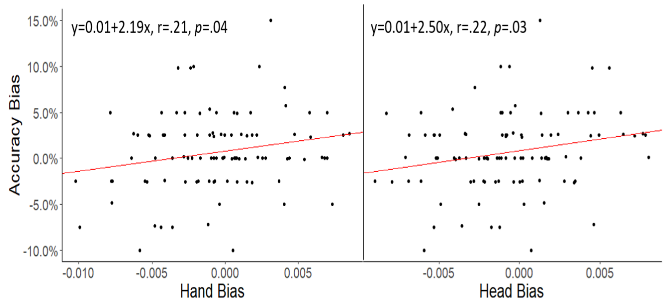

Fig. 5: Scatter plots of trials where avatars carried guns. Bias is defined as White - Black bias. Each participant's shooting accuracy bias (y-axis) is plotted against their (left) hand bias or (right) head bias score (x-axis). A negative score indicates more variance in motion explained for Black avatar trials and a positive score indicates more variance in motion for White avatar trials.

higher-order significant race $\times$ decision interaction, $Q = 5.29$, $p = .02$. No significant differences between avatar race were found when participants made correct decisions to not shoot at avatars carrying phones, ($\hat{\Psi} = -0.03\%$, CI = $(-0.16\%, .10\%)$, $p = .64$). However a significant difference in p95 was found when participants incorrectly chose to shoot at unarmed avatars, ($\hat{\Psi} = .66\%$, CI = $(.09\%, 1.22\%)$, $p = .02$). More variance was explained in p95 when incorrectly shooting at Black avatars, ($M = 98.66\%$, $SE = .18\%$, n=76) compared to White avatars ($M = 98.01\%$, $SE = .23\%$, n=93). Moreover, there was no significant difference between p95 for the correct and incorrect decisions made for White avatars, ($\hat{\Psi} = -0.31\%$, CI = $(-0.77\%, .14\%)$, $p = .17$), but there was a significant difference for Black avatars ($\hat{\Psi} = -1.00\%$, CI = $(-1.36\%, -0.63\%)$, $p < .001$).

Participants who incorrectly shot unarmed Black avatars had significantly simpler hand movements, which may be attributed to decisiveness and lack of hesitation in making a shooting decision, compared to shooting at unarmed White avatars.

Phone Data Bias Scores: No significant bias predictions were found.

4.5.2 Head Data

The p95 data was determined to be the variance accounted for in the first four principal components (4PCs): percentage variance in 4PCs ($M = 97.95\%$, $SE = .01\%$), and 3PCs, ($M = 93.49\%$, $SE = .04\%$)).

Gun Data Motion Paths: No significant main effects or interactions were found when investigating head movements of participants in trials where the avatar was holding a gun.

Gun Data Bias Scores: Similarly to the hand data, the racial head variance bias score was able to predict racial accuracy bias using the following formula: $accuracy\ bias = 0.01 + 2.50 \times head\ bias$, $F(1,97) = 5.11$, $r = .22$, $p = .03$. Participants with more variance explained in their head motions on White avatar trials compared to Black avatar trials also tended to have higher shooting accuracy on White avatars compared to Black avatars. See Figure 5.

The racial head variance bias score was also able to predict race IAT d-score; note that this base model was significantly improved when also including the order of the IAT test, $R^2 = .11$, $p = .003$. See Table 2. Participants with more variance explained in their head motions on White avatar trials compared to Black avatar trials also tended to have higher pro-White d-scores. See Figure 6.

Phone Data Motion Paths: When an avatar was holding a phone, a significant main effect of decision, $Q = 14.71$, $p = .001$, and a trending race $\times$ SES interaction, $Q = 3.94$, $p = .05$, was found. These were qualified by the higher-order significant three-way race $\times$ SES $\times$ decision interaction, $Q = 4.39$, $p = .04$. See Figure 7. Two post-hoc tests were run and significance is reported at $p < .025$. For high SES avatars no significant difference was found when comparing the variance explained between correct and incorrect decisions ($\hat{\Psi} = .07\%$, CI = $(-.33\%, .47\%)$, $p = .71$) however a significant difference was found for White avatars ($\hat{\Psi} = .90\%$, CI = $(.22\%, 1.58\%)$, $p = .01$).

	R^2	B	SE B	β	p
Step 1	.05				.03
Intercept		-0.18	0.04		<.001
Head Bias		21.27	9.51	.22	0.03
Step 2	0.11				0.003
Intercept		-0.27	0.05		<.001
Head Bias		23.27	9.26	.24	.01
Order		0.19	0.07	.26	0.009

Table 2: Step 1: The basic linear regression considering head-bias and d-score. The model was significantly improved when IAT test order was added as a predictor. The R^2, B, standard error of B, β, and $p-values$ are reported.

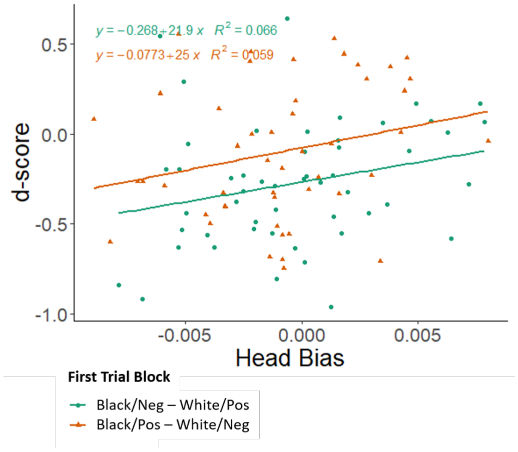

Fig. 6: A scatter plot of trials where avatars carried guns. Each participant's d-score (y-axis) is plotted against their head bias score (x-axis). Positive d-score indicate a pro-White bias and negative d-scores indicate a pro-Black bias. Negative head bias scores indicate more variance in head motion explained for Black avatar trials and positive scores indicate more variance in head motion explained for White avatar trials.

Phone Data Bias Scores: No significant bias predictions were found.

5 Discussion

One of the primary goals of the present research was to improve upon the previous methodology used in desktop-based shooter bias simulation studies by providing a robust test of shooter bias within a fully immersive VE. While prior psychological research has replicated shooter bias in several different populations across different desktop-based environments [11, 12, 26, 23, 37], and even with video targets [13], to date no published research has investigated shooter bias using VR technology. By doing so, we were able to collect more complex movement data than has been available before, adding nuance to the understanding of shooting behaviors and not just shooting decisions.

In contrast with past desktop-based research on shooter bias, in the present study we did not find any significant effects for the latency data. We argue that the long length of time given to participants to make a shooting decision may have washed out any latency-related effects. Additionally, with the increased realism of the VE, participants may have been less inclined to focus simply on quickly making the shooting decision; in other words, shooting a realistic avatar in front of them may have made the shooting decision feel more consequential. Note also that although participants received points for correct shooting decisions, there was no incentive based on how quickly they made those decisions. It may also be that greater variance was introduced by requiring participants to lift their arm and aim at the avatar rather than simply press a button on a keyboard; the greater variance could have

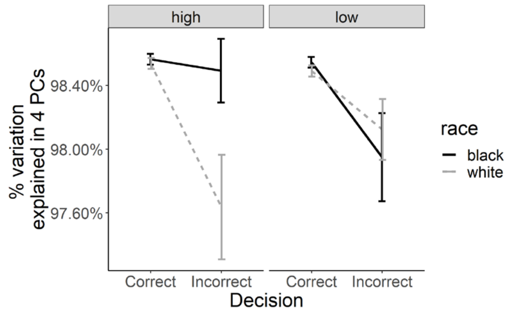

Fig. 7: The percentage of the variation explained by the first four principal components (y-axis) of the participants' head tracker data for trials when the avatar holds a phone, with mean and standard error bars, grouped by high SES (left) and low SES (right) with the decision (correct: not shoot, incorrect: shoot) along the x-axis.

obscured latency-related effects.

Because of the move to VR, we did not make specific predictions about accuracy or latency effects. However, based on patterns observed in desktop-based tests of shooter bias, we would have expected that that participants would be less accurate with Black avatars compared to White avatars, particularly when avatars were dressed in low SES clothing. This is because a signal of high SES (business suits) could disrupt automatic stereotype processing and therefore lead to greater accuracy for high SES Black avatars (equivalent to White avatars). Our accuracy data experienced ceiling effects in which most participants had high accuracy. Making the task more challenging, for example by shortening the time to respond, may reveal additional differences in accuracy similar to previous shooter bias studies. What we actually observed in our data was that participants were most accurate in their decisions when presented with high SES Black avatars, and to a greater extent than when interacting with high SES White avatars. This finding highlights a general tendency we observed throughout our outcome variables; participants either genuinely held a pro-Black bias or were attempting to correct for an undesired pro-White bias. For example, we observed that participants tended to not mistakenly shoot unarmed Black avatars (significantly greater accuracy than with unarmed White avatars) and to err on the side of not shooting armed Black avatars (though not significantly lower than armed White avatars). Participants' implicit bias scores were surprisingly pro-Black on average. It is possible that our sample of participants was simply less racially biased than samples typically used in shooter bias research. More likely however, we think that the VR scenario of actually having to point a gun at someone and pull the trigger coupled with widespread press and public outcry over recent instances of police brutality inflicted upon Black individuals made participants hyper-aware of their own potential for bias. This awareness was likely compounded by participants being primed by the experiment before taking the IAT. Indeed, one reason we did not observe latency effects may have been that participants were actively controlling their decisions rather than acting as quickly as possible. The present results therefore highlight how shooter bias might be observed differently in an immersive VE than in a standard desktop simulation.

In particular, tracking different types of variables such as movement might be more informative when participants are attempting to control their biased responding. From our motion path data, we observed that hand movement was less complex (more variance explained) when correctly shooting an armed avatar rather than making an incorrect decision to not shoot an avatar with a gun. Variance in motion paths could be caused by a multitude of factors. One possibility is that greater hand movement complexity could be due to indecision or hesitation. Indeed, hand movement was more complex when participants incorrectly shot unarmed White avatars compared to unarmed Black avatars, potentially indicating more hesitation or indecision for unarmed White compared to Black targets. If motion complexity is a behavioral symptom of cognitive indecision, then measurement of motion paths may reveal racial bias where none (or the opposite pattern) was present using only typical desktop outcome variables.

Similarly, head motion data support the assertion that movement complexity is associated with bias. In our data we observed that pro-White head motion bias (less movement complexity for White trials compared to Black trials) was positively related to pro-White implicit bias scores. Mirroring this, pro-White head motion bias was also associated with greater accuracy within White avatar trials, whereas pro-Black head motion bias (less movement complexity for Black trials compared to White trials) was positively related to accuracy within Black avatar trials. The analysis of motion complexity bias is novel to the literature on shooter bias, and as such requires replication. Based on the pattern of results for hand and head motion complexity however, we posit that body movement (as mapped in a VE) could be used as an indicator of racial bias, even when participants might be trying to control said bias. More research is necessary to validate the use of motion complexity as a behavioral indicator of cognitive indecision/bias, however the present research highlights the importance of using VR to study the complex phenomenon of shooter bias.

Interestingly, although signals of SES (clothing) moderated the effect of avatar race for the accuracy of participants' shooting decisions, we did not find effects of SES on motion path data. Because race is a basic social category and tends to be processed within 200ms of stimulus presentation [30], perhaps avatar race influenced less controlled variables such as the complexity of a motion path. While signals of SES can also be processed quickly (within 60s [29]), the somewhat slowed processing relative to race may make SES a stronger determinant of the final shooting decision rather than the process of shooting. This argument is consistent with work showing that although race was processed immediately, individuating information ultimately did impact behavioral outcomes [30]. In the present study then, perhaps hand and head motion is sensitive to the immediate processing of race information, whereas the ultimate decision of whether or not to shoot was also impacted by clothing signalling SES (individuating information). Without the use of motion tracking in VR this argument would be impossible to assess.

5.1 Implications

Biased treatment of people with Black and Brown skin at the hands of law enforcement is a critically important societal issue. Large scale participation in social movements such as Black Lives Matter demonstrate the public's desire to see real change in the way that police interact with and use force against Black suspects. Importantly, we are not just at a moment of societal unrest regarding this issue, but have also reached a time in which police departments are recognizing the need for reform and turning to researchers for guidance in how to enact change. For example, the Oakland, CA Police Department has partnered with the SPARQ Lab (Stanford University - [3]) both to conduct research on their policing practices and to develop training to reduce observed racial bias within their department. The Madison, WI Police Department has partnered with researchers in order to test shooter bias effects within samples of real police officers [13]. At a national level, the U.S. Department of Justice has funded researchers in a national initiative to improve relationships between community and police, one goal of which is to address implicit bias in policing [18, 2]. Thus, the national conversation has turned from "what is the problem" to "what can be done." It is at this juncture that VR researchers can provide critical research and tools to support the mission of ending racial bias in police use of force.

Although police departments seem more willing than ever to engage in implicit bias training, these types of trainings are often criticized for being ineffective [31]. Training conducted in VR engages the participants in an experiential setting, with sights, sounds, and emotions similar to real life [9]. The immersive nature as well as embodiment in an avatar likely increases self-involvement, one component of effective implicit bias interventions [31]. Thus we argue that VR offers a critical opportunity for training to remediate implicit bias, and the development of these systems will further the understanding of human behavior and response to VR. Indeed, VR research has already shown that cross-race body swap illusions can be effective at reducing implicit racial bias [6, 44]. Importantly, police and military already utilize VR for weapons training and combat simulations, [58] and mental health calls [60] as well as more mundane procedures like traffic stops [48]. Reasons for implementing VR training in law enforcement include increased safety, reduced cost, opportunity for correction of officer behavior, and learning [48].

Importantly, the present research highlights the disconnect between controlled, explicit decisions (e.g., the decision to shoot) and less controlled processes (e.g., the movement involved in the decision making process). Our sample of participants surprisingly demonstrated an average pro-Black implicit bias and very high accuracy in shooting decisions regardless of avatar race. Thus, we believe our participants were actively attempting to inhibit their racial bias, and yet it surfaced in the complex motion paths of their hands and heads. We argue that focusing solely on using VR to train individual police officers is therefore not enough. Anti-Black racism is culturally pervasive and it would be naive to think it can simply be trained out of anyone. Structural, policy and procedural-level changes are necessary to really reduce disproportionate use of force by police. The present research demonstrates the utility of VR in making the argument that structural change is necessary. Regardless of whether individual police officers “pass” an implicit association test, as many of our participants did, measures of complex motion in VR showcase that bias is still present. We argue that using VR as an assessment tool may provide greater incentive for police departments to invest in suggested structural changes that have been gaining public support in the U.S. and across the globe [5].

Finally, though participants were given a self-avatar, the self-avatar appearance (age, race, gender, etc.) was not modified, and the subjective level of embodiment did not effect accuracy, latency, or motion path data. This suggests, not surprisingly, that the extent to which participants are embodied in an avatar that matches their gender and race identities is unrelated to assessments of bias. However, embodiment in an avatar representing different identities might show very different patterns. The present research sets the foundation for further investigating the effects of self-avatars that are visually different from their owners on shooter bias and other socially complicated scenarios. Head and hand motion path data may provide nuanced and implicit metrics to better understand and detect the psychological effects of both matched and manipulated self-avatars on their owners.

5.2 Limitations and Future Directions

The present research utilizes a novel methodology to test a robust psychological phenomenon and lays the foundation for developing a virtual reality application for training away implicit bias. However the work is not without limitations. The primary limitation is the sample, both in size and composition. Although considerable time and resources were deployed to sample one hundred participants, the size does limit broad, population-level generalizability. Moreover, we did not sample police officers specifically. Desktop-based shooter bias studies with police officers have shown mixed results [13]. Future research should include police officers within a VR methodology in order to determine whether the novel results we obtained for complex motion paths replicate in a police sample.

The motion path data were exploratory and therefore need additional validation. Our finding of a significant relationship between accuracy bias and IAT d-score is promising, however future replication work is essential for establishing the reliability and validity of motion path data as an indicator of bias. Further, participants were not screened for handedness, but were required to hold a virtual gun in their right-hand. It is possible that holding the gun in the non-dominant hand added noise into the data and we recommend future replications control for handedness.

Additionally, within the virtual environment, targets were 3-dimensional but did not walk or move beyond an animation to draw an item from their pocket and point it at the participant. Participants also earned points for correctly shooting armed avatars and lost points for incorrectly shooting unarmed avatars. These gamification choices were made for ethical and safety considerations for our participants. Future work should consider evaluating the ecological validity of VR for shooter bias scenarios using more realistic environments. Increasing the realism and intensity of the situation could affect plausibility of the threatening environment and therefore produce more realistic participant behavior [55]. Future research aimed at testing or training police officers should consider increasing plausibility with more dynamic targets as greater movement likely interferes with shooters’ ability to accurately process the information available and make the correct shooting decision. Similarly, greater scene complexity, including bystanders or multiple threatening targets, might also impact both shooting decisions and motion paths while better mimicking real-world situations encountered by police. Inducing greater stress through higher stakes consequences of shooting decisions (threat of actual harm rather than just losing points) might lead to more intuition- or stereotype-based shooting decisions [64]. In the present research, we utilized identifiably Black and White male avatars dressed in casual clothes and suits. Future research could consider other race or social identity groups (see [59]), as well as other methods for manipulating SES. Contrary to predictions, we did not find any effects of race or SES on latency to shoot. Future research could consider incentivizing faster shooting responses in order to observe greater variation in latency.

Though we have identified several incremental future directions for this line of research that would advance the understanding of basic science processes, we argue that the most compelling and urgent future direction is to use VR to reduce disparate use of force against Black individuals. This major societal issue is caused by a combination of individual (implicit bias) and structural (policy-level) factors. Importantly, we argue that VR has the potential to impact both of these paths leading to police bias. First, mounting evidence shows that VR can be used to reduce implicit bias [6, 44]. Future work should test whether raising awareness of individual bias, as measured by motion paths, could reduce future instances of bias. Second, VR research can be used as evidence in support of policy change by lawmakers and police officials, as well as a way of testing the efficacy of structural policy changes. For example, policy reform movements such as 8Can’tWait [5] propose policies requiring de-escalation and a warning before shooting. Training police officers to use and practice these tactics through VR would enable researchers to assess the impact of such changes on shooting decisions. Credible, scientific evidence about the efficacy or non-efficacy of policy changes would support efforts to engage in wide-scale reform. We call on VR researchers to utilize our science for the public good by applying our research skills and resources to the longstanding and far reaching problem of racist violence.

6 Conclusion

Using a novel methodological framework, we investigated “shooter bias” within a fully immersive virtual environment. Consistent with past desktop-based research, we found that signals of SES disrupted automatic racial stereotypes, leading participants to be most accurate in their shooting decisions when encountering high SES Black avatars. Importantly, by utilizing head and hand motion tracking within VR we were able to identify a potentially new way of assessing racial bias in shooting behavior. Indeed, differential motion complexity for White versus Black avatars was associated both with measures of implicit bias and shooting accuracy within avatar racial groups. The research comes at a precipitous time, in which there is substantial public motivation to reduce racial bias in policing, funding for research to do so, and engagement from law enforcement offices. This research provides the necessary foundation for critically important future research assessing nuanced racial bias among police officers, as well as strategies for reducing shooter bias.

Acknowledgments

The authors wish to thank the Davidson Research Initiative (DRI), and the Davidson Faculty, Study, and Research (FS&R) grant for supporting this work. We further thank student contributor Evan Blanpied and the reviewers and Evan Suma Rosenburg for helpful and thoughtful feedback.

REFERENCES

[1] Implicit bias police training simulator. https://www.apexofficer.com/implicit-bias.

[2] National initiative for building community trust and justice. https://trustandjustice.org/, 2019.

[3] Stanford university sparq lab: Criminal justice projects, 2019.

[4] George floyd: What happened in the final moments of his life. *BBC*, May 2020.

[5] *#8 Can't Wait*, Accessed 2020-07-03. Campaign Zero, http://https://8cantwait.org/.

[6] D. Banakou, P. D. Hanumanthu, and M. Slater. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience*, 10:601, 2016.

[7] D. Banakou, S. Kishore, and M. Slater. Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias. *Frontiers in psychology*, 9:917, 2018.

[8] M. Botvinick and J. Cohen. Rubber hands 'feel'touch that eyes see. *Nature*, 391(6669):756, 1998.

[9] F. P. Brooks. What's real about virtual reality? *IEEE Computer graphics and applications*, 19(6):16–27, 1999.

[10] J. Correll, B. Park, C. M. Judd, and B. Wittenbrink. The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83(6):1314–1329, 2002. doi: 10.1037/0022-3514.83.6.1314

[11] J. Correll, B. Park, C. M. Judd, B. Wittenbrink, M. S. Sadler, and T. Keesee. Across the thin blue line: police officers and racial bias in the decision to shoot. *Journal of personality and social psychology*, 92(6):1006, 2007.

[12] J. Correll, B. Wittenbrink, B. Park, C. M. Judd, and A. Goyle. Dangerous enough: Moderating racial bias with contextual threat cues. *Journal of experimental social psychology*, 47(1):184–189, 2011.

[13] W. T. Cox, P. G. Devine, E. A. Plant, and L. L. Schwartz. Toward a comprehensive understanding of officers' shooting decisions: No simple answers to this complex problem. *Basic and Applied Social Psychology*, 36(4):356–364, 2014. doi: 10.1080/01973533.2014.923312

[14] J. F. Dovidio, K. Kawakami, and S. L. Gaertner. Implicit and explicit prejudice and interracial interaction. *Journal of personality and social psychology*, 82(1):62, 2002.

[15] P. S. Forscher, C. K. Lai, J. R. Axt, C. R. Ebersole, M. Herman, P. G. Devine, and B. A. Nosek. A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*, 117(3):522, 2019.

[16] R. Fribourg, F. Argelaguet, A. Lécuyer, and L. Hoyet. Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2062–2072, 2020.

[17] O. Gillath, A. J. Bahns, F. Ge, and C. S. Crandall. Shoes as a source of first impressions. *Journal of Research in Personality*, 46(4):423–430, 2012.

[18] P. A. Goff. Center for policing equity: What we do. https://https://policingequity.org/, 2019.

[19] J. S. Goldkamp. Minorities as victims of police shootings: Interpretations of racial disproportionality and police use of deadly force. *The Justice System Journal*, pp. 169–183, 1976.

[20] M. Gonzalez-Franco and T. C. Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018.

[21] A. G. Greenwald, D. E. McGhee, and J. L. Schwartz. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6):1464, 1998.

[22] A. G. Greenwald, B. A. Nosek, and M. R. Banaji. Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of personality and social psychology*, 85(2):197, 2003.

[23] A. G. Greenwald, M. A. Oakes, and H. G. Hoffman. Targets of discrimination: Effects of race on responses to weapons holders. *Journal of Experimental Social Psychology*, 39(4):399–405, 2003.

[24] L. James, B. Vila, and K. Daratha. Results from experimental trials testing participant responses to white, hispanic and black suspects in high-fidelity deadly force judgment and decision-making simulations. *Journal of Experimental Criminology*, 9(2):189–212, 2013.

[25] A. Jazeera. International reaction to george floyd killing. https://www.aljazeera.com/news/2020/06/international-reaction-george-floyd-killing.html, 2020.

[26] K. B. Kahn and P. G. Davies. What influences shooter bias? the effects of suspect race, neighborhood, and clothing on decisions to shoot. *Journal of Social Issues*, 73(4):723–743, 2017.

[27] K. Kilteni, A. Maselli, K. P. Kording, and M. Slater. Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in human neuroscience*, 9:141, 2015.

[28] A. L. Kobler. Figures (and perhaps some facts) on police killing of civilians in the united states, 1965–1969. *Journal of Social Issues*, 31(1):185–191, 1975.

[29] M. W. Kraus, J. W. Park, and J. J. Tan. Signs of social class: The experience of economic inequality in everyday life. *Perspectives on Psychological Science*, 12(3):422–435, 2017.

[30] J. T. Kubota and T. Ito. Rapid race perception despite individuation and accuracy goals. *Social neuroscience*, 12(4):468–478, 2017.

[31] C. K. Lai, M. Marini, S. A. Lehr, C. Cerruti, J.-E. L. Shin, J. A. Joy-Gaba, A. K. Ho, B. A. Teachman, S. P. Wojcik, S. P. Koleva, et al. Reducing implicit racial preferences: I. a comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, 143(4):1765, 2014.

[32] A. Lele. Virtual reality and its military utility. *Journal of Ambient Intelligence and Humanized Computing*, 4(1):17–26, 2013.

[33] D. S. Ma, J. Correll, B. Wittenbrink, Y. Bar-Anan, N. Sriram, and B. A. Nosek. When fatigue turns deadly: The association between fatigue and racial bias in the decision to shoot. *Basic and Applied Social Psychology*, 35(6):515–524, 2013.

[34] M. K. MacLin and V. Herrera. The criminal stereotype. *North American Journal of Psychology*, 8(2):197–208, 2006.

[35] J. B. McConahay. Modern racism, ambivalence, and the modern racism scale. 1986.

[36] B. J. Mohler, S. H. Creem-Regehr, W. B. Thompson, and H. H. Bülthoff. The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment. *Presence: Teleoperators and Virtual Environments*, 19(3):230–242, 2010.

[37] S. Moore-Berg, A. Karpinski, and E. A. Plant. Quick to the draw: How suspect race and socioeconomic status influences shooting decisions. *Journal of Applied Social Psychology*, 47(9):482–491, 2017.

[38] J. Nix, B. A. Campbell, E. H. Byers, and G. P. Alpert. A bird's eye view of civilians killed by police in 2015: Further evidence of implicit bias. *Criminology & Public Policy*, 16(1):309–340, 2017.

[39] B. A. Nosek, A. G. Greenwald, and M. R. Banaji. Understanding and using the implicit association test: Ii. method variables and construct validity. *Personality and Social Psychology Bulletin*, 31(2):166–180, 2005.

[40] B. K. Payne. Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of personality and social psychology*, 81(2):181, 2001.

[41] T. C. Peck, M. Doan, K. A. Bourne, and J. J. Good. The effect of gender body-swap illusions on working memory and stereotype threat. *IEEE transactions on visualization and computer graphics*, 24(4):1604–1612, 2018.

[42] T. C. Peck and M. Gonzalez-Franco. Avatar embodiment. a standardized questionnaire. *Frontiers in Virtual Reality*, 2021. Preprint at https://www.frontiersin.org/articles/10.3389/frvir.2020.575943.

[43] T. C. Peck, J. J. Good, and K. A. Bourne. Inducing and mitigating stereotype threat through gendered virtual body-swap illusions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.

[44] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition*, 22(3):779–787, 2013. doi: 10.1016/j.concog.2013.04.016

[45] T. C. Peck and A. Tutar. The impact of a self-avatar, hand collocation, and hand proximity on embodiment and stroop interference. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1964–1971, 2020.

[46] E. A. Plant and P. G. Devine. Internal and external motivation to respond without prejudice. *Journal of personality and social psychology*, 75(3):811, 1998.

[47] E. A. Plant, B. M. Peruche, and D. A. Butz. Eliminating automatic racial bias: Making race non-diagnostic for responses to criminal suspects. *Journal of Experimental Social Psychology*, 41(2):141–156, 2005.

[48] E. L. Reyes. How do police use vr? very well. https://www.policefoundation.org/.

[49] G. D. Robin. Justifiable homicide by police officers. *J. Crim. L. Criminology & Police Sci.*, 54:225, 1963.

[50] L. A. Rudman, R. D. Ashmore, and M. L. Gary. " unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, 81(5):856–868, Nov. 2001. doi: 10.1037/0022-3514.81.5.856

[51] M. S. Sadler, J. Correll, B. Park, and C. M. Judd. The world is not black and white: Racial bias in the decision to shoot in a multiethnic context. *Journal of Social Issues*, 68(2):286–313, 2012.

[52] N. Salmanowitz. The impact of virtual reality on implicit racial bias and mock legal decisions. *Journal of Law and the Biosciences*, 5(1):174–203, 2018.

[53] K. Scott, D. S. Ma, M. S. Sadler, and J. Correll. A social scientific approach toward understanding racial disparities in police shooting: Data from the department of justice (1980–2000). *Journal of Social Issues*, 73(4):701–722, 2017. doi: 10.1111/josi.12243

[54] J. J. Sim, J. Correll, and M. S. Sadler. Understanding police and expert performance: When training attenuates (vs. exacerbates) stereotypic bias in the decision to shoot. *Personality and social psychology bulletin*, 39(3):291–304, 2013.

[55] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

[56] M. Slater and M. V. Sanchez-Vives. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3:74, 2016.

[57] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)*, pp. 67–76. IEEE, 2016.

[58] J. Strickland. How virtual reality military applications work. `https://science.howstuffworks.com/virtual-military.htm`, 2007.

[59] C. Unkelbach, J. P. Forgas, and T. F. Denson. The turban effect: The influence of muslim headgear and induced affect on aggressive responses in the shooter bias paradigm. *Journal of Experimental Social Psychology*, 44(5):1409–1413, 2008. doi: 10.1016/j.jesp.2008.04.003

[60] L. R. Valmaggia, L. Latif, M. J. Kempton, and M. Rus-Calafell. Virtual reality in the psychological treatment for mental health problems: an systematic review of recent evidence. *Psychiatry research*, 236:189–195, 2016.

[61] E. J. Vanman, J. L. Saltz, L. R. Nathan, and J. A. Warren. Racial discrimination by low-prejudiced whites: Facial movements as implicit measures of attitudes related to behavior. *Psychological Science*, 15(11):711–714, 2004.

[62] D. S. Vincent, A. Sherstyuk, L. Burgess, and K. K. Connolly. Teaching mass casualty triage skills using immersive three-dimensional virtual reality. *Academic Emergency Medicine*, 15(11):1160–1165, 2008.

[63] N. Yee and J. Bailenson. The proteus effect: The effect of transformed self-representation on behavior. *Human communication research*, 33(3):271–290, 2007.

[64] R. Yu. Stress potentiates decision biases: A stress induced deliberation-to-intuition (sidi) model. *Neurobiology of stress*, 3:83–95, 2016.