Sameer Kishore
Bernhard Spanlang
Guillermo Iruretagoyena
Virtual Bodyworks S.L. Barcelona, Spain

Shivashankar Halan
Key Lime Interactive,
Doral, Florida,
United States

Dalila Szostak
Jigsaw
New York, NY
United States

Mel Slater*
Event Lab
Faculty of Psychology
University of Barcelona Barcelona, Spain

and

Institute of Neurosciences of the University of Barcelona Barcelona, Spain

A Virtual Reality Embodiment Technique to Enhance Helping Behavior of Police Towards a Victim of Police Racial Aggression

Abstract

There is an alarming level of violence by police in the US towards African Americans. Although this may be rooted in explicit racial bias, the more intractable problem is overcoming implicit bias, bias that is non-conscious but demonstrated in actual behavior. If bias is implicit, it is difficult to change through explicit methods that attempt to change attitudes. We carried out a study using virtual reality (VR) with 38 officers in a US police department, who took part in an interrogation of an African American suspect alongside an officer who was racially abusive towards the suspect. Seventeen of the participants witnessed the interview again from a third person perspective (Observer) and 21 from the embodied perspective of the suspect, now a victim of the interrogation (Victim condition), having been assigned randomly to these two groups. Some weeks later, all witnessed aggression by an officer towards an African American man in a virtual cafe scenario. The results show that the actions of those who had been in the Victim condition were coded as being more helpful towards the victim than those in the Observer condition. We argue that such VR exposures operate at the experiential and implicit level rather than the explicit, and hence are more likely to be effective in combating aggression rooted in implicit bias.

1 Introduction

Violence by police against African Americans in the United States has been at an alarming level, with recurring internationally reported incidents (Peeples, 2019) most recently resulting in a massive resurgence of the Black Lives Matter movement. Evidence suggests that unarmed Black suspects are more likely to be shot by police than armed White suspects (Scott, Ma, Sadler, & Correll, 2017). Apart from the human misery this causes, including long-term psychological trauma for survivors and their families, this is perceived as an assault on the entire community, which ultimately can lead to more violence. Recent results—for example, Banakou et al. (2020)—that exploit the power of Virtual Reality to give people experiences of being another person suggest that it may be possible to diminish implicit racial bias. This paper describes an experimental study that utilizes such findings in the context of police racial bias.

Throughout this paper we use the terminology "Black" to denote African Americans and "White" to denote people of (mainly) European heritage,

Presence, *Vol. 28, No. 1, Winter 2019*, 1–23
https://doi.org/10.1162/PRES_a_00339

*Corresponding to melslater@ub.edu.

without these words signifying anything other than labels. Also, this largely accords with the terminology used in the research literature.

The issue of police violence against Black people is a severe one. Ross (2015) reports that the probability in the United States of being Black and unarmed and shot by the police is nearly 3.5 times the probability of being shot by the police if White and unarmed. There is considerable variation across different counties with this ratio being 20 to 1 in some areas. Willits and Makin (2017) found that police are faster to use force against male suspects in general, and particularly Black male suspects. Police racial discrimination is likely to be diminished in more racially diverse police forces, since one of the causes (the perception of threat by certain groups) is diminished (Legewie & Fagan, 2016). Additionally though, there is an increasing tendency for police forces to become militarized (Baumgart, 2016) which exacerbates the problem. The consequences of police bias and violence are many, including long-term psychological trauma amongst survivors and in the community generally (DeVylder et al., 2017, 2016; Ross, 2017).

Two major contributors to racial bias have been considered in this context. The first is institutional, the set of norms and behaviors that accumulate over a long time that establish unspoken, perhaps even non-conscious, deep-seated prejudices within an organization that locks behavior into ritualized patterns of response to encountered situations. Our research is not directly aimed at such institutional racism, which is best combatted by thorough, repeated, conscious, and highly explicit reviews of such norms and behavior directed from the very highest levels of authority in the organization, and indeed in the wider society. An example would be the root and branch attempt to reform the police in the UK following the murder of Black teenager Stephen Lawrence, and the failure of police to adequately prosecute the case against the (White) perpetrators for many years (Bourne, 2001). However, institutional racism does rely on the racial bias of the individuals within the organization for its perpetuation and transmission through the generations of recruits, and policies such as the "war on drugs" lead to collateral damage to Black communities (Fellner, 2009).

Racial bias may often be non-conscious (implicit) rather than explicit, matching what happens in the general society—where even though individuals may report a low level of explicit racial bias, tests that go beyond opinions, that directly test implicit bias, tend to show that there is a strong degree of implicit bias even amongst the most liberal individuals (Dovidio & Gaertner, 2004). Such findings usually rely on the "implicit association test" (Greenwald, McGhee, & Schwartz, 1998) which is a reaction time test based on the idea that if people pair (for example) White faces with good attributes and Black faces with bad attributes faster than White faces with bad attributes and Black faces with good attributes, then this indicates a bias towards White. The implicit association test for racial bias (IAT) has been shown to correlate with behavior in reality (Rooth, 2010). Hence, whereas institutional racism encodes implicit bias at the organizational level, implicit bias in the sense measured by the IAT elicits it at the individual level. The two are very tightly interrelated since the set of negative associations with Black at the individual level are daily reinforced through the norms, attitudes, and behaviors of the institution and the wider society. Hence by tackling one, we also address the other.

There are many interventions discussed in the literature designed to reduce implicit racial bias of White against Black people. Seventeen interventions were described and evaluated by Lai et al. (2014) of which the most successful involved examples that countered expectations based on stereotypical behavior, and the least effective involved imaginal perspective taking or the invocation of egalitarian values. In a subsequent study Lai et al. (2016) found that not one of the interventions discussed was successful in inducing a sustained reduction of racial bias, even for hours, let alone weeks or months.

Unlike imaginal perspective, our own work has relied upon virtual embodiment, where a person's body is visually substituted by another using virtual reality. When a person wears a head-tracked wide field-of-view head-mounted display, the scenario can be programmed so that when they look down towards their own body or towards a reflective surface, they would see a life-sized virtual body in place of their own. If real-time motion capture is used, then that body can be programmed to

move synchronously and in accordance with the real movements of the person. We refer to this as visuomotor synchrony. The first-person perspective view over the virtual body together with at least visuomotor synchrony is sufficient to give the person the powerful perceptual illusion that the virtual body is their own (even though they know that this is not the case)—for example, Banakou and Slater (2017).

Research on body ownership was inspired by the Rubber Hand Illusion (RHI) (Armel & Ramachandran, 2003; Botvinick & Cohen, 1998). In the RHI, a participant sits by a table on which there is a rubber hand that is in an anatomically plausible position with respect to the person's posture. The experimenter touches and strokes the rubber hand while tapping and stroking the person's corresponding hidden real hand synchronously in time and location. After approximately 20 s of stimulation, proprioception typically shifts to the rubber hand, which feels to the participant as if it were their real hand. Petkova and Ehrsson (2008) applied this technique of visuotactile synchrony to the whole body using a pair of cameras mounted on top of a manikin pointing down to its body that streamed to a head-mounted display worn by the participants, so that they saw the manikin body replacing their own while seeing tapping on the manikin body and feeling it synchronously on their own body. It was shown in Slater, Spanlang, Sanchez-Vives, and Blanke (2010) that the body ownership illusion can be realized in virtual reality, even amongst men with respect to a female virtual child body. Rather than visuotactile synchrony, the RHI can also be achieved with visuomotor synchrony in virtual reality (Sanchez-Vives, Spanlang, Frisoli, Bergamasco, & Slater, 2010) or with a rubber hand (Kalckert & Ehrsson, 2012). Visuomotor synchrony (with first-person perspective) also leads to body ownership over a full virtual body (for example, Banakou, Groten, & Slater, 2013; Banakou & Slater, 2014) and may have advantages over visuotactile synchrony (Kokkinara & Slater, 2014).

As well as the illusion of ownership over the virtual body, the evidence suggests that the type of the virtual body can influence the person's perception, attitudes, behaviors, physiology, and cognition. This is the idea first referred to as the Proteus Effect (Yee & Bailenson, 2007; Yee, Bailenson, & Ducheneaut, 2009). For example, putting an adult in the body of a child leads to overestimation of object sizes and greater self-identification with child-like attributes compared with being embodied in an adult-shaped body of the same height as the child body (Banakou et al., 2013), a result that has been replicated (Tajadura-Jiménez, Banakou, Bianchi-Berthouze, & Slater, 2017).

What happens when White people are given a Black virtual body? Groom, Bailenson, and Nass (2009) did not study the body ownership aspects of this, but found that with such embodiment implicit racial bias increased. However, more recent studies that have started from the concept of body ownership have found that the body ownership illusion occurs independently of the skin color of the virtual bodies. Moreover, on the average, implicit racial bias decreases. The difference in these later studies is that in the study of Groom et al. (2009) the social situation depicted was one that would typically generate negative affect, whereas all the recent ones have been neutral or positive in this regard. In the context of the RHI, Farmer, Tajadura-Jiménez, and Tsakiris (2012) and Maister, Sebanz, Knoblich, and Tsakiris (2013) found that White participants with a Black rubber arm generally showed reduced implicit racial bias against Black people. In Peck, Seinfeld, Aglioti, and Slater (2013) embodiment of White people in a Black virtual body using virtual reality with visuomotor synchrony, resulted in a reduction in implicit racial bias compared to three control conditions. Both these sets of results were reviewed in Maister, Slater, Sanchez-Vives, and Tsakiris (2015), and a neural network model that well-simulates these results is presented in Bedder et al. (2019). Further work has shown that the reduction in implicit racial bias lasts for at least one week after the VR exposure (Banakou, Hanumanthu, & Slater, 2016). Banakou et al. (2020) have replicated these results when the surrounding situation is affectively neutral or positive and also confirmed the finding of Groom et al. (2009) that implicit racial bias of White against Black increases when the embodiment takes place in a situation that generates negative affect.

All of the above studies have used the racial IAT as a measure of implicit bias. A different approach was taken

in Hasler Lev-Tov, Spanlang, and Slater (2017) where White participants were embodied in a White or Black virtual body, and interacted with a White or Black virtual partner where they carried out a task together, in a counterbalanced design. The measure was how much participants mimicked the postures and gestures of their virtual partner. It was found that the level of mimicry was greater when the virtual partner had the same skin color as the skin color of their own *virtual* body. This showed that embodiment can lead to rapid reconfiguration of self-identification to the extent that unconscious activity such as mimicry could be influenced. Mimicry is important since it is a sign of rapport (Chartrand & Bargh, 1999). Our conclusion is that embodying White participants in a Black virtual body, with first-person perspective over the body (i.e., it visually substitutes their own body) and visuomotor synchrony is likely to lead to a (possibly sustained) reduction in implicit racial bias.

The experiment described in this paper was carried out in an inner-city police department in the United States. The question we considered was how the embodied perspective of a police officer as a Black victim of an intimidatory interrogation clearly based on racial bias by the interrogator, might influence their later behavior as a bystander to intimidatory and violent behavior of a police officer towards a Black person who simply entered a cafe to order coffee. Our major question was how much the participant would actually intervene to engage in helping behavior towards the victim of the police behavior in the cafe.

2 Materials and Methods

2.1 Ethics

The research project was reviewed by a senior researcher at Jigsaw and the research team at the police department and approved by the director of the police division. Participants who volunteered to take part were given an information sheet to read, and a consent form. The consent form was also read aloud to each participant. The participants gave written and informed consent. The experiment was carried out in accordance with relevant guidelines and regulations and conformed to the Helsinki Declaration. There was no financial or other compensation to the participants.

2.2 The Sample

Recruitment was carried out by the police force and was not under our control. Participants took part voluntarily. In a pre-questionnaire, participants had self-identified as either male or female and we will use those terms throughout. Our ideal would have been to recruit White male officers of approximately the same rank. However, this was not possible. The initial sample consisted of 37 patrol officers and 10 promoted officers. We were not allowed to record information about race, but the vast majority of those who took part were White and male. There were 8 females, all of whom were patrol officers. Patrol officers are those who directly interact most often with the public and are typically the first on the scene of an incident. They are responsible for patrolling particular areas on the lookout for situations where there are problems or crimes.

Further information about the sample is given about the sample in the section Covariates below.

2.3 Procedures

All participants first started in VR in a bedroom, embodied as a White male police avatar dressed in sports clothes, where they were asked to perform a set of simple exercises in front of a mirror to get them accustomed to their virtual body. Moving their body and seeing the movements directly by looking at their virtual body and also reflected in the mirror, was designed to provide the initial multisensory integration necessary to induce body ownership over the virtual body (Blanke, Slater, & Serino, 2015; Ehrsson, 2012). Here there was synchrony between their movement proprioception, observing the movement directly from first-person perspective on their virtual body, and also in the mirror, a technique that has been used in multiple studies as discussed above.

Following this, they were transported to a briefing room inside a police station, where a virtual senior

officer explained that there had been a robbery in a store, and there were two suspects seen running away in different directions. The photos of the suspects were on a display and one was White and the other Black. The participant and another (virtual) police officer were told that they were to interrogate the Black suspect. The experiment then had three phases, during which all the virtual officers were portrayed as White and male: Interrogation, Replay, and Cafe. In the Interrogation phase, the participant embodied as a male police officer was standing next to the other virtual officer, facing a Black suspect, who was seated at a table with his hands handcuffed behind. The other officer opened a line of abusive and intimidatory questioning, with the participant invited to ask questions of the suspect from time to time. (Since the suspect becomes a victim during the interrogation, we use the terms "suspect" and "victim" interchangeably in this paper.) The other officer had a mild Southern United States accent.

The other officer's behavior and questioning were based on detailed discussions with several police officers at various ranks, including a sergeant and a detective from two other inner-city police departments, who had 17 years and more than 10 years' experience, respectively. Several rounds of corrections were made in terms of the script and scene setting before arriving at the final version that was used.

During the course of the entire interrogation, the participant's movements and audio were recorded. After the interview session ended, participants then experienced one of two conditions in a replay of the scenario.

In the Victim condition they found themselves back in time, before the interrogation, in a cell, embodied as the victim, and were asked to carry out various exercises where they saw their virtual body reflected in a mirror in their cell. Then an officer appeared outside the cell, asked them to back away, and then proceeded to handcuff the participant in preparation to take them to the interrogation. They were not actually handcuffed, but a strap was loosely fastened between the hand controllers that the participants were holding. Next, they were in the interrogation room, and re-experienced the entire interrogation, but now from the embodied perspective of the victim. The participant's movements and audio that had been recorded earlier were played back on the virtual police officer that the participants had previously embodied (see Figure 1).

Alternatively, in the Observer condition, they were embodied in the body of a police officer, and they watched the victim in the cell on a virtual TV monitor, and then witnessed the interrogation through a window outside the interrogation room. Hence this experiment was between-groups with two conditions: Victim and Observer.

Three to four weeks later, the participants returned, and all had the same experience, that we refer to as the Cafe condition. After going through the same embodiment exercises in the bedroom as in the previous session, participants were in a cafe with the same abusive officer as the interrogation scene. The virtual officer engaged the participant in some small talk. After a while, a male Black person entered the cafe and stood by the counter next to a White woman. Immediately on seeing the person enter, the officer said "Damn it!" and started to aggressively question this newly entered Black customer, accusing him of wanting to steal the handbag of a woman standing next to the counter. The confrontation between the officer and the customer ramped up, with the customer continually protesting that he had done nothing wrong and simply wanted to get some coffee, and the officer becoming more and more aggressive. The victim occasionally looked towards the participant, and at one point asked pointedly to the participant: "Are you seeing this?" Finally, the officer pulled out his gun and approached towards the customer, and the scenario ended (see Figure 2). All scenarios (Interrogation, Cell, and Cafe) are shown in the Supplemental Video available at https://www.youtube.com/watch?v=mSSBX5ZteBI.

2.4 Virtual Reality Setup

The scenario was displayed on an HTC VIVE head-mounted display with the Deluxe Audio strap headphones. The program was executed on an Alienware 17 R4 laptop computer with i7 Intel CPU and a GTX1080 Nvidia Geforce graphics card. The VIVE head- and hand-tracking controllers were used for

Figure 1. *The interrogation and cell scenes (A) The briefing room where the commanding officer (left) explains the situation and shows pictures of the suspects to the participant (reflection seen in mirror) and another officer. (B) The other officer aggressively interrogates the Black suspect. (C) The participant is embodied as the suspect inside the cell in the Victim condition. (D) First-person perspective as the suspect during the interrogation. (E) The Observer view of the interrogation.*

upper-body tracking and 3 additional VIVE trackers for hip and foot position and orientation tracking.

The program was implemented using Unity3D (unity.com). The tracker information was streamed into Unity using the SteamVR Plugin from the Unity Asset store (assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647#content). FinalIK was used to map the tracker information to the avatar's body movements (assetstore.unity.com/packages/tools/animation/final-ik-14290).

The animations were constructed using the VIVE trackers for direct capture into Unity and some animations with the NaturalPoint Motive motion capture software (www.optitrack.com/software/motive/). The motion capture was cleaned and edited in Autodesk Motionbuilder (www.autodesk.com/products/motionbuilder/overview).

Lipsync was implemented by using the viseme information derived from the audio by the Oculus lipsync plugin for Unity (developer.oculus.com/downloads/package/oculus-lipsync-unity/) which was mapped to the Adobe Fuse avatars facial blendshapes (helpx.adobe.com/beta/fuse/faq.html#).

The police station and cafe interior were also purchased on the Unity Asset store and adapted for the study (assetstore.unity.com/packages/3d/environments/urban/office-and-police-station-pack-modular-74311).

The audio dialogues were recorded by staff at Jigsaw. The mapping of the tracking information was based on FinalIK, but there was some custom scripting to account for the heights of the participants and their orientation with respect to the tracking space. The audio replays of female participants were modified by reducing the pitch of their voice by 2 semitones when it was being played back because all the embodied characters were male.

2.5 Covariates

In addition to the independent variable Condition (Observer = 0, Victim = 1) there are two important covariates: sex and rank. Table 1A shows the breakdown by Condition and Sex, with the means and standard deviations of age for those who completed the

Figure 2. *The Cafe scene (A) The victim is standing by the bar after having just walked in. (B) The officer confronts the victim. (C) The victim complains to the officer that he has not done anything.*

Interrogation and Replay scenarios. Hence there were initially 47 participants, 24 as Victim and 23 as Observer.

Fewer officers (38) were able to complete the Cafe scenario than the original interview scenario. Hence the results for the Cafe scenario refer only to these 38. The breakdown in shown in Table 1B.

Participants in the Observer and Victim experimental groups were similar across various demographic variables, as shown in Supplemental Table S1.

Table 1. *Distribution of the Sample by Condition and Sex and Mean ± SD of Age in the Interrogation Scenario*

Condition	Male	Female	Total	Patrol
(A) Interrogation and Replay Scenario				
Observer	19	4	**23**	16
	34.7 ± 8.65	30.2 ± 5.00	34.0 ± 8.22	
Victim	20	4	**24**	21
	30.7 ± 8.19	33.5 ± 6.14	31.2 ± 7.84	
Total	39	8	**47**	37
	32.7 ± 8.54	31.9 ± 5.46	32.6 ± 8.06	
(B) Cafe Scenario				
Observer	14	3	**17**	12
	33.4 ± 8.87	31.7 ± 5.03	33.1 ± 8.22	
Victim	17	4	**21**	19
	30.2 ± 6.81	33.5 ± 6.14	30.9 ± 6.84	
Total	31	7	**38**	31
	31.7 ± 7.84	32.7 ± 5.31	31.9 ± 7.39	

Notes. The Patrol column shows the number of patrol officers. The number of promoted officers is Total − Patrol.

For the female officers, the only difference compared to the males is that when they heard their own voice during the Replay scenario, it was deepened compared to their real voice, as mentioned above.

The second covariate is the rank of the participants. The majority of officers were Patrol Officers, and the rest ranged in rank from Sergeant to Deputy Police Chief. The rank variable is 0 for the promoted ranks and 1 for the patrol officers. The distribution is shown in Table 1.

2.6 Response Variables

There were two main categories of response variable. The first were questionnaire-based assessing participant's level of presence, body ownership, and agency, which were assessed while the participant was in the virtual reality after each of the Interrogation, Replay, and Cafe scenarios. The second set of response variables were behavioral data based on ratings of participant actions during the Cafe scenario carried out by two independent coders. This is our major response variable.

2.6.1 Presence and Body Ownership. Presence is a fundamental illusion that typically occurs for participants in a virtual reality (Sanchez-Vives & Slater, 2005). It consists of two components, Place Illusion (PI) and Plausibility (Psi) (Slater, 2009). PI is a perceptual illusion of being in the place depicted by the VR. It depends on the extent to which sensorimotor contingencies for perception (O'Regan & Noë, 2001) are reproduced within the VR, through head-tracking and body movement. Psi is a cognitive illusion that the events perceived to be happening are really happening. When both PI and Psi are strong, then people tend to behave realistically in the VR. Additionally, copresence refers to the sense of being with other real people.

Presence was assessed on the two dimensions, PI and Psi, and co-presence. We use the convention that "1" refers to the initial Interrogation exposure, "2" refers to the re-exposure in the Interrogation scenario (either as the suspect or observer), and "3" refers to the cafe scenario. Table 2A shows the questions used. These were administered verbally, while the participants were

Table 2. *Presence and Body Ownership Questions*

Variable name	Question
(A) Presence	
there[1,2,3]	I had the sensation of [being in the interview room | the observation room | the cafe].
real[1,2,3]	There were moments when the [interview room | the observation room | the cafe] seemed more real to me than the room in which the events were really taking place.
happening[1,2,3]	I had the sensation that the events taking place in the virtual world were really happening even though I knew that they were only virtually happening.
copresence[1,2,3]	I felt that I was actually with the other people I saw.
(B) Body ownership	
mirror[1,2,3]	Although the body that I saw in the mirror did not look like me I had the sensation that it was my body.
down[1,2,3]	Although the body I saw when I looked down towards myself did not look like me, I had the sensation it was my body.
mymovements[1,2,3]	The body I had moved in sync with my movements.
twobodies[1,2,3]	I felt as though I had two completely separate bodies – my real body and the virtual one

Note. Each item was scored on a 1–7 Likert scale where 1 meant disagreement and 7 agreement with the corresponding statement.

wearing the head-mounted display, immediately after the corresponding scenario.

The method involving embodiment as one of the officers conducting the interrogation and then as the victim requires a high level of body ownership in the virtual body of the officer and also, and more importantly, in the body of the victim for those in the Victim condition. The questions are shown in Table 2B, based on several previous papers (for example, Banakou & Slater, 2014). Note that the "*twobodies*" question is a control. Participants should not have felt that they had two bodies, only the virtual one, hence we would require lower scores on the *twobodies* questions than the others.

2.6.2 Responses to the Cafe Scenario. After the Cafe scenario, participants answered a written questionnaire to assess their responses about their comfort, sense of safety, and whether they could and should have stopped the confrontation. The questionnaire is shown in Supplemental Table S2A and was previously used in Rovira, Swapp, Southern, Zhang, and Slater (2013) and Slater et al. (2013).

Two raters independently observed and scored video recordings of the graphics and verbal output of the Cafe scenario and were unaware of the experimental conditions. The Cafe scenario was broken down into a sequence of individual events (phrases or actions) initiated by the perpetrator or victim. The raters were required to rank the degree of helping behavior of the participant towards the victim for each such event on the scale shown in Table 3. There were 29 such events, the first being when the perpetrator said "Damn it!" after the entry of the victim into the cafe, and the last being when the perpetrator pulled out a gun and shouted "Freeze, on the ground now!"

There was one critical event when the victim turned towards the participant and said: "This is crazy; are you seeing this?" At this point nothing more happened for 5 seconds to give time for the participant to react. The

Table 3. *Coding for Each Event in the Cafe Sequence after the Victim Entered*

Action	Code
speaks and gestures in support of officer or against victim	−3
gestures in favor of officer or against victim	−2
speaks in favor of officer or against victim	−1
says something	0
speaks in favor of victim or against officer	1
gestures in favor of victim or against officer	2
speaks and gestures in support of victim or against officer	3
no response	leave blank

response to this event was coded in the same way as all the others, and the raters were not told that there was anything special about this event. It was designed in order to provide a specific moment where the participant might be likely to intervene. We refer to the scores on this as *seeingthis*. Cohen's Kappa measures the extent of agreement between the two raters. Here the agreement expected by chance is 39%, the actual agreement is 78%, and $\kappa = 0.64$, which is judged as in the "substantial" range.

2.7 Statistical Methods

The presence, body ownership, and agency questions (see Table 2) are presented descriptively. For the experiment to be meaningful, we require that the overall scores be high, and comparable with previous work in this area. We are not making statistical inferences about these variables, but we needed them to be high as a necessary condition for the experiment. Where we provide significance levels, these are only to illustrate the strength of a relationship or a difference, and are not meant inferentially to a wider population.

A Bayesian model was used that includes analysis of all response variables simultaneously. This avoids the need for ad hoc control of the significance level in multiple comparisons that are necessary in null hypothesis significance testing since all posterior probability statements are based on the joint posterior distribution of all the model parameters. Also, there is increasing disquiet about null hypothesis significance testing (McShane, Gal, Gelman, Robert, & Tackett, 2019), problems that the Bayesian method avoids, and there has been a very pronounced increase in the use of Bayesian methods in, for example, psychology (Van De Schoot, Winter, Ryan, Zondervan-Zwijnenburg, & Depaoli, 2017).

The main response variables are the *seeingthis* question evaluated by the 2 raters, and their coding over the behavioral responses of the participants in the Cafe scenario. Each coder rated each action as one of 4 possible outcomes: Negative (negative codes in Table 3), Neutral (code 0 in Table 3), Positive (positive codes in Table 3), and Null (the no response in Table 3 where the coders did not make any entry). For each participant ($i = 1, 2, \ldots, n = 38$), therefore, the results for each coder would be a vector $r_i = (r_{i1}, r_{i2}, r_{i3}, r_{i4})$, where r_{i1} is the number of Negative codes, r_{i2} the number of Neutral codes, r_{i3} the number of Positive codes, and r_{i4} the number of Null codes, for the ith participant. Moreover, $\sum_{j=1}^{4} r_{ij} = N = 29$ since each coder responded to 29 events. Hence, we consider the frequency vector r_i as an observation on a multinomial distribution, with $N = 29$ as the number of trials, $\pi_i = (\pi_{i1}, \pi_{i2}, \pi_{i3}, \pi_{i4})$, as the probabilities of classifying a response as Negative, Neutral, Positive, or Null, with $\sum_{k=1}^{4} \pi_{ik} = 1$.

The analysis shows how the probabilities π_i might be influenced by the experimental condition (Observer, Victim) and the covariates, sex and rank. We use the Multinomial Logit Model (MLM) (Greene, 2012) for analysis. We define linear predictors, that is, linear terms

in the independent variables, with the Null condition taken as the baseline. The principle of the MLM is that the log odds of the condition against the baseline condition are equal to linear predictors (η) involving the independent variables.

$$\log\left(\frac{\pi_{i1}}{\pi_{i4}}\right) = \eta_{i1} = \beta_{Neg,0} + \beta_{Neg,1}C_i + \beta_{Neg,2}R_i + \beta_{Neg,3}(C_i \times R_i) + \beta_{Neg,4}S_i$$
$$\log\left(\frac{\pi_{i2}}{\pi_{i4}}\right) = \eta_{i2} = \beta_{Neu,0} + \beta_{Neu,1}C_i + \beta_{Neu,2}R_i + \beta_{Neu,3}(C_i \times R_i) + \beta_{Neu,4}S_i$$
$$\log\left(\frac{\pi_{i3}}{\pi_{i4}}\right) = \eta_{i3} = \beta_{Pos,0} + \beta_{Pos,1}C_i + \beta_{Pos,2}R_i + \beta_{Pos,3}(C_i \times R_i) + \beta_{Pos,4}S_i, \quad (1)$$

for $i = 1, 2, \ldots, n$, where

C_i is Condition, where $C_i = 0$ (Observer condition) and $C_i = 1$ (Victim condition).
R_i is rank, where $R_i = 0$ (Promoted) and $R_i = 1$ (Patrol).
S_i is sex, where $S_i = 0$ (female) and $S_i = 1$ (male).
$C_i \times R_i$ is the interaction term between Condition and Rank.

Taking exponentials in (Eqn 1) and since $\sum_{k=1}^{4} \pi_{ik} = 1$,

$$\pi_{ik} = \frac{\exp(\eta_{ik})}{1 + \sum_{j=1}^{3} \exp(\eta_{ij})}, k = 1, 2, 3, 4,$$

with the baseline (Null) parameters all set to 0 so that $e^{\eta_{i4}} = 1$.

The prior distributions for the $\beta_{k,j}$ parameters are set to be weakly informative (Gelman, Jakulin, Pittau, & Su, 2008; Lemoine, 2019), $\beta_{k,j} \sim Normal(0, 10)$ (normal distribution, mean 0, and standard deviation 10), $k =$ *Neg*, *Neu*, *Pos*, $j = 0, 1, \ldots, 4$. The prior 95% credible intervals are therefore −20 to 20. The same model is used separately for both coders.

The response variables *seeingthis1* and *seeingthis2* are ordinal variables (for the two different coders), and can be modelled by a standard Bayesian ordered logistic model. The linear predictor for the model is similar to that above with different coefficients, but the same for both variables:

$$\gamma_0 + \gamma_1 C_i + \gamma_2 R_i + \gamma_3 (C_i \times R_i) + \gamma_4 S_i. \quad (2)$$

The cut points for the logistic model are $c_0, c_1, \ldots, c_6$, and the prior distributions are

$$\gamma_j \sim Normal(0, 10),\ j = 0, 1, \ldots, 4$$
$$c_j \sim Normal(0, 4),\ j = 1, \ldots, 6.$$

The cut points are parameters so that if X is an ordinal variable with possible values $1, 2, \ldots, 7$ based on an underlying latent continuous variable Υ then, $X = i$ when $c_i \leq \Upsilon < c_{i+1}$, $i = 0, 1, \ldots, 6$, where $c_0 = -\infty$ and $c_7 = \infty$. In this case X represents the *seeingthis* variables, and the scale is −3 to 3 instead of 1 to 7, although in the analysis it is convenient to convert the scores to be 1 to 7.

The Stan system[1] was used for the analysis (Carpenter et al., 2016; Stan Development Team, 2011–2019) with the RStudio[2] interface. The simulation was run with 4000 iterations.

The simulation converged with all Rhat = 1, and trace results showing no sign of autocorrelation. The Stata 16 system[3] was used to produce some graphs.

3 Results

3.1 Presence and Body Ownership

Figures 3A–3C show the box plots for the results of the presence questions. It is clear that presence was high in all scenarios and under both conditions. All medians are above "4" (the middle) score, and in most cases the entire interquartile range (IQR) is above the middle score. It is noteworthy that the copresence scores are particularly high, with all the IQRs above the midpoint.

Figures 3D–3F show the results for the body ownership and agency questions. The scores for *mirror*,

[1] https://mc-stan.org
[2] https://rstudio.com
[3] https://www.stata.com

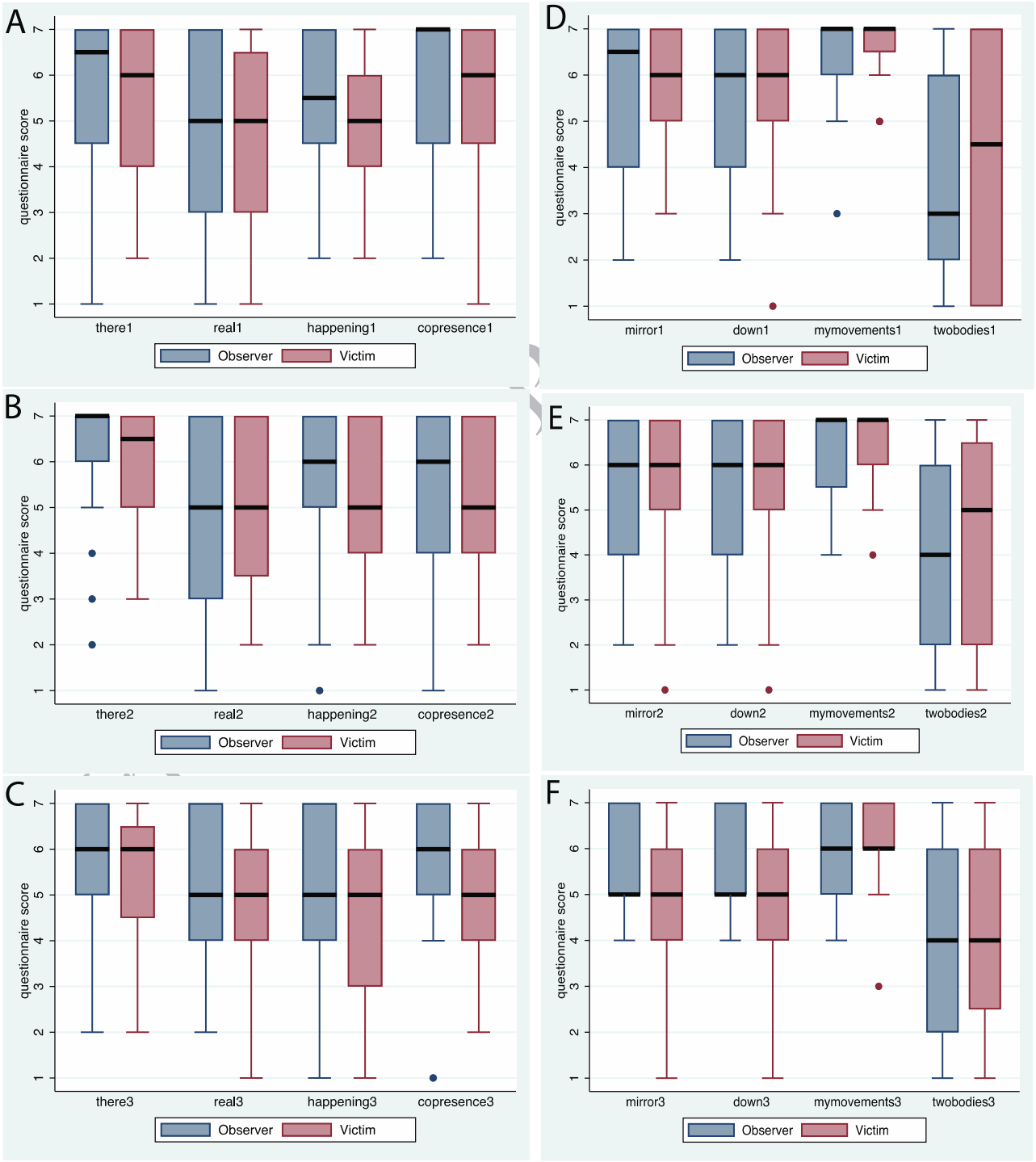

Figure 3. *Box plots of presence, body ownership, and agency by Condition (A) Presence in the Interrogation, (B) Presence in the Replay, and (C) Presence in the Cafe. (D) Body ownership in the Interrogation, (E) Body ownership in the Replay, and (F) Body ownership in the Cafe. The thick horizontal lines are the medians and the boxes are the interquartile ranges. The whiskers extent from max(lowest value, lower quartile − 1.5∗IQR) to min(highest value, upper quartile + 1.5∗IQR). Points outside this range are shown individually.*

Table 4. *Totals of the r = (Negative, Positive, Neutral) Vectors by Condition and Rank, and for Each Rater*

	Rater 1				Rater 2			
	Negative	Neutral	Positive	Null	Negative	Neutral	Positive	Null2
Promoted								
Observer (n = 5)	0	5	68	101	1	6	60	107
Victim (n = 2)	0	4	12	42	0	5	11	42
Patrol								
Observer (n = 12)	2	11	47	259	3	8	46	262
Victim (n = 19) (n = 38)	1	22	122	406	0	15	117	419

Note. For each rater the sum of each row divided by n is 29.

down, and *mymovements* are all high, with all medians and IQRs greater than the mid-point. As expected, the scores for the control question *twobodies* have much lower medians and also greater variance. For example, comparing mirror1 with twobodies1 using a Wilcoxon matched-pairs signed-rank test $z = 4.075$, $P < 0.00005$. The corresponding values for mirror2 and twobodies2 are $z = 2.898$, $P = 0.0033$, and for mirror3 and twobodies3 $z = 2.034$, $P = 0.041$. (These are not meant as inferential tests, but just to show the strength of the difference.) Overall non-control question scores are high and comparable with previous research.

The *mymovements* questions are a test of the system, which was designed to map participant movements onto the movements of their virtual body. Even though the tracking was with the head, two hands, two feet, and a body-mounted tracker, with inverse kinematics to infer movements, the results appear to be successful. The *mymovements* scores have the greatest medians and smallest variance.

Most importantly the ownership scores are high for the Victim condition in the Replay scenario, meaning that participants had strong feelings of body ownership when embodied as the suspect. There appears to be no difference in the level of body ownership between males and females even though the virtual body was male (Supplemental Figure S1), in line with previous findings (Slater et al., 2010).

3.2 Subjective Responses to the Cafe Scenario

Supplemental Table S2B gives the results for the post-Cafe scenario questionnaires. Both Observer and Victim groups reported high levels of discomfort, concern for the victim, worry that others might turn up to make the situation worse, awareness of the victim looking at them for help, and felt that they should do something to stop the argument. On the other hand, they did not feel the need to move away, or the need to get out, and their mind did not wander thinking of other things.

3.3 Behavioral Responses to the Cafe Scenario

Table 4 shows the assignments by the raters of numbers of responses in the Negative, Neutral, Positive, and Null categories. The results are very similar between the two raters. Apart from Null, amongst the Patrol group the greatest selection is Positive in the Victim

Table 5. *Summary of the Posterior Distributions of the Parameters of the Model for Coder 1*

Parameter	Coefficient of:	Mean	SD	2.5%	97.5%	Prob > 0
Negative						
$\beta_{Neg,0}$		−13.1	5.02	−24.2	−5.0	0.000
$\beta_{Neg,1}$	Victim	−3.3	5.66	−15.7	6.7	0.289
$\beta_{Neg,2}$	Patrol	5.0	3.90	−0.7	14.2	0.943
$\beta_{Neg,3}$	Victim × Patrol	1.8	5.71	−8.5	13.9	0.600
$\beta_{Neg,4}$	Sex (Male)	4.2	3.75	−0.9	13.3	0.917
Neutral						
$\beta_{Neu,0}$		−1.5	0.62	−2.8	−0.4	0.004
$\beta_{Neu,1}$	Victim	0.6	0.73	−0.8	2.0	0.814
$\beta_{Neu,2}$	Patrol	−0.3	0.58	−1.4	0.9	0.303
$\beta_{Neu,3}$	Victim × Patrol	−0.3	0.82	−1.9	1.3	0.344
$\beta_{Neu,4}$	Sex (Male)	−0.6	0.40	−1.3	0.2	0.086
Positive						
$\beta_{Pos,0}$		1.5	0.25	1.0	1.9	1.000
$\beta_{Pos,1}$	Victim	−0.9	0.37	−1.6	−0.2	0.005
$\beta_{Pos,2}$	Patrol	−1.6	0.23	−2.1	−1.2	0.000
$\beta_{Pos,3}$	Victim × Patrol	1.5	0.41	0.7	2.3	1.000
$\beta_{Pos,4}$	Sex (Male)	−0.8	0.19	−1.2	−0.5	0.000
seeingthis						
γ_0		4.7	2.29	0.3	9.3	0.981
γ_1	Victim	−1.1	1.87	−4.7	2.6	0.271
γ_2	Patrol	−2.5	1.36	−5.5	−0.1	0.022
γ_3	Victim × Patrol	2.8	2.02	−1.2	6.7	0.917
γ_4	Sex (Male)	0.2	0.86	−1.5	1.9	0.593

Notes. The columns 2.5% and 97.5% show the 95% equal tails credible intervals. Prob > 0 is the posterior probability that the parameter is positive.

condition (taking into account the differing numbers of participants in the Observer and Victim conditions).

Table 5 summarizes the posterior distributions of the parameters of the model (Eqns 1–2) for coder 1. The results for the other coder are very similar and the equivalent table is given in Supplemental Table S3.

By way of explanation, the probabilities relate to the corresponding parameters. For example, $P(\beta_{Pos,3}|Data) = 1.000$ has to be interpreted according to the model (Eqn 1). Since $\beta_{Pos,3}$ is the coefficient of $(C_i \times R_i)$ it means that when $R_i = 1$ (Patrol Officers) and they experienced the Victim condition ($C_i = 1$) then it is certain (up to rounding errors) that the log-odds of being rated Positive compared to a Null rating are greater than 0 (and hence the ordinary – non-log - odds are > 1). Since the mean of the distribution of $\beta_{Pos,3}$ is 1.5, this implies that (other things being equal) the log-odds of being rated Positive compared to a Null rating has a mean of 1.5 for Patrol Officers. This translates to odds of $e^{1.5} = 4.5$, to 1. We now consider the various results.

For the Negative evaluations (where the officer was coded as having carried out an action that was against the victim) Patrol officers were likely to have their actions evaluated as Negative (probability = 0.943), other things being equal. The same is true of male officers

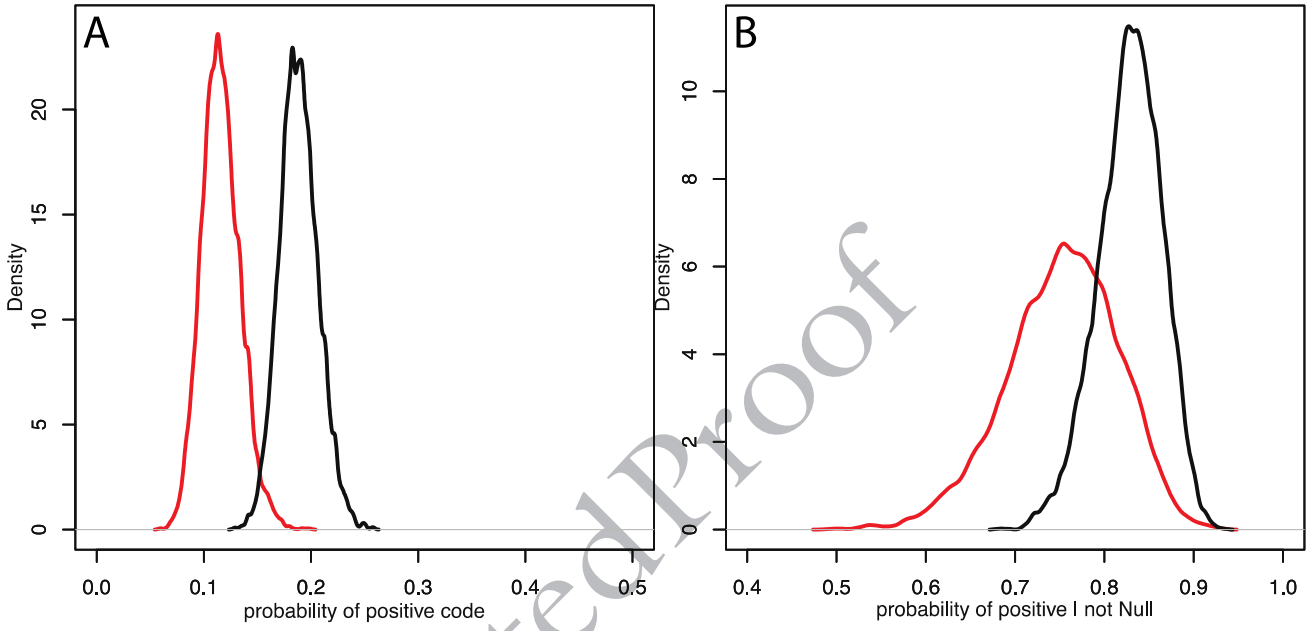

Figure 4. *Predicted posterior distribution of p_2 the probability of classifying an event as Positive in the case of male Patrol officers. The blue curve is for the Observer condition and the black curve for the Victim condition.*

(probability = 0.917). These results are relative to Null evaluations. Male officers are less likely to have Neutral evaluations (probability = 1 − 0.086 = 0.914). Patrol officers in the Victim condition have probability 1 (up to rounding error) of making a positive evaluation. For the *seeingthis* variable amongst the Patrol officers in the Victim condition there is a probability of 0.917 of there being a higher rating.

From the model (Eqns 1 and 2) using the posterior distributions of the parameters, 8000 pseudorandom observations were simulated. From these we can find the predicted posterior distributions of the π. Figure 4 shows the probability distributions of the π_3 parameter in the case of male patrol officers. Figure 4A shows the absolute distributions demonstrating that the probability of a Positive rating is overall greater for the Victim than the Observer condition. For example, the posterior probability $P(\pi_3 > 0.15 \mid \text{Observer, Male, Patrol}) = 0.036$, whereas $P(\pi_3 > 0.15 \mid \text{Victim, Male, Patrol}) = 0.989$. Figure 4B shows the distributions but restricted to the non-Null choices. This shows the greater probability of the parameters π_3 conditional on not-Null. For example, for the Observer condition the probability that the conditional $\pi_3 > 0.80$ is 0.233, whereas for the Victim condition the corresponding probability is 0.794. Considering $\pi_3 > 0.85$ these values are 0.046 and 0.284, respectively.

3.4 Model Checking

The predicted posterior distributions can be found for (r_1, r_2, r_3, r_4) and *seeingthis*. In each case there is a posterior probability distribution for each individual $i = 1, 2, \ldots, 38$. We can obtain point estimates for each individual by taking the means of the distributions, and then compare those with the observed values.

Table 6 shows the correlations between the observed and point estimates for the predicted values. The classical 95% confidence intervals for the correlations are given for interpretation of the sizes of these correlations (these are not meant as significant tests). Apart from r_1 and r_2 for Rater 1 with moderate effect sizes, the remaining correlations are all in the high medium range.

Table 6. *Pearson Correlations between the Means Per Individual of the Predicted Posterior Distributions and the Observed Values for Raters 1 and 2*

	Rater 1		Rater 2	
Variable	Pearson correlation	95% confidence interval	Pearson correlation	95% confidence interval
r_1 (Negative)	0.25	−0.07 to 0.53	0.40	0.08 to 0.64
r_2 (Neutral)	0.26	−0.07 to 0.53	0.36	0.05 to 0.61
r_3 (Positive)	0.51	0.22 to 0.71	0.48	0.19 to 0.69
r_4 (Null)	0.48	0.19 to 0.69	0.48	0.19 to 0.69
seeingthis	0.39	0.08 to 0.63	0.43	0.13 to 0.66

Notes. The 95% classical confidence intervals for the correlation are given. All n = 38.

4 Discussion

The principle finding of the study is that those who had been in the Victim condition showed greater support for the victim in the Cafe scenario than those who had been in the Observer condition, according to the scores of two independent coders. This holds amongst the Patrol Officers. The posterior probability of this is high (the coefficient of the Victim × Patrol term for the Positive response in Table 5). The same holds for the *seeingthis* response, where the victim had looked towards the participant exclaiming "Are you seeing this?" and nothing further happened for 5 s in order to clearly allow for a response. There is a high posterior probability of the Victim × Patrol term being positive (Table 5). Recalling that the prior 95% credible intervals were −20 to 20 and considering the very narrow posterior credible intervals, these are strong results, as shown in Figure 4.

The essence of the method is that participants experience two different embodied perspectives of the same event. However, the second time that the event is experienced also portrays any actions that participants carried out during the first time, so they would be at the receiving end of their own prior behavior. This corresponds to an embodied experiential version of the "Golden Rule." In the so-called negative form this is: "What is hateful to you, do not do to your fellow" or in the positive form "Treat others as you would want them to treat you." Singer (1963) argues that the positive and negative forms are equivalent and some boundaries and scope of this rule are discussed in detail in Reinikainen (2005) and Weiss (1941). In this design, the Golden Rule is not a rule or a moral imperative but something that is experienced personally by the participants.

Another example of this method was reported in Neyret et al. (2020) in the context of sexual harassment of a virtual woman. In this case, male participants were first embodied in a male body amongst a group of virtual men who harassed a lone woman in a bar setting. In the second exposure, the participants either re-experienced the scenario from the embodied perspective of the woman (Woman condition) or as another of the men (Group condition). A control group experienced only an empty bar setting. One week later, all participants took part in a virtual obedience scenario where they were required by a group of virtual male experimenters to administer shocks to a virtual female, based on one of the Obedience paradigms of Stanley Milgram (1974). It was found that those who had Group condition replay gave the greatest number of shocks, those in Woman condition the least number of shocks, and the control group a number of shocks between these two.

The closest to this in the psychological literature is self-distancing theory (Kross & Ayduk, 2017) where people are, for example, trained to recall a traumatic or anxiety producing event from a third person perspective as a "fly on the wall" rather than from an immersed first-person perspective. In many studies it has been shown

that compared to a first-person immersive point of self-distancing results in participants focusing ". . . less on recounting the emotionally arousing features of their past experience and focus more on reconstruing it in ways that provide them with a sense of insight and closure" (Kross & Ayduk, 2017, p. 85). In relation to a past event, participants are given instructions such as "Now take a few steps back. Move away from the situation to a point where you can now watch the event unfold from a distance and see yourself in the event." This leads to a reduction of negative emotions surrounding the event in question compared to an immersed condition. This method has also been successful in relation to aggression. Mischkowski, Kross, and Bushman (2012) found that subsequent to being provoked as part of the experimental procedure, participants who reflected on this from a self-distancing perspective were less likely to respond aggressively compared to those who considered the provocation from an immersed perspective.

VR offers a way through which self-distancing can be put into practice in a highly tangible manner: people experience an event, and then do not have to recall it from a psychologically self-distanced perspective but actually live through it again from the embodied perspective of another. This relies on decades of research into presence in VR where it has been repeatedly shown that people tend to respond to virtual situations and events as if they were real. This is also why the use of VR in psychological therapy has been one of its major research areas (Carl et al., 2019; Freeman et al., 2017; Rizzo & Kim, 2005), since if people did not respond realistically in VR, the therapy could not be successful.

The fact that such an experiment could take place in a busy inner-city police department is itself noteworthy. The police themselves considered the topic to be important enough to warrant the considerable time that they made available, and other arrangements that they had to make. The limitations of this study, however, are inherent in the circumstances in which it could be carried out. It was not easy to find a cooperative police department, and when we were able to, we had no control over the sample. Hence our sample was heterogeneous, men and women, patrol officers and promoted officers, and although most officers were White, only very few were Black. However, although the sample size was low, we started from non-informative prior distributions with large variances, and the resulting posterior distributions are narrowly focused. For example, the critical prior 95% credible intervals of −20 to 20 narrowed to 0 to 3.5 for the interaction term for Positive codings, and to −1.2 to 6.8 for the *seeingthis* interaction term, and all the others are similarly narrow. Hence, we can have confidence in these findings. Moreover, the results were very similar for the two independent coders.

It might be argued that the results can be explained by demand characteristics with participants behaving how they thought they ought to behave. However, this was a between-groups experiment so that each participant experienced only one experimental condition. If the results were due to demand characteristics, we would expect to find no difference between the results of the two conditions. In any case, our experience is that in VR people react automatically to the events and situations in which they find themselves, and if they think about their responses afterwards, it is already too late. So even if they had the intention to behave according to accepted norms, they would not have had time to prepare and act out such deliberate responses. The Cafe scenario was a fast moving and rapidly escalating one, it is unlikely that participants could deliberately control their responses. An example of this was provided by Peck, Good, and Seitz (2021) who found that in a shooter bias scenario although there was no difference between latency to shoot at Black or White targets, deeper analysis of participant movement revealed differences that would be highly unlikely to be under conscious control.

The Cafe scenario also relates to the bystander problem, that is, the extent to which people intervene to help victims of aggression or accidents and other negative events. Bystander behavior has been extensively studied starting with the work of Darley and Latané (1968); see also Latane and Darley (1968), who put forward the hypothesis that the greater the number of bystanders to an event, the less the likelihood that any would intervene—since verified in many other studies. For a comprehensive review see Fischer et al. (2011). A complementary approach is based on social identity theory, where the actions of the bystanders are dependent on their degree

of social identity with the victim exemplified in the work of Levine (1999) and Levine, Prosser, Evans, and Reicher (2005) with results reproduced in a VR study in the context of a violent incident (Slater et al., 2013). An alternative theory based on studies of brain activation during bystander incidents has been put forward by Hortensius and de Gelder (2018) where bystanders suffer a conflict between apathy and intervention which is ultimately resolved through distress influenced by the presence of other bystanders mediated by the personality of the individual. The Cafe scenario does not quite fit into any of these paradigms since typically bystanders are strangers, and have not had previous interactions with the perpetrator, and in our case the bystander was a police officer and the perpetrator portrayed as one, thus raising issues of loyalty. Our approach is related to those that use an imaginal perspective taking strategy to enhance empathy and helping behavior towards others, which has a long history in social psychology (for example, Chandler, 1973; Oswald, 1996). However, our approach has an important different component; participants are first involved in an aggressive event, and then re-experience it from the embodied perspective of the victim. This double exposure, experiencing the situation from two different viewpoints and its experiential rather than imaginal basis, is what marks this out as a different approach.

We cannot be sure that our findings are due to changes in implicit or explicit bias since we did not measure these. It is possible that the difference in behavior between the Observer and Victim conditions might be for other reasons; for example, dismay at the effects of aggressive policing, or even the belief that this particular aggressive officer was a "bad apple," an atypical and badly behaved one whose actions should be challenged. Our data offers no further information on mechanism other than the "Golden Rule" and self-distancing discussions mentioned above. It will be important in further research using this method to also include specific measures of bias, and interviews with participants to elicit their own understanding of their actions.

The fact that we were not allowed to record the race of the officers involved is a drawback, although the vast majority of officers were White. We cannot know from our sample whether results were different for Black officers; even if we knew which particular individuals were Black, the sample would be too small for inference. LeCount (2017) carried out a large-scale survey across the US assessing racial bias amongst police and other citizens. They found, for example, that White police were nine times more likely than non-police to agree with the idea that Black people are more violent than White people. However, such differences between White police and the general population were not manifest amongst Black police. This study focused on explicit attitudes which may not hold for implicit bias. Avenanti, Sirigu, and Aglioti (2010) carried out a pain observation study where both White and Black participants showed enhanced implicit bias against the out-group (Black) compared to a neutral group although there were no differences in explicit bias. This finding was replicated by Azevedo et al. (2013), although the evidence pointed to the possibility that the effect of out-group bias even amongst members of the out-group might be stronger the more salient the out-group population in the surrounding society. Our experiment was carried out in a multiracial urban environment, so the saliency of African Americans is not in doubt.

Returning to the issues raised in our opening paragraphs, the fact that racial bias is implicit does not make it any less serious in its consequences, especially in a demanding and dangerous profession such as policing. Dovidio and Gaertner (2004) noted with respect to US society that "A critical aspect of the aversive racism framework is the conflict between Whites' denial of personal prejudice and the underlying unconscious negative feelings toward and belief about Blacks." They argue that in spite of the largely liberal and tolerant explicit attitudes of Whites towards Black people and their support for racial equality, everyday behavior does not bear this out. A recent case highlighted this in an extreme way when a White woman called the police on a Black man after he asked her to put her dog on a leash in Central Park in New York. According to the reports,[4] she told the police specifically that the man was "African American," thus potentially putting him in danger. Implicit

[4] https://www.bbc.com/news/world-us-canada-52759502

racial bias is perhaps even more difficult to combat than explicit bias precisely because it is non-conscious, rooted in automatic ways of thinking and behaving. Welsh, Chanin, and Henry (2020) carried out structured interviews with 52 officers in the SDPD about their attitudes and behaviors during traffic stops. They found that the officers followed an explicit policy of "colorblind policing," where race was in no way part of the behavior of police during routine traffic stops. However, their actual behavior belied this claim, not in a deliberate attempt to lie, but in genuine almost non-conscious misunderstanding and obfuscation of their own procedures and motivations. For example, on a traffic stop police would have "hunches" that the people in the car were engaged in criminal activity based on their clothes, way of speaking, tattoos, and so on—of course not race—but not mentioning the supposed correlations between these types of signal and the race of the supposed suspects.

When attitudes and behaviors are so implicitly situationally rooted, they cannot be combatted simply by explicit education. Here we put forward the Golden Rule method as one that combats implicit racial bias through an implicit means—simply giving officers the chance to experience their own actions and those of their colleagues from the point of view of those on whom it has an effect. Hence to diminish racially biased behavior and attitudes for police carrying out traffic stops we would advise that in VR police carry out traffic stops and then become the drivers and re-experience their own behavior over again from the point of view of the driver and passengers. Here there is no lecturing about what is right or wrong, but simply experience.

5 Data Availability

Data are available at https://doi.org/10.5281/zenodo.4603687.

Acknowledgements

The project was funded by Jigsaw. We thank Key Lime Interactive employees Shivashankar Halan and Mariano Rodriguez for help in preparation of the experiment. We thank Maddy Hoffman for help in the preparation of the paper. MS is supported by the ERC Advanced Grant MoTIVE # 742989.

Competing Interests Statement

The implementation of this work was carried out by Virtual Bodyworks S.L. under contract to Jigsaw. MS and BS are Founders of Virtual Bodyworks. No other author has competing interests. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Contributions

Concept and design of the experiment, statistical analysis and first draft of the paper: MS.

Preparation of the data: SK, GI, BS, DS.

Carrying out of the experiment: SK, MS.

Design and implementation of the virtual reality scenarios: SK, BS, GI.

Obtained the funding: MS.

Contributed to and reviewed the paper: DS, BS, SK.

REFERENCES

Armel, K. C., & Ramachandran, V. S. (2003). Projecting sensations to external objects: Evidence from skin conductance response. *Proceedings of the Royal Society B*, *270*, 1499–1506. https://doi.org/10.1098/rspb.2003.2364, PubMed: 12965016

Avenanti, A., Sirigu, A., & Aglioti, S. M. (2010). Racial bias reduces empathic sensorimotor resonance with other-race pain. *Current Biology*, *20*(11), 1018–1022. https://doi.org/10.1016/j.cub.2010.03.071, PubMed: 20537539

Azevedo, R. T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., & Aglioti, S. M. (2013). Their pain is not our pain: Brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Human Brain Mapping*, *34*(12), 3168–3181. https://doi.org/10.1002/hbm.22133, PubMed: 22807311

Banakou, D., Beacco, A., Neyret, S., Blasco-Olivera, M., Seinfeld, S., & Slater, M. (2020). Virtual body ownership and its consequences for implicit racial bias are dependent

on social context. *Royal Society Open Science*, *7*, 201848. https://doi.org/10.1098/rsos.201848

Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *PNAS*, *110*, 12846–12851. https://doi.org/10.1073/pnas.1306779110, PubMed: 23858436

Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in theirimplicit racial bias. *Frontiers in Human Neuroscience*, *10*, 601. https://doi.org/10.3389/fnhum.2016.00601, PubMed: 27965555

Banakou, D., & Slater, M. (2014). Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *PNAS*, *111*(49), 17678–17683. https://doi.org/10.1073/pnas.1414936111

Banakou, D., & Slater, M. (2017). Embodiment in a virtual body that speaks produces agency over the speaking but does not necessarily influence subsequent real speaking. *Scientific Reports*, *7*(1), 14227. https://doi.org/10.1038/s41598-017-14620-5

Baumgart, Z. (2016). *Crime, arrests, legitimacy or race? Militarization of American police from 1990 to 2007*. Madison, WI: The University of Wisconsin-Madison.

Bedder, R. L., Bush, D., Banakou, D., Peck, T., Slater, M., & Burgess, N. (2019). A mechanistic account of bodily resonance and implicit bias. *Cognition, 184*, 1–10. https://doi.org/10.1016/j.cognition.2018.11.010, PubMed: 30553934

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron*, *88*(1), 145–166. https://doi.org/10.1016/j.neuron.2015.09.029, PubMed: 26447578

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756–756. https://doi.org/10.1038/35784, PubMed: 9486643

Bourne, J. (2001). The life and times of institutional racism. *Race & Class*, *43*(2), 7–22. https://doi.org/10.1177/0306396801432002

Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B., Emmelkamp, P., et al. (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of Anxiety Disorders, 61*, 27–36. https://doi.org/10.1016/j.janxdis.2018.08.003, PubMed: 30287083

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., et al. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, *20*.

Chandler, M. J. (1973). Egocentrism and antisocial behavior: The assessment and training of social perspective-taking skills. *Developmental Psychology*, *9*(3), 326. https://doi.org/10.1037/h0034974

Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910. https://doi.org/10.1037//0022-3514.76.6.893, PubMed: 10402679

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies—Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*, 377–383. https://doi.org/10.1037/h0025589, PubMed: 5645600

DeVylder, J. E., Cogburn, C., Oh, H. Y., Anglin, D., Smith, M. E., Sharpe, T., et al. (2017). Psychotic experiences in the context of police victimization: Data from the survey of police–public encounters. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/sbx038, PubMed: 28369639

DeVylder, J. E., Oh, H., Nam, B., Sharpe, T., Lehmann, M., & Link, B. (2016). Prevalence, demographic variation and psychological correlates of exposure to police victimisation in four US cities. *Epidemiology and Psychiatric Sciences*, 1–12. https://doi.org/10.1017/S2045796016000810, PubMed: 27834166

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in Experimental Social Psychology*, *36*, 1–52. https://doi.org/10.1007/978-0-387-09467-0_19

Ehrsson, H. H. (2012). The concept of body ownership and its relation to multisensory integration. In B. E. Stein (Ed.), *The new handbook of multisensory processes* (pp. 775–792). Cambridge, MA: MIT Press.

Farmer, H., Tajadura-Jiménez, A., & Tsakiris, M. (2012). Beyond the colour of my skin: How skin colour affects the sense of body-ownership. *Consciousness and Cognition*, *21*(3), 1242–1256. https://doi.org/10.1016/j.concog.2012.04.011, PubMed: 22658684

Fellner, J. (2009). Race, drugs, and law enforcement in the United States. *Stanford Law and Policy Review*, *20*, 257.

Fischer P, Krueger JI, Greitemeyer T, Vogrincic C, Kastenmüller A, & Frey D. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, *137*, 517 *Stan. L. & Pol'y Rev.* 537. https://doi.org/10.1037/a0023304, PubMed: 21534650

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, *1 Stan. L. & Pol'y Rev.* 8. https://doi.org/10.1017/S003329171700040X, PubMed: 28325167

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360 *Stan. L. & Pol'y Rev.* 1383. https://doi.org/10.1214/08-AOAS191

Greene, W. (2012). *Econometric analysis*. New York: Prentice Hall.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464. https://doi.org/10.1037//0022-3514.74.6.1464, PubMed: 9654756

Groom, V., Bailenson, J. N., & Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence*, *4*, 231–248. https://doi.org/10.1080/15534510802643750

Hasler Lev-Tov, B., Spanlang, B., & Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PLOS One*, *12*(4), e0174965. https://doi.org/10.1371/journal.pone.0174965

Hortensius, R., & de Gelder, B. (2018). From empathy to apathy: The bystander effect revisited. *Current Directions in Psychological Science*, *27*(4), 249–256. https://doi.org/10.1177/0963721417749653, PubMed: 30166777

Kalckert, A., & Ehrsson, H. H. (2012). Moving a rubber hand that feels like your own: A dissociation of ownership and agency. *Frontiers in Human Neuroscience*, *6*, 40. https://doi.org/10.3389/fnhum.2012.00040, PubMed: 22435056

Kokkinara, E., & Slater, M. (2014). Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. *Perception*, *43*(1), 43–58. https://doi.org/10.1068/p7545, PubMed: 24689131

Kross, E., & Ayduk, O. (2017). Self-distancing: Theory, research, and current directions. *Advances in Experimental Social Psychology*, *55*, 81–136).

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

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., et al. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 100–1016. https://doi.org/10.1037/xge0000179, PubMed: 27454041

Latane, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, *10*(3), 215–221. http://dx.doi.org/10.1037/h0026570, PubMed: 5704479

LeCount, R. J. (2017). *More black than blue? Comparing the racial attitudes of police to citizens*. Paper presented at the Sociological Forum. https://doi.org/10.1111/socf.12367

Legewie, J., & Fagan, J. (2016). Group threat, police officer diversity and the deadly use of police force. *Columbia Public Law Research Paper No. 14-512*. Retrieved from https://ssrn.com/abstract=2778692

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928. https://doi.org/10.1111/oik.05985

Levine, M. (1999). Rethinking bystander nonintervention: Social categorization and the evidence of witnesses at the James Bulger murder trial. *Human Relations*, *52*, 1133–1155.

Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, *31*, 443–453. https://doi.org/10.1177/0146167204271651, PubMed: 15743980

Maister, L., Sebanz, N., Knoblich, G., & Tsakiris, M. (2013). Experiencing ownership over a dark-skinned body reduces implicit racial bias. *Cognition*, *128*, 170–178. https://doi.org/10.1016/j.cognition.2013.04.002, PubMed: 23680793

Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2015). Changing bodies changes minds: Owning another body affects social cognition. *Trends in Cognitive Sciences*, *19*(1), 6–12. https://doi.org/10.1016/j.tics.2014.11.001, PubMed: 25524273

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, *73*(sup1), 235–245.

Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.

Mischkowski, D., Kross, E., & Bushman, B. J. (2012). Flies on the wall are less aggressive: Self-distancing "in the heat

of the moment" reduces aggressive thoughts, angry feelings and aggressive behavior. *Journal of Experimental Social Psychology*, *48*(5), 1187–1191.

Neyret, S., Oliva, R., Beacco, A., Navarro, X., Valenzuela, J., & Slater, M. (2020). An embodied perspective as a victim of sexual harassment in virtual reality reduces action conformity in a later Milgram obedience scenario. *Scientific Reports*, *10*, 6207. https://doi.org/10.1038/s41598-020-62932-w

O'Regan, J. K., & Noë, A. (2001). What it is like to see: A sensorimotor theory of perceptual experience. *Synthese, 129*, 79–103. https://doi.org/10.1023/A:101269922

Oswald, P. A. (1996). The effects of cognitive and affective perspective taking on empathic concern and altruistic helping. *The Journal of Social Psychology*, *136*(5), 613–623. https://doi.org/10.1080/00224545.1996.9714045, PubMed: 8942318

Peck, T. C., Good, J. J., & Seitz, K. (2021). Evidence of racial bias using immersive virtual reality: Analysis of head and hand motions during shooting decisions. *IEEE Transactions on Visualization and Computer Graphics*, *27*(5), 2502–2512. https://doi.org/10.1109/TVCG.2021.3067767, PubMed: 33750702

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition, 22*, 779–787. https://doi.org/10.1016/j.concog.2013.04.016, PubMed: 23727712

Peeples, L. (2019). What the data say about police shootings. *Nature, 573*, 24–26. https://doi.org/10.1038/d41586-019-02601-9

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PLOS One*, *3*, e3832. https://doi.org/10.1371/journal.pone.0003832, PubMed: 19050755

Reinikainen, J. (2005). The golden rule and the requirement of universalizability. *The Journal of Value Inquiry*, *39*(2), 155–168. https://doi.org/10.1007/s10790-006-8363-y

Rizzo, A., & Kim, G. J. (2005). A SWOT analysis of the field of virtual reality rehabilitation and therapy. *PRESENCE: Teleoperators and Virtual Environments*, *14*, 119–146. https://doi.org/10.1162/1054746053967094

Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, *17*(3), 523–534. https://doi.org/10.1016/j.labeco.2009.04.005

Ross, C. T. (2015). A multi-level Bayesian analysis of racial bias in police shootings at the county-level in the United States, 2011–2014. *PLOS One*, *10*(11), e0141854. https://doi.org/10.1371/journal.pone.0141854, PubMed: 26540108

Ross, C. T. (2017). Digging deeper: Population-level racial disparities, exposure to police victimisation and psychological trauma. *Epidemiology and Psychiatric Sciences*, 1–3. https://doi.org/10.1017/S2045796017000130, PubMed: 28397648

Rovira, A., Swapp, D., Southern, R., Zhang, J. J., & Slater, M. (2013). *The impact of enhanced projector display on the responses of people to a violent scenario in immersive virtual reality*. Paper presented at the Virtual Reality (VR), 2013 IEEE. https://doi.org/10.1109/VR.2013.6549350

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience, 6*, 332–339. https://doi.org/10.1038/nrn1651, PubMed: 15803164

Sanchez-Vives, M. V., Spanlang, B., Frisoli, A., Bergamasco, M., & Slater, M. (2010). Virtual hand illusion induced by visuomotor correlations. *PLOS One*, *5*, e10381. https://doi.org/10.1371/journal.pone.0010381, PubMed: 20454463

Scott, K., Ma, D. S., Sadler, M. S., & Correll, J. (2017). 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. https://doi.org/10.1111/josi.12243

Singer, M. G. (1963). The golden rule. *Philosophy*, *38*(146), 293–314.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B*, *364*, 3549–3557. https://doi.org/10.3389/frobt.2016.00074, PubMed: 19884149

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., et al. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PLOS One*, e52766. https://doi.org/10.1371/journal.pone.0052766, PubMed: 23300991

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLOS One*, *5*(5), e10564. https://doi.org/10.1371/journal.pone.0010564, PubMed: 20485681

Stan Development Team. (2011–2019). *Stan modeling language users guide and reference manual 2.25.*

Tajadura-Jiménez, A., Banakou, D., Bianchi-Berthouze, N., & Slater, M. (2017). Embodiment in a child-like talking virtual body influences object size perception, self-identification,

and subsequent real speaking. *Scientific Reports*, *7*(1), 9637. https://doi.org/10.1038/s41598-017-09497-3

Van De Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217. https://doi.org/10.1037/met0000100

Weiss, P. (1941). The golden rule. *The Journal of Philosophy*, *38*(16), 421–430.

Welsh, M., Chanin, J., & Henry, S. (2020). Complex color-blindness in police processes and practices. *Social Problems*. https://doi.org/10.1093/socpro/spaa008

Willits, D. W., & Makin, D. A. (2017). Show me what happened: Analyzing use of force through analysis of body-worn camera footage. *Journal of Research in Crime and Delinquency*, 0022427817701257. https://doi.org/10.1177/0022427817701257

Yee, N., & Bailenson, J. N. (2007). The Proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research, 33*, 271–290. https://doi.org/10.1111/j.1468-2958.2007.00299.x

Yee, N., Bailenson, J. N., & Ducheneaut, N. (2009). The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research*, *36*(2), 285–312. https://doi.org/10.1177/0093650208330254