[1] Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018).
Robots and Racism. Proceedings of the ACM/IEEE International Conference on Human Robot Interaction, Chicago pp. 196-204. | DOI: 10.1145/3171221.3171260

Robots And Racism

Christoph Bartneck, Kumar Yogeeswaran,
Qi Min Ser, Graeme Woodward
University of Canterbury, HIT Lab NZ
Christchurch, Canterbury, New Zealand
christoph@bartneck.de

Robert Sparrow
Department of Philosophy, Monash University
Victoria, Australia
robert.sparrow@monash.edu

Siheng Wang
Guizhou University of Engineering Science
Guizhou, China
siheng.wang@canterbury.ac.nz

Friederike Eyssel
University of Bielefeld
Bielefeld, Germany
friederike.eyssel@uni-bielefeld.de

ABSTRACT

Most robots currently being sold or developed are either stylized with white material or have a metallic appearance. In this research we used the shooter bias paradigm and several questionnaires to investigate if people automatically identify robots as being racialized, such that we might say that some robots are "White" while others are "Asian", or "Black". To do so, we conducted an extended replication of the classic social psychological shooter bias paradigm using robot stimuli to explore whether effects known from human-human intergroup experiments would generalize to robots that were racialized as Black and White. Reaction-time based measures revealed that participants demonstrated 'shooter-bias' toward both Black people and robot racialized as Black. Participants were also willing to attribute a race to the robots depending on their racialization and demonstrated a high degree of inter-subject agreement when it came to these attributions.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **Empirical studies in HCI**;

KEYWORDS

shooter bias, racism, robot, implicit, explicit, prejudice

ACM Reference Format:
Christoph Bartneck, Kumar Yogeeswaran, Qi Min Ser, Graeme Woodward, Robert Sparrow, Siheng Wang, and Friederike Eyssel. 2018. Robots And Racism. In *Proceedings of 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3171221.3171260

HRI '18, March 5–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-4953-6/18/03…$15.00
https://doi.org/10.1145/3171221.3171260

1 INTRODUCTION

When forming first impressions about other people, we often rely on social cues to categorize on the basis of age, gender, and race [15]. Previous research has shown that people use these social categories even in impression formation about nonhuman entities [11, 13]. For instance, manipulation of a robot's body shape [1] or hairstyle in a gender stereotypical fashion elicits the perception of gender in robots [11, 23, 33].

Seeing that robots can be perceived as having gender on the basis of simple social cues such as stereotypic hairstyles or body shapes, and may give rise to significant social and ethical issues [35], such as whether it is ethical to design a female sex robot that does not consent to sex or that consents all the time. The question thus arises whether they might also be perceived to have race if presented with cues stereotypic of various racial identities. That is, do people automatically identify robots as being racialized, such that we might say that some robots are "White" while others are "Asian" or "Black", and are there socioethical concerns therein?

Because race corresponds with complicated patterns of social relationships, economic injustice, and political power [29, 37], the perception of race in the design space of robots has potential implications for HRI. In particular, an abundance of social psychological research shows that people have implicit racial biases which significantly affect their behavior [6, 17] (for a recent review, see [44]). For example, people are quicker to categorize negative words after subliminal or supraliminal exposure to faces of Black people, whereas after similar exposure to White faces people are quicker to categorize positive words [9, 14]. Neuroscientific studies further reveal the automaticity and impact of such biases. For example, people's brain activity reflects greater vigilance when viewing Black faces relative to White faces [5, 27]. Such implicit tendencies have been found to impact people's behavior in a variety of contexts including people's nonverbal behavior in intergroup interactions (e.g., [7, 9]), job discrimination [19, 30, 30, 43, 43], voting patterns [8, 18], and medical [16] and economic decisions [36].

Determining whether people perceive robots to have race, and if so, whether the same race-related prejudices extend to robots, is thus an important matter. To investigate these questions, we adapted the shooter bias paradigm [4] – a well established method for investigating the automaticity of race-based categorization and of biased behavioral responding. In this paradigm, participants are asked to play the role of a police officer whose job it is to shoot

[1] Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018).
Robots and Racism. Proceedings of the ACM/IEEE International Conference on Human Robot Interaction, Chicago pp. 196-204. | DOI: 10.1145/3171221.3171260

Figure 1: Example stimuli used in the experiment. Top row is robot and human holding an object, bottom row is robots and humans holding a gun. The left four images are our new robot photos while the right four images are the original images used in [4].

(by pressing a button) individuals carrying a gun, while refraining from shooting people carrying harmless objects such as a soda can, wallet, or a cell phone. The task is carried out using image-based stimuli on a computer, with multiple trials depicting the full manipulation of the individuals' race (Black versus White) crossed with the objects in hand. These trials occur on the screen in rapid succession to mirror the rapid context in which police officers are expected to make decisions.

Motivated by contemporary concerns regarding police brutality towards Black Americans, the shooter bias procedure was specifically designed to test the automaticity and severity of anti-Black racism. The effect (shooter bias) refers to people's tendency to more readily shoot Black agents over White agents [3, 4] (for a recent meta-analysis, see [21]). Thus, if people also exhibit shooter bias towards robots, this would provide strong evidence of both the automaticity with which we attribute race to robots, as well as a significant degree to which this may influence human-robot interactions.

1.1 Research Questions

Here we explored two overarching research questions:

- Do people ascribe race to robots?
- If so, does the ascription of race to robots affect people's behavior towards them?

Using a close adaptation of the shooter bias paradigm in which we manipulated the appearance of a Nao robot, we looked at one specific racially-motivated behavior: whether anti-Black racism extends to robots when racialized as Black. Based on the aforementioned literature, such bias would be evidenced by participants being:

(1) faster to shoot armed agents when they are Black (versus White);
(2) faster to not shoot unarmed agents when they are White (versus Black); and
(3) more accurate in their discernment of White (versus Black) aggressors

In addition, we explored whether people's self-reported attitudes and stereotypes about racial groups correlate with their implicit discrimination on the shooter bias task.

2 EXPERIMENT 1

Experiment 1 extended the shooter bias paradigm by Correll et al. [4] to realise a 2 (agent: human vs. robot) × 2 (race/racialization: Black vs. White) × 2 (object in hand: gun vs. benign object) within-subjects design.

2.1 Participants

A total of 192 participants from the USA were recruited from Crowdflower [24], an online platform similar to Amazon's Mechanical Turk. Of these participants, 29 failed a basic attention check question asking them to indicate the response '3' to a specific question, leaving a sample of 163 participants (80 male, 83 female). Participants' age varied from 18 to 73 years ($M = 33.09, SD = 11.48$). A majority of these participants identified as "White or European American" ($N = 126$), with the remaining participants identifying as "Hispanic or Latino American" ($N = 21$), "African American" ($N = 7$), "Asian American" ($N = 3$), "Native American" ($N = 2$), and "Other" ($N = 4$). Participants received $1.50 USD for completing the experiment. In addition, following the reward system of the original study [4], they were informed that the top three scorers in the study would receive a prize bonus of $30, $15, and $10 respectively, and that the top 30% scorers would be able to enter a lucky draw where five winners would receive a bonus of $10 each.

[1] Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018).
Robots and Racism. Proceedings of the ACM/IEEE International Conference on Human Robot Interaction, Chicago pp. 196-204. | DOI: 10.1145/3171221.3171260

2.2 Stimuli

We took photos of the Nao robot holding a gun, a remote control, a candy bar and a soda can, and did so against a green backdrop so that we could easily separate Nao and Photoshop it into the eight backgrounds from the original experiment [4]. To increase the realism of the scene, we edited the image so that Nao cast a shadow in the background (see Figure 1). To give the Nao a racialized appearance (as Black and as White), we calibrated its color based on sampling the skin tones in a professional photograph of a Black woman and a White woman (see Figure 2). We manipulated the Nao in this way to match the stimuli (wherein melanation was the most salient identifier of the Black versus White human exemplars) of the original experiment (see Figure 1). In total, we had 64 different pictures consisting of 8 backgrounds × 2 racializations × 4 objects (one gun and three benign object). We used the exact same size of the pictures as in the original study. In addition we had the 64 original images adding up to a total of 128 stimuli.

Figure 2: The professional photograph used for the calibration of the NAO's racialization. We based the racialization on the skin tones of the Black (left) and White (right) women.

To ensure that participants did not ascribe a race to the robots because they were completed the shooter bias task and the self-report measures about racial attitudes and stereotypes, we conducted another study recruiting a separate sample ($N = 53$) from Crowdflower. Participants in this new study were only asked to define the race of the robot with several options including "Does not apply". Data revealed that only 11.3% of participants and 7.5% of participants selected "Does not apply" for the black and white racialized robots, respectively. By comparison, 52.8% of participants indicated that the robot racialized as Black was indeed perceived as Black, while 67.9% indicated that the robot racialized as White was indeed perceived as White. These results suggest that robots may be ascribed a race even when no questions about race are included prior to making such judgements.

2.3 Procedure

Participants were recruited using CrowdFlower[1] where they were first provided with an information sheet which they could open and read, before being asked to give consent for their participation in the study. Participants then completed the self-report measures outlined below before being directed to a web link of Inquisit Web[2] which allowed us to record latency and error rates with millisecond precision.

[1]https://www.crowdflower.com/
[2]http://www.millisecond.com/

On the Inquisit Web Player, participants were informed that they would be flashed several images of either humans or robots holding guns, and they were to shoot the ones holding a gun using either the 'A' or 'L' key (counterbalanced based on subject numbers), and to choose not to shoot the ones holding other objects using the other keys provided. Instructions on the monetary bonus rewards were also flashed again to reinforce their motivation to try their best.

When participants were ready, they could move on to the training session. This comprised of 20 practice trials of random conditions. Each trial included a fixation for 500ms, flashing a random number of empty backgrounds with replacement (between 1 to 4) for a random amount of duration (between 500 to 1000ms), followed by the image for a duration of 850ms. This was to ensure participants do not sink into a routine. Participants then received feedback (3000ms) on each trial as to how they performed.

After the practice trials, they were allowed to rest, and to continue to the main study when they were ready. The main study was comprised of 128 trials (16 backgrounds * 8 conditions) split into two blocks, where participants could take a break after 64 trials. It is important to note that the 4 objects held in the hand was folded into just two conditions: gun and no-gun.

2.4 Measures

2.4.1 Demographics. Participants completed a demographic questionnaire including questions about their age, race, gender, nationality, and political ideology. In addition, participants also completed an attention check question ("This is an attention check. Please check 3.") embedded in the middle of the questionnaire, as well as several self-report measures as detailed below.

2.4.2 Robot Race. As a manipulation check, participants were shown two images of robots racialized as Black or White and were asked to indicate whether they thought the robot had a race. They also had the option to tick "Does not apply".

2.4.3 Attitudes toward Black and White Americans. Participants completed a measure of their personal attitudes toward Black Americans and White Americans using a feeling thermometer where 0=Very Cold and 100=Very Warm [2]. The scale was presented as a 11-point Likert scale varying in 10 degree increments (i.e., 0=0, 1=10, 2=20, 3=30 . . . 10=100).

2.4.4 Personal Stereotypes about Black and White Americans. In line with [5], participants were asked to report the extent to which they personally felt that Black Americans and White Americans were aggressive and dangerous using a 10-point Likert scale (1=Not at all; 11=Very Much). These items showed high internal consistency and were therefore collapsed to form a single index of stereotyping ($\alpha_{BlackAmericans} = .90$; $\alpha_{WhiteAmericans} = .87$).

2.4.5 Cultural Stereotypes. Likewise, based on previous research [5], participants reported the extent to which they thought other Americans believed that Black and White Americans were aggressive and dangerous using a 11-point Likert scale (1=Not at all; 11=Very Much). Once again these items showed high internal consistency and were therefore collapsed into a single index of cultural stereotypes ($\alpha_{BlackAmericans} = 0.91$; $\alpha_{WhiteAmericans} = 0.88$).

[1] Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018).
Robots and Racism. Proceedings of the ACM/IEEE International Conference on Human Robot Interaction, Chicago pp. 196-204. | DOI: 10.1145/3171221.3171260

2.4.6 Shooter Bias. We assessed both participants' reaction times (RT) to decide whether to shoot/not shoot and accuracy (correct identification of aggressors versus non-aggressors) while completing the shooter bias task. The reaction time is measured as the time between the end of the stimuli being shown on the screen and the time when a key was pressed. Following the procedures outlined in [4], we calculated the average latency and error for different types of trials for the variables of: agent (human vs. robot), race/racialization (Black vs. White), and object in hand (gun vs. benign object).

2.5 Procedure

Participants were recruited using CrowdFlower[3] where they were first provided with an information sheet which they could open and read, before being asked to give consent for their participation in the study. Participants then completed the self-report measures outlined above before being directed to a web link of Inquisit Web[4] which allowed us to record latency and error rates with millisecond precision.

2.6 Results

Following the procedures outlined in [4], we calculated the average latency for different types of trials for the variables of: agent (human vs. robot), racialization (Black vs. White), and object in hand (gun vs. benign object) (see 1).

2.6.1 Manipulation Check: Attribution of Race to Robots. The vast majority of participants associated a race to the robot (see Table 2). Only 11% of the participants selected the "Does Not Apply" option. The robot racialized as Black was otherwise perceived to be Black while the robot racialized as White was perceived to be White. We can therefore conclude that our manipulation of stylization did serve to alter the race of the robot in the eyes of most participants.

2.6.2 Explicit Bias. Participants showed significantly more positive attitudes toward White Americans ($M = 7.41; SD = 2.09$) than Black Americans ($M = 7.01; SD = 2.24$), $t(162) = -2.16, p = 0.03, d = 0.18$. Similarly, participants reported that others would have stronger stereotypes about Black Americans ($M = 6.15; SD = 2.19$) as aggressive and dangerous relative to White Americans ($M = 4.59; SD = 2.16$), $t(162) = -7.12, p < 0.001, d = 0.72$. No significant differences, however, were found in participants' personal stereotypes of Black Americans ($M = 5.03; SD = 2.05$) and White Americans ($M = 4.83; SD = 2.12$), $t(162) = -1.13, p = 0.26, d = 0.10$.

2.6.3 Shooter Bias. A $2 \times 2 \times 2$ within-subjects analysis of variance (ANOVA) on latency revealed a significant 2-way interaction between racialization $\times$ object in hand, ($F(1, 162) = 31.31, p < 0.001, \eta^2 = 0.16$). Paired sample t-tests revealed that participants were quicker to shoot an armed Black agent ($M = 601.22; SD = 68.89$) than an armed White agent ($M = 605.88; SD = 71.18$), ($t(162) = -2.27, p = 0.02, d = 0.08$), and simultaneously faster to not shoot at an unarmed White agent ($M = 655.15; SD = 66.81$) than an unarmed Black agent ($M = 666.77; SD = 68.13$), ($t(162) = -5.98, p < 0.001, d = 0.17$). The means and standard deviations for all conditions are shown in Table 1. Additionally, there was a significant 2-way interaction between agent and object, ($F(1, 162) = 12.44, p < 0.01, \eta^2 = 0.07$). Decomposing this interaction further, it appeared that participants were significantly quicker to shoot an armed human ($M = 600.94; SD = 70.91$) than an armed robot, ($M = 606.15; SD = 70.07$), ($t(162) = -2.17, p = 0.03, d = 0.07$), and also significantly slower to refrain from shooting an unarmed human ($M = 663.15; SD = 69.00$) than unarmed robot, ($M = 658.06; SD = 66.67$), ($t(162) = 2.59, p = 0.01, d = 0.08$). All other interactions were non-significant.

We also conducted an ANOVA on accuracy rates. Once again, there was a significant 2-way interaction between racialization $\times$ object in hand, ($F(1, 162) = 4.32, p < 0.04, \eta^2 = 0.03$). Paired sample t-tests revealed that participants were marginally more accurate when a White agent had a gun ($M = 0.866; SD = 0.195$) compared to when a Black agent had a gun ($M = 0.856; SD = 0.188$), ($t(162) = 1.78, p = 0.077, d = 0.08$). Additionally, there was a significant 2-way interaction between agent and object in hand ($F(1, 162) = 55.90, p < 0.001, \eta^2 = 0.26$). Examining this interaction further, data revealed that participants were significantly more accurate when deciding to shoot armed robots ($M = 0.878; SD = 0.199$) over armed humans ($M = 0.845; SD = 0.186$), ($t(162) = -4.992, p < 0.01, d = 0.17$). Participants were also significantly more accurate at determining not to shoot at unarmed humans ($M = 0.819; SD = 66.67$) over unarmed robots ($M = 0.768; SD = 0.223$), ($t(162) = 6.48, p < 0.01, d = 0.25$). All other interactions were non-significant.

2.6.4 Relationship between Shooter Bias and Explicit Bias. To explore the relationship between indicators of implicit (shooter) and explicit anti-Black bias, we calculated a set of difference scores regarding the latencies derived from the shooter bias paradigm. This score was correlated with the difference of the average scores related to personal and cultural stereotypes about Whites subtracted from mean scores regarding Blacks, respectively.

Furthermore, we calculated an attitude difference score, subtracting personal attitudes towards Whites from personal attitudes towards Blacks. This was then correlated with the measure of implicit bias (the latency difference score) as well. Pearson correlation analyses revealed no significant statistical relationships between the level of personal stereotypes towards Blacks relative to Whites and implicit bias, ($r(163) = -0.05, p = 0.56$). Similarly, implicit bias was not correlated with the cultural stereotypes difference score, ($r(163) = -0.05, p = 0.50$). Personal attitudes were also not significantly correlated with implicit bias, ($r(163) = -0.03, p = 0.73$).

3 EXPERIMENT 2

Experiment 1 gave participants sufficient latency and thereby assumed that any shooter bias would reveal itself in latency differences as opposed to error rates since participants would make few mistakes. However, in subsequent work, [4] reduced the latency to better examine error rates as a potential index of shooter bias. Therefore in Experiment 2, we decided to make the task considerable harder by shortening the amount of time during which the participants were expected to respond. To be in line with the original work by [4], we reduced the response window from 850ms to

[3] https://www.crowdflower.com/
[4] http://www.millisecond.com/

[1] Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018).
Robots and Racism. Proceedings of the ACM/IEEE International Conference on Human Robot Interaction, Chicago pp. 196-204. | DOI: 10.1145/3171221.3171260

Table 1: Means and Standard Deviations for Reaction Times and Accuracy across all conditions

		Study 1: 850ms				Study 2: 630ms			
		Reaction Times		Accuracy Rate		Reaction Times		Accuracy Rate	
		Black	White	Black	White	Black	White	Black	White
Armed agent	Human	598 (74)	604 (74)	0.839 (0.189)	0.851 (0.199)	487 (75)	492 (77)	0.709 (0.181)	0.703 (0.180)
	Robot	605 (71)	608 (73)	0.873 (0.205)	0.882 (0.204)	493 (89)	492 (83)	0.682 (0.202)	0.693 (0.204)
Unarmed agent	Human	670(71)	658 (71)	0.821 (0.227)	0.817 (0.220)	524 (85)	515 (91)	0.493 (0.255)	0.513 (0.226)
	Robot	664 (70)	652 (67)	0.774 (0.234)	0.763 (0.227)	522 (84)	516 (86)	0.461 (0.221)	0.467 (0.216)

Table 2: Results of the manipulation check of experiment 1: attribution of racial identity to the robot racialized as Black and as White.

	Black		White	
	Count	%	Count	%
Does Not Apply	17	10.43	18	11.04
Black	**114**	**69.94**	0	0.00
White	2	1.23	**108**	**66.26**
Latino	21	12.88	15	9.20
Native	2	1.23	6	3.68
East Asian	1	0.61	9	5.52
Indian	5	3.07	2	1.23
Arab	0	0	3	1.84
Pacific	1	1.23	2	1.23

630ms, but maintained the same exact procedures and measures as before.

3.1 Participants

Similar to Experiment 1, we recruited a total of 172 participants from the USA using Crowdflower. Of these participants, 93 identified as male, 73 as female, and 6 as "Other". A total of 116 identified as "White, Caucasian, or European American", while 19 identified as "Black or African American", 18 as "Hispanic or Latino American", 9 as "Asian American", and 8 as "Other" (7 chose not to identify with any group). Participants ranged in age from 18 to 67 years ($M = 32.94; SD = 11.67$).

3.2 Measures

All measures were identical to that used in Experiment 1. The only difference was that instead of giving participants a latency of 850 ms, they had only 630 ms to respond in the shooting paradigm. This specific response window was chosen to mirror the original work by [4] in order to increase the number of error rates participants would likely make during the task.

Note that similar to Experiment 1, there was high internal consistency between ratings of personal stereotypes for both Black Americans and White Americans allowing for these to form a indexes of personal stereotypes for both groups ($\alpha_{B}lackAmericans = 0.86$ for and $\alpha_{W}hiteAmericans = 0.84$). There was also high internal consistency between ratings of perceived stereotypes of others about both Black Americans and White Americans allowing for these to be averaged to form indexes of cultural stereotypes for both groups ($\alpha_{B}lackAmericans = 0.89$ and $\alpha_{W}hiteAmericans = 0.87$).

3.3 Results

3.3.1 Manipulation Check: Attribution of Race to Robots. Similar to Experiment 1, our manipulation check revealed that participants racialized the two robots even when provided with the option to select "Does not apply" at the top of the list (see Table 3).

Table 3: Results of the manipulation check of Experiment 2: Attribution of racial identity to the robot racialized as Black and as White.*Note that total percentages are based on total number of completed responses to the specific question. A total of 5-6 participants did not respond to the specific questions.*

	Black		White	
	Count	%	Count	%
Does Not Apply	21	16.7	25	20.0
Black	**80**	**63.5**	2	1.6
White	4	3.2	**80**	**64.0**
Latino	16	12.7	9	7.2
Native	0	0	3	2.4
East Asian	1	0.8	4	3.2
Indian	1	0.8	1	0.8
Arab	2	1.6	1	0.8
Pacific	1	0.8	0	0

3.3.2 Explicit Bias. Unlike Experiment 1, no significant differences were found in participants' attitudes toward Black Americans ($M = 7.51; SD = 1.86$), ($t(126) = -2.16, p = 0.03, d = 0.04$) and White Americans ($M = 7.40; SD = 2.03$). There were similarly no significant differences in participants' personal stereotypes of Black Americans ($M = 5.14; SD = 2.01$) and White Americans ($M = 5.10; SD = 2.00$), ($t < 1, p = 0.84, d = 0.01$).

However, similar to Experiment 1, participants reported that others would have stronger stereotypes about Black Americans ($M = 6.41; SD = 2.12$) as aggressive and dangerous relative to White Americans ($M = 4.73; SD = 2.16$), ($t(126) = 7.01, p < 0.001, d = 0.77$).

3.3.3 Shooter Bias. As Experiment 2 utilized a shortened the response window of 630ms, there were a number of trials on which errors were made. While this was expected, some participants made errors on more than three-fourths of the trials indicating that these participants were not carefully attending to the stimulus. This left a sample of 131 participants for the subsequent analyses.

With a shortened response window in Experiment 2, previous work [4] would suggest that shooter bias would more clearly emerge

[1] Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018).
Robots and Racism. Proceedings of the ACM/IEEE International Conference on Human Robot Interaction, Chicago pp. 196-204. | DOI: 10.1145/3171221.3171260

in accuracy rates. To find out, we calculated the average accuracy rate for different types of trials including: agent (human vs. robot), racialization (Black vs. White), and object in hand (gun vs. benign object). A $2\times2\times2$ within-subjects ANOVA comparing mean accuracy rates showed no significant main or interaction effects.

We then also examined mean differences in latency similar to Experiment 1. To do so, we again conducted a $2 \times 2 \times 2$ within-subjects ANOVA. Similar to Experiment 1, there was a significant 2-way interaction between racialization × object in hand, ($F(1, 130) = 9.945, p = 0.002$, $\eta^2 = 0.07$) Paired sample t-tests revealed that participants were faster to refrain from shooting at an unarmed White agent ($M = 515.66; SD = 87.02$) than an unarmed Black agent ($M = 523.05; SD = 83.24$), ($t(130) = 3.720, p < 0.001, d = 0.10$). No other interaction effects were significant.

These results are in contrast with previous work which found no effect of racialization and object on latency with a shortened response window, but found an effect on accuracy rates instead. Here, we instead found a racialization by object interaction on latency, but not on accuracy rates. This may be because the original [4] work had participants come into a lab to complete the study while we ran the study online, but future work is needed to examine this more closely.

3.3.4 Relationship between Shooter Bias and Explicit Bias. As in Experiment 1, we computed Pearson correlation analyses to explore the relationship between indicators of implicit and explicit anti-Black bias using difference scores. These analyses revealed no significant relationship between implicit bias and explicit cultural stereotypes, ($r(126) = 0.01, p = 0.91$), nor between explicit personal attitudes and implicit bias, ($r(126) = -0.03, p = 0.73$). Implicit bias and personal stereotypes, however, were negatively correlated, ($r(126) = -0.21, p = 0.02$).

4 DISCUSSION

4.1 Summary of Findings

The present research examined the role of racialized robots on participants' responses on the shooter bias task, a task widely used in social psychological intergroup research to uncover automatic prejudice towards Black men relative to White men. We conducted two online experiments to replicate and extend the classic study on shooter bias toward Black agents [4]. To do so, we adapted the original research materials by [4] and sought to explore the shooter bias effect in the context of social robots that were racialized either as of Black or White agents. Similar to previous work, we explored the shooter bias using different response windows and focused both on error rates and latencies as indicators of an automatic bias.

4.1.1 Shooter Bias. Experiment 1 revealed that participants were quicker to shoot an armed Black agent than an armed White agent, and simultaneously faster to refrain from shooting an unarmed White agent than an unarmed Black agent regardless of whether it was a human or robot. These findings illustrate the shooter bias toward both human and robot agents. This bias is both a clear indication of racism towards Black people, as well as the automaticity of its extension to robots racialized as Black.

Experiment 2 used a shorter response window in order to examine shooter bias via accuracy instead of latency. Similar to Experiment 1, participants were faster to not shoot at an unarmed White agent than an unarmed Black agent. However, there was no evidence of shooter bias on error rates.

4.1.2 Relationship between Explicit and Implicit Bias. Since the original study [4], subsequent work has shown that the bias appears to be influenced by harmful stereotyping of Black American men as dangerous [3, 5]. It has also been indicated that people with chronic beliefs in interpersonal threat can demonstrate shooter bias toward outgroups even when those groups are not believed to be particularly dangerous (e.g., artificial groupings and Asians in the USA; [22]). The protocol has also been extended to test other forms of racism, such as Islamophobia. For example, Essien et al. [10] found shooter bias toward Muslims with a set of German participants completing the task involving Muslim vs. White German agents. This Islamophobic bias has been replicated with Middle Eastern participants and found to be exacerbated by traditional Muslim clothing [32].

However, correlation analyses revealed no significant relationships between shooter bias and the explicitly measured constructs (personal attitudes, personal stereotypes, and cultural stereotypes). The lack of correlation between implicit and explicit measures might question the construct validity of the shooting bias paradigm. At the same time, it might simply reflect the dissociation between two types of attitudes, one that is directly measured whereas the other is indirectly measured. In that sense, these two distinct attitudes need not necessarily be statistically related. Future research should look into this further to explore whether this is systematic or rather a random null finding.

4.1.3 Other Findings. Thus, further research is necessary, taking into account a variety of improvements. For one, the studies could be conducted in the lab setting and could make use of different experimental stimuli. While the images of the robots we used for these studies were images of humanoid robots, they were also very clearly images of machines. There is a clear sense, then, in which these robots do not have – indeed cannot have – race in the same way had by people. Nevertheless, our studies demonstrated that participants were strongly inclined to attribute race to these robots, as revealed both by their explicit attributions and evidence of shooter bias. The level of agreement amongst participants when it came to their explicit attributions of race was especially striking. Participants were able to easily and confidently identify the race of robots according to their racialization and their performance in the shooter bias task was informed by such social categorization processes. Thus, there is also a clear sense in which these robots – and by extension other humanoid robots – do have race.

4.2 Broader Implications

This result should be troubling for people working in social robotics given the profound lack of diversity in the robots available and under development today [40]. As Riek and Howard [28] have pointed out a Google image search result for "humanoid robots" shows predominantly robots with gleaming white surfaces or that have a metallic appearance (see Figure 3). There are currently very

[1] Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018).
Robots and Racism. Proceedings of the ACM/IEEE International Conference on Human Robot Interaction, Chicago pp. 196-204. | DOI: 10.1145/3171221.3171260

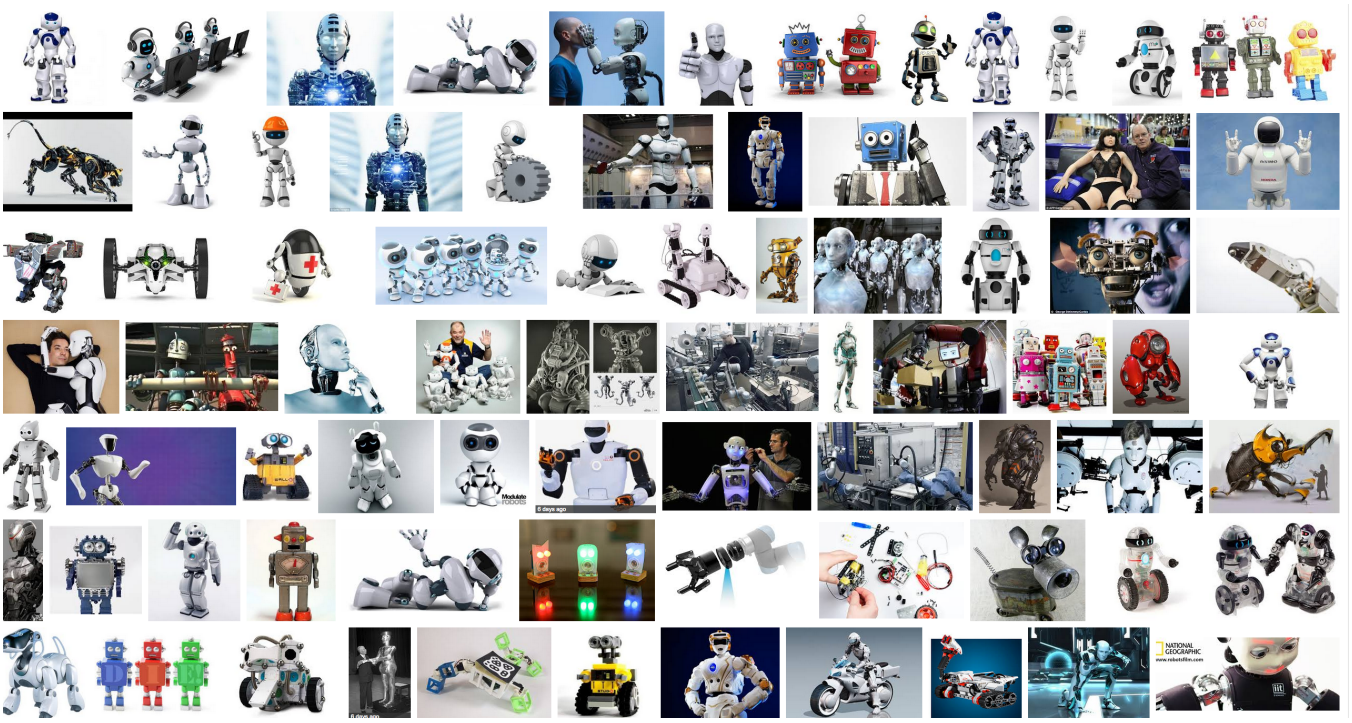

Figure 3: Results of a Google Image Search on the term "Robot".

few humanoid robots that might plausibly be identified as anything other than White or Asian. Most of the main research platforms for social robotics, including Nao, Pepper, and PR2, are stylized with white materials and are presumably White. There are some exceptions to this rule, including some of the robots produced by Hiroshi Ishiguro's team, which are modelled on the faces of particular Japanese individuals and are thereby clearly – if they have race at all – Asian. Another exception is the Bina 48 robot that is racialized as Black (although it is again worth noting that this robot was created to replicate the appearance and mannerisms of a particular individual rather than to serve a more general role). This lack of racial diversity amongst social robots may be anticipated to produce all of the problematic outcomes associated with a lack of racial diversity in other fields. We judge people according to societal stereotypes that are associated with these social categories. Social stereotypes do, for example, play out at times in the form of discrimination [15]. If robots are supposed to function as teachers, friends, or carers, for instance, then it will be a serious problem if all of these roles are only ever occupied by robots that are racialized as White.

A design consideration for the field is how to represent/replicate racial identity. There are, for instance, numerous social contexts in multi-racial societies, wherein a person's race plays a key role in people's attitudes, beliefs, and behaviours toward them. If robots aren't perceived to have race, then, presumably they will be unable to replicate important aspects of human interactions in these contexts. Robots will be unable to serve as certain sorts of role models, be taken seriously on certain topics, or relate effectively to people when issues related to race would otherwise be anticipated to come to the forefront.For these reasons, the question of whether robots have race or not should be of vital concern to those working in the field of social robotics. Existing literature that has explored social categorization effects in the context of HRI, however, is sparse.

In German context, [12], where prejudice against Turkish people is prevalent, German participants have shown less preference for a robot that was introduced as a Turkish product compared to the very same robot that had been presented as an ingroup prototype. In that study, robot group membership was conveyed through the name of the robot and the location of its production - which was manipulated as being German vs. Turkish, respectively.

Recent research by Eyssel and Loughnan [13] has even investigated the effect of visual cues for robot group membership in an experiment in which they manipulated the racialization of a robot. This work revealed that White American participants differentially attributed dimensions of mind perception (i.e., perceived agency and experience) to the robot racialized as Black vs. White. However, this was only the case for participants who scored high in self-reported modern-racist-attitudes. Ingroup bias, that is, favoring the White American ingroup over the outgroup was not demonstrated in the case of robots with a race.

At the very least, our results suggest that people's responses to the race of robots is a topic highly deserving of further study. Furthermore, it might be worthwhile to explore the notion of discrimination towards robots beyond the race category, replicating our findings with robots that may be categorized as Muslim by manipulating the s' headgear (i.e., a turban). Indeed, recent research by [39] revealed that German participant were more inclined to shoot at a Muslim s. Our study is inline with previous research that

[1] Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018).
Robots and Racism. Proceedings of the ACM/IEEE International Conference on Human Robot Interaction, Chicago pp. 196-204. | DOI: 10.1145/3171221.3171260

has shown that people perceive robots as racialized and judge them accordingly [11, 13]. These effects were shown with rather liberal university students, a sample we could also explore in further studies that should ideally be conducted in the laboratory setting. The work by Unkelbach also revealed an effect of participants' mood, with happy people being more likely to shoot Muslim s in particular, while angry participants showed aggressive tendencies across the board. Follow-up studies on shooter bias with a focus on robotic s could also explore the role of mood as a mediating variable to shed more light on the psychological underpinnings of the effect. Similarly, it might be worthwhile to explore as to whether participants with a high proclivity to anthropomorphize nonhuman entities [41] would show a stronger shooter bias towards humanlike robots.

More generally, we believe our findings also make a case for more diversity in the design of social robots so that the impact of this promising technology is not blighted by a racial bias. The development of an Arabic looking robot [20] as well as the significant tradition of designing Asian robots in Japan are encouraging steps in this direction. Especially since these robots were not intentionally designed to increase diversity, but they were the result of a natural design process. Finally, we hope that our paper might serve as a prompt for reflection on the social and historical forces that have brought what is now quite a racially diverse community of engineers to, almost entirely and seemingly without recognising it, design and manufacture robots that, our research suggests, are easily identified by those outside this community as being White.

4.3 Limitations and future work

We may speculate that different levels of anthropomorphism might result in different outcomes. If the robot would be indistinguishable from humans then we would expect to find the same results as the original study while a far more machine like robot might have yet to be determined effects. One may also speculate about the racialization approach we used. To best replicate the original shooter bias stimuli, we opted to utilized human-calibrated racialization of the NAO rather than employ the NAO's default appearance (white plastic) against it stylized with black materials. We are currently preparing a study that systematically varies anthropomorphism and color using the methodology described above and we hope to be able to report on it in due time.

It is important to note that the Nao robot did not wear any clothes while the people in the original study did. Strangely, the people in the original study did not cast a shadow. Given the powerful functions of Adobe Photoshop, we were able to include a more realistic montage of the Nao robot in the background by casting shadows. Future studies should include multiple postures of the Nao robot holding the gun and the objects.

A further potential limitation of the present work is the degree of embodiment (image-based depictions of the NAO), and whether this extends to more realistic human-robot interaction scenarios. However, previous work has suggested that interactions with embodied robots are not fundamentally different from interactions with virtual ones. For instance, several studies found no difference in behavioral and attitudinal responses for virtual agents and physical robots [25, 26, 38, 42]. Rosenthal-von der Pütten et al. [31] and Slater et al. [34], indicated that virtual robots, too, are recognised as social agents. The images used in this study therefore are unlikely to elicit fundamentally different responses compared to using actual robots. By replicating Correll's original study that used still images we are able to compare our results and thereby ensure that we receive similar results. This chance to validate our results outweighed the benefits of using a physical robot. Moreover, if we had used a physical robot it would also be consistent to ask the participants to physically shoot the robot. Using a keyboard to symbolically shoot a physical robot would not be plausible. Having participants use a real gun to shoot a robot would, however, be a far more dramatic task than the one proposed by Correll and colleagues and may yield very different results.

We also need to acknowledge that we used categories in the race attribution question that experts consider to be inconsistent. More specifically we mixed categories of ethnicity and race. While we believe that most people would be able to understand the meaning of these categories, we cannot deny that the categories offered could be considered as inconsistent. We reported the categories as we had asked them. For us, the main question was if the participants choose anything but the "Does not apply" option. We therefore believe that the information obtained still has value.

5 ACKNOWLEDGEMENT

We would like to thank Megan Strait who shepherded this paper and provided in depth feedback and suggestions.

REFERENCES

[1] Jasmin Bernotat, Friederike Eyssel, and Janik Sachse. 2017. *Shape It – The Influence of Robot Body Shape on Gender Perception in Robots.* Springer International Publishing, Cham, 75–84. https://doi.org/10.1007/978-3-319-70022-9_8
[2] Jean M Converse and Stanley Presser. 1986. *Survey questions: Handcrafting the standardized questionnaire.* Vol. 63. Sage, Thousand Oaks, California.
[3] Joshua Correll, Sean M Hudson, Steffanie Guillermo, and Debbie S Ma. 2014. The police officer's dilemma: A decade of research on racial bias in the decision to shoot. *Social and Personality Psychology Compass* 8, 5 (2014), 201–213. https://doi.org/10.1111/spc3.12099
[4] Joshua Correll, Bernadette Park, Charles M. Judd, and Bernd Wittenbrink. 2002. The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology* 83, 6 (2002), 1314–1329. https://doi.org/10.1037//0022-3514.83.6.1314
[5] Joshua Correll, Geoffrey R Urland, and Tiffany A Ito. 2006. Event-related potentials and the decision to shoot: The role of threat perception and cognitive control. *Journal of Experimental Social Psychology* 42, 1 (2006), 120–128. https://doi.org/10.1016/j.jesp.2005.02.006
[6] Nilanjana Dasgupta. 2004. Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research* 17, 2 (2004), 143–169. https://doi.org/10.1023/B:SORE.0000027407.70241.15
[7] Nilanjana Dasgupta and Luis M Rivera. 2006. From automatic antigay prejudice to behavior: the moderating role of conscious beliefs about gender and behavioral control. *Journal of personality and social psychology* 91, 2 (2006), 268–280. https://doi.org/10.1037/0022-3514.91.2.268
[8] Thierry Devos and Debbie S Ma. 2013. How "American" is Barack Obama? The role of national identity in a historic bid for the White House. *Journal of Applied Social Psychology* 43, 1 (2013), 214–226. https://doi.org/10.1111/jasp.12069
[9] John F Dovidio, Kerry Kawakami, Craig Johnson, Brenda Johnson, and Adaiah Howard. 1997. On the nature of prejudice: Automatic and controlled processes. *Journal of experimental social psychology* 33, 5 (1997), 510–540. https://doi.org/10.1006/jesp.1997.1331
[10] Iniobong Essien, Marleen Stelter, Felix Kalbe, Andreas Koehler, Jana Mangels, and Stefanie Meliß. 2017. The shooter bias: Replicating the classic effect and introducing a novel paradigm. *Journal of Experimental Social Psychology* 70 (2017), 41–47. https://doi.org/10.1016/j.jesp.2016.12.009
[11] Friederike Eyssel and Frank Hegel. 2012. (S)he's got the look: Gender-stereotyping of social robots. *Journal of Applied Social Psychology* 42, 9 (2012), 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

[1] Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018).
Robots and Racism. Proceedings of the ACM/IEEE International Conference on Human Robot Interaction, Chicago pp. 196-204. | DOI: 10.1145/3171221.3171260

[12] Friederike Eyssel and Dieta Kuchenbrandt. 2012. Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology* 51 (2012), 724–731. https://doi.org/10.1111/j.2044-8309.2011.02082.x

[13] Friederike Eyssel and Steve Loughnan. 2013. 'It don't matter if you're Black or White'? Effects of robot appearance and user prejudice on evaluations of a newly developed robot companion. *Lecture Notes in Computer Science* 8239 (2013), 422–431. https://doi.org/10.1007/978-3-319-02675-6_42

[14] Russell H Fazio, Joni R Jackson, Bridget C Dunton, and Carol J Williams. 1995. Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of personality and social psychology* 69, 6 (1995), 1013–1027. https://doi.org/10.1037/0022-3514.69.6.1013

[15] Susan T Fiske. 1998. *Stereotyping, prejudice, and discrimination.* McGraw Hill, Boston, MA, 357–411.

[16] Alexander R Green, Dana R Carney, Daniel J Pallin, Long H Ngo, Kristal L Raymond, Lisa I Iezzoni, and Mahzarin R Banaji. 2007. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of general internal medicine* 22, 9 (2007), 1231–1238. https://doi.org/10.1007/s11606-007-0258-5

[17] Anthony G Greenwald and Mahzarin R Banaji. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review* 102, 1 (1995), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

[18] Anthony G Greenwald, Colin Tucker Smith, N Sriram, Yoav Bar-Anan, and Brian A Nosek. 2009. Implicit race attitudes predicted vote in the 2008 US presidential election. *Analyses of Social Issues and Public Policy* 9, 1 (2009), 241–253. https://doi.org/10.1111/j.1530-2415.2009.01195.x

[19] Jerry Kang, Nilanjana Dasgupta, Kumar Yogeeswaran, and Gary Blasi. 2010. Are ideal litigators white? Measuring the myth of colorblindness. *Journal of Empirical Legal Studies* 7, 4 (2010), 886–915. https://doi.org/10.1111/j.1740-1461.2010.01199.x

[20] N. Mavridis, A. AlDhaheri, L. AlDhaheri, M. Khanii, and N. AlDarmaki. 2011. Transforming IbnSina into an advanced multilingual interactive android robot. In *2011 IEEE GCC Conference and Exhibition (GCC)*. IEEE, New York, NY, USA, 120–123. https://doi.org/10.1109/IEEEGCC.2011.5752467

[21] Yara Mekawi and Konrad Bresin. 2015. Is the evidence from racial bias shooting task studies a smoking gun? Results from a meta-analysis. *Journal of Experimental Social Psychology* 61, November 2015 (2015), 120–130. https://doi.org/10.1016/j.jesp.2015.08.002

[22] Saul L Miller, Kate Zielaskowski, and E Ashby Plant. 2012. The basis of shooter biases: Beyond cultural stereotypes. *Personality and Social Psychology Bulletin* 38, 10 (2012), 1358–1366. https://doi.org/10.1177/0146167212450516

[23] Jahna Otterbacher and Michael Talias. 2017. S/He's Too Warm/Agentic!: The Influence of Gender on Uncanny Reactions to Robots. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. ACM, New York, NY, USA, 214–223. https://doi.org/10.1145/2909824.3020220

[24] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70, May 2017 (2017), 153 – 163. https://doi.org/10.1016/j.jesp.2017.01.006

[25] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. 2007. Comparing a computer agent with a humanoid robot. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*. ACM/IEEE, Arlington, USA, 145–152. https://doi.org/10.1145/1228716.1228736

[26] Juliane Reichenbach, Christoph Bartneck, and Julie Carpenter. 2006. Well done, Robot! The importance of praise and presence in human-robot collaboration. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*. IEEE, Hatfield, UK, 86–90. https://doi.org/10.1109/ROMAN.2006.314399

[27] Jennifer A Richeson, Abigail A Baird, Heather L Gordon, Todd F Heatherton, Carrie L Wyland, Sophie Trawalter, and J Nicole Shelton. 2003. An fMRI investigation of the impact of interracial contact on executive function. *Nature neuroscience* 6, 12 (2003), 1323–1328. https://doi.org/10.1038/nn1156

[28] Laurel D Riek and Don Howard. 2014. A code of ethics for the human-robot interaction profession. In *Proceedings of We Robot*. Social Science Research Network. https://ssrn.com/abstract=2757805

[29] Michael Root. 2000. How We Divide the World. *Philosophy of Science* 67 (2000), S628–S639. https://doi.org/10.1086/392851

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

[31] Astrid M Rosenthal-von der Pütten, Nicole C Krämer, Laura Hoffmann, Sabrina Sobieraj, and Sabrina C Eimler. 2013. An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics* 5, 1 (2013), 17–34. https://doi.org/10.1007/s12369-012-0173-8

[32] Timothy P Schofield, Timothy Deckman, Christopher P Garris, C Nathan DeWall, and Thomas F Denson. 2015. Brief report: Evidence of ingroup bias on the shooter task in a Saudi sample. *SAGE Open* 5, 1 (2015), 1–6. https://doi.org/10.1177/2158244015576057

[33] M. Siegel, C. Breazeal, and M. I. Norton. 2009. Persuasive Robotics: The influence of robot gender on human behavior. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, New York, 2563–2568. https://doi.org/10.1109/IROS.2009.5354116

[34] Mel Slater, Angus Antley, Adam Davison, David Swapp, Christoph Guger, Chris Barker, Nancy Pistrang, and Maria V Sanchez-Vives. 2006. A virtual reprise of the Stanley Milgram obedience experiments. *PloS one* 1, 1 (2006), e39. https://doi.org/10.1371/journal.pone.0000039

[35] Robert Sparrow. 2017. Robots, Rape, and Representation. *International Journal of Social Robotics* 9, 4 (2017), 465–477. https://doi.org/10.1007/s12369-017-0413-z

[36] Damian A Stanley, Peter Sokol-Hessner, Mahzarin R Banaji, and Elizabeth A Phelps. 2011. Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences* 108, 19 (2011), 7710–7715. https://doi.org/10.1073/pnas.1014345108

[37] Paul C Taylor. 2000. Appiah's uncompleted argument: WEB Du Bois and the reality of race. *Social Theory and Practice* 26, 1 (2000), 103–128.

[38] Sam Thellman, Annika Silvervarg, Agneta Gulz, and Tom Ziemke. 2016. Physical vs. Virtual Agent Embodiment and Effects on Social Interaction. In *Intelligent Virtual Agents: 16th International Conference, IVA 2016*. Springer International Publishing, Los Angeles, USA, 412–415. https://doi.org/10.1007/978-3-319-47665-0_44

[39] Christian Unkelbach, Joseph P. Forgas, and F. Denson, Thomas. 2008. The turban effect: The influence of Muslim headgear and induced affect on aggressive responses in the shooter bias paradigm. *Journal of Experimental Social Psychology* 44 (2008), 1409–1413. https://doi.org/10.1016/j.jesp

[40] Astrid Marieke von der Pütten and Nicole C. Krämer. 2012. A Survey on Robot Appearances. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '12)*. ACM, New York, NY, USA, 267–268. https://doi.org/10.1145/2157689.2157787

[41] Adam Waytz, John Cacioppo, and Nicholas Epley. 2014. Who Sees Human? The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives in Psychological Science* 5, 3 (2014), 219–232. https://doi.org/10.1177/1745691610369336

[42] Ricarda Wullenkord, Marlena R Fraune, Friederike Eyssel, and Selma Šabanović. 2016. Getting in Touch: How imagined, actual, and physical contact affect evaluations of robots. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, New York, USA, 980–985. https://doi.org/10.1109/ROMAN.2016.7745228

[43] Kumar Yogeeswaran and Nilanjana Dasgupta. 2010. Will the "real" American please stand up? The effect of implicit national prototypes on discriminatory behavior and judgments. *Personality and Social Psychology Bulletin* 36, 10 (2010), 1332–1345. https://doi.org/10.1177/0146167210380928

[44] Kumar Yogeeswaran, Thierry Devos, and Kyle Nash. 2016. *Understanding the Nature, Measurement, and Utility of Implicit Intergroup Biases*. Cambridge University Press, Cambridge, 241–266. https://doi.org/10.1017/9781316161579.011