Research Article

Psychological Science
1–10

DOI: 10.1177/0956797619827529
www.psychologicalscience.org/PS

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

David J. Johnson[1,2] and John Paul Wilson[3]
[1]Department of Psychology, Michigan State University; [2]Department of Psychology, University of Maryland at College Park; and [3]Department of Psychology, Montclair State University

Abstract
Recent research has shown that race can influence perceptions of men's size and strength. Across two studies (Study 1: *N* = 1,032, Study 2: *N* = 303) examining men and women from multiple racial groups (Asian, Black, and White adults), we found 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, open data, preregistration

Received 8/1/18; Revision accepted 10/29/18

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, & Navarrete, 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 not only are relevant to our understanding of social perception but may also inform thinking about social issues such as police use of force. Because officers can use lethal force only when a person poses a threat to the officer or other people (*Tennessee v. Garner*, 1985), impressions of threat posed by a person are critical. For example, in recent cases in which lethal force was used against unarmed Black men, these men were often described by officers as large and physically imposing, even if they were only average sized (Hayes, 2014).

In the current work, we examined 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 that Black men were rated as larger and stronger than White men, controlling for size and strength. They concluded that 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

Corresponding Author:
David J. Johnson, University of Maryland at College Park, Department of Psychology, 4094 Campus Drive, College Park, MD 20742
E-mail: djjohnson@smcm.edu

Table 1. Racial Differences in Size and Strength

Study	Outcome	Men			Women		
		N	d_{WB}	d_{WA}	N	d_{WB}	d_{WA}
Fryar, Gu, Ogden, & Flegal (2016)	Height	3,982	0.06	0.59	4,209	–0.01	0.58
Fryar et al. (2016)	Bicep size	3,845	–0.18	0.63	3,920	–0.44	0.57
Silva et al. (2010)	Skeletal muscle	469	–0.41		1,280	–0.30	0.20
Jackson, Ellis, McFarlin, Sailors, & Bray (2009)	Lean mass	932	–0.36	0.24	566	–0.38	1.02
Chen, Liu, & Yu (2012)	Bench press	132		0.25	52		0.65
Chen et al. (2012)	Arm curl	172		1.10	54		–0.34

Note: Cohen's *d*s are shown for the difference between White and Black adults (d_{WB}) and for the difference between White and Asian adults (d_{WA}). Positive values indicate that White adults scored higher on the outcome.

group, however, for which 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 the downstream effects of racial bias 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 toward 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, & Snyder, 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, 2012), and are seen as more adultlike than White girls (Epstein, Blake, & González, 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 (K. L. 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 researchers have argued that many stereotypes originate from real group-level differences (Jussim, Cain, Crawford, Harber, & Cohen, 2009; Jussim, Crawford, & Rubinstein, 2015). This account requires evidence of differences between groups. Data from the nationally representative Centers for Disease Control and Prevention (CDC) National Health and Nutrition Examination Survey (Fryar, Gu, Ogden, & Flegal, 2016) shows that although White and Black adults are roughly the same height, Asian adults are moderately shorter than Whites (*d*s ~ 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 ~ –0.30) but larger biceps than Asians (*d*s ~ 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).

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 impact decisions only modestly, 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. In an exploratory analysis, we found 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.

Method

Raters. Raters were 1,088 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 taken that semester. Given the demographics of our participant pool and our interest in racial stereotypes about Asian and Black individuals relative to Whites, we limited our analyses to raters from those three racial groups (N = 1,032). This sample consisted of 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 years (SD = 2.7).

Targets. Targets were 1,660 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 = 1,545). This sample consisted of 102 Asian targets (56.9% women), 135 Black targets (56.3% women), and 1,308 White targets (56.7% women). The average age of the targets was 19.7 years (SD = 1.6).

In addition to taking targets' photographs, we measured their upper body strength 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) taken that semester. They rated each target on four dimensions: (a) strength, (b) toughness, (c) their likelihood of beating an opponent, and (d) 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 (men: α = .94, women: α = .93), so we focused on strength ratings to ensure our results were comparable with those of past research (Wilson et al., 2017). Attractiveness ratings are discussed elsewhere (D. J. Johnson & Wilson, 2019). 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 raters' differences in judgments.

Analytic model. Strength judgments were analyzed using multilevel regression in the *lme4* package in the R programming environment (Version 1.1-13; Bates, Mächler, Bolker, & Walker, 2015). Race was dummy coded with White as the reference group for both raters and targets. Raters' sex was effects coded (women = −.5, men = .5), and all continuous measures were centered and standardized so that a β of 0.50 indicates that as the predictor increases by 1 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, 2016) to test whether the experimental design was sufficient to detect a small to moderate effect of targets' race for Asian and Black men and women (i.e., β = 0.30). 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 are reported in the Supplemental Material available online.

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 (β = 0.157, 95% confidence interval, or CI = [0.115, 0.199], p < .001) and larger biceps (β = 0.276, 95% CI = [0.234, 0.318], p < .001) were rated as stronger. Height did not predict perceived strength when analyses controlled for upper body strength and bicep circumference (β = 0.004, 95% CI = [−0.036, 0.044], p < .832). We also tested whether targets' race impacted raters' strength judgments. Black men were rated as stronger than White men (β = 0.495, 95% CI = [0.354, 0.636], p < .001), and Asian men were rated as weaker than White men (β = −0.312, 95% CI = [−0.473, −0.150], p < .001). These differences occurred when analyses controlled for targets' objective strength, suggesting that

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

Variable	β	*df*	*SE*	*p*
Targets' height	0.004	664	0.021	.837
Targets' upper body strength	0.157	665	0.021	< .001
Targets' bicep circumference	0.276	667	0.022	< .001
Targets' race (Asian)	–0.312	663	0.083	< .001
Targets' race (Black)	0.495	663	0.072	< .001
Raters' sex (male)	–0.136	985	0.035	< .001
Raters' race (Asian)	–0.111	1002	0.056	.046
Raters' race (Black)	–0.140	987	0.053	.008
Raters' strength	–0.069	997	0.015	< .001
Raters' height	0.000	972	0.018	.997

Note: Race was dummy coded with White as the reference group.

raters continued to use race despite the presence of individuating information. Finally, although some raters' characteristics did predict strength judgments, these effects were small compared with the effects of targets' characteristics (βs < |0.15|; see Table 2).

Female targets. We next 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, results showed that women with more upper body strength (β = 0.098, 95% CI = [0.072, 0.124], $p < .001$) and larger biceps (β = 0.211, 95% CI = [0.184, 0.237], $p < .001$) were rated as stronger. Unlike for male targets, height predicted judgments of women targets' strength when accounting for the physical measures (β = 0.088, 95% CI = [0.063, 0.113], $p < .001$). Did targets' race impact raters' strength judgments for female targets? When strength was controlled for, Black women were rated as stronger than White women (β = 0.335, 95% CI = [0.245, 0.425], $p < .001$), and Asian women were rated as weaker than White women (β = –0.501, 95% CI = [–0.604, –0.399], $p < .001$). Thus, raters relied on both individuating information and race when making strength judgments for women as well as men.

Table 3. Raters' Perceptions of Female Targets' Strength (Study 1)

Variable	β	*df*	*SE*	*p*
Targets' height	0.088	873	0.013	< .001
Targets' upper body strength	0.098	870	0.013	< .001
Targets' bicep circumference	0.211	865	0.013	< .001
Targets' race (Asian)	–0.501	887	0.052	< .001
Targets' race (Black)	0.335	865	0.046	< .001
Raters' sex (male)	–0.131	873	0.041	.001
Raters' race (Asian)	0.011	898	0.062	.862
Raters' race (Black)	–0.043	884	0.065	.511
Raters' strength	–0.021	881	0.017	.206
Raters' height	–0.021	868	0.021	.314

Note: Race was dummy coded with White as the reference group.

Do raters use race or individuating information more? Strength judgments for men and women targets were influenced by targets' 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 (targets' race)? We tested this by comparing the variance in strength judgments explained by the physical variables with targets' race (Table 4). In Study 1, physical features explained three to five 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 raters' characteristics explained. Raters' characteristics explained little variability in strength judgments (2% or less), less than the impact of targets' race and far less than the impact of physical features.

Do perceivers' or targets' characteristics drive strength judgments? As a final exploratory test, we examined whether variability in strength judgments was driven primarily by targets' or raters' characteristics (see Table 5). For both male and female targets, more of the variability in ratings of strength was due to differences between targets and not between raters (the same was true for size). 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.

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 preregistered replication.

Study 2: Confirmatory Analysis

Study 2 replicated and extended the results of Study 1. We measured the impact of physical features on strength

Table 4. Variance Explained by Targets' Race, Targets' Physical Features, and Raters' Characteristics

		Men			Women		
Study	Variable	Targets' race	Targets' physical features	Raters' charac-teristics	Targets' race	Targets' physical features	Raters' charac-teristics
Study 1	Strength	.024	.121	.013	.023	.071	.008
Study 2	Strength	.070	.124	.006	.034	.108	.008
Study 2	Height	.022	.126	.017	.014	.175	.020

Note: The table shows pseudo-R^2 values, which were calculated using the methods from Nakagawa and Schielzeth (2013). Values represent the increase in variance explained by adding the set of predictors to the model including all other variables. Targets' race was dummy coded.

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 raters' biases; these did not replicate and are discussed in the Supplemental Material.

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 deviated from our analysis plan in only one respect; we collected data from five Black male undergraduates from Montclair State University in addition to data from our primary participant pool of Michigan State University students to try to meet our sample-size goal.

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. Given the demographics of our participant 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 years (SD = 1.6).

Table 5. Relative Contributions of Raters' and Targets' Variance to Judgments

		Men		Women	
Study	Variable	Targets	Raters	Targets	Raters
Study 1	Strength	.396	.183	.235	.218
Study 2	Strength	.492	.133	.357	.131
Study 2	Height	.361	.173	.414	.124

Note: The table shows intraclass correlations for targets and raters. Larger values indicate that characteristics of the raters or targets drove judgments of strength or height.

Targets. Targets were selected from the 1,660 undergraduate photographs collected in Study 1. To maximize diversity, we selected all nonblurry 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 1,308 White targets.[1] 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 nth 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 years (SD = 0.6). Because we were interested in raters' subjective judgments of height, we digitally edited photos of targets 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: (a) strength, (b) height, and (c) 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 preregistration plan, we do not focus on attractiveness ratings in the current report. To test the role of raters' individual differences in judgments, we gathered the same information about raters as was collected in Study 1.

Power analysis. As in 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 targets' race for Asian and Black men and women (i.e., β = 0.30). 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

Descriptive statistics and correlations for all variables are reported in the Supplemental Material.

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 (β = 0.164, 95% CI = [0.078, 0.251], $p < .001$) and larger biceps (β = 0.290, 95% CI = [0.200, 0.380], $p < .001$) were rated as stronger. Taller men were not rated as stronger when analyses controlled for these physical measures (β = −0.079, 95% CI = [−0.165, 0.008], $p = .076$). Race also impacted strength judgments when analyses controlled for objective strength. Black men were rated as stronger than White men (β = 0.349, 95% CI = [0.151, 0.546], $p < .001$). Asian men were also rated as weaker than White men (β = −0.415, 95% CI = [−0.642, −0.188], $p < .001$). Although a target's race and physical features both influenced raters' judgments, individuating information explained nearly twice as much variance as race (Table 4).

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 (β = 0.382, 95% CI = [0.318, 0.446], $p < .001$). Upper body strength and bicep circumference did not predict height, *p*s > .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 (β = 0.116, 95% CI = [−0.030, 0.262], $p = .123$). However, Asian men were rated as shorter than White men (β = −0.306, 95% CI = [−0.4747, −0.138], $p < .001$). Again, physical features better explained raters' judgments; they accounted for more than five times the amount of variance than raters' race (Table 4).

Female targets.

Strength judgments. Consistent with the results for male targets, raters' judgments of women's strength were predicted by their objective strength (see Table 7). Women with more upper body strength (β = 0.086, 95% CI = [0.029, 0.142], $p = .003$) and larger biceps (β = 0.321, 95% CI = [0.260, 0.382], $p < .001$) were rated as stronger. Taller women were not rated as stronger when analyses controlled for these physical measures (β = 0.021, 95% CI = [−0.032, 0.073], $p = .441$). Race also influenced raters' strength judgments of women. Black women were descriptively rated as stronger than White women (β = 0.115, 95% CI = [−0.012, 0.242], $p = .078$), although this effect was not significant according to our preregistered alpha level (.05). Asian women were rated as significantly weaker than White women (β = −0.385, 95% CI = [−0.521, −0.249], $p < .001$). As with male targets, female targets' physical features explained much more variability (three times) than their race (see Table 4).

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 = 0.425, 95% CI = [0.360, 0.489], $p < .001$). Upper

Table 6. Raters' Perceptions of Male Targets' Strength and Height (Study 2)

	Strength				Height			
Variable	β	*df*	*SE*	*p*	β	*df*	*SE*	*p*
Targets' height	−0.079	157	0.044	.076	0.382	157	0.033	< .001
Targets' upper body strength	0.164	157	0.044	< .001	0.015	157	0.033	.650
Targets' bicep circumference	0.290	157	0.046	< .001	0.055	157	0.034	.107
Targets' race (Asian)	−0.415	157	0.116	< .001	−0.306	157	0.086	< .001
Targets' race (Black)	0.349	157	0.101	< .001	0.116	157	0.075	.123
Raters' sex (male)	0.042	286	0.057	.465	−0.038	287	0.064	.552
Raters' race (Asian)	−0.041	286	0.054	.445	0.030	286	0.060	.623
Raters' race (Black)	−0.143	286	0.055	.010	−0.180	286	0.061	.004
Raters' strength	−0.001	286	0.024	.970	0.057	286	0.026	.031
Raters' height	−0.047	286	0.029	.109	−0.075	287	0.033	.023

Note: Race was dummy coded with White as the reference group.

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

	Strength				Height			
Variable	β	*df*	*SE*	*p*	β	*df*	*SE*	*p*
Targets' height	0.021	195	0.027	.441	0.425	195	0.033	< .001
Targets' upper body strength	0.086	195	0.029	.003	0.018	195	0.036	.614
Targets' bicep circumference	0.321	195	0.031	< .001	0.059	195	0.038	.126
Targets' race (Asian)	–0.385	195	0.069	< .001	–0.311	195	0.085	< .001
Targets' race (Black)	0.115	195	0.065	.078	–0.077	195	0.080	.335
Raters' sex (male)	–0.058	287	0.056	.305	–0.132	286	0.053	.013
Raters' race (Asian)	0.114	287	0.053	.033	0.094	287	0.050	.060
Raters' race (Black)	0.008	287	0.054	.887	–0.100	286	0.051	.050
Raters' strength	0.031	287	0.023	.178	0.078	287	0.022	< .001
Raters' height	–0.062	287	0.029	.034	–0.056	287	0.027	.038

Note: Race was dummy coded with White as the reference group.

body strength and bicep circumference did not predict height, *p*s > .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 = –0.077, 95% CI = [–0.233, 0.079], *p* = .335). However, Asian women were rated as shorter than White women (β = –0.311, 95% CI = [–0.479, –0.144], *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 raters' 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 even when analyses control 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 targets' race and actual physical features with the correlation between targets' 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 Material) revealed that correlations between race and perceived physical features were very similar in size to the correlations based on nationally representative data, indicating that 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 group stereotypes improved accuracy (other than for Black men) is because raters' judgments tracked targets' actual strength and size only moderately. The photographs may not have provided sufficient individuating information to make accurate judgments of size and strength. If such information were perfectly discernable, 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 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 those of past work (Holbrook et al., 2016; Wilson et al., 2017). However, Holbrook et al. (2016) did not ask participants to rate actual targets, and Wilson et al. (2017; Study 1B) had raters make height judgments 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 that 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, rather than inaccurate beliefs (Ashmore & Del Boca, 1981; 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 impact judgments only modestly when individuating information is provided (Jussim et al., 2015; Jussim et al., 1996; Kunda & Thagard, 1996; Madon et al., 1998), our results showed that 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 and 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 raters' characteristics might moderate racial bias in judgments (see the Supplemental Material). Neither raters' 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 perceivers' 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 use of 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 connect 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 in which suspects engage in ambiguously aggressive behaviors while explicitly manipulating targets' 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 suspects' race, size, and strength from police reports. These approaches would better address the degree to which perceptual biases translate into disparate outcomes on the basis of suspects' 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.

Action Editor

Michael Inzlicht served as action editor for this article.

Author Contributions

Both authors conceptualized the experiments and collected the data. D. J. Johnson analyzed the data and wrote the first draft of the manuscript. Both authors edited and reviewed the manuscript and approved the final draft for submission.

Acknowledgments

We thank Joseph Cesario, Bill Chopik, and David Clark for their helpful comments on earlier drafts of this article.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619827529

Open Practices

All data and analysis scripts, as well as various measures used in these studies, have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/8bjcs/. We did not apply for the Open Materials badge because we cannot post the photographs of students used in these studies. The design and analysis plans for Study 2 were preregistered prior to data collection (https://osf.io/bmpcd/). Study 1 was not preregistered because it was an exploratory study. Study 2 was preregistered as a direct replication. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619827529. This article has received the badges for Open Data and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Notes

1. We selected 130 white targets (65 men, 65 women), but one photo was unintentionally not rated.
2. We thank Lee Jussim for providing this helpful suggestion in peer review.

References

Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1–35). Hillsdale, NJ: Erlbaum.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, *67*(1). doi:10.18637/jss.v067.i01

Chen, G., Liu, L., & Yu, J. (2012). A comparative study on strength between American college male and female students in Caucasian and Asian populations. *Sport Science Review*, *21*, 153–165.

Chen, Y. Y., & Geiselman, R. E. (1993). Effects of ethnic stereotyping and ethnically-related cognitive biases on eyewitness recollections of height. *American Journal of Forensic Psychology*, *11*, 13–19.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to "prejudice." *Journal of Personality and Social Psychology*, *88*, 770–789.

Crenshaw, K. (2012). From private violence to mass incarceration: Thinking intersectionally about women, race, and social control. *UCLA Law Review*, *59*, 1418–1472.

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of Blacks. *Journal of Personality and Social Psychology*, *34*, 590–598.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*, 383–386.

Epstein, R., Blake, J. J., & González, T. (2017). *Girlhood interrupted: The erasure of Black girls' childhood*. Washington, DC: The Center on Poverty and Inequality, Georgetown Law.

Fessler, D., Holbrook, C., Tiohkin, L., & Snyder, J. K. (2014). Sizing up Helen: Nonviolent physical risk-taking enhances the envisioned bodily formidability of women. *Journal of Evolutionary Psychology*, *12*, 67–80.

Fryar, C. D., Gu, Q., Ogden, C. L., & Flegal, K. M. (2016). Anthropometric reference data for children and adults: United States, 2011–2014. *Vital and Health Statistics*, *3*(39).

Geiselman, R. E., Lam, L., Lee, T., & Chen, Y. Y. (1995). Recall of perpetrator height and weight by Asian and Caucasian eyewitnesses: Evidence of ethnic stereotyping. *Journal of Police and Criminal Psychology*, *1*, 31–34.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, *78*, 81–91.

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, *10*, 47–66.

Hayes, M. (2014, December 1). Dontre Hamilton's autopsy shows police officer shot him from behind. Retrieved from *Buzzfeed*: http://www.buzzfeed.com/mikehayes/dontre-hamilton-autopsy#.ckLlbLqP2

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 Personality and Social Psychology*, *113*, 513–529.

Holbrook, C., Fessler, D. M. T., & Navarrete, C. D. (2016). Looming large in others' eyes: Racial stereotypes illuminate dual adaptations for representing threat versus prestige as physical size. *Evolution and Human Behavior*, *37*, 67–78.

Jackson, A. S., Ellis, K. J., McFarlin, B. K., Sailors, M. H., & Bray, M. S. (2009). Body mass index bias in defining obesity of diverse young adults: The Training Intervention and Genetics of Exercise Response (TIGER) Study. *British Journal of Nutrition*, *102*, 1084–1090.

Johnson, D. J., & Wilson, J. P. (2019). *Racial biases in attractiveness*. Manuscript in preparation.

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, *102*, 116–131.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54–69.

Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review*, *98*, 54–73.

Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 199–227). New York, NY: Psychology Press.

Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015). Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science*, *24*, 490–497.

Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 281–388). Cambridge, MA: Academic Press.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*, 284–308.

Madon, S., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The accuracy and power of sex, social class, and ethnic stereotypes: A naturalistic study in person perception. *Personality and Social Psychology Bulletin*, *24*, 1304–1318.

Mok, T. A. (1998). Getting the message: Media images and stereotypes and their effect on Asian Americans. *Cultural Diversity and Mental Health*, *4*, 185–202.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142.

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 575–584.

Silva, A. M., Shen, W., Heo, M., Gallagher, D., Wang, Z., Sardinha, L. B., & Heymsfield, S. B. (2010). Ethnicity-related skeletal muscle differences across the lifespan. *American Journal of Human Biology*, *22*, 76–82.

Smith-Evans, L., George, J., Graves, F. G., Kaufmann, L. S., & Frohlich, L. (2014). *Unlocking opportunity for African American girls: A call to action for educational equity*. Washington, DC: National Women's Law Center.

Tennessee v. Garner, 471 U.S. 1 (1985).

Waytz, A., Hoffman, K. M., & Trawalter, S. (2015). A superhumanization bias in Whites' perceptions of Blacks. *Social Psychological & Personality Science*, *6*, 352–359.

Wilkins, C. L., Chan, J. F., & Kaiser, C. R. (2011). Racial stereotypes and interracial attraction: Phenotypic prototypicality and perceived attractiveness of Asians. *Cultural Diversity & Ethnic Minority Psychology*, *17*, 427–431.

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, *113*, 59–80.