Racial Bias in Perceptions of Size and Strength:
The Impact of Stereotypes and Group Differences

David J. Johnson\textsuperscript{1,2} and John Paul Wilson\textsuperscript{3}

\textsuperscript{1}Michigan State University
\textsuperscript{2}University of Maryland at College Park
\textsuperscript{3}Montclair State University

THIS HAS NOT BEEN COPYEDITED
PLEASE DO NOT QUOTE DIRECTLY

Author Note: We thank Joseph Cesario, Bill Chopik, and David Clark for their helpful comments on earlier drafts of this paper. Study materials, data, and analytic scripts can be found at https://osf.io/hkpwa. The preregistration for Study 2 can be found at https://osf.io/bmpcd.

Corresponding Author: David J. Johnson, email: djjohnson@smcm.edu. Author contributions:
Conceptualization D.J.J. & J.P.W.; Data collection D.J.J & J.P.W.; Formal analysis D.J.J;
Writing - Original Draft D.J.J; Writing - Reviewing & Editing, D.J.J. & J.P.W.

Word Count: 2000 (Introduction, Discussion, and Footnotes Only)
Abstract

Recent research has shown race can influence perceptions of men’s size and strength (Wilson, Hugenberg, & Rule, 2017). Across two studies (Study 1 $N = 1032$; Study 2 $N = 303$) examining men and women from multiple racial groups (Asian, Black, and White adults) we show that although race does impact judgments of size and strength, raters’ judgments primarily track targets’ objective physical features. In some cases, racial stereotypes actually improved group-level accuracy, as these stereotypes aligned with racial group differences in size and strength according to nationally representative data. We conclude that individuals primarily rely on individuating information when making physical judgments, but do not completely discount racial stereotypes, which reflect a combination of real group-level differences and culturally transmitted beliefs.

*Keywords:* person perception, race bias, stereotype accuracy
Racial Bias in Perceptions of Size and Strength: The Impact of Stereotypes and Group Differences

How much does race impact perceptions of size and strength? Although much work has highlighted that perceivers stereotype Black men as more threatening (Cottrell & Neuberg, 2005) and larger and more muscular (Holbrook, Fessler, and Navarrette, 2016) than White men, only recently has it been shown that Black men are judged as larger and stronger than White men (Wilson, Hugenberg, & Rule, 2017). Such findings are not only relevant to our understanding of social perception, but may also inform thinking about social issues such as police use of force. Because officers can only use lethal force when a person poses a threat to the officer or others (Tennessee v. Garner, 1985), impressions of threat posed by a person are critical. For example, in recent cases where lethal force was used against unarmed Black men, these men are often described by officers as large and physically imposing, even if they were only average sized (Hayes, 2014).

The current work examines the joint influence of race and individuating information on size and strength judgments. We did so by testing how much race impacts judgments relative to what people should use when making judgments: physical information. We also tested whether culturally transmitted stereotypes or group-level differences explained race effects while investigating whether such effects extended to women and other racial groups.

Stereotypes as a Source of Bias

Perceptual biases in size and strength might occur from socially transmitted stereotypes. Wilson and colleagues (2017) found Black men were rated larger and stronger than White men, controlling for size and strength. They concluded these distorted perceptions reflect stereotypes of Black men as threatening (Cottrell & Neuberg, 2005), larger and more muscular (Holbrook et
al., 2016), and superhuman (Waytz, Hoffman, & Trawalter, 2015). Black men are not the only group, however, where stereotypes reflect threat relevant information. Asian men are stereotyped as feminine (Wilkins, Chan, & Kaiser, 2011), short (Chen & Geiselman, 1993; Geiselman, Lam, Lee, & Chen, 1995), and weak (Mok, 1998). Thus, Asian men might be seen as less threatening than other men.

Racial biases in perceptions of size and strength have primarily been tested with men. This may be due to a focus on their downstream effects on behavior, such as police use of force (Wilson et al., 2017) and criminal sentencing (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). However, it may also reflect a lack of stimuli depicting women with known physical characteristics. Biases towards men have partially been explained by invoking evolved mechanisms that manage errors in threat detection (Haselton & Buss, 2000; Haselton & Nettle, 2006). Insofar as this error management system favors hypervigilance, women stereotyped as threatening may also be subject to size bias (Fessler, Holbrook, Tiohkin, & Synder, 2014).

Moreover, stereotypes about same-race men and women often overlap. For example, Black women are stereotyped as confrontational and assertive (Smith-Evans, George, Graves, Kaufmann, & Frohlich, 2014), are incarcerated at higher rates than White women (Crenshaw & Ritchie, 2015), and are seen as more adult-like than White girls (Epstein, Blake, & Gonzalez, 2017). These stereotypes convey threat information and reflect the interdependence of sex and race as social categories, with Blacks associated with men and Asians with women (Johnson, Freeman, & Pauker, 2012).

**Group Differences as a Source of Bias**

Perceptual biases in size and strength might also reflect group differences. In fact, some have argued that many stereotypes originate from real group-level differences (Jussim, Cain,
Crawford, Harber, & Cohen, 2009; Jussim, Crawford, & Rubenstein, 2015). This account requires evidence of differences between groups. Data from the nationally representative Centers for Disease Control’s (CDC) National Health and Nutrition Examination Survey (NHNES; Fryar, Ogden, & Flegal, 2016) shows that although White and Black adults are roughly the same height, Asian adults are moderately shorter than Whites ($d_s \sim 0.60$). CDC data measuring bicep circumference (Fryar et al., 2016)—a valid indicator of strength (Sell et al., 2009)—shows that White adults have smaller biceps than Blacks ($d_s \sim -0.30$) but larger biceps than Asians ($d_s \sim 0.60$). In general, across nationally representative and convenience samples, Black adults are more muscular and stronger than Whites, who are more muscular and stronger than Asians (see Table 1).

### Table 1. Racial Differences in Size and Strength.

<table>
<thead>
<tr>
<th>Study</th>
<th>Outcome</th>
<th>N (Men)</th>
<th>$d_{WB}$</th>
<th>$d_{WA}$</th>
<th>N (Women)</th>
<th>$d_{WB}$</th>
<th>$d_{WA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fryar, Ogden, &amp; Flegal (2016)</td>
<td>Height</td>
<td>3982</td>
<td>0.06</td>
<td>0.59</td>
<td>4209</td>
<td>-0.01</td>
<td>0.58</td>
</tr>
<tr>
<td>Fryar, Ogden, &amp; Flegal (2016)</td>
<td>Bicep Size</td>
<td>3845</td>
<td>-0.18</td>
<td>0.63</td>
<td>3920</td>
<td>-0.44</td>
<td>0.57</td>
</tr>
<tr>
<td>Silva et al. (2010)</td>
<td>Skeletal Muscle</td>
<td>469</td>
<td>-0.41</td>
<td>---</td>
<td>1280</td>
<td>-0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>Jackson, Ellis, McFarlin, Sailors, &amp; Bray (2009)</td>
<td>Lean Mass</td>
<td>932</td>
<td>-0.36</td>
<td>0.24</td>
<td>566</td>
<td>-0.38</td>
<td>1.02</td>
</tr>
<tr>
<td>Chen, Liu, &amp; Yu (2009)</td>
<td>Bench Press</td>
<td>132</td>
<td>---</td>
<td>0.25</td>
<td>52</td>
<td>---</td>
<td>0.65</td>
</tr>
<tr>
<td>Chen, Liu, &amp; Yu (2009)</td>
<td>Arm Curl</td>
<td>172</td>
<td>---</td>
<td>1.10</td>
<td>54</td>
<td>---</td>
<td>-0.34</td>
</tr>
</tbody>
</table>

*Note. $d_{WB}$ = Cohen’s $d$ for the difference between White and Black adults. $d_{WA}$ = Cohen’s $d$ for the difference between White and Asian adults. Positive coefficients indicate White adults scored higher on the outcome.*

If stereotypes about physical characteristics are partially accurate, stereotypes could be a consequence of group-level differences, rather than the source of bias. Of course, even if these
stereotypes are accurate at the group-level they may bias judgments if used when more relevant individuating information is available. Fortunately, stereotypes typically only modestly impact decisions, whereas individuating information is much more important (Jussim, Eccles, & Madon, 1996; Madon et al., 1998; for reviews, see Jussim et al., 2009; Kunda & Thagard, 1996). On the other hand, stereotypes may improve judgment accuracy through constructive accuracy (Jussim, 1991). When individuating information is imperfect, perceivers may reach accurate perceptions by relying on stereotypes that reflect base rate differences between groups. Thus, just because race impacts decisions does not inherently mean it decreases accuracy.

We tested the stereotype and group differences accounts in two studies by measuring the impact of race on size and strength judgments while controlling for objective variation in these traits. We used male and female targets from multiple racial groups. Both studies provided evidence that judgments of physical features mostly track individuating information but are also influenced by race. We found exploratory evidence that stereotype effects improved group-level accuracy in judgments for some but not all groups.

**Study 1: Exploratory Analysis**

Study 1 tested whether race or physical features better explained variation in strength judgments, and whether those effects generalized to Asian and female targets.

**Methods**

**Raters**

Raters were 1088 undergraduates from Michigan State University who completed the study for course credit. At the end of each semester, a set of raters judged photographs of targets photographed that semester. Based on the demographics of our subject pool and our interest in racial stereotypes about Asian and Black individuals relative to Whites, we limited our analyses
raters from those three racial groups \(N = 1032\). This left 78 Asian raters (41.0% women), 89 Black raters (58.4% women), and 865 White raters (51.9% women). The average age of the raters was 19.8 \((SD = 2.7)\).

**Targets**

Targets were 1660 Michigan State University undergraduates photographed over multiple semesters (2013 – 2015) for course credit. A full body photo was taken of each participant in front of a wall with a marker for height. Men were photographed without their shirts; women were photographed wearing a standard black t-shirt. All participants provided consent to have their photographs rated for experimental purposes. We limited our analyses to targets from White, Black, and Asian groups \((N = 1545)\). This left 102 Asian targets (56.9% women), 135 Black targets (56.3% women), and 1308 White targets (56.7% women). The average age of the targets was 19.7 \((SD = 1.6)\).

In addition to having their photographs taken, targets’ upper body strength was measured with an inverted hand dynamometer using the procedures outlined by Sell and colleagues (2009). Participants’ bicep circumference and height were also recorded. These measures were used to control for the objective strength of targets.

**Procedure**

Raters completed the task online. Each rater saw 40 random photographs of targets (20 men, 20 women) photographed that semester. They rated each target on four dimensions: 1) strength, 2) toughness, 3) their likelihood of beating an opponent, and 4) their attractiveness. Ratings were made on a 7-point scale. Participants were instructed to rate each target relative to other targets of the same sex. Ratings on the first three judgments were strongly correlated \(\alpha_{Men} = .94, \alpha_{Women} = .93\), so we focused on strength ratings to remain comparable to past research.
RACIAL BIAS IN SIZE AND STRENGTH

(Wilson et al., 2017). Attractiveness ratings are discussed elsewhere (Johnson & Wilson, 2018).

Raters also reported their own race, sex, height, and strength relative to same-sex others on a 100-point scale. These variables were used to test the role of rater differences in judgments.

Analytic Model

Strength judgments were analyzed using multilevel regression using the lme4 package in R (Version 1.1-13; Bates, Maechler, Bolker, & Walker, 2015). Race was dummy coded with Whites as the reference group for both raters and targets. Rater sex was effects coded (women = -.5, men = .5), and all continuous measures were centered and standardized so that a \( \beta \) of .50 indicates that as the predictor increases by one standard deviation the outcome increases by half a standard deviation. Random intercepts were included for both raters and targets to control for nonindependence (Judd, Westfall, & Kenny, 2012). Preliminary analyses revealed a lack of measurement invariance in perceived strength between men and women targets, so we analyzed judgments separately for each sex. We used the simr package in R (Version 1.04; Green & MacLeod) to test whether the experimental design was sufficient to detect a small-to-moderate effect of target race for Asian and Black men and women (i.e., \( \beta = .30 \)). This analysis revealed that in all cases we had at least 90% power to detect these effects.

Results

Descriptive statistics and correlations for all variables across all studies are reported in the Supplemental Materials.

Male Targets

We first verified that raters relied on targets’ actual physical features when making strength judgments about men. Table 2 lists the regression coefficients for the multilevel model. As predicted, men with more upper body strength (\( \beta = .157, 95\% \text{ CI} = [.115, .199], p < .001 \)) and
larger biceps ($\beta = .276, 95\% \text{ CI} = [.234, .318], p < .001$) were rated as stronger. Height did not predict perceived strength when controlling for upper body strength and bicep circumference ($\beta = .004, 95\% \text{ CI} = [-.036, .044], p < .832$). We also tested whether target race impacted raters’ strength judgments. As predicted, Black men were rated as stronger than White men ($\beta = .495, 95\% \text{ CI} = [.400, .736], p < .001$), and Asian men were rated as weaker than White men ($\beta = -.312, 95\% \text{ CI} = [-.613, -.225], p < .001$). These differences occurred when controlling for targets’ objective strength, suggesting raters continued to use race despite the presence of individuating information. Finally, although some rater characteristics did predict strength judgments, these effects were small compared to the effects of target characteristics ($\beta$s < .15; see Table 2).

Table 2. Perceptions of Strength for Male Targets (Study 1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>$df$</th>
<th>SE</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Height</td>
<td>0.004</td>
<td>664</td>
<td>0.021</td>
<td>.832</td>
</tr>
<tr>
<td>Target Upper Body Strength</td>
<td>0.157</td>
<td>665</td>
<td>0.021</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Target Bicep Circumference</td>
<td>0.276</td>
<td>667</td>
<td>0.022</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Target Race (Asian)</td>
<td>-0.312</td>
<td>663</td>
<td>0.083</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Target Race (Black)</td>
<td>0.495</td>
<td>663</td>
<td>0.072</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Rater Sex (Male)</td>
<td>-0.136</td>
<td>985</td>
<td>0.035</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Rater Race (Asian)</td>
<td>-0.112</td>
<td>1002</td>
<td>0.056</td>
<td>.046</td>
</tr>
<tr>
<td>Rater Race (Black)</td>
<td>-0.140</td>
<td>987</td>
<td>0.053</td>
<td>.008</td>
</tr>
<tr>
<td>Rater Strength</td>
<td>-0.069</td>
<td>997</td>
<td>0.015</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Rater Height</td>
<td>0.000</td>
<td>972</td>
<td>0.018</td>
<td>.997</td>
</tr>
</tbody>
</table>

*Note.* Race was dummy coded with Whites as the reference group.

**Female Targets**

We first tested whether raters relied on targets’ actual physical features when making strength judgments about women. Table 3 lists the regression coefficients for the multilevel model. Consistent with the analysis of male targets, women with more upper body strength ($\beta = .098, 95\% \text{ CI} = [.072, .124], p < .001$) and larger biceps ($\beta = .211, 95\% \text{ CI} = [.184, .237], p < .001$) were rated as stronger. Unlike male targets, height predicted judgments of women targets’
strength when accounting for the physical measures ($\beta = .088$, 95% CI = [.063, .113], $p < .001$).

Did target race impact raters’ strength judgments for female targets? When strength was controlled for, Black women were rated as stronger than White women ($\beta = .335$, 95% CI = [.245, .425], $p < .001$), and Asian women were rated as weaker than White women ($\beta = -.501$, 95% CI = [-.604, .399], $p < .001$). Thus, raters relied on both individuating information and race when making strength judgments for women as well as men.

### Table 3. Perceptions of Strength for Female Targets (Study 1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>$df$</th>
<th>SE</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Height</td>
<td>0.088</td>
<td>873</td>
<td>0.013</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Target Upper Body Strength</td>
<td>0.098</td>
<td>870</td>
<td>0.013</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Target Bicep Circumference</td>
<td>0.211</td>
<td>865</td>
<td>0.013</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Target Race (Asian)</td>
<td>-0.501</td>
<td>887</td>
<td>0.052</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Target Race (Black)</td>
<td>0.335</td>
<td>865</td>
<td>0.046</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Rater Sex (Male)</td>
<td>-0.132</td>
<td>873</td>
<td>0.041</td>
<td>.001</td>
</tr>
<tr>
<td>Rater Race (Asian)</td>
<td>0.011</td>
<td>898</td>
<td>0.062</td>
<td>.862</td>
</tr>
<tr>
<td>Rater Race (Black)</td>
<td>-0.043</td>
<td>884</td>
<td>0.065</td>
<td>.511</td>
</tr>
<tr>
<td>Rater Strength</td>
<td>-0.021</td>
<td>881</td>
<td>0.017</td>
<td>.206</td>
</tr>
<tr>
<td>Rater Height</td>
<td>-0.021</td>
<td>868</td>
<td>0.021</td>
<td>.314</td>
</tr>
</tbody>
</table>

*Note. Race was dummy coded with Whites as the reference group.*

### Do Raters Use Race or Individuating Information More?

Strength judgments for men and women targets were influenced by target race and by objective measures of strength, but which set of variables better explains judgments? That is, do raters use individuating information (physical strength) more than categorical information (target race)? We tested this by comparing the variance in strength judgments explained by the physical variables versus target race (Table 4). In Study 1, *physical features explained 3–5 times more variance than race*. Although raters did not fully discount the race of a target when individuating information was present, race played a much smaller role than objective markers of strength.

We also examined how much variability in strength judgments rater characteristics
explained. Rater characteristics explained little variability in strength judgments (2% or less), less than the impact of target race, and far less than the impact of physical features.

Table 4. Variance Explained by Race, Target Physical Features, and Rater Characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>Variable</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Race</td>
<td>Physical</td>
</tr>
<tr>
<td>Study 1</td>
<td>Strength</td>
<td>.024</td>
<td>.121</td>
</tr>
<tr>
<td>Study 2</td>
<td>Strength</td>
<td>.070</td>
<td>.124</td>
</tr>
<tr>
<td>Study 2</td>
<td>Height</td>
<td>.022</td>
<td>.126</td>
</tr>
</tbody>
</table>

*Note.* Pseudo-$R^2$ values were calculated using the methods from Nakagawa & Schielzeth (2013). Values represent the increase in variance explained by adding the set of predictor variables to the model including all other variables. Race = dummy coded target race. Physical = objective target measurements. Rater = rater variables.

Do Perceiver or Target Characteristics Drive Strength Judgments?

As a final exploratory test, we examined whether variability in strength judgments was driven primarily by target or rater characteristics (see Table 5). For both male and female targets, more of the variability in ratings of strength (and size) was due to differences between targets and not between raters. There was more nonindependence in strength for male than female targets. This is likely because men were photographed shirtless, making it easier to see differences in musculature.

Table 5. Relative Contributions of Rater and Target Variance to Judgments

<table>
<thead>
<tr>
<th>Study</th>
<th>Variable</th>
<th>Men</th>
<th></th>
<th></th>
<th>Women</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Target</td>
<td>Rater</td>
<td>Target</td>
<td>Rater</td>
<td>Target</td>
<td>Rater</td>
</tr>
<tr>
<td>Study 1</td>
<td>Strength</td>
<td>.396</td>
<td>.183</td>
<td>.235</td>
<td>.218</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td>Strength</td>
<td>.492</td>
<td>.133</td>
<td>.357</td>
<td>.131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td>Height</td>
<td>.361</td>
<td>.173</td>
<td>.414</td>
<td>.124</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Table values refer to the intra-class correlations for targets and raters. Larger values indicate that characteristics of the rater or target drive judgments of strength or size.

Discussion
Raters’ judgments were influenced by targets’ objective strength and race. While the effects of race were small-to-moderate, judgments were primarily driven by targets’ objective strength. One limitation of this study was that it was exploratory; stimuli were collected for unrelated purposes. Additionally, the majority of raters and targets were White women, limiting generalizability. We addressed these limitations with a pre-registered replication.

**Study 2: Confirmatory Analysis**

Study 2 replicated and extended the results of Study 1. We measured the impact of physical features on strength and height judgments. We focused on height judgments because height can be measured directly rather than through a proxy (e.g., bicep circumference for strength). Finally, we attempted to replicate rater biases; these did not replicate and are discussed in the Supplemental Materials.

**Method**

**Preregistration**

Study 2 was preregistered on the Open Science Framework: https://osf.io/bmpcd/. We detailed our hypotheses, study design, sampling plan, and analysis plan in advance. We only deviated from our analysis plan in one respect; we collected data from five Black male undergraduates from Montclair State University in addition to our primary subject pool of Michigan State University students to try to meet our rater goals.

**Raters**

Raters were 303 undergraduates who completed the study for course credit. We wanted to recruit a racially diverse sample (i.e., Asian, Black, and White raters) while maintaining similar numbers of raters in each group. Based on the demographics of our subject pool, we made a realistic goal to recruit 50 participants from each unique combination of race and sex. We
determined a priori to stop data collection by the end of the fall 2017 semester. We met our goal for all racial groups except for Black men ($N = 33$). Our sample consisted of 106 Asian raters (51.9% women), 89 Black raters (62.9% women), and 108 White raters (50.9% women). The average age of the raters was 19.6 ($SD = 1.6$).

**Targets**

Targets were selected from the 1660 undergraduates collected in Study 1. To maximize diversity we selected all non-blurry photos of Asian and Black targets. This left 92 (of 102) Asian targets (60.9% women) and 133 (of 135) Black targets (55.6% women). We then selected 129 White targets (50.4% women) from our sample of 1308 White targets. These targets were chosen to maximize variability in perceived strength. We achieved this by averaging ratings of perceived strength for each target in Study 1 across raters. For each sex, targets were sorted by strength, and every $n$th person was chosen to obtain 65 targets. This selection process created a normal distribution of perceived strength that spanned the entire range of values. The average age of the targets was 19.3 ($SD = 0.6$). Because we were interested in raters’ subjective judgments of height, photos of targets were digitally edited to remove the height marker.

**Procedure**

Raters completed the task in the lab. Each rater saw 100 randomly selected targets (50 men, 50 women). They rated each target on three dimensions: 1) strength, 2) height, and 3) attractiveness. Ratings were made on a 7-point scale. Participants were instructed to rate each target relative to other targets of the same sex. Per our pre-registration plan, we did not focus on attractiveness ratings in the current report. To test the role of rater individual differences in judgments, we gathered the same information about raters as was collected in Study 1.

---

1We selected 130 white targets (65 men, 65 women) but one photo was unintentionally not rated.
Power Analysis

Similar to Study 1, we conducted a power analysis based on our experimental design. We again tested whether the experimental design was sufficient to detect a small-to-moderate effect of target race for Asian and Black men and women (i.e., \( \beta = .30 \)). All analyses had over 90% power to detect these effects, except for our analyses of perceived strength for men. Our power to detect an effect of race was somewhat lower for Asian men (72% 95% CI [.70, .76] and Black men (.88, 95% CI [.85, .89]).

Results

Male Targets

**Strength judgments.** We tested whether raters’ judgments of men’s strength were predicted by targets’ objective strength. Table 6 lists the regression coefficients for the multilevel model. As predicted, men with more upper body strength (\( \beta = .164, 95\% \text{ CI} = [.078, .251], p < .001 \)) and larger biceps (\( \beta = .290, 95\% \text{ CI} = [.200, .380], p < .001 \)) were rated as stronger. Taller men were not rated as stronger when controlling for these physical measures, (\( \beta = -.079, 95\% \text{ CI} = [-.165, .008], p = .076 \)). Race also impacted strength judgments when controlling for objective strength. Black men were rated as stronger than White men (\( \beta = .349, 95\% \text{ CI} = [.151, .546], p < .001 \)). Asian men were also rated as weaker than White men (\( \beta = -.415, 95\% \text{ CI} = [-.642, -.188], p < .001 \)). Although a target’s race and physical features both influenced raters’ judgments, individuating information explained nearly twice as much variance as race (Table 4).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Strength</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>df</td>
</tr>
<tr>
<td>Target Height</td>
<td>-.079</td>
<td>157</td>
</tr>
<tr>
<td>Target Upper Body Strength</td>
<td>0.164</td>
<td>157</td>
</tr>
<tr>
<td>Target Bicep Circumference</td>
<td>0.290</td>
<td>157</td>
</tr>
</tbody>
</table>
Target Race (Asian)  -0.415  157  0.116  < .001  -0.306  157  0.086  < .001
Target Race (Black)   0.349  157  0.101  < .001  0.116  157  0.075  .123
Rater Sex (Male)       0.042  286  0.057  .465  -0.038  287  0.064  .552
Rater Race (Asian)    -0.041  286  0.054  .445  0.030  286  0.060  .623
Rater Race (Black)    -0.143  286  0.055  .010  -0.180  286  0.061  .004
Rater Strength        -0.000  286  0.024  .970  0.057  286  0.026  .031
Rater Height          -0.005  286  0.029  .109  -0.075  287  0.033  .023

Note. Race was dummy coded with Whites as the reference group.

**Height judgments.** Were raters’ height judgments predicted by targets’ actual height?

Table 6 lists the regression coefficients for the multilevel model. As expected, taller men were rated as taller ($\beta = .382$, 95% CI = [.318, .446], $p < .001$). Upper body strength and bicep circumference did not predict height, $ps > .10$. We found mixed support for the hypothesis that race influenced raters’ height judgments, controlling for targets’ actual height. Black men were not rated as taller than White men ($\beta = .116$, 95% CI = [-.030, .262], $p = .123$). However, Asian men were rated as shorter than White men ($\beta = - .306$, 95% CI = [-.4747, -.138], $p < .001$). Again, physical features better explained raters’ judgments; they accounted for more than five times the amount of variance than rater race (Table 4).

**Female Targets**

**Strength judgments.** Consistent with the analysis of male targets, raters’ judgments of women’s strength were predicted by their objective strength (see Table 7). Women with more upper body strength ($\beta = .086$, 95% CI = [.029, .142], $p = .003$) and larger biceps ($\beta = .321$, 95% CI = [.260, .382], $p < .001$) were rated as stronger. Taller women were not rated as stronger when controlling for these physical measures, ($\beta = .021$, 95% CI = [-.032, .073], $p = .441$). Race also influenced raters’ strength judgments of women. Black women were descriptively rated as stronger than White women ($\beta = .115$, 95% CI = [-.020, .250], $p = .077$), although this effect was not significant according to our pre-registered alpha level (.05). Asian women were rated as
significantly weaker than White women (β = -.385, 95% CI = [-.521, -.249], p < .001). As with male targets, physical features explained much more variability than target race (three times; see Table 4).

Table 7. Perceptions of Strength and Height for Female Targets (Study 2)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Strength</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>df</td>
</tr>
<tr>
<td>Target Height</td>
<td>0.021</td>
<td>195</td>
</tr>
<tr>
<td>Target Upper Body Strength</td>
<td>0.086</td>
<td>195</td>
</tr>
<tr>
<td>Target Bicep Circumference</td>
<td>0.321</td>
<td>195</td>
</tr>
<tr>
<td>Target Race (Asian)</td>
<td>-0.385</td>
<td>195</td>
</tr>
<tr>
<td>Target Race (Black)</td>
<td>0.115</td>
<td>195</td>
</tr>
<tr>
<td>Rater Sex (Male)</td>
<td>-0.058</td>
<td>287</td>
</tr>
<tr>
<td>Rater Race (Asian)</td>
<td>0.114</td>
<td>287</td>
</tr>
<tr>
<td>Rater Race (Black)</td>
<td>0.008</td>
<td>287</td>
</tr>
<tr>
<td>Rater Strength</td>
<td>0.031</td>
<td>287</td>
</tr>
<tr>
<td>Rater Height</td>
<td>-0.062</td>
<td>287</td>
</tr>
</tbody>
</table>

Note. Race was dummy coded with Whites as the reference group.

**Height judgments.** Paralleling the results for men, raters’ height judgments of women were predicted by targets’ actual height. Taller women were rated as taller (β = .425, 95% CI = [.360, .489], p < .001). Upper body strength and bicep circumference did not predict height, ps > .10. We found mixed support for the hypothesis that race influenced raters’ height judgments. Black women were not rated as taller than White women (β = -.077, 95% CI = [-.233, .079], p = .335). However, Asian women were rated as shorter than White women (β = -.311, 95% CI = [-.479, -.144], p < .001). These findings were consistent with the effects of race observed for male targets. Again, physical features better explained raters’ judgments; they accounted for more than five times the amount of variance than rater race (Table 4).

**Does Race Increase or Decrease Accuracy in Perceptions of Size and Strength?**

Judgments of size and strength were both predicted by race and physical information.
However, just because race impacts judgments when controlling for physical features does not mean it decreases accuracy. Rather, when individuating information is difficult to parse, relying on stereotypes may increase accuracy insofar as those stereotypes reflect valid information about group differences. We tested this in an exploratory analysis\(^2\) comparing the correlation between target race and *actual* physical features to the correlation between target race and *perceptions* of physical features. Raters’ judgments accurately track real group differences when the correlation between race and perceptions of size or strength is similar in size to the correlation between race and actual differences in size or strength.

This exploratory analysis (see the Supplemental Materials) revealed that correlations between race and perceived physical features were very similar in size to the correlations based on nationally representative data, indicating raters’ judgments of the relationship between race and physical features were accurate at the group-level. In addition, perceptions of strength and height were *less* accurate when the effect of stereotypes was removed for every group except Black men. For Black men, stereotypes caused people to overestimate the relationship between race and strength and size. The reason these group stereotypes improved accuracy (other than for Black men) is because raters’ judgments only moderately tracked targets’ actual strength and size. The photographs may not have provided sufficient individuating information to make accurate judgments of size and strength. If such information were perfectly legible stereotypes would decrease accuracy rather than increase it.

**Discussion**

Strength judgments were influenced primarily by targets’ physical features rather than their race. We observed similar effects for height judgments. Asian targets were rated as shorter

\(^2\)We thank Lee Jussim for providing this helpful suggestion in peer review.
than Whites but Blacks were not rated as taller than Whites. This is consistent with height bias partially originating from group-level size differences. Whereas Asian adults are more than half a standard deviation shorter than Whites, Black adults are the same height as Whites (Fryar et al. 2016). Our findings that raters did not judge Black men as taller than White men may seem inconsistent with past work (Holbrook et al., 2016; Wilson et al., 2017). However, in Holbrook et al. (2016) participants did not rate actual targets, and in Wilson et al. (2017; Study 1B) raters’ height judgments were made from facial photographs and did not track targets’ actual height. Thus, when less individuating information is provided, race may bias decisions more.

**General Discussion**

We examined across two studies how objective physical information and race impacted perceptions of size and strength. Although both explained variability in judgments, physical information explained much more variability. Race did impact judgments, but actually improved group-level accuracy in some cases. The main exception to this was that stereotypes exaggerated the relationship between Black men and size or strength.

**Stereotypes or Group Differences as a Source of Bias?**

We explored two possible sources of racial bias in size and strength judgments. One is that racial stereotypes about size and strength might bias judgments of physical features in the service of error management. The other is that judgments might reflect accurate group-level differences in physical features. These explanations are not mutually exclusive. The overlap between racial stereotypes and group-level differences suggests physical differences might lead to different stereotype content across groups.

While culturally transmitted stereotypes can explain biases in size and strength, group-level differences alone might explain some biases. For example, the stereotype that Asian adults
are short is *sufficient* but not *necessary* to explain height biases because height differences exist between Asian and White adults. Similarly, height differences between White and Black adults are essentially zero and we did *not* observe height bias for Black adults relative to Whites. If one conceptualizes stereotypes simply as a set of beliefs about a group (Ashmore & Del Boca, 1981), rather than inaccurate beliefs (Jussim et al. 2009), the shorter average height of Asians would be an accurate part of that stereotype.

Even if stereotypes are accurate at the group-level, individuals should discount this information when given individuating information. In fact, researchers argue that stereotypes should not be used when one has “vividly clear, relevant individuating information about a member of a group” (Jussim et al., 2009, p. 213). Consistent with past research demonstrating that stereotypes only modestly impact judgments when individuating information is provided (Jussim et al., 1996, 2015; Kunda & Thagard, 1996; Madon et al., 1998), raters’ judgments primarily tracked objective physical features from photographs.

However, these photographs did not provide perfect information about targets, as evidenced by the moderate relationship between actual size and strength to perceptions of size and strength. Raters also relied on group stereotypes, and this information improved the degree to which raters’ judgments corresponded with actual group differences, with the exception of Black men. The fact that racial stereotypes exaggerated the relationship between race and size and strength for Black men suggests that racial stereotypes are shaped by both group level differences and culturally transmitted information.

Racial stereotypes were not limited to men. Asian women were rated as weaker and shorter than White women, and Black women were rated as stronger. We also explored whether rater characteristics might *moderate* racial bias in judgments (see the Supplemental Materials).
Neither rater sex nor race consistently moderated bias in size and strength judgments. This is consistent with work showing that appearance-based appraisals are less driven by perceiver characteristics (Hehman, Sutherland, Flake, & Slepian, 2017).

**Limitations and Future Directions**

The current studies focused exclusively on perceptual judgments of size and strength. While biases in these judgments are informative on their own we cannot make conclusions about their downstream effects on behavior. However, other work has demonstrated that size and strength judgments predict perceptions of the appropriateness of police force (Wilson et al., 2017). This is relevant because racial stereotypes improved the accuracy of perceptions for Black women and Asians but exaggerated the relationship between race and physical features for Black men. Although the impact of race was small, these biases may have implications for real world police-civilian interactions.

One way to bridge this work to actual police-civilian interactions would be to create ecologically valid designs to test the impact of race on decisions of police use of force. In an experimental approach, researchers could create videos where suspects’ engage in ambiguously aggressive behaviors while explicitly manipulating target race, size, and strength (e.g., Duncan, 1976). An alternative approach would rely on body-worn camera footage from actual police-civilian interactions and information about suspect race, size, and strength from police reports. These approaches would better address the degree to which perceptual biases translate into disparate outcomes based on suspect race.

**Conclusion**

Size and strength judgments primarily track physical differences rather than the race of an individual. The impact of race on judgments was consistent with actual group-level
differences for some groups, but was exaggerated for Black men. This shows the importance of testing racial biases in light of actual group differences and culturally transmitted beliefs.
References


Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of*


