Supplemental Material for

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]Michigan State University

[2]University of Maryland at College Park

[3]Montclair State University

Table of Contents

Descriptive Statistics for all Variables

Table S1. Descriptive Statistics for all Variables

		Study 1				Study 2			
		Men		Women		Men		Women	
Variable	Person	M (SD)	Range	M (SD)	Range	M (SD)	Range	M (SD)	Range
Height	Target	71 (2.7)	60-81	66 (2.7)	53-73	70 (2.7)	60-78	66 (2.9)	57-73
Upper B.	Target	40 (12.6)	0-78	20 (5.7)	2-38	39 (13.1)	11-78	19 (5.9)	5-34
Bicep	Target	346 (38.0)	200-490	291 (35.1)	190-480	347 (39.2)	270-460	293 (43.2)	190-480
Asian	Target	44		58		36		56	
Black	Target	59		76		59		74	
White	Target	567		741		64		65	
Sex	Target	670		875		159		195	
Asian	Rater	46		32		51		55	
Black	Rater	37		52		33		56	
White	Rater	416		449		53		55	
Sex	Rater	499		533		137		166	
Strength	Rater	54 (20.1)	0-100	46 (20.0)	0-100	53 (21.2)	10-100	41 (22.0)	2-90
Height	Rater	70 (4.6)	8-94	65 (4.1)	17-87	70 (4.1)	56-85	63 (4.4)	49-74
Strength	Outcome	4 (1.5)	1-7	4 (1.3)	1-7	4 (1.5)	1-7	4 (1.3)	1-7
Height	Outcome	NA	NA	NA	NA	4 (1.3)	1-7	4 (1.3)	1-7

Note. Upper B. refers to upper body strength in kg. Bicep refers to bicep circumference in cm.

Correlations Between Variables (Study 1)

Table S2. Correlations Between Target Variables (Study 1)

Variable	1	2	3	4	5	6
1. Target Height	--	**.16**	**.15**	**-.21**	-.07	**.08**
2. Target Upper Body Strength	**.07**	--	**.33**	-.06	**-.11**	**.24**
3. Target Bicep Circumference	**.09**	**.24**	--	**-.10**	**.13**	**.35**
4. Target Race (Asian)	**-.11**	**-.14**	**-.18**	--	NA	**-.12**
5. Target Race (Black)	.03	-.06	**.19**	NA	--	**.16**
6. Rated Strength	**.13**	**.17**	**.28**	**-.18**	**.13**	--

Note. Correlations for male targets are above the diagonal; correlations for female targets are below the diagonal. N_{Male} = 666, N_{Female} = 872. Correlations with strength ratings control for nonindependence among raters and targets. Correlations with race are relative to Whites. Bolded correlations are significant at $p < .05$.

Table S3. Correlations Between Rater Variables (Study 1)

Variable	1	2	3	4	5	6
1. Rater Sex (Male)	--	.06	-.03	**.18**	**.55**	**-.08**
2. Rater Race (Asian)		--	**-.09**	-.03	**-.11**	-.03
3. Rater Race (Black)			--	-.02	**-.06**	**-.04**
4. Rater Strength				--	**.18**	**-.08**
5. Rater Height					--	**-.05**
6. Strength Ratings						--

Note. N = 1020. Correlations with strength ratings control for nonindependence among raters and targets. Bolded correlations are significant at $p < .05$.

Correlations Between Variables (Study 2)

Table S4. Correlations Between Target Variables (Study 2)

Variable	1	2	3	4	5	6	7
1. Target Height	--	.12	.08	**-.46**	-.15	.03	**.43**
2. Target Upper Body Strength	**.15**	--	**.27**	-.11	-.20	**.21**	.07
3. Target Bicep Circumference	.13	**.38**	--	-.02	**.30**	**.40**	**.13**
4. Target Race (Asian)	**-.24**	**-.30**	**-.31**	--	NA	**-.20**	**-.33**
5. Target Race (Black)	.01	-.13	**.27**	NA	--	**.25**	.03
6. Rated Strength	**.12**	**.25**	**.44**	**-.31**	**.14**	--	**.23**
7. Height Ratings	**.43**	.07	**.13**	**-.27**	**-.02**	**.23**	--

Note. Correlations for male targets are above the diagonal; correlations for female targets are below the diagonal. N_{Male} = 157, N_{Female} = 195. Correlations with strength and height ratings control for nonindependence among raters and targets. Correlations with race are relative to Whites. Bolded correlations are significant at $p < .05$.

Table S5. Correlations Between Rater Variables (Study 2)

Variable	1	2	3	4	5	6	7
1. Rater Sex (Male)	--	.02	-.10	**.27**	**.62**	.00	-.04
2. Rater Race (Asian)		--	**-.46**	.03	**-.12**	.01	**.05**
3. Rater Race (Black)			--	**-.13**	-.07	**-.05**	**-.09**
4. Rater Strength				--	**.34**	.00	.04
5. Rater Height					--	-.03	**-.06**
6. Strength Ratings						--	**.23**
7. Height Ratings							--

Note. N = 303. Correlations with strength and height ratings control for nonindependence among raters and targets. Bolded correlations are significant at $p < .05$.

Rater Predictions for Study 2

In addition to the focal pre-registered hypotheses about race in Study 2, we examined whether we could replicate some of the smaller biases in strength judgments due to individual differences in rater characteristics. Based on the data from Study 1, we also pre-registered the hypotheses that male raters, as well as raters who reported being stronger, would rate targets that were men as weaker (or shorter). We did not replicate these findings. There was no evidence than raters who were men rated targets as weaker (β = .042, 95% CI = [-.070, .154], p = .465), nor did stronger raters rate targets as weaker (β = -.000, 95% CI = [-.047, .046], p = .970). We also tested whether raters who were men and who reported themselves as stronger would rate men as shorter. Neither hypothesis was supported; in fact, stronger raters rated targets as *taller* (β = .057, 95% CI = [.005, .109], p = .031).

Despite improving the internal validity of our research design in Study 2 by recruiting a more racially balanced sample of participants and targets, increasing the number of targets, and moving the task into the laboratory, we did not replicate the findings from Study 1 that rater characteristics impacted judgments of size or strength. This is perhaps somewhat unsurprising given that in both studies the variance explained by rater characteristics was quite small (1%) and outweighed considerably by the physical characteristics of the targets. Similarly, the intra-class correlations revealed that perceptions of size and strength are driven primarily by target characteristics rather than rater characteristics (see Table 2). This is consistent with work showing that appearance-based appraisals are less driven by perceiver characteristics (Hehman, Sutherland, Flake, & Slepian, 2017).

Does Rater Sex or Race Moderate Judgments of Size and Strength?

As an exploratory analysis we tested whether racial bias in judgments of size and strength depended on rater characteristics. Specifically, we tested whether these biases were moderated by rater sex, race, and their interactions. We estimated the models specified in the main text but added additional coefficient terms that allowed racial bias to vary by participant sex (participant sex and target race interaction), participant race (participant race and target race interaction), and participant race and sex (three-way interaction). Because these analyses were exploratory, we only discuss results that were significant at an alpha level of .05 *and* had an effect size of greater than .2 (in standardized regression units). This approach is sensible given that the degrees of freedom are inflated for these models. This is because our analyses only model variation by participant and target (i.e., in the intercepts) and not variation in the fixed effects (i.e., random slopes). While more complex models would better account for this variability, our goal was not to test whether the effects of race or sex vary by participant or target and so we did not include random effects in the model.

The coefficients for these exploratory models can be seen in Tables S6 – S9. Tables S6 and S7 detail the results from Study 1, while Tables S8 and S9 detail the results from Study 2. When using the aforementioned thresholds, racial biases varied based on rater characteristics in only two cases. Racial bias in strength judgments for Asian men was weaker for Asian raters, regardless of sex (see Table S6). Racial bias in height judgments for Asian men was also weaker for raters who were men, regardless of their race (see Table S8). Furthermore, including these interactions resulted in a negligible increase in variance explained; in no model did the variance explained increase by more than .02%. Thus, we saw little evidence that rater characteristics moderated the effects of racial biases on perceptions of size and strength.

Table S6. Study 1 Multilevel Regression Summary Table for Male Targets

Variable	β	*df*	*SE*	*p*
Target Height	-0.004	644	0.021	.837
Target Upper Body Strength	0.157	644	0.021	< .001
Target Bicep Circumference	0.276	667	0.022	< .001
Target Race (Asian)	0.326	680	0.083	< .001
Target Race (Black)	0.500	679	0.072	< .001
Rater Sex (Male)	-0.148	1022	0.037	< .001
Rater Race (Asian)	-0.123	1046	0.057	.031
Rater Race (Black)	-0.145	1023	0.054	.007
Rater Strength	-0.070	997	0.015	< .001
Rater Height	0.001	971	0.018	.943
TA × RM	0.107	15909	0.046	.020
TB × RM	0.015	15902	0.041	.710
TA × RA	0.207	15850	0.083	.013
TB × RA	-0.096	15884	0.078	.215
TA × RB	0.018	15842	0.075	.815
TB × RB	0.011	15904	0.065	.861
RM × RA	0.500	1042	0.112	.407
RM × RB	-0.148	1016	0.107	.552
TA × RM × RA	-0.147	15833	0.166	.376
TB × RM × RA	0.128	15894	0.155	.411
TA × RM × RB	0.249	15843	0.150	.097
TB × RM × RB	-0.034	15888	0.130	.796

Note. Race was dummy coded with Whites as the reference group. T = Target, R = Rater, A = Asian, B = Black, M = Male

Table S7. Study 1 Multilevel Regression Summary Table for Female Targets

Variable	β	*df*	*SE*	*p*
Target Height	0.088	873	0.013	< .001
Target Upper Body Strength	0.098	870	0.013	< .001
Target Bicep Circumference	0.211	864	0.013	< .001
Target Race (Asian)	-0.522	935	0.053	< .001
Target Race (Black)	0.339	912	0.047	< .001
Rater Sex (Male)	-0.146	896	0.043	.001
Rater Race (Asian)	-0.002	925	0.064	.973
Rater Race (Black)	-0.053	912	0.066	.425
Rater Strength	-0.022	881	0.017	.194
Rater Height	-0.021	868	0.021	.326
TA × RM	-0.055	19900	0.047	.240
TB × RM	0.095	19784	0.040	.018
TA × RA	0.197	19925	0.086	.023
TB × RA	-0.094	19896	0.072	.191
TA × RB	0.113	19775	0.090	.208
TB × RB	0.047	19890	0.078	.547
RM × RA	0.117	922	0.125	.350
RM × RB	0.000	913	0.133	.997
TA × RM × RA	0.036	19922	0.173	.835
TB × RM × RA	-0.131	19820	0.143	.361
TA × RM × RB	0.200	19828	0.181	.269
TB × RM × RB	-0.060	19908	0.156	.699

Note. Race was dummy coded with Whites as the reference group. T = Target, R = Rater, A = Asian, B = Black, M = Male

Table S8. Study 2 Multilevel Regression Summary Table for Male Targets

	Strength				Height			
Variable	β	*df*	*SE*	*p*	β	*df*	*SE*	*p*
Target Height	-0.079	156	0.044	.076	0.382	156	0.033	< .001
Target Upper Body Strength	0.165	156	0.044	< .001	0.015	156	0.033	.651
Target Bicep Circumference	0.290	156	0.046	< .001	0.055	156	0.034	.104
Target Race (Asian)	-0.488	165	0.118	< .001	-0.322	176	0.088	< .001
Target Race (Black)	0.364	165	0.102	< .001	0.127	175	0.077	.100
Rater Sex (Male)	0.035	327	0.084	.676	-0.101	327	0.094	.281
Rater Race (Asian)	-0.060	337	0.057	.286	0.014	338	0.063	.821
Rater Race (Black)	-0.168	337	0.058	.004	-0.173	338	0.065	.008
Rater Strength	-0.001	286	0.024	.963	0.057	286	0.026	.032
Rater Height	-0.048	286	0.029	.104	-0.076	286	0.033	.021
TA × RM	0.133	12325	0.048	.005	0.210	12329	0.053	< .001
TB × RM	0.017	12327	0.041	.684	0.033	12331	0.046	.473
TA × RA	0.167	12330	0.035	< .001	0.071	12335	0.039	.070
TB × RA	-0.049	12328	0.030	.099	0.001	12331	0.034	.971
TA × RB	0.069	12325	0.036	.057	0.004	12329	0.040	.916
TB × RB	0.014	12325	0.031	.649	-0.035	12327	0.035	.315
RM × RA	0.007	340	0.111	.951	0.046	341	0.124	.709
RM × RB	-0.069	338	0.115	.550	-0.014	339	0.128	.914
TA × RM × RA	-0.153	12329	0.070	.028	-0.092	12334	0.078	.238
TB × RM × RA	-0.014	12329	0.060	.821	0.070	12333	0.067	.295
TA × RM × RB	-0.074	12326	0.072	.303	-0.062	12329	0.081	.444
TB × RM × RB	0.067	12325	0.062	.283	-0.028	12328	0.070	.689

Note. Race was dummy coded with Whites as the reference group. T = Target, R = Rater, A = Asian, B = Black, M = Male

Table S9. Study 2 Multilevel Regression Summary Table for Female Targets

	Strength				Height			
Variable	β	*df*	*SE*	*p*	β	*df*	*SE*	*p*
Target Height	0.021	194	0.027	.441	0.425	194	0.033	< .001
Target Upper Body Strength	0.085	194	0.029	.004	0.018	195	0.036	.618
Target Bicep Circumference	0.321	194	0.031	< .001	0.058	195	0.038	.128
Target Race (Asian)	-0.448	228	0.072	< .001	-0.339	214	0.087	< .001
Target Race (Black)	0.112	227	0.067	.098	-0.039	214	0.082	.630
Rater Sex (Male)	-0.070	359	0.085	.411	-0.140	361	0.080	.080
Rater Race (Asian)	0.114	374	0.057	.046	0.106	377	0.054	.048
Rater Race (Black)	-0.069	378	0.059	.238	-0.105	380	0.055	.058
Rater Strength	0.031	286	0.023	.187	0.077	286	0.022	< .001
Rater Height	-0.064	286	0.029	.027	-0.058	286	0.027	.033
TA × RM	0.102	14618	0.050	.042	0.102	14614	0.048	.032
TB × RM	0.048	14630	0.046	.296	-0.017	14626	0.044	.699
TA × RA	0.055	14616	0.036	.132	0.039	14612	0.035	.264
TB × RA	-0.040	14624	0.034	.233	-0.060	14619	0.032	.064
TA × RB	0.152	14617	0.038	< .001	0.058	14613	0.036	.108
TB × RB	0.056	14622	0.035	.109	-0.044	14619	0.033	.193
RM × RA	-0.002	378	0.112	.986	-0.037	381	0.105	.727
RM × RB	-0.077	379	0.116	.508	-0.081	382	0.109	.461
TA × RM × RA	-0.010	14615	0.073	.894	0.051	14611	0.069	.458
TB × RM × RA	0.049	14623	0.068	.469	0.083	14618	0.065	.199
TA × RM × RB	-0.095	14617	0.075	.206	-0.072	14613	0.072	.319
TB × RM × RB	-0.052	14625	0.070	.464	0.127	14621	0.067	.057

Note. Race was dummy coded with Whites as the reference group. T = Target, R = Rater, A = Asian, B = Black, M = Male

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 valid information about group differences may increase accuracy. We tested this in an exploratory analysis by comparing the correlation between target race and *actual* physical features to the correlation between target race and *perceptions* of physical features. When the correlation between race and perceptions is larger (smaller) than the correlation between race and actual differences, individuals' overestimate (underestimate) real group differences.

We used the conceptual model outlined in Madon et al. (1998; see also Jussim, 1991) to estimate the correlation between race and perceived physical features (Figure S1). This model states that individuating information (Path A) and stereotypes (Path B) independently impact perceptions of size and strength. Using path tracing rules, the correlation between race and *actual* size or strength (r_{RA}) can be decomposed as: r_{RA} = Path B + r_{RP} (Path A), where r_{RP} is the correlation between target race and *perceived* size or strength.

This decomposition makes it possible to compare the correlation between race and perceived physical features with and without the effect of race stereotypes (i.e., setting Path B to 0). If the predicted correlation between race and perceived size or strength (r_{RP}) is closer to the actual correlation between race and size or strength (r_{RA}) when the effect of stereotypes is set to zero race stereotypes *decrease* accuracy. However, if the predicted correlation between race and perceived size or strength (r_{RP}) is farther from the actual correlation between race and size or strength (r_{RA}) when the effect of stereotypes is set to zero race stereotypes *increase* accuracy.

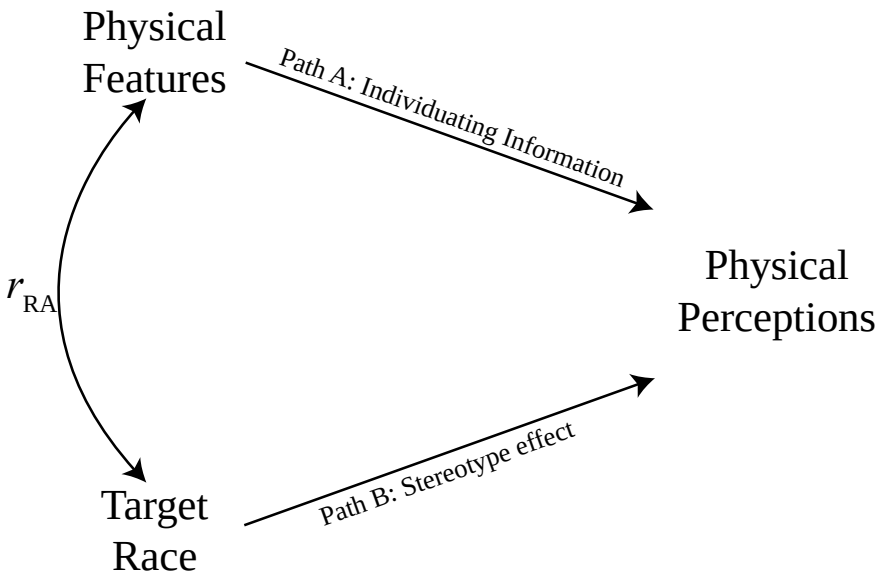

Figure S1. Conceptual model relating targets' physical features and race to raters' perceptions of strength and size. r_{RA} refers to the correlation between target race and *actual* size or strength. r_{RP} (not photographed) refers to the correlation between target race and *perceived* size or strength.

We estimated this model by regressing perceptions of physical features on individuating information and target race within a multilevel regression controlling for targets and raters. We estimated different models comparing each race and sex category to the White target category (e.g., Asian men to White men). All predictors were standardized. Results are reported in Table S10, alongside the predicted correlation between race and perceived strength, calculated two ways. r_{RP} reflects the correlation between race and perceived strength given the effect of stereotypes observed in the sample.[1] r_{RP0} reflects a correlation assuming raters do *not* use group membership (i.e., the stereotype effect is 0). These correlations are compared to the correlation between race and physical features (r_{RA}) based on nationally representative data from the NHNES (N = 11,196 adults). We chose to compare the correlations between race and perceived strength to this value over the sample value for two reasons. First, raters only saw a small subset

[1] These expected correlations differ somewhat from the observed correlations between race and physical perceptions (see the Supplemental Materials). Larger discrepancies between predicted and observed correlations indicate poorer model fit in a structural equation modeling framework.

of the targets in both studies, so the sample correlation between race and physical features differed for each rater. More importantly, we assumed raters' stereotypes would mirror real world experiences (rather than implicit knowledge of our sample's characteristics). Such experiences are better captured by representative data than the nonrepresentative sample in our study.

Table S10. Correlations between Race and Physical Features or Physical Perceptions (Study 2)

		Strength					Height				
Sex	Race	Race	Bicep	r_{RP}	r_{RA}	r_{RP0}	Race	Height	r_{RP}	r_{RA}	r_{RP0}
Male	Asian	-.18	.38	-.28	-.28	-.10	-.13	.43	-.29	-.36	-.15
Male	Black	.14	.35	.18	.10	.04	.08	.38	.07	-.03	-.01
Female	Asian	-.18	.41	-.30	-.29	-.12	-.16	.43	-.31	-.34	-.14
Female	Black	.05	.32	.12	.20	.06	-.03	.45	-.01	.03	.01

Note. White is the reference category for all race effects. Race, Bicep, and Height reflect the standardized independent contribution of these variables to ratings of strength or height. r_{RP} = predicted correlation between race and perceived strength. r_{RA} = actual correlation between race and bicep circumference or height based on data on all adults over 19 from the NHNES (2011 – 2016). r_{RP0} = predicted correlation between race and perceived strength *if a person did not use race stereotypes*. If stereotypes aid accuracy r_{RP} should be closer to r_{RA} than r_{RP0}.

Consistent with the main results, both race and individuating information independently predicted perceptions of strength and size. Overall, predicted correlations between race and perceived physical features (r_{RP}) were remarkably close to the correlations from national data (r_{RA}), 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 (r_{RP} = .18) relative to the actual correlation from nationally representative data (r_{RA} = .10). The same pattern occurred for the relationship between race and height (r_{RP} = .07 vs. r_{RA} = -.03).

In sum, for all groups other than Black men, relying on valid information about group

differences improved participants' accuracy when making judgments of size and strength. The reason these group stereotypes improve accuracy is because raters' judgments only moderately reflected individuating information ($.30 < \beta s < .45$). It is likely the photographs did not provide the perfect information needed to make accurate judgments of size and strength. If such information were perfectly clear any remaining stereotype effect would decrease accuracy rather than increase it.

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

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

Madon, S., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. 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.