**Supplemental Materials**

**Racial Bias in Judgments of Physical Size and Formidability: From Size to Threat**

**by J. P. Wilson et al., 2017, *Journal of Personality and Social Psychology***

**http://dx.doi.org/10.1037/pspi0000092**

This supplement contains two sets of additional materials that contribute to the overall body of knowledge surrounding the present work that we did not include in the manuscript for brevity. First, we describe a brief meta-analysis of gender differences in the race biases that we observed across all studies from the main text. Second, we briefly summarize three additional studies excluded from the manuscript for the reasons noted below.

**Meta-Analysis of Gender Differences Across Studies**

We conducted a meta-analysis to explore the possibility that perceivers’ gender would impact their judgments about targets’ size and formidability. We proposed that such a gender difference might occur because of previous findings showing that men tend to be more competitive in intergroup situations (McDonald, Navarrete, & Van Vugt, 2012; Van Vugt, De Cremer, & Janssen, 2007). Men, then, should attune more to features of the environment that are relevant to conflict, including the size of potential adversaries. Specifically, we hypothesized that the race bias that we observed in our studies would be larger for men than women.

We tested this by aggregating the point estimates for the two-way interaction between participant gender and target race. The results showed a small but significant overall effect that confirmed our hypothesis: Men showed more race bias than women across all studies, yielding a mean weighted effect size for the interaction of *r* = .05, 95% CI [.003, .11], *p* = .04. However, we also observed significant heterogeneity among the effect sizes (*Q* = 78.90, *p* < .001), possibly because some studies involved judgments directly related to physical conflict whereas other studies involved more direct assessments of physical size (see Table S1).

Because the male warrior hypothesis (McDonald, Navarrette, & Van Vugt, 2012) primarily applies to situations of potential conflict, we reasoned that gender differences might be most likely to appear in the studies measuring inferences about conflict. That is, we might not expect this gender difference to arise in studies where perceivers merely judge another person’s physical size. Attuning to conflict, however (as in judgments that involve assessing a target’s ability to cause physical harm), might lead to a larger bias among male perceivers. We thus coded each study as to whether it measured physical conflict, reporting weighted meta-analytic effect sizes for each type of study (conflict vs. size).

This analysis showed that gender strongly moderated the (Black > White) race bias for conflict-related variables, *r* = .24, 95% CI [.13, .30], *p* < .001, but not for size-related variables, *r* = -.05, 95% CI [- 12, .02], *p* = .14. We acknowledge that this assessment is post hoc and merely exploratory, and that there are other potential explanations for these findings. For example, women may be inferring that virtually all male targets are capable of physically harming them. Somewhat consistent with this possibility, inspection of the descriptive statistics (Table S2) shows that female participants tended to rate both groups of targets as more capable of harm than male participants, but that this relative difference was smaller for Black targets than White targets. Thus, reduced race bias tended to be driven more by elevated judgments of White targets’ harm potential than by reduced judgments of Black targets’ harm potential (e.g., Studies 2, 3, and 4). This view of the data casts some doubt on the male warrior hypothesis as the sole mechanism of the gender difference. However, these descriptive statistics also show that we did not observe ceiling effects in our data. For example, female participants’ mean harm capability ratings never exceeded 4.83 (on a 7-point scale) for any target group. We also point out that we observed a strong gender difference in Study 5, in which participants assessed force justification and did not implicitly consider their own physical capabilities (as in Study 2, for example). In this study, women were only slightly less likely than men to see force as justified for White targets, but they were much less likely than men to see force as justified for Black targets. As such, we see mixed evidence for the male warrior hypothesis and grant the possibility that the gender difference may not be theoretically meaningful. We ultimately see this as an open question and, accordingly, will seek to confirm this pattern in future work.

**Additional Correlations**

In Study 6, we reported the correlations between all of the measures of formidability. We also reported that target race did not qualify the relationship between Afrocentricity and the various measures of formidability. However, it still could be of interest whether there were descriptively different relationships between these variables. As such, in Tables S3 and S4 we report these correlations separately for White and Black targets. As the reader can see, these relationships were often significant for Black targets but not White targets. However, the difference between these Afrocentricity-formidability correlations based on race never achieved significance. These descriptive differences do suggest that further research should determine whether the role of Afrocentricity in informing formidability judgments is indeed stronger for Black participants. This would be a sensible finding.

**Additional Studies**

We conducted four additional studies not reported in the main text. We excluded these studies from the manuscript for various reasons and include them here for completeness and transparency. We excluded Study S1 in part because the target sample size was small and appears to have been underpowered to detect effects using the cross-classified analyses employed elsewhere in the main text. Similarly, we excluded Study S3 because we had used only a single race-ambiguous target. We excluded Study S2 because it was redundant with Study 3 but did not include the measures of prejudice. We excluded Study S4 because we decided that it was more appropriate to focus on force justification than force decisions, given that a decision to use force is quite far removed from the experimental context that we created here. In all four excluded studies, the results either directionally (Study S1, S3), marginally (Study S1), or significantly (Study S2, Study S4) supported our hypotheses.

**Study S1.** To test the hypothesis that people generally perceive Black men as larger than White men, we asked participants to estimate the height and weight of a series of standardized Black and White men’s faces in Study S1.

***Method.***

*Participants.*We recruited 60 US residents (33 male, 27 female; *M*age= 31.6 years, *SD* = 9.7) from Amazon’s Mechanical Turk (MTurk) for a study on person perception. We excluded three Black participants, leaving 57 participants in total. After providing informed consent, the participants learned that they would be asked to view a series of faces for whom they would guess each person’s height or weight.

*Stimuli.*We presented participants with color photographs of 42 male faces (21 White, 21 Black) from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015), all of whom exhibited a neutral expression. We standardized the faces for interpupillary distance using Psychomorph (Tiddeman, Burt, & Perrett, 2001) and sized each image to 640 × 480 pixels (240 pixels/inch).

*Procedure.* The participants viewed the images in blocks, such that they estimated all 42 targets’ height or weight (order counterbalanced between participants) individually in random order before moving on to the next judgment. The target image appeared above a slider scale for each height (weight) rating. The scale for the weight ratings ranged from 120 to 300 lbs, with the possible responses in increments of 1 lb. The scale for the height ratings ranged from 60 in. (5 ft 0 in.) to 78 in. (6 ft 6 in.), with the possible responses in increments of 1 in. We selected these ranges to include plausible values for men of average size (McDowell, Fryar, Ogden, & Flegal, 2008).

***Results and discussion*.** Participants estimated Black targets (*M* Height = 70.37 in., *SD* = 1.55; *M* Weight = 182.08 lbs, *SD* = 13.89) to be only descriptively heavier and taller than White targets (*M* Height = 70.10 in., *SD* = 1.47; *M* Weight = 176.53 lbs, *SD* = 13.80). The effect of race on weight judgments was non-significant, *B* = 5.50, *SE* = 3.52, 95% CI [-1.61, 12.60], *t*(41.77) = 1.56, *p* = .13, *d* = 0.48, as was the effect of race on height judgments, *B* = 0.29, *SE* = 0.40, 95% CI [-0.53, 1.10], *t*(45.29) = 0.71, *p* = .48, *d* = 0.21. One clear weakness of this study is that we do not know the actual height or weight of the targets. Thus, the Black targets in this stimulus set might have actually been taller and heavier than the White targets and the participants simply perceived their size accurately (see Burton & Rule, 2013; Coetzee, Chen, Perrett, & Stephen, 2010). Another weakness is the low number of targets. One difference between this study and most other studies reported in the main text is that we used more stimuli in those studies (see Westfall, Kenny, & Judd, 2014, for a discussion of stimulus power). Furthermore, although demographic data suggest that Black and White American men are quite similar in size, on average (Fryar, Gu, & Ogden, 2012), we conducted additional studies using more targets for whom we knew the height and weight (reported in the main text). This latter point is critical – although we can speculate that this study was likely underpowered, a more substantial unknown aspect is whether we succeeded in using targets of similar size between the two groups.

**Study S2.** We conducted Study S2 as an initial attempt to investigate the relationship between muscularity bias and harm capability bias. Because we did not include a measure of explicit prejudice in Study S2, however, we replicated it including such a measure in Study 3, which then replaced this original version.

***Method*.**

*Participants***.** We recruited 95 US residents from MTurk for a study on person perception. We excluded three Black participants’ data, leaving 92 participants in total (48 male, 44 female; *M*age= 34.9 years, *SD* = 11.3).

*Stimuli and Procedure.* We used the 90 athlete faces described in the bulk of the studies in the text. As in Study 3, we asked participants to rate both the muscularity and harm capability of each target in random order, organizing the trials for each judgment into separate counterbalanced blocks.

***Results and discussion.***

*Replication of muscularity and harm capability differences.*We first analyzed participants’ mean scores for the Black and White targets on each judgment to confirm that the bias in muscularity and harm perceptions that we found in Study 1C and Study 2 replicated here, respectively. Indeed, we again observed that participants rated the Black targets (*M* = 3.66, *SD* = 0.74) as more muscular than the White targets (*M* = 3.29, *SD* = 0.62), *B* = 0.41, *SE* = 0.14, 95% CI [0.13, 0.68], *t*(100.78) = 2.99, *p* = .004, *d* = 0.60, and that they rated the Black targets (*M* = 4.67, *SD* = 1.18) as more capable of harm than the White targets (*M* = 4.03, *SD* = 1.26), *B* = 0.67, *SE* = 0.13, 95% CI [0.40, 0.93], *t*(165.47) = 5.00, *p* < .001, *d* = 0.78.

*Relationship between size and threat.*To analyze the relationship between muscularity and harm perceptions, we subtracted each participant’s mean score for the White targets from his or her mean score for the Black targets to create Black-White difference scores in which positive values indicated greater muscularity (harm) for Black over White targets, whereas negative values indicated greater muscularity (harm) for White over Black targets. The targets’ mean muscularity and harm bias difference scores significantly correlated: *r*(88) = .37, *p* < .001. This finding replicated the pattern that we reported in Study 3, increasing our confidence that the tendency to show racial bias in these two measures is strongly related.

**Study S3.** In Study S3, we examined possible top-down effects of race information on formidability estimates prior to conducting the research that we included as Study 7. Although we found effects in the predicted direction, they did not reach conventional standards of statistical significance on any of the measures that we included, likely because of the low power associated with only using one target stimulus per participant. We therefore relegated this study to the supplemental material to reduce the length of the main text.

***Method.***

*Participants.*We recruited 164 US residents from MTurk, excluding nine Black participants for a total of 155 participants (101 male, 54 female; *M*age= 30.9 years, *SD* = 9.6).

*Materials and Procedure*. We asked participants to read a vignette describing an attempted armed robbery at a convenience store. Within this vignette, they viewed an image of the suspect supposedly captured by a closed-circuit video camera in the convenience store in which the crime occurred. This low-resolution photograph depicted the full body of a man wearing a hooded sweatshirt facing away from the camera whose race was not apparent (see Figure S1).

Participants read one of eight individual vignettes, each of which subtly indicated the race of the suspect by referring to him using one of four stereotypically White (Neil, Brett, Brendan, Todd) or Black (Jamal, Rasheed, Tremayne, Kareem) names (Bertrand & Mullainathan, 2003). The participants read only one vignette, after which we asked them to rate the target on all of height, weight, muscularity, strength, and harm capability using the scales employed in Studies 1 and 2 of the main manuscript.

***Results and discussion*.** We subjected each of the five formidability dimensions to independent-samples *t*-tests that compared the mean ratings for each name grouped by race. Although no comparison reached statistical significance, each yielded a small non-significant effect in the predicted direction (see Table S5).

This supplementary study did not provide strong evidence that racially stereotypical names bias estimates of physical size and formidability in a top-down manner. However, we likely did not have enough power to observe an effect (*M* = 28%, *SD* = 10%), as we used a between-subjects design with only one target per participant. We therefore modified our approach to addressing the question in Study 7 by employing 16 targets of each race and a within-subjects design.

**Study S4.** Finally, in Study S4, we asked whether participants would show racial bias in a task very much like Study 5, but in which they were asked to register a hypothetical decision about how to use force. The method of this study was very similar to all of the other studies using the athlete faces, and did not investigate speeded decisions or decisions in an ecologically valid situation. As such, we elected not to report this study in the main text. However, we included it here for the sake of completeness.

***Method.***

*Participants.*We recruited 120 US residents from MTurk for a study on person perception, but an additional three participants completed the study without collecting compensation for a total of 123 participants. We excluded nine Black participants and an additional 20 participants who gave the same response on every trial, however, leaving 96 participants in the final sample (52 male, 44 female; *M*age= 34.6 years, *SD* = 12.0).

*Stimuli and Procedure.* We again used the 90 athlete faces as in the studies above. We asked participants to imagine that they were police officers who were faced with a suspect who was potentially dangerous. We told participants that it was their job to detain each target person, with the goal of avoiding the use of a weapon like a taser or a gun. However, we noted that some people might be more physically difficult to detain and, as such, participants should respond “Yes” or “No” to the question “Would you need to use a weapon to subdue this person?” As mentioned above, some participants provided the same response to every single question and were eliminated from analysis. Stimuli were presented sequentially in random order.

***Results and discussion.***

We tested for potential racial differences in the number of force decisions made by each participant. Here, we simply conducted a paired-samples *t*-test on the mean number of times that each participant chose to use force for each target race. As expected, we found that participants were more likely to use force for Black targets (*M* = 19.73, *SD* = 12.94) than for White targets (*M* = 14.40, *SD* = 8.67), *t*(95) = 4.44, *p* < .011, 95% CI [2.95, 7.72], *d* = 0.45.

*Figure S1.* Target body used in Study S3



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S1 | |  |  | |  |  | | |  | | |  | | |  | | |  | |
| *Parameter Estimates and Effect Size Estimates for Gender Interaction Among Studies in Main Manuscript* | | | | | | | | | | | | | | | | |  | |
| Study (DV) | | *B* | *SE* | | *df* | | *p* | | | *r* | | | Conflict | | |
| 1a (Height) | | -0.30 | 0.37 | | 53 | | .42 | | | -.11 | | | 0 | | |
| 1a (Weight) | | -0.49 | 2.14 | | 44 | | .82 | | | -.03 | | | 0 | | |
| 1b (Height) | | 0.03 | 0.15 | | 28 | | .85 | | | .04 | | | 0 | | |
| 1b (Weight) | | -1.67 | 1.03 | | 28 | | .12 | | | -.29 | | | 0 | | |
| 1c (Muscularity) | | -0.11 | 0.09 | | 53 | | .25 | | | -.16 | | | 0 | | |
| 1d (Strength) | | -0.31 | 0.16 | | 56 | | .05 | | | -.26 | | | 0 | | |
| 1e (Strength) | | -0.07 | 0.16 | | 60 | | .70 | | | -.06 | | | 0 | | |
| 2 (Harm Cap.) | | 0.32 | 0.13 | | 166 | | .01 | | | .19 | | | 1 | | |
| 3 (Muscularity) | | 0.08 | 0.12 | | 107 | | .49 | | | .07 | | | 0 | | |
| 3 (Harm Cap.) | | 0.33 | 0.14 | | 107 | | .02 | | | .22 | | | 1 | | |
| 4 (Muscularity) | | -0.07 | 0.06 | | 237 | | .28 | | | -.07 | | | 0 | | |
| 4 (Harm Cap) | | 0.23 | 0.09 | | 237 | | .01 | | | .18 | | | 1 | | |
| 5 (Force Justify) | | 0.76 | 0.24 | | 74 | | <.01 | | | .35 | | | 1 | | |
| 7 (Height) | | 0.14 | 0.17 | | 117 | | .41 | | | .08 | | | 0 | | |
| 7 (Weight) | | -0.42 | 1.26 | | 117 | | .74 | | | -.03 | | | 0 | | |
| *Note. B* = estimate from cross-classified linear mixed model, *SE* = standard error of estimate, *df* = degrees of freedom from cross-classified linear model, *r* = effect size of interaction estimate (positive values indicate larger race bias for male participants than female participants), Conflict: 1 = Yes, 2 = No. | | | | | | | | | | | | | | | | | | |
| Table S2 |  | | |  | | | |  | | |  | | |
| *Means and Standard Deviations (in Parentheses) by Participant Gender and Target Race.* | | | | | | | | | | | | | | | | | | | | |
| Study (DV) | Male Ps | | | | | | | Female Ps | | | | | | | | | | | | |
|  | Black | | | White | | | | Black | | | | | | White | | | | | | |
| 1a (Height) | 70.28 (1.34) | | | 69.94 (1.16) | | | | 70.14 (1.61) | | | | | | 69.82 (1.56) | | | | | | |
| 1a (Weight) | 188.36 (15.95) | | | 178.07 (15.27) | | | | 187.58 (17.41) | | | | | | 175.72 (14.20) | | | | | | |
| 1b (Height) | 71.73 (1.57) | | | 70.79 (1.51) | | | | 72.24 (1.98) | | | | | | 71.36 (2.07) | | | | | | |
| 1b (Weight) | 180.47 (18.98) | | | 177.41 (18.91) | | | | 182.42 (19.92) | | | | | | 176.01 (19.76) | | | | | | |
| 1c (Muscularity) | 3.54 (0.75) | | | 3.28 (0.71) | | | | 3.58 (0.74) | | | | | | 3.21 (0.59) | | | | | | |
| 1d (Strength) | 4.45 (0.61) | | | 4.11 (0.67) | | | | 4.65 (0.85) | | | | | | 4.00 (0.82) | | | | | | |
| 1e (Strength) | 4.91 (0.62) | | | 4.43 (0.81) | | | | 4.96 (0.76) | | | | | | 4.41 (1.04) | | | | | | |
| 2 (Harm Cap.) | 4.24 (1.08) | | | 3.58 (1.02) | | | | 4.65 (1.29) | | | | | | 4.31 (1.34) | | | | | | |
| 3 (Muscularity) | 3.71 (0.80) | | | 3.27 (0.77) | | | | 3.69 (0.76) | | | | | | 3.34 (0.69) | | | | | | |
| 3 (Harm Cap.) | 4.30 (1.10) | | | 3.54 (1.05) | | | | 4.83 (1.25) | | | | | | 4.41 (1.33) | | | | | | |
| 4 (Muscularity) | 3.85 (0.82) | | | 3.61 (0.78) | | | | 3.53 (0.84) | | | | | | 3.22 (0.77) | | | | | | |
| 4 (Harm Cap) | 4.04 (1.46) | | | 3.76 (1.34) | | | | 4.16 (1.32) | | | | | | 4.11 (1.28) | | | | | | |
| 5 (Force Justify) | 4.48 (1.21) | | | 3.64 (1.10) | | | | 3.48 (1.40) | | | | | | 3.40 (1.37) | | | | | | |
| 7 (Height) | 71.78 (2.00) | | | 71.33 (1.82) | | | | 71.72 (1.28) | | | | | | 71.41 (1.50) | | | | | | |
| 7 (Weight) | 190.52 (20.08) | | | 188.76 (18.48) | | | | 193.00 (18.37) | | | | | | 190.82 (17.71) | | | | | | |

Table S3

*Relationships between the Mean Afrocentricity Ratings and Various Perceptions of Their Formidability for White Targets*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 1. Afrocentricity a |  | .86\*\*\* | .26† | .02 | .20 | .21† | .27† | .20 | .14 | .13 | -.05 | -.05 | .21 |
| 2. Feature-Specific a | .85\*\* |  | .43\*\* | .02 | .23 | .28† | .29† | .22 | .17 | .14 | -.05 | -.01 | .31\* |
| 3. Skin Tone a | .20 | .35\* |  | .12 | .31\* | .30† | .38\* | .30\* | .34\* | .33\* | .04 | .06 | .36 |
| 4. Estimated Height b | .09 | .08 | .20 |  | .76\*\*\* | .69\*\*\* | .68\*\*\* | .83\*\*\* | .44\*\* | .33\* | .24 | .46\*\* | -.17 |
| 5. Estimated Weight b | .19 | .16 | .24 | .81\*\*\* |  | .96\*\*\* | .88\*\*\* | .97\*\*\* | .69\*\*\* | .57\*\*\* | .05 | .37\* | .31\* |
| 6. Muscularity c, d, e | .17 | .20 | .19 | .78\*\*\* | .95\*\*\* |  | .91\*\*\* | .96\*\*\* | .69\*\*\* | .56\*\*\* | .02 | .29† | .38\* |
| 7. Strength f | .25 | .23 | .31\* | .74\*\*\* | .87\*\*\* | .90\*\*\* |  | .94\*\*\* | .80\*\*\* | .71\*\*\* | .05 | .25† | .28† |
| 8. Formidability Composite h | .19 | .18 | .25 | .88\*\*\* | .97\*\*\* | .97\*\*\* | .94\*\*\* |  | .71\*\*\* | .60\*\*\* | .09 | .37\* | .22 |
| 9. Harm Capability d, e, g | .12 | .14 | .30† | .46\*\*\* | .70\*\*\* | .69\*\*\* | .80\*\*\* | .71\*\*\* |  | .94\*\*\* | .05 | .14 | .15 |
| 10. Force Justification h | .11 | .11 | .32\* | .38\* | .61\*\*\* | .58\*\*\* | .72\*\*\* | .62\*\*\* | .95\*\*\* |  | -.05 | .02 | .13 |
| 11. Actual Height |  |  |  |  |  |  |  |  |  |  |  | .71\*\*\* | -.35\* |
| 12. Actual Weight |  |  |  |  |  |  |  |  |  |  |  |  | -.13 |
| 13. fWHR a |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Note*. Values above the diagonal (*df* = 43) represent bivariate correlations; values below the diagonal (*df* = 40) represent partial correlations controlling for the targets’ actual height and weight and fWHR. | | | | | | | | | | | | | |

Table S4

*Relationships between the Mean Afrocentricity Ratings and Various Perceptions of Their Formidability for Black Targets*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 1. Afrocentricity a |  | .79\*\*\* | .77\*\*\* | .10 | .34\* | .37\* | .40\*\* | .34\* | .51\*\*\* | .46\*\* | -.19 | -.01 | .08 |
| 2. Feature-Specific a | .80\*\*\* |  | .72\*\*\* | .17 | .41\*\* | .44\*\* | .46\*\* | .42\*\* | .46\*\* | .38\* | -.10 | -.05 | .15 |
| 3. Skin Tone a | .81\*\*\* | .75\*\*\* |  | .10 | .18 | .19 | .25† | .20 | .32\* | .31\* | -.01 | -.07 | -.07 |
| 4. Estimated Height b | .14 | .20 | .10 |  | .54\*\*\* | .31\* | .45\*\* | .68\*\*\* | .45\*\* | .40\*\* | .16 | .10 | -.01 |
| 5. Estimated Weight b | .31\* | .43\*\* | .30† | .81\*\*\* |  | .93\*\*\* | .92\*\*\* | .97\*\*\* | .84\*\*\* | .73\*\*\* | -.02 | .32\* | .48\*\* |
| 6. Muscularity c, d, e | .34\* | .47\*\* | .35\* | .78\*\*\* | .90\*\*\* |  | .94\*\*\* | .91\*\*\* | .79\*\*\* | .66\*\*\* | -.12 | .25 | .56\*\*\* |
| 7. Strength f | .40\*\* | .47\*\* | .39\* | .74\*\*\* | .88\*\*\* | .91\*\*\* |  | .94\*\*\* | .85\*\*\* | .72\*\*\* | -.05 | .23 | .54\*\*\* |
| 8. Formidability Composite h | .31\* | .42\*\* | .30\* | .88\*\*\* | .97\*\*\* | .89\*\*\* | .93\*\*\* |  | .83\*\*\* | .72\*\*\* | -.00 | .26† | .44\*\* |
| 9. Harm Capability d, e, g | .51\*\* | .48\*\* | .44\*\* | .46\*\*\* | .76\*\*\* | .70\*\*\* | .80\*\*\* | .77\*\*\* |  | .94\*\*\* | .05 | .14 | .15 |
| 10. Force Justification h | .45\*\* | .38\* | .39\* | .38\* | .67\*\*\* | .60\*\*\* | .67\*\*\* | .65\*\*\* | .95\*\*\* |  | .02 | .28† | .28† |
| 11. Actual Height |  |  |  |  |  |  |  |  |  |  |  | .63\*\*\* | .09 |
| 12. Actual Weight |  |  |  |  |  |  |  |  |  |  |  |  | .14 |
| 13. fWHR a |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Note*. Values above the diagonal (*df* = 43) represent bivariate correlations; values below the diagonal (*df* = 40) represent partial correlations controlling for the targets’ actual height and weight and fWHR. | | | | | | | | | | | | |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table S5  *Means, Standard Deviations, and Test Statistics Comparing Participants’ Ratings Along Multiple Measures of Physical Formidability (Height, Weight, Muscularity, Strength) and Harm Capability for a Single Target Purported to be Black or White* | | | | | | |
| Measure | Race | *M* | *SD* | *t* | *p* | Cohen's *d* |
| Height (in.) | Black | 72.19 | 2.30 | 1.71 | .09 | 0.28 |
|  | White | 71.59 | 2.03 |  |  |  |
|  |  |  |  |  |  |  |
| Weight (lbs) | Black | 206.11 | 24.01 | 1.55 | .12 | 0.25 |
|  | White | 199.91 | 25.71 |  |  |  |
|  |  |  |  |  |  |  |
| Muscularity | Black | 3.46 | 1.30 | 1.18 | .24 | 0.19 |
|  | White | 3.20 | 1.43 |  |  |  |
|  |  |  |  |  |  |  |
| Strength | Black | 4.57 | 1.37 | 1.51 | .13 | 0.24 |
|  | White | 4.26 | 1.14 |  |  |  |
|  |  |  |  |  |  |  |
| Harm Capability | Black | 5.16 | 1.40 | 0.90 | .37 | 0.15 |
|  | White | 4.97 | 1.22 |  |  |  |
| *Note*. *df* = 154 for each comparison;muscularity, strength, and harm rated on a 1-7 Likert scale. | | | | | | |

References

McDonald, M. M., Navarrete, C. D., & Van Vugt, M. (2012). Evolution and the psychology of intergroup conflict: The male warrior hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 670-679.

Van Vugt, M., De Cremer, D., & Janssen, D. P. (2007). Gender differences in cooperation and competition the male-warrior hypothesis. *Psychological Science*, *18*, 19-23.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*, 2020 – 2045.