

Human freezing responses to virtual characters in immersive virtual reality are impacted by body expression, group affiliation and threat proximity

Manuel Mello^{a,b,c,✉,*}, Lennie Dupont^{a,b,*}, Tahnée Engelen^d, Adriano Acciarino^e, Aline W. de Borst^{f,g},
Beatrice de Gelder^h

a SCNLab, Fondazione Santa Lucia, IRCCS, Rome, Italy

b Sapienza, Università degli Studi di Roma & CLNS@Sapienza, Istituto Italiano di Tecnologia, Rome, Italy

c Cognitive Neuroscience Research Unit, Department of Psychology, City University of London

d Cognitive and Computational Neuroscience Laboratory (LNC2), Inserm U960, Department of Cognitive Studies, Ecole Normale Supérieure, Paris, France

e Saint Camillus International Medical University (UniCamillus), Rome, Italy.

f Biological Psychology and Neuropsychology, Psychology and Human Movement, University of Hamburg, Hamburg, Germany

g UCLIC, University College London, London, United Kingdom

h Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

* These authors contributed equally to this work

✉ Manuel Mello: manuel.mello@uniroma1.it, manuel.mello@city.ac.uk

Keywords

Freezing, Embodiment, Body expressions, Group affiliation, Threat proximity, Virtual Reality

Abstract

Social threat requires fast adaptive reactions. One prominent threat-coping behavior present in humans is freezing, of which heart rate deceleration and reduced postural mobility are two key components. Previous studies focused mainly on freezing reactions in rodents, but now virtual reality offers unique possibilities for controlled and ecologically valid lab-based experiments. Using immersive virtual reality, this study examined how several understudied aspects of social threat, i.e., emotional body expressions, group affiliation, and distance from the potential threat, affect freezing behavior in humans. Reduced heart rate and postural mobility were observed in participants when they faced aggressive-looking and proximal avatars. Freezing was also observed for ingroup aggression when participants were embodied in a black-skinned virtual body and faced black-skinned aggressive and proximal avatars. Our results, based on a highly ecological virtual reality paradigm, provide novel evidence on the social factors that elicit freezing behavior in humans.

Introduction

The ability to detect threat and to react adaptively is a major evolutionary endowment in many species (Blanchard et al., 1986). Human and non-human studies of defensive behavior have discovered different stages of reactivity to incoming threats. Freezing has been defined as a threat-anticipatory state whereby an individual is hyperattentive to a potentially threatening stimulus to enhance its processing (Terburg et al., 2018; Livermore et al., 2021). Research on freeze-like reactions to threat in humans is still limited. Previous work has investigated freezing using threat-related social stimuli, such as facial expressions (Roelofs et al., 2010; Stins et al., 2011) and affective films (Hagenaars et al., 2014), but also computer-based tasks (e.g., a gun shooting task, Gladwin et al., 2016). These studies have examined freezing in humans by measuring heart rate and postural mobility, as bradycardia, a reduction in one's heart rate, and reduced postural mobility are two principal components of freezing (Roelofs, 2017; Livermore et al., 2021).

Observing an aggressive body expression triggers fear and prepares the observer for adaptive action (de Gelder et al., 2004, 2006; Grezes et al., 2007; Pichon et al., 2012), and a threat-anticipatory freezing

state plays a key role in the detection and processing of such social threat (Roelofs, 2017; Livermore et al., 2021). So far, a major obstacle in research on human defensive behavior relates to the difficulty of rendering threatening situations in a realistic manner. The use of virtual reality (VR) now opens unique chances (Parsons et al., 2017; Monti and Aglioti, 2018) as it allows participants to experience a threatening event in a controlled laboratory environment “as if” it was actually happening to them (Tieri et al., 2015; de Borst et al., 2020). Here, we took advantage of VR techniques to create avatars showing an aggressive or neutral body expression and we investigated their effect on freezing. We predicted a reduction of heart rate and postural mobility when facing aggressive- compared to neutral-looking approaching avatars.

The potential of VR studies is substantially increased by capitalizing on the phenomenon of body ownership (BO), i.e., the feeling that the body and its parts belong to the “self” (Berlucchi and Aglioti, 2010). BO illusion paradigms, e.g., the full-body illusion (Maselli and Slater, 2013), have demonstrated that the feeling of BO is a crucial aspect to consider when studying reactions to threatening stimuli (Tieri et al., 2015; Fusaro et al., 2016). When the virtual body is experienced in first-person perspective (1PP) and/or visuomotor synchrony between the real and the artificial body is provided, strong BO feelings can be experienced for avatars representing a different ethnic group (Peck et al., 2013; Banakou et al., 2016), a different age group (Banakou et al., 2013), or a different sex (de Borst et al., 2020; Mello et al., 2021). We used this opportunity to examine the effect of ethnic group affiliation on freezing by having white people embody, in two different sessions, either a white-skinned or a black-skinned virtual body (VB) and confront either white-skinned or black-skinned virtual characters. Group affiliation based on ethnicity is a prominent aspect of social life. It has been linked to cognitive bias and implicit attitudes and to behavioral and physiological reactivity associated with threat perception when facing outgroup individuals (Boyer et al., 2015; Amodio and Cikara, 2021). In this study, we based our predictions on the group affiliation literature and expected a reduction of heart rate and postural mobility when participants embodied a white-skinned VB and encountered a black-skinned avatar and similarly when they embodied a black-skinned VB and faced white-skinned avatars. Moreover, we tested whether embodiment in a VB of a different ethnicity would lead to a reduction in participants’ implicit biases

towards people of that specific ethnicity, measured via the implicit association test (IAT) (Peck et al., 2013; Banakou et al., 2016).

Another dimension of social threat we focussed on was social distance. Freezing reactions in humans are highly dependent on the proximity of the threat. Animal studies have shown that different behavioral strategies are adopted by an individual depending on whether the potential threat is distal or proximal (Blanchard et al., 1986, Fanselow, 1994). When a menace is detected and it enters an individual's proximal space, a freezing state of fear bradycardia and motor immobility helps to allocate attentional resources to the threat. When the threat is most imminent, defensive behavior may switch from passive (e.g., freezing) to active (e.g., fight or flight) (Fanselow, 1994; Lang et al., 1997). Recent studies have confirmed this proximity-related pattern of behavioral strategies when facing a threat in humans (Mobbs et al., 2007; Löw et al., 2015; Wendt et al., 2017; Rosén et al., 2019). Specifically, when the threat is proximal but inescapable, attentive freezing predominates. On the other hand, if participants are given the opportunity to escape the approaching threat, the defensive behavior switches from passive freezing to active avoidance (Löw et al., 2015; Wendt et al., 2017). While some of these studies have used paradigms that maximize experimental control at the expense of ecological validity, we balanced these two factors in our immersive VR design by utilizing approaching virtual avatars. We predicted that proximal avatars would elicit an enhanced freezing response as compared to distal avatars. Furthermore, we hypothesized an enhanced freezing response to aggressive avatars and outgroup members when they reached close proximity.

Materials and methods

Participants. Thirty healthy volunteers participated in this study (17 men, 13 women; age range 20-31 years, $M = 24.8$; $SD = 5.13$). Participants were all white/Caucasian except for one white/Asian and had normal or corrected-to-normal vision. Four participants were excluded from the analyses because of technical problems during the electrocardiography (ECG) and postural mobility recordings. The sample size was based on that of similar previous studies (e.g., Gladwin et al., 2016) and on a sample size

estimation using MorePower 6.0 software, which indicated that a sample of ~30 participants was sufficient to identify a medium effect (e.g., a three-way interaction) with at least 80% power.

General procedure and design. The experiment was divided into two similar sessions with a minimum of five days in between. In one session the participants were embodied in a black-skinned VB and in the other, they were embodied in a white-skinned VB (Fig. 1). The order of embodiment type was randomized across participants. At the start of each session, participants completed the race Implicit Association Test (race IAT, Greenwald et al., 1998). Before starting the experiment, participants were asked to stand barefoot on a force platform and ECG electrodes for measuring heart rate (HR) were attached to their chest. Next, they went through an immersive embodiment training using the first virtual environment (Fig. 1A). They viewed either a sex-matched white-skinned or a black-skinned avatar reflected in a mirror from a first-person perspective (1PP) and performed several movements (e.g., move the arms) intended to induce embodiment through visuomotor synchrony (e.g., Peck et al., 2013).

After this training, the experimental task started during which participants were immersed in a second virtual environment (Fig. 1B) and observed avatars walking towards them one by one. The approaching avatars were either black-skinned or white-skinned and displayed either a neutral (arms down) or an aggressive (raised arms) body posture (Fig. 1B). Thus, the experimental variables consisted of the within-subject factors *Embodiment type* (black-skinned, white-skinned), *Avatar skin color* (black-skinned, white-skinned), *Avatar posture* (aggressive, neutral), and *Avatar Proximity* (far, intermediate, close; see next section for the rationale of this latter factor). The experiment was divided into three blocks. A session consisted of 84 trials (28 per block), each lasting 6 seconds, with a 12-second intertrial interval (ITI). The order of trials was randomized across blocks and across participants. After the main task, participants completed the race IAT again and filled out an embodiment questionnaire (Peck et al., 2013). Finally, participants were debriefed about the study and rewarded for their participation.



Figure 1. (A): Embodiment environment and an example from the embodiment training task: “*Slightly lift your arms up and wave to yourself in the mirror*”. Left: white-skinned embodiment; right: black-skinned embodiment. **(B):** Stimuli and environment from main experiment. From left to right: black-skinned avatar with aggressive body posture (raised arms), white-skinned avatar with aggressive body posture (raised arms), black-skinned avatar with neutral body posture (arms down), white-skinned avatar with neutral body posture (arms down).

Data collection, pre-processing, and analysis.

Virtual embodiment questionnaire. The ownership scores for the VBs were obtained by combining the two questions that relate to body ownership, Q1 (*Mybody*) and Q4 (*Notme*), according to Peck et al. (2013): $(Q1 + 6 - Q4)/2$ (Table 1). A paired sample t-test was performed to test for differences in body ownership in black-skinned vs. white-skinned VB conditions using a significance threshold of $p = 0.05$.

Item	Type	Statement
------	------	-----------

Q1	<i>Mybody</i>	I felt as if the body I saw in the virtual world might be my body.
Q2	<i>Nervous</i>	I became nervous when the other avatars approached me.
Q3	<i>Control</i>	I felt like I controlled the avatar as if it was my own body.
Q4	<i>Notme</i>	I felt like the avatar was not me.
Q5	<i>Liked</i>	I liked being able to control the movements of the avatar.
Q6	<i>Greet</i>	I wanted to say hello to the avatars as they walked towards me.

Table 1: Embodiment questionnaire items.

Baselined heart rate and postural mobility. Participants' ECG was recorded using an EGI system (Electrical Geodesics, Inc., Eugene, USA), via two active alligator-connected electrodes, one placed at the centre of the sternum and the other placed at the ribs under the left breast. Raw ECG data were pre-processed in Matlab (see Supplementary information) and HR changes to the experimental conditions were computed as differential values between instantaneous HR (IHR) in three subsequent windows of 2000 ms, covering the 6000 ms trial duration (representing *Avatar proximity*), and 2000 ms of baseline before stimulus onset. A preliminary data inspection showed the model residuals to have a slightly negatively skewed and leptokurtic distribution, therefore the baselined HR values were normalized with a Lambert W x F function using the package LambertW in R (Georg, 2015; Fig. S1). The normalized HR values were analysed with a linear mixed-effects approach in RStudio using the lmer() function from the lme4 package (Bates et al., 2015), which allows to model dependency in the data (within-subjects factors) and to account for missing values. All the fitted models included by-subject intercepts. Graphical inspection of model residuals and fitted vs. predicted values revealed that normality of model residuals, homoscedasticity and linearity assumptions were met (Fig. S2).

Centre of pressure (COP) excursions were recorded with an AMTI AccuSway Force Platform (Advanced Mechanical Technology, Inc., Watertown, USA). The anterior-posterior (AP) and medial-lateral (ML) time series define the COP path relative to the origin of the force platform coordinate system (Prieto et al., 1996; see also Supplementary information). COP raw data were filtered with a low-pass Butterworth filter (cut-off frequency: 5 Hz, filter order: 4). Postural mobility (body sway) was

calculated as the standard deviation of the COP in the AP and ML planes in three subsequent windows of 2000 ms, covering the 6000 ms post-stimulus onset (representing *Avatar proximity*). 0.001% (~25 trials) of the data was manually removed as they represented clear technical artifacts. Scores were analysed with a generalized linear mixed-effects approach in RStudio using the `glmer()` function from the `lme4` package (Bates et al., 2015), which allows to model dependency in data with non-normal distributed model residuals and to account for missing values. The data were assumed to follow a gamma distribution. All the fitted models included by-subject intercepts.

For both baselined HR and postural mobility measures, model complexity was gradually increased by inserting fixed effects – *Embodiment type*, *Avatar skin color*, *Avatar posture*, and *Avatar proximity* – and their interactions to check for the model that best fitted the data. The different statistical models were compared using the `anova()` function from the `stats` package (R Core Team, 2019). AIC, BIC, and Chi-square statistics informed us on which model best fitted the data compared to the previous ones in the hierarchy. When relevant, Tukey-corrected post-hoc comparisons were performed with the `R` package `lsmeans` (Russel and Lenth, 2016). For all results, a threshold of $p < 0.05$ for statistical significance was used.

Race IAT. The IAT measures the strength of the automatic concept-attribute associations that are thought to underlie implicit biases and stereotypes (Greenwald et al., 1998). The critical blocks consisted in answering with the same keyword button to “bad” words (e.g., awful, nasty) *and* images of black-skinned individuals and with another keyword button (placed in the opposite direction, symmetrically) to “good” words (e.g., love, happy) *and* images of white-skinned individuals (Knutson et al., 2007; see also Table S1). Vice versa for non-critical blocks. Individual IAT D1 scores were extracted as the mean difference in reaction times between non-critical and critical blocks of the task, divided by the pooled standard deviation of the two types of blocks. These scores represent the strength of the implicit concept-attribute association that sees black people as “bad” and white people as “good”. We created an index subtracting the individual D1 scores for the IAT taken before the embodiment to those obtained with the IAT taken at the end of each session – we call this $\Delta D1$. We then compared the $\Delta D1$ scores across sessions (black and white embodiment) with a paired sample t-test.

Results

Embodiment questionnaire. Participants' ownership scores did not differ depending on the type of VB during the two sessions ($t = -0.91, p = 0.37$) indicating that they experienced similar embodiment in a white-skinned VB ($M = 2.99, SD = 1.38$) and in a black-skinned VB ($M = 3, SD = 1.42$). These results are further supported by the significant positive correlation found between ownership scores for a white-skinned avatar and those for a black-skinned avatar (Pearson's $r = 0.43, p = 0.03$), clearly indicating within-subject coherence in eliciting embodiment using immersive virtual reality.

Baselined HR. The linear model run with baselined heart rate values as outcome yielded main effects of *Embodiment type*, *Avatar skin color*, *Avatar posture*, and *Avatar proximity*. Overall, HR was significantly lower when participants embodied a black-skinned VB, compared to a white-skinned one (estimate = 0.65; $t = 2.96, p = 0.003$); when they faced a black-skinned, compared to a white-skinned avatar (estimate = 0.58; $t = 2.44, p = 0.01$); when they faced an aggressive, compared to a neutral avatar (estimate = 0.51; $t = 2.36, p = 0.01$); and for close compared to both intermediate (estimate = 0.53; $t = 2.44, p = 0.01$) and far avatar proximity (estimate = 0.42; $t = 1.96, p = 0.04$). *Embodiment type* further interacted with *Avatar skin color* and *Avatar posture*. In the first interaction, the effect represented by a reduced heart rate when facing a black-skinned avatar, compared to a white-skinned one, was greater when participants embodied a black-skinned VB (estimate = -0.77; $t = -3.07, p = 0.002$). This effect was driven by a greater reduction in heart rate when facing a black-skinned avatar and embodying a black-skinned VB. Similarly, the effect represented by a reduced heart rate when facing an aggressive avatar, compared to a neutral one, was greater when participants embodied a black-skinned VB (estimate = -0.53; $t = -2.12, p = 0.03$). This effect was also driven by a greater reduction in heart rate when facing an aggressive avatar and embodying a black-skinned VB. Importantly, we found a significant three-way interaction: *Embodiment type* interacted with *Avatar skin color*, and *Avatar posture* (estimate = 0.8; $t = 2.26, p = 0.02$). This interaction is explained by the reduction in heart rate when facing a black-skinned, aggressive avatar, compared to a black-skinned, neutral avatar, and when participants embodied a black-skinned VB themselves (Fig. 2).

Baselined heart rate

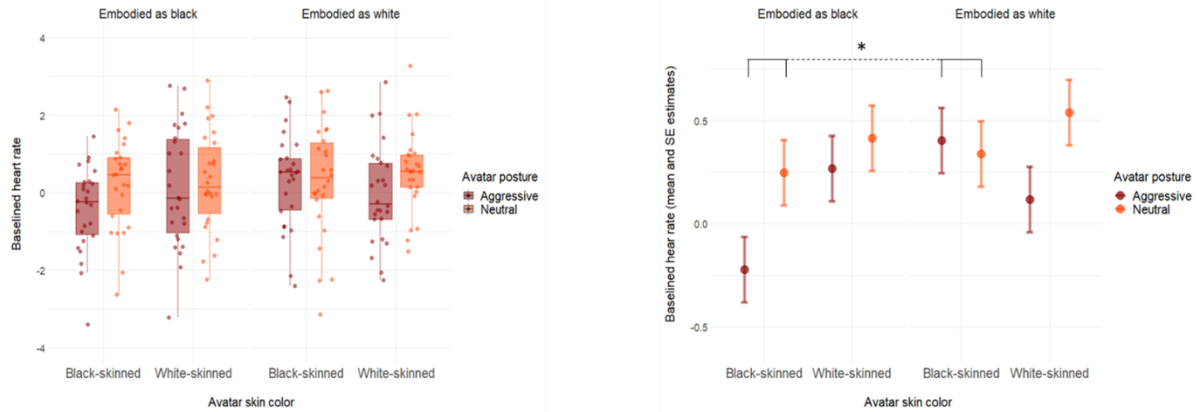


Figure 2: *Left:* baselined heart rate data averaged across participants. *Right:* corresponding mean and SE model estimates showing a three-way interaction effect (*Embodiment type*Avatar skin color*Avatar posture*).

Postural mobility. The generalized linear model run with the SDs in the AP plane (AP_SD) as outcome yielded main effects of *Avatar skin color* and *Avatar proximity*. AP_SD decreased when facing a black-skinned, compared to a white-skinned, avatar (estimate = -1.34; $t = -2.56$, $p = 0.01$). Moreover, a decrease in AP_SD was observed for close avatar proximity compared to both far (estimate = -2.03; $t = -3.29$, $p = 0.001$) and intermediate (estimate = -1.53; $t = -2.45$, $p = 0.01$) avatar proximity. The factor *Avatar proximity* further interacted with both *Embodiment type* and *Avatar skin color*. As concerns the former interaction, the effects represented by a reduction in AP_SD for close versus far avatar proximity (estimate = -1.41; $t = -2.1$, $p = 0.03$) and for close versus intermediate avatar proximity (estimate = -2.56; $t = -3.8$, $p < 0.001$) were greater when participants embodied a white-skinned compared to a black-skinned VB. This effect was driven by an overall decrease in AP_SD when embodying a black-skinned VB (Fig. 3A). In fact, post hoc tests revealed that the embodiment in a black-skinned VB compared to white-skinned VB, was associated with reduced AP_SD for far avatar proximity (estimate = 1.99; $z = 4.66$, $p < 0.001$) and intermediate avatar proximity (estimate = 3.13; $z = 7.38$, $p < 0.001$). Finally, the same proximity effects, a reduction in AP_SD for close versus far avatar proximity (estimate = 1.69; t

= 2.57, $p = 0.01$) and for close versus intermediate avatar proximity (estimate = 1.54; $t = 2.42$, $p = 0.01$) were greater when participants faced a black-skinned compared to a white-skinned avatar. These effects were mainly driven by a reduced AP_SD when participants faced a black-skinned avatar in the last 2 seconds of the condition (close proximity) (Fig. 3B). All other main effects and interactions were not significant.

COP SDs in the AP plane

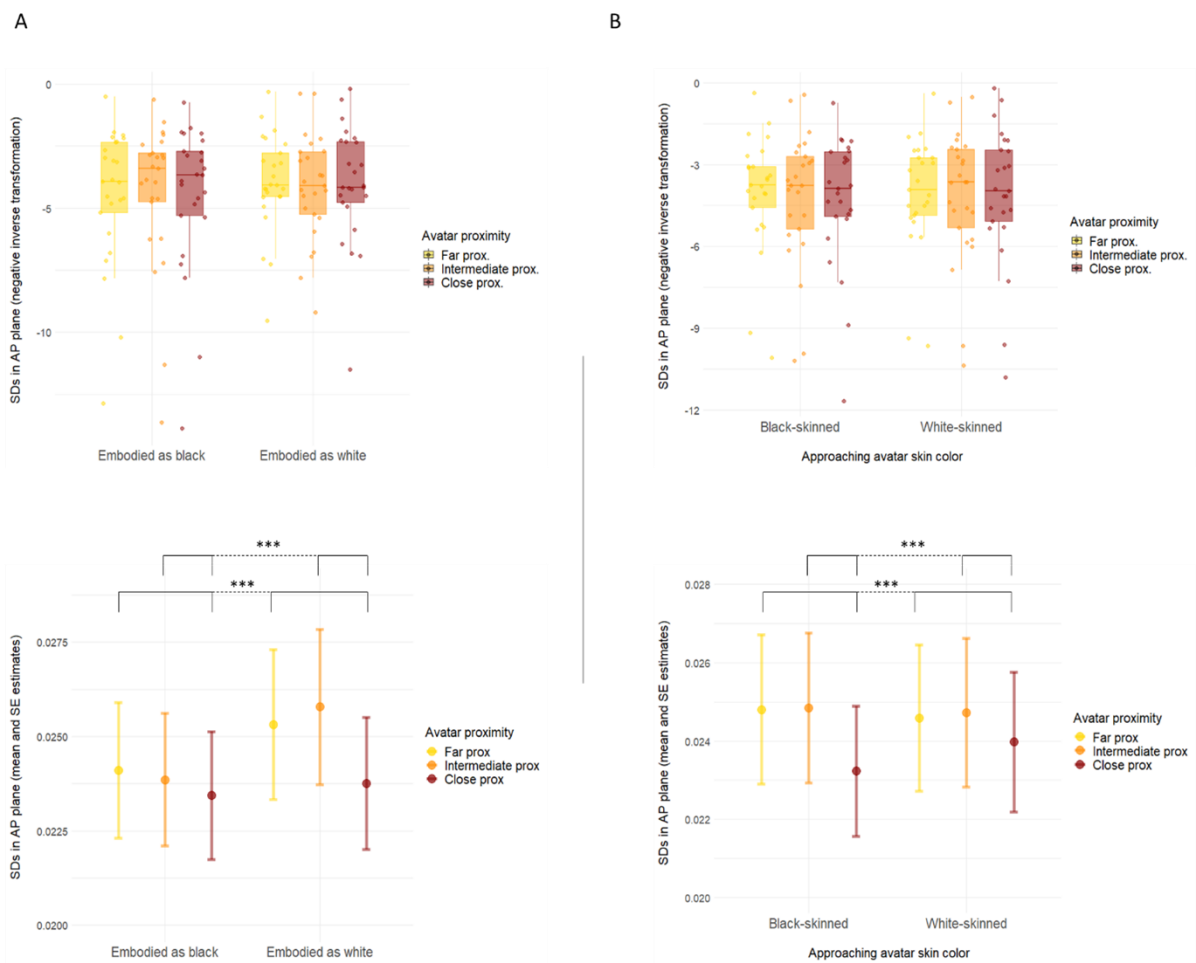


Figure 3. (A): Top: body sway in the AP plane data averaged across participants. **Bottom:** corresponding mean and SE model estimates showing a two-way interaction effect (*Embodiment type*Avatar proximity*). **(B): Top:** body sway in the AP plane data averaged across participants. **Bottom:** corresponding mean and SE model estimates showing a two-way interaction effect (*Avatar skin color*Avatar proximity*).

The generalized linear model run with the SDs in the ML plane (ML_SD) as outcome yielded main effects of *Embodiment type*, *Avatar skin color*, and *Avatar posture*. We observed a reduction in ML_SD when embodying a black-skinned versus white-skinned VB (estimate = -2.08; $t = -3.84$, $p < 0.001$); when facing a black-skinned versus white-skinned avatar (estimate = -1.31; $t = -2.36$, $p = 0.01$); and when facing an aggressive versus neutral avatar (estimate = -1.32; $t = -2.37$, $p = 0.01$). We also found a set of two-way interactions – *Embodiment type* and *Avatar skin color*, *Embodiment type* and *Avatar posture*, *Avatar skin color* and *Avatar posture*, and *Avatar skin color* and *Avatar proximity* – that are better explained by the three-way and four-way interactions we observed. *Embodiment type* interacted with *Avatar skin color* and *Avatar posture*. ML_SD in response to a black-skinned, aggressive avatar versus a black-skinned, neutral avatar had an inverse relationship when embodying a black-skinned compared to a white-skinned VB. In fact, ML_SD decreased when facing a black-skinned, aggressive avatar (compared to neutral) and embodying a black-skinned VB, while it increased when facing a black-skinned, aggressive avatar (compared to neutral) and embodying a white-skinned VB (estimate = -3.09; $t = -3.1$, $p = 0.002$) (Fig. 4). Supporting this, post hoc tests revealed that, when facing a black-skinned, aggressive avatar, participants' ML_SD was significantly lower when embodying a black-skinned VB, compared to a white-skinned one (estimate = 1.77; $z = 5.56$, $p < 0.001$). Lastly, this three-way interaction effect was modulated by *Avatar proximity*, being greater for close versus far avatar proximity (four-way interaction: estimate = -3.91; $t = 2.71$, $p = 0.006$).

COP SDs in the ML plane

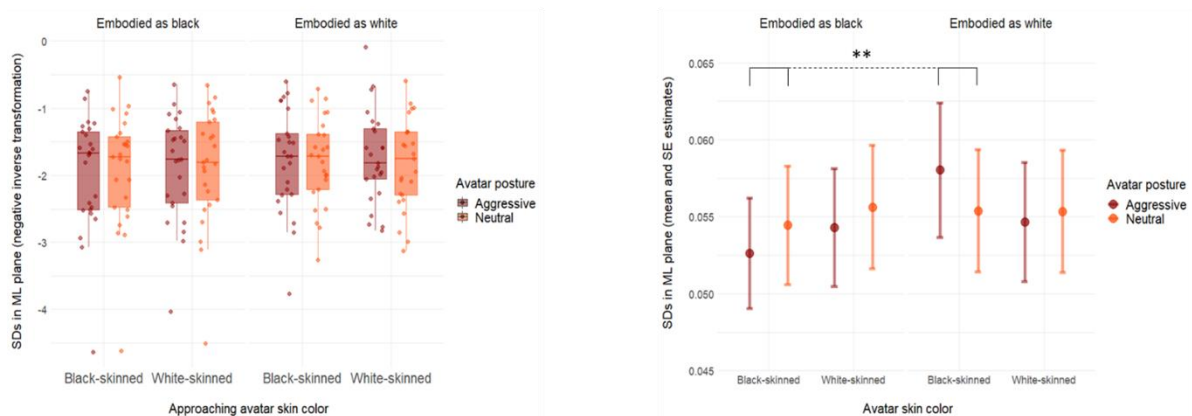


Figure 4. *Left:* body sway in the ML plane data averaged across participants. *Right:* corresponding mean and SE model estimates showing a three-way interaction effect (*Embodiment type*Avatar skin color*Avatar posture*). *Note:* a four-way interaction effect (*Embodiment type*Avatar skin color*Avatar posture*Avatar proximity*) was found, but a three-way effect is shown for simplicity.

Race-IAT. Reaction times differed significantly for critical ($M = 0.69$, $SD = 0.13$) compared to non-critical blocks ($M = 0.77$, $SD = 0.15$) (averaged across sessions). Participants were faster in categorizing “bad” words *with* images of black-skinned individuals and “good” words *with* images of white-skinned individuals compared to the opposite ($t = -8.16$, $p < 0.001$). This indicates that an implicit outgroup bias was present in our sample. $\Delta D1$ scores did not differ across embodiment conditions.

Discussion

This study set out to test the effect of understudied aspects of social threat, namely body expression, group affiliation and proximity, on freezing. We hypothesized that aggressive and outgroup avatars approaching the participants would elicit a reduction in heart rate and body sway and that this would be enhanced when the avatars are closest to the participant. Our results partly confirm this hypothesis. In fact, we found that people freeze more when facing aggressive-looking avatars and when avatars are in close proximity, independent of their characteristics. In contrast to our prediction, freezing was stronger for ethnic ingroup members, i.e., for black-skinned aggressive avatars when participants embodied a black-skinned VB. Below, we discuss these findings against the relevant literature.

Our findings on freezing when facing aggressive avatars are in line with previous work reporting freezing for social threats in humans (Roelofs et al., 2010; Hagenaaers et al., 2014; Gladwin et al., 2016). For instance, Roelofs and colleagues (2010) showed that the perception of angry facial expressions, compared to happy and neutral expressions, was associated with both reduced body sway and bradycardia. We report that a freezing state is triggered by threatening whole-body expressions as well. The notion that seeing expressive whole-body postures and movements triggers motor preparation was

at the core of the earliest studies on body expressions (e.g., de Gelder et al., 2004) but no direct physiological or behavioral evidence was so far available. Consistent with this perspective, Borgomaneri and colleagues (2015) showed that watching fearful body expressions, compared to happy and neutral, suppresses TMS-induced intracortical facilitation in the motor cortex and the authors interpreted this as an index of freezing response when observing a cue indicating a potential threat in the environment. We expand this at the behavioral and physiological levels by showing that facing an aggressive body posture is also associated with freezing reactivity.

The present work also clarifies the impact of the proximity of the social threat on freezing as we found an overall reduction of heart rate and body sway in the AP plane when participants faced a virtual avatar in close proximity. This situation represented an inescapable threat as participants were instructed to stand still on the force platform and to not try and react to the approaching stimuli. Our finding is consistent with previous studies showing that facing approaching inescapable threats results in augmentation of freezing, as represented by fear bradycardia, and increased skin conductance and startle reflex (Löw et al., 2015; Wendt et al., 2017). For instance, Wendt and colleagues (2017) showed that while facing approaching inescapable threats is associated with attentive freezing behavior, escapable threats elicit instead active avoidance. Thus, our results support the finding that, when people are not given the opportunity to actively avoid a threat, passive freezing is the predominant defensive strategy implemented by humans in the face of proximal threats. However, our paradigm could not dissociate passive from active defensive behavior, as escapable threats were not included. Future studies should address this limitation by implementing a paradigm in which a freezing-like reaction to virtual threatening stimuli is assessed in the face of both escapable and inescapable threats.

The other main aspect of our findings concerns the role of group affiliation, studied here by manipulating skin color. We showed that behavioral reactions and physiological activation differ as a function of the skin color of the threatening avatar but also as a function of the type of embodiment. We observed enhanced freezing (reduction of heart rate and body sway in the ML plane) to black-skinned aggressive avatars when the participants embodied a black-skinned VB themselves. This effect was also greater for proximal avatars (only concerning the body sway in the ML plane). This finding is surprising considering results from studies on group affiliation and on implicit bias. First, the literature

on group affiliation bias and perception of threat (Amodio and Cikara, 2021) predicts that encountering an avatar belonging to a different group (by virtue of skin color) under the above conditions exacerbates freezing. In fact, an approach based on the role of group dynamics, with ‘group’ here specified as skin color, would predict that behavior is consistent with the participants’ group affiliation such that a black-skinned aggressor is more threatening when participants embody a white avatar and vice versa. However, the group affiliation logic seems to fail here as our results show the opposite pattern and suggest that ingroup-outgroup dynamics and their influence on defensive behavior are not simply a matter of skin color (of both the offender and the offended).

Beyond simple group affiliation-based threat reactions (Boyer et al., 2015), the literature suggests various other explanations which we present as speculations. The fact that the minority (black-skinned) avatar condition tends to be perceived by the participants as more threatening when they embody a black-skinned VB finds support in studies on internalized stereotypes, where for example black individuals consider black neighbourhoods as “less safe” if they believe the stereotype that black communities are dangerous (Bailey et al., 2011). A different possibility is that ingroup violence is perceived as more realistic and therefore more threatening because aggression from the same group member is likely to be more serious than random and less personal aggression from a stranger. Supporting this, and along with our main findings on black-skinned embodiment, our results (Figure 2) show that the difference in heart rate between white-skinned aggressive and white-skinned neutral conditions is bigger when embodying a white-skinned avatar, compared to a black-skinned one. However, the direct comparison between the conditions of interest – white embodiment, white-skinned aggressive avatar compared to black embodiment, white-skinned aggressive avatar – was not significant (Tukey-corrected, post-hoc comparison: $p > 0.05$).

Finally, it is worth noting that previous studies have shown that the embodiment of white people in a black-skinned VB is associated with reduced negative biases towards black people (Peck et al., 2013; Banakou et al., 2016). But we did not find such a reduction in IAT-measured implicit biases when embodying a black-skinned VB. We cannot exclude that a longer embodiment phase (see *General procedure and design*) would have made a critical difference in eliciting the expected reduction in IAT-measured outgroup biases when embodying a black-skinned VB. Still, our results show that this

condition caused the illusion of body ownership in the participants. Alternatively, the long exposure to repeated stimuli during the main experimental task might have weakened the initial strength of the embodied perspective-taking process.

Taken together our findings show that group affiliation dynamics, skin color, implicit attitudes and behavior have complex relations with one another such that an explanation focusing on just one of them is unlikely to explain the data. More studies are needed to shed light on the dispositional and contextual factors that contribute to defensive behavior both in within-group and in between-group situations and to obtain a better understanding of the within-group dynamics that may have been neglected in favour of ingroup-outgroup difference.

Author Contributions

M. M., L. D., T. E., A. A., A. W. d. B., B. d. G.: Conceptualization, Design; M. M., L. D., A. A.: Data Collection, Data Curation; M. M., L. D., T. E.: Formal Analysis; M. M., L. D.: Writing – Original Draft, Visualization; M. M., L. D., T. E., A. A., A. W. d. B., B. d. G.: Writing – Review & Editing; B. d. G.: Funding Acquisition.

Acknowledgments

This work was supported by a European Research Council (ERC) Grant [European Union's Seventh Framework Programme (FP7/2007e2013)/ERC Grant agreement number 295673] to B. de Gelder. No conflicts of interest, financial or otherwise, are declared by the authors.

References

- Amodio, D. M., & Cikara, M. (2021). The social neuroscience of prejudice. *Annual Review of Psychology*, 72, 439-469. <https://doi.org/10.1146/annurev-psych-010419-050928>

- Bailey, T. K. M., Chung, Y. B., Williams, W. S., Singh, A. A., & Terrell, H. K. (2011). Development and validation of the Internalized Racial Oppression Scale for Black individuals. *Journal of Counseling Psychology*, 58(4), 481. <https://doi.org/10.1037/a0023585>
- Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, 110(31), 12846-12851. <https://doi.org/10.1073/pnas.1306779110>
- Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). 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. <https://doi.org/10.3389/fnhum.2016.00601>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- Berlucchi, G., & Aglioti, S. M. (2010). The body in the brain revisited. *Experimental Brain Research*, 200(1), 25-35. <https://doi.org/10.1007/s00221-009-1970-7>
- Blanchard, R. J., Flannelly, K. J., & Blanchard, D. C. (1986). Defensive behaviors of laboratory and wild *Rattus norvegicus*. *Journal of Comparative Psychology*, 100(2), 101. <https://doi.org/10.1037/0735-7036.100.2.101>
- Borgomaneri, S., Vitale, F., Gazzola, V., & Avenanti, A. (2015). Seeing fearful body language rapidly freezes the observer's motor cortex. *Cortex*, 65, 232-245. <https://doi.org/10.1016/j.cortex.2015.01.014>
- Boyer, P., Firat, R., & van Leeuwen, F. (2015). Safety, threat, and stress in intergroup relations: A coalitional index model. *Perspectives on Psychological Science*, 10(4), 434-450. <https://doi.org/10.1177/1745691615583133>
- De Borst, A. W., Sanchez-Vives, M. V., Slater, M., & de Gelder, B. (2020). First-person virtual embodiment modulates the cortical network that encodes the bodily self and its surrounding space during the experience of domestic violence. *eNeuro*, 7(3). <https://doi.org/10.1523/ENEURO.0263-19.2019>

- De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, 7(3), 242-249. <https://doi.org/10.1038/nrn1872>
- De Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjikhani, N. (2004). Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Sciences*, 101(47), 16701-16706. <https://doi.org/10.1073/pnas.0407042101>
- Fanselow, M. S. (1994). Neural organization of the defensive behavior system responsible for fear. *Psychonomic Bulletin & Review*, 1(4), 429-438. <https://doi.org/10.3758/BF03210947>
- Fusaro, M., Tieri, G., & Aglioti, S. M. (2016). Seeing pain and pleasure on self and others: behavioral and psychophysiological reactivity in immersive virtual reality. *Journal of Neurophysiology*, 116(6), 2656-2662. <https://doi.org/10.1152/jn.00489.2016>
- Georg, G. M. (2015). The lambert way to gaussianize heavy-tailed data with the inverse of Tukey'sh transformation as a special case. *The Scientific World Journal*, 2015. <https://doi.org/10.1155/2015/909231>
- Gladwin, T. E., Hashemi, M. M., van Ast, V., & Roelofs, K. (2016). Ready and waiting: Freezing as active action preparation under threat. *Neuroscience Letters*, 619, 182-188. <https://doi.org/10.1016/j.neulet.2016.03.027>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464. <https://doi.org/10.1037//0022-3514.74.6.1464>
- Grezes, J., Pichon, S., & De Gelder, B. (2007). Perceiving fear in dynamic body expressions. *Neuroimage*, 35(2), 959-967. <https://doi.org/10.1016/j.neuroimage.2006.11.030>
- Hagens, M. A., Roelofs, K., & Stins, J. F. (2014). Human freezing in response to affective films. *Anxiety, Stress & Coping*, 27(1), 27-37. <https://doi.org/10.1080/10615806.2013.809420>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons, M. Balaban, & R. Simons (Eds). *Attention and orienting: Sensory and motivational processes* (97-135). Psychology Press.

- Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, 69(1), 1-33. <https://doi.org/10.18637/jss.v069.i01>
- Livermore, J. J., Klaassen, F. H., Bramson, B., Hulsman, A. M., Meijer, S. W., Held, L., ... & Roelofs, K. (2021). Approach-avoidance decisions under threat: The role of autonomic psychophysiological states. *Frontiers in Neuroscience*, 15, 353. <https://doi.org/10.3389/fnins.2021.621517>
- Löw, A., Weymar, M., & Hamm, A. O. (2015). When threat is near, get out of here: Dynamics of defensive behavior during freezing and active avoidance. *Psychological Science*, 26(11), 1706-1716. <https://doi.org/10.1177/0956797615597332>
- Maselli, A., & Slater, M. (2013). The building blocks of the full body ownership illusion. *Frontiers in Human Neuroscience*, 7, 83. <https://doi.org/10.3389/fnhum.2013.00083>
- Mello, M., Fusaro, M., Tieri, G., Aglioti S. M. (2021). Wearing same- and opposite-sex virtual bodies and seeing them caressed in intimate areas. *Quarterly Journal of Experimental Psychology*. Epub ahead of print. <https://doi.org/10.1177/17470218211031557>
- Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., ... & Frith, C. D. (2007). When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science*, 317(5841), 1079-1083. <https://doi.org/10.1126/science.1144298>
- Monti, A., & Aglioti, S. M. (2018). Flesh and bone digital sociality: On how humans may go virtual. *British Journal of Psychology*, 109(3), 418-420. <https://doi.org/10.1111/bjop.12300>
- Parsons, T. D., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain Sciences*, 7(4), 42. <https://doi.org/10.3390/brainsci7040042>
- 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(3), 779-787. <https://doi.org/10.1016/j.concog.2013.04.016>
- Pichon, S., de Gelder, B., & Grezes, J. (2012). Threat prompts defensive brain responses independently of attentional control. *Cerebral Cortex*, 22(2), 274-285. <https://doi.org/10.1093/cercor/bhr060>

- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Roelofs, K. (2017). Freeze for action: neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1718), 20160206. <https://doi.org/10.1098/rstb.2016.0206>
- Roelofs, K., Hagensaars, M. A., & Stins, J. (2010). Facing freeze: social threat induces bodily freeze in humans. *Psychological Science*, 21(11), 1575-1581. <https://doi.org/10.1177/0956797610384746>
- Rosén, J., Kastrati, G., Repling, A., Bergkvist, K., & Åhs, F. (2019). The effect of immersive virtual reality on proximal and conditioned threat. *Scientific reports*, 9(1), 1-8. <https://doi.org/10.1038/s41598-019-53971-z>
- Stins, J. F., Roerdink, M., & Beek, P. J. (2011). To freeze or not to freeze? Affective and cognitive perturbations have markedly different effects on postural control. *Human Movement Science*, 30(2), 190-202. <https://doi.org/10.1016/j.humov.2010.05.013>
- Terburg, D., Scheggia, D., Del Rio, R. T., Klumpers, F., Ciobanu, A. C., Morgan, B., ... & van Honk, J. (2018). The basolateral amygdala is essential for rapid escape: a human and rodent study. *Cell*, 175(3), 723-735. <https://doi.org/10.1016/j.cell.2018.09.028>
- Tieri, G., Tidoni, E., Pavone, E. F., & Aglioti, S. M. (2015). Body visual discontinuity affects feeling of ownership and skin conductance responses. *Scientific Reports*, 5(1), 1-8. <https://doi.org/10.1038/srep17139>