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

John Paul Wilson[1], Kurt Hugenberg[2], & Nicholas O. Rule [3]

Montclair State University[1] Miami University[2] University of Toronto[3]

Please address correspondence to:

John Paul Wilson
Department of Psychology
Montclair State University
1 Normal Ave
Montclair, NJ 07043
wilsonjoh@montclair.edu

Author Note

This research is supported by the Natural Sciences and Engineering Research Council of Canada to NOR and NSF award BCS-1423765 awarded to KH.

Submitted January 6, 2017

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

Abstract

Black men tend to be stereotyped as threatening and, as a result, may be disproportionately targeted by police even when unarmed. Here, we found evidence that biased perceptions of young Black men's physical size may play a role in this process. The results of seven studies showed that people have a bias to perceive young Black men as bigger (taller, heavier, more muscular) and more physically threatening (stronger, more capable of harm) than young White men. Both bottom-up cues of racial prototypicality and top-down information about race supported these misperceptions. Furthermore, this racial bias persisted even among a target sample from whom upper-body strength was controlled (suggesting that racial differences in formidability judgments are a product of bias rather than accuracy). Biased formidability judgments in turn promoted participants' justifications of hypothetical use of force against Black suspects of crime. Thus, perceivers appear to integrate multiple pieces of information to ultimately conclude that young Black men are more physically threatening than young White men, believing that they must therefore be controlled using more aggressive measures.

Keywords: person perception, race bias, intergroup relations, motivated perception

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

On April 30th, 2014, an unarmed Black man named Dontre Hamilton was shot 14 times and killed by a White police officer in Milwaukee, Wisconsin. The officer later testified that Hamilton had a "muscular build" and "most definitely would have overpowered…me or pretty much any officer I can think of, to tell you the truth. He was just that big, that muscular…" (Hayes, 2014). This account is contradicted by the autopsy, in which the medical examiner reported that Hamilton was 5'7" and 169 pounds.

Similarly tragic examples have mounted in recent years. For example, Tamir Rice was a 12-year-old Black boy shot and killed by a Cleveland police officer in November, 2014, while he played with a toy gun in a park. A representative of the Cleveland Police later explained the shooting by saying "Tamir Rice is in the wrong. He's menacing. He's 5-feet-7, 191 pounds. He wasn't that little kid you're seeing in pictures. He's a 12-year-old in an adult body," (Schultz, 2015). In the wake of the 2012 shooting of Trayvon Martin (an unarmed Black teen in Florida), images circulated depicting Martin as older and larger than he was. In one notorious example, people widely shared a photograph of a man with facial tattoos in what was purported to be an up-to-date representation of Martin. In fact, it was a rap musician known as Game who was in his 30s in the photograph (Capehart, 2013). In each of these cases, people attempting to explain the events post hoc included appeals to the size and threat posed by the targets. Although it is impossible to know to what extent these single anecdotes may represent a broader phenomenon, the Hamilton, Rice, and Martin shootings join a long history of other tragedies that suggest a troubling bias in perceptions of the physical formidability and threat posed by Black men and boys (Geller, 1982).

With these and many more cases of unarmed Black men being shot by police, protests have swept across the United States and public trust in law enforcement is at its lowest in decades (Jones, 2015). Americans have thus struggled to understand these sustained patterns of force decisions by police. One clear theme in the officers' descriptions of many of the shooting victims is their physical size and overall formidability. As a possible explanation for this tendency, we proposed that the stereotype of young Black men as physically threatening (Cottrell & Neuberg, 2005; Harris-Lacewell, 2001), less innocent (Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014), and perhaps even physically "superhuman" (Waytz, Hoffman, & Trawalter, 2014) may create conditions that prepare perceivers to show distorted perceptions of Black men's physical size and formidability. The effect of these stereotypes would likely be magnified by the functionality of hypervigilance to threat in the service of error management (Haselton & Buss, 2000). From this perspective, perceivers should show a bias toward perceiving danger when none exists, such as in decisions to shoot unarmed Black men in first-person shooter simulations (Correll, Park, Judd, & Wittenbrink, 2002; Plant, Goplen, & Kunstman, 2011). Heightened sensitivity to possible danger likely is not limited to binary armed/not-armed decisions, however. In fact, perceivers much more commonly make judgments not about whether someone is armed, but about how large and physically formidable a person is, and how that formidability may translate into potential aggression.

In the present research, we examined possible racial biases in such judgments. Put simply, we hypothesized that people would misperceive young Black men as physically larger and more formidable than young White men of veridically comparable size. Further, we proposed that this race-based bias in perception of physical formidability could lead to race-based biases in decisions about the use of force. We tested this across seven studies

demonstrating that people perceive young Black men as taller, heavier, more muscular, more physically formidable and more capable of physical harm than young White men of the same actual size; and that this bias in physical size perception can influence the decision to use force against them.

Race Stereotypes in Person Perception

A long history of psychological research has found that, compared to Whites, Black people are subject to automatic negative stereotypes and prejudice (Devine, 1989). Beyond broad negativity, however, stereotypes of Blacks, and particularly young Black men, often focus on violence, threat, and crime. For example, Black men are more likely than White men to be misremembered as carrying a weapon (Allport & Postman, 1947), are more likely to facilitate the visual recognition of a weapon (Payne, 2001), are more likely to be shot mistakenly in a virtual crime scenario while holding an innocent object such as a soda can (Correll et al., 2002; Correll, Wittenbrink, Park, Judd, & Goyle, 2011; cf. James, Vila, & Daratha, 2013), and are more likely to activate concepts related to crime (Eberhardt, Goff, Purdie, & Davies, 2004). In ambiguous contexts, Black men are more likely than White men to be seen as threatening or aggressive (e.g., Duncan, 1976; Hugenberg & Bodenhausen, 2003, 2004; Sagar & Schofield, 1980). Indeed, representations of Black individuals, and particularly young Black men, in popular media reliably depict them as aggressors (e.g., Dixon & Linz, 2000; Chiricos & Escholz, 2002), and these representations have real implications for intergroup relations (Johnson, Trawalter, & Dovidio, 2000). For instance, the physical features associated with African-Americans activate race-related stereotypes regardless of the target's actual race (Blair, Judd, Sadler, & Jenkins, 2002), leading to consequences as severe as exacerbated criminal sentences and even execution (Blair, Judd, & Chapleau, 2004; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006).

Although such research has established the role of race and race-signifying features in signaling the stereotype of Black men as dangerous in the minds of Whites, few studies have directly linked race to perceptions of physical size and formidability. Indeed, much of this research (though certainly not all—see Blair et al., 2002) has focused on the connection between race and stereotypes in broad and general terms, with little attention to biased perceptions of the targets' actual threat-relevant physical characteristics. Such knowledge is critical for understanding decisions to engage in physical force against a potentially dangerous person. For example, in the police shootings of unarmed Black civilians described above, police officers decided to shoot based on a subjective perception that their lives were in danger. This perception would be consistent with the assessment that a "reasonable" officer witnessing the incident would believe the action to be justified, a criterion required among law enforcement agents before they may use lethal force (Klinger & Brunson, 2009). In many instances, this is relatively clear – a suspect brandishes a weapon in the direction of police. But in other cases, such as that involving Dontre Hamilton, the officer relies on a qualitative assessment of the target's physical size and strength to gauge the magnitude of threat and the need for force. These circumstances are more ambiguous and, thus, more susceptible to errors resulting from systematic biases.

From Threat Stereotypes to Perceptual Bias

Despite the robust cultural stereotype of Black men as threatening, little work has investigated how these stereotypes may translate into biased perceptions of their physical formidability. From an evolutionary perspective, people are particularly motivated to attend closely to and minimize threats to their personal safety, meaning that humans have likely evolved mechanisms to detect threat, including threats from other humans (Neuberg, Kenrick, & Schaller, 2011).

Theorists have argued that this threat detection system works like a smoke detector device (Nesse, 2005). Smoke detectors monitor the environment for signs of fire. Sometimes they accurately detect a fire, but sometimes they may either miss the presence of a fire (false negative) or erroneously signal a fire when none exists (false positive). Although false positives can be distressing, they are much less damaging to the perceiver than false negatives. Human psychological error management processes thus strive to minimize the more costly of these two error types (Haselton & Nettle, 2006). As a result, the desire for self-protection motivates people to vigilantly attend to signs of potential threats. For example, White perceivers show greater sensitivity to signs of anger in outgroup Black faces compared to ingroup White faces (e.g., Maner et al., 2005), and White perceivers appear more attentionally vigilant to the faces of Blacks than Whites (e.g., Trawalter, Todd, Baird, & Richeson, 2008). It is an empirical question as to whether these error management processes are limited to outgroup perceptions, however (Olsson, Ebert, Banaji, & Phelps, 2005). To the extent that threat-related stereotypes exist regarding a particular group (e.g., young black men), targets of that specific group may be subject to biased perceptions regardless of the group membership of the perceiver. In fact, there is some evidence that Black participants show the shooter bias with a magnitude similar to that of White participants (Correll et al., 2002), though this question has not been tested with adequate power to reliably find participant ethnicity differences.

Here, we report the results of a series of studies evidencing another perceptual bias arising from hypervigilance to threat cues: biased judgments of physical size and formidability. When assessing the formidability and potential threat of other people, one must consider information about their physical size, strength, and ability to inflict physical harm. Recent research has shown that people make such assessments with some accuracy. For example,

perceivers' estimates of men's physical strength correlate reliably with their actual physical strength when looking at images of the entire person (r = .71), of the targets' bodies (r = .66), or of even just the targets' faces (r = .45; Sell, Cosmides, Tooby, Sznycer, von Rueden, & Gurven, 2009). However, as is often the case in person perception (e.g., Gilovich, Medvec, & Savitsky, 2000, Study 3; Klein & Kunda, 1993), such accuracy occurs simultaneously with robust biases.

Indeed, a variety of contextual cues can bias judgments of others' formidability. For instance, perceivers who are physically incapacitated envision antagonists as larger (Fessler & Holbrook, 2013), perceivers viewing pictures of hands holding weapons conceptualize the person wielding them as larger than people holding non-weapon objects (Fessler, Holbrook, & Snyder, 2012), and people inflate terrorists' estimated physical attributes when they are depicted as being successful or powerful (Holbrook & Fessler, 2013). Generally speaking, people misperceive targets as larger and more formidable as a function of how much they believe their own capabilities to be inferior (Fessler, Holbrook, & Gervais, 2014). These findings support the idea that perceivers connect others' status, power, or physical threat with others' size in ways that distort mental representations of other people. All of these findings suggest that characteristics independent of a target's actual physical size can influence threat-related judgments. Indeed, recent research has provided initial support for our hypothesis. Holbrook, Fessler, and Navarrete (2016) found that participants imagined Black men to be larger than White men. However, unlike the current research, this recent work examined biases in abstract conceptualizations rather than in actual (i.e., sensory) perceptions. This is a very important distinction. Holbrook et al. (2016) provided evidence that people tend to envision Black men as more physically formidable and aggressive on the basis of text vignettes, but they did not investigate whether this bias extends to person perception. From their data, one might conclude

that people tend to imagine Black men as posing a more formidable threat than White men. The present research tests whether such biases will remain when perceivers view actual individuals. Observing a racial bias in perceived physical formidability from actual photographs would establish that such processes occur not just in the "mind's eye" but in the everyday assessment of other people.

By showing that the bias exists with actual images of the faces and bodies of targets, this provides a number of key advantages. First, this provides a much more conservative test of the hypothesis. Indeed, past research has reliably shown that the effects of stereotypes can be dramatically attenuated or even eliminated in the presence of strong individuating information (e.g., Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, Ortiz, 1982). As noted above, perceivers often show a robust ability to extract objective markers of physical formidability from the faces and bodies of men, yet this past race-based size bias has only been demonstrated in imagined settings, which leaves open the key question of whether the purported bias actually affects person perception. Demonstrating that the race-based size bias exists even in the presence of actual target faces and bodies is important to extend the basic size bias to ecological settings. Second, by demonstrating a race-based bias in the size of Black versus White men using actual faces and bodies, we can also draw a direct link between the existing phenomenon and the well established literature on biases in threat and motivated perception.

Threat and Motivated Perception

Related research on how motives can influence perception bolsters the possible link between error management and biased judgments of the properties of one's physical environment. Cole, Balcetis, and Dunning (2013), for example, found that individuals estimate threat-related objects (including aggressive men) as physically closer to them than non-

threatening objects. In other work, Cesario and Navarette (2014) found that White participants surrounded by ingroup members (who, thus, presumably felt safe) imagined that Detroit (a city with a very high Black population) was further away than they did when they were alone. The individual sense of threat versus safety therefore appears to influence the perceived proximity (and consequent ability to harm) of a threatening individuals or groups.

Perceivers may exhibit overly sensitive perceptions of threat signals from people belonging to groups stereotyped as threatening. For instance, they judge targets that can harm them as closer than they actually are, likely because it serves their goal of avoiding or confronting the threat. However, judgments of physical distance are only one (exogenous and indirect) component of a target's threat value. Here, we were interested in assessing judgments of *endogenous* target characteristics that are directly associated with physical threat: targets' physical size and formidability. We posit that distorted size perceptions may arise from a combination of related factors: stereotypes linking Blacks to size and superhumanness (e.g., Holbrook et al., 2016; Waytz et al., 2014), and associations linking young Black men to threat (e.g., Hugenberg & Bodenhausen, 2003; 2004; Olsson et al., 2005), which could result in a "looming" effect in perceptual judgment. Both factors could act either in combination or in isolation to produce effects whereby young Black male targets "loom" in perceptual judgment, relative to young White men.

Importantly, if perceivers merely tend to assume that young Black men are more threatening and larger than White men in the abstract, then one might suppose that these assumptions would not bear out when judging photographs of them. That is, individuals'sensory perceptions of the targets might be accurate, despite their cognitive assumptions. We suspected that this would not be the case. Instead, we predicted that perceivers' assumptions about targets'

size and threat would influence their perceptions of Black versus White men's physical formidability, thereby showing evidence of racially biased perceptions. From an error management perspective, it would be functional for perceivers to misperceive the size of potential adversaries, as it would facilitate acting to dispense the threat. We were interested in testing this possible bias, which would present novel, critical evidence that group differences in physical size judgments transcend mere cognitive associations (stereotypes) about the groups and are based—at least in part—on a tendency to misperceive physical size from concrete visual evidence. Furthermore, assessments of targets along evaluatively neutral dimensions (such as physical size) may then feed into evaluatively meaningful assessments, such as whether police are justified in using physical force to subdue a target. Specifically, we predicted that race-based stereotypes and threat would manifest in biased judgments of individuals' physical size, which would then result in biased judgments of these same individuals' ability to inflict physical harm in an altercation, thus promoting racial bias in justifications of hypothetical use of force by police.

A Bias or a Kernel of Truth?

Finally, it is important to consider whether race differences in judgments of size and formidability might reflect bias or accuracy. In fact, although we have framed this research using language suggestive of racial bias in judgments of size and formidability, some researchers have vigorously argued that a "kernel of truth" might underlie stereotypes (Jussim, Crawford, & Rubinstein, 2015; Jussim, Crawford, Anglin, Chamberse, Stevens, & Cohen, in press). Such a perspective would posit that stereotypes regarding the threat value of young Black men at least partially stem from accurate representations borne from a combination of genetic differences and socioeconomic disparities (e.g, food quality, environmental health challenges). These partially

accurate stereotypes could then lead to differences in the perceived physical formidability of Black versus White male targets, which could themselves result from true group differences.

Although the current work is not equipped to completely rule out such an explanation, past biometric research indicates that, at least in terms of height and weight, Black and White men in America are nearly equivalent in size. The 2012 Center for Disease Control report on summarizing the *2007-2010 National Health and Nutrition Examination Survey* data places the average height of non-Hispanic White men (20 years or older) at 177.4 cm and 90.4 kgs, and of non-Hispanic Black men at 176.4 cm and 90.4 kgs (Fryar, Gu, & Ogden, 2012). Thus, although our data cannot rule out the possibility of an objective size difference between Black and White men, what data are available indicate that the objective size difference actually operates in the *opposite* direction of what we predict here (with White men being slightly taller than Black men), making ours an especially conservative test of a stereotypic bias hypothesis. Further, in our studies, we attempt to account for the actual physical size of the stimuli, equating stimuli height and weight. Notably, the CDC data do not include objective measures of actual physical muscularity, however, in Study 1E in the current work we also demonstrate that the race-based size bias occurs when controlling for targets' objective upper body strength (i.e., maximum bench press weight), which we discuss in more detail below. Thus, as with most stereotypes, we cannot rule out a kernel of truth hypothesis in the current work, however, any size bias in judgment that is observed does not appear to be borne out objective differences as measured in cross-sectional, nationally representative health examinations.

Second, when considering a kernel of truth hypothesis, it is also possible for there to be simultaneous group-level accuracy in stereotypes, which when applied to individuals, can create an unfair bias in social judgment. Under the circumstance that Black men in America actually

were larger than White men (which again, the data indicate they are not), believing Blacks to be larger than Whites would be an accurate group-level stereotype. Yet, when a Black and a White exemplar are actually the same size, applying this stereotype, and mis-perceiving the Black exemplar as larger than the White exemplar would be an example of how a group stereotype can bias social judgments about individuals, in spite of potential group level accuracy.

The Current Research

To investigate a possible racial bias in perceptions of physical size and formidability, we asked participants to make judgments of Black and White male targets from photos. We began in Study 1 by assessing the estimated height, weight, muscularity, and strength of Black and White targets (Studies 1A, 1B, 1C, 1D, and 1E). We then explained the relationship between misperceptions of size and threat in Study 2 by examining perceptions of targets' capacity for causing physical harm, and repeated this while accounting for racial prejudice by measuring White participants' racial bias in Study 3. We probed multiple potential mechanisms for these effects by testing effects among both White and Black perceivers in Study 4. We then extended the link between formidability and size perceptions to judgments of the appropriateness of hypothetical police force in interactions with these targets in Study 5, and sought to test whether the effects of race occur from feature-based racial prototypicality and/or category-based beliefs about targets' race. Specifically, we examined whether variability in how Afrocentric a person looks predicts how formidable he seems in Study 6 and investigated whether category-based race information (simply believing that the same target is either Black or White) can bias size perceptions even in the absence of any actual physical differences in the stimuli in Study 7. The current data collectively provided consistent, novel evidence demonstrating the existence of a stereotype of young Black men as larger, stronger, and more physically formidable than young

White men, that this stereotype distorts perceptions of size and physical formidability, which in turn can distort decisions about the use of force against Black crime suspects.

Study 1

To test our hypothesis that a target's race distorts perceptions of his physical size, we first sought to establish whether young Black men are judged as physically larger and more formidable than young White men. In Study 1A, we thus asked participants to estimate the height and weight of a series of young Black and White men from photos of their faces. We repeated this in Study 1B, replacing the faces with individuals for whom we knew their actual height and weight to address the possibility that *actual* racial differences in size might explain *perceived* racial differences in size. In Study 1C, we extended the investigation by asking participants to match each face in Study 1B to a putative body that varied in apparent muscularity and physical formidability, allowing us to link size perceptions to characteristics more closely related to physical threat. Finally, we directly assessed estimates of targets' physical strength from their faces to more closely bridge perceptions of size to perceptions of physical ability and harm capability in Study 1D, and repeated this in Study 1E while controlling for measures of the targets' actual physical strength.

Study 1A

To explore whether people generally perceive Black men as larger than White men, we asked participants to estimate the height and weight of a series of Black and White men's faces in Study 1A.

Method.

Participants. In each study, we followed a heuristic of recruiting at least 60 participants per sample, which would provide greater than 90% power to detect a small-medium effect size (*d*

= .40). In this study, we recruited 125 US residents from Amazon's Mechanical Turk (MTurk) because we asked two separate samples to judge either height or weight. We did not analyze the data from 14 participants who identified as Black,[1] leaving 111 participants in total (63 male, 48 female, M_{age} = 34.3 years, SD = 11.3). We randomly assigned participants to estimate either the targets' height (n = 55) or weight (n = 56). After providing informed consent, the participants read that they would view a series of faces for whom they would guess each person's height or weight.

Stimuli. We presented participants with color photographs of 200 male faces (100 White, 100 Black) borrowed from previous person perception research (e.g., Eberhardt et al., 2004). Each face exhibited a neutral expression and each image measured approximately 150 × 200 pixels (72 pixels/inch) in size.

Procedure. The participants viewed each of the 200 images in the same block in random order. For each judgment, we presented the target image above a slider scale for each height (weight) rating. The scale ranged from 120 to 300 lbs for the weight ratings, with the possible responses in increments of 1 lb. Height values ranged from 60 inches (5 feet 0 inches) to 78 inches (6 feet 6 inches), with the possible responses in increments of 1 inch. We selected these ranges to include plausible values for men appearing to be of average size (McDowell, Fryar, Ogden, & Flegal, 2008).

Results and discussion. Because repeated-measures designs in which participants respond to samples of stimuli can yield unacceptable false positive rates when analyzed with

[1] Excluding Black participants did not meaningfully change the results of any of the studies reported here. We directly addressed the issue of Black participants' biases in Study 4.

simple ANOVAs (Judd, Westfall, & Kenny, 2012), we used mixed-model analyses that cross participants and stimuli as random factors, effect-coding race (Black = .5, White = -.5).

As predicted, we observed a significant difference in height such that participants estimated Black targets (M = 70.36 in., SD = 1.47) as taller than White targets (M = 69.89 in., SD = 1.35), B = 0.48, SE = 0.23, 95% CI [0.02, 0.93], t(122.95) = 2.06, p = .04, d = 0.37, and a significant difference in weight such that participants estimated Black targets (M = 188.04 lbs, SD = 16.42) as heavier than White targets (M = 177.10 lbs, SD = 14.75), B = 10.94, SE = 1.91, 95% CI [7.17, 14.70], t(229.50) = 5.73, p < .001, d = .76.

Although the data support the hypothesis that people perceive Black men as larger than White men, we did not know the actual height or weight of the targets. It is therefore possible that the Black targets in this stimulus set were actually taller and heavier than the White targets, and that participants simply accurately perceived them as such (see Burton & Rule, 2013; Coetzee, Chen, Perrett, & Stephen, 2010). Thus, we repeated the study accounting for actual target size in Study 1B.

Study 1B

We intended to replicate Study 1A in Study 1B using a new stimulus set consisting of targets whose height and weight were publicly available: high school football players being recruited to play college football. A number of media sources track the recruitment of elite high school athletes, some of which (e.g., Rivals.com) post both facial photographs of top recruits alongside their height and weight. Moreover, the recruits consist of young men aged 16–19 years, a group particularly prone to the same violent confrontations in which race-based misperceptions of formidability might be most applicable (e.g., Brunson & Miller, 2006). This stimulus set therefore satisfied both our theoretical goal of investigating the role of race in threat-

motivated biases in size perception, as well as our applied goal of understanding how race may lead people to perceive young Black men as taller and heavier than young White men.

Method.

Participants. We recruited 30 non-Black US residents (17 male, 13 female; M_{age} = 34.1 years, *SD* = 11.1) from MTurk for a study on person perception. We deviated from the sample size heuristic in this study due to experimenter miscommunication. After providing informed consent, participants read that they would view a series of faces for whom they would guess each target's height and weight.

Stimuli. We presented the participants with color photographs of 90 male faces (45 White, 45 Black) downloaded from a popular college football recruiting website (www.rivals.com) in early 2015. This site lists acclaimed high school athletes being recruited to play college football (organized by position), including their height, weight, and facial photograph. To gather these stimuli, a research assistant collected the first 50 White and 50 Black players listed under the "Quarterback" position, skipping targets wearing helmets or with otherwise obscured faces. The website lists players in order of its own ranking system; targets are not sorted by size or any other attribute. We cropped each image so that the body was not visible (see Figure 1).

Although the stimulus set originally consisted of 100 faces, independent-samples *t*-tests of the targets' actual body size showed that the White targets were marginally taller and significantly heavier than the Black targets, on average. We therefore eliminated five photos of each race to match the groups' average height and weight. In our final stimulus set, White targets (*M* = 73.8 in., *SD* = 2.5) were not significantly taller than Black targets (*M* = 73.2 in., *SD* = 1.7), $t(89) = 1.34$, $p = .18$, $d = 0.28$; nor were White targets (*M* = 195.7 lbs, *SD* = 17.1) significantly

heavier than Black targets (M = 189.6 lbs, SD = 18.6), $t(89) = 1.64$, $p = .11$, $d = 0.35$. However, because the White targets were still descriptively larger than the Black targets, we accounted for targets' actual height and weight in our analyses.

Procedure. Participants rated all 90 faces on estimated height and weight in separate counterbalanced blocks within which targets appeared in random order. Otherwise, the procedure was identical to Study 1A except that we increased the range of the height scale from Study 1A to a maximum of 82 inches because a few of the targets were actually taller than 78 inches. We did not inform participants that the targets were athletes.

Results and discussion. We again analyzed the data using mixed models that cross participants and stimuli as random factors with target race, actual height and weight, and perceived height and weight as fixed factors.

Height. As in Study 1A, participants judged Black targets (M = 71.93 in., SD = 1.73) as taller than White targets (M = 71.01 in., SD = 1.74), $B = 1.01$, $SE = 0.25$, 95% CI [0.51, 1.51], $t(93.48) = 4.03$, $p < .001$, $d = 0.84$. Although the targets' actual height did not independently predict how tall they were perceived, $B = 0.02$, $SE = 0.07$, 95% CI [-0.12, 0.16], $t(86) = 0.29$, $p = .77$, their actual weight marginally predicted estimates of their height, $B = 0.014$, $SE = 0.008$, 95% CI [-0.002, 0.031], $t(86) = 1.745$, $p = .08$.

Weight. Participants also perceived the Black targets (M = 181.25 lbs, SD = 19.04) as heavier than the White targets (M = 176.85 lbs, SD = 18.92), $B = 5.35$, $SE = 2.17$, 95% CI [1.04, 9.66], $t(103.05) = 2.46$, $p = .015$, $d = 0.49$. Here, the targets' actual weight positively predicted their estimated weight, $B = 0.33$, $SE = 0.07$, 95% CI [0.19, 0.47], $t(86) = 4.55$, $p < .001$, and their actual height negatively predicted their estimated weight, $B = -1.81$, $SE = 0.61$, 95% CI [-3.02, -0.59], $t(86) = -2.95$, $p < .01$.

Thus, Black targets appeared taller and heavier than White targets even when controlling for their actual height and weight. We therefore replicated the results of Study 1A with a new stimulus set. Critically, we used stimuli that were statistically equivalent in actual size (although the *White* targets were descriptively larger). Furthermore, this difference occurred beyond the significant correspondence that we observed between targets' actual and perceived weight. Participants did not accurately perceive the targets' height, however, perhaps suggesting greater legibility for weight than for height from the face (at least for this relatively homogeneous sample; see also Coetzee et al., 2010; Re & Rule, 2016). Studies 1A and 1B therefore provided initial evidence for race-based biases in size perception.

Study 1C

In Studies 1A and 1B, we found that perceivers overestimated young Black men as taller and heavier than young White men, possibly representing a misperception of their overall physical formidability. In Study 1C, we sought to directly link participants' race-based misperceptions to their perceived formidability by asking participants to match each face with a sample body from an array (e.g., Fessler & Holbrook, 2013). Because perceivers may be more likely to assess others' formidability by evaluating their general size, this method may provide more externally valid estimates of perceived physical formidability than numerical estimates of height and weight do.

Method.

Participants. We recruited 60 US residents from MTurk for a study on person perception. We excluded five Black participants from the analysis, leaving 55 participants (28 male, 27 female; M_{age} = 33.9 years, SD = 11.2).

Stimuli. We used the same 90 athlete faces from Study 1B. The body-matching stimuli consisted of arrays of seven male bodies varying in overall body size and muscularity, adapted from the UCLA Body Matrices II (Frederick & Peplau, 2007) and used in previous research on perceptions of physical formidability (e.g., Fessler & Holbrook, 2013); see Figure 2. The bodies vary in gross size as a function of their muscularity; we will henceforth refer to this measure as "muscularity" for the sake of simplicity.

Procedure. On each trial, participants viewed a randomly selected target face below which they saw the numbered body array depicted in Figure 2. We instructed participants to select the body that appeared to be the best match for each target face by selecting a response between 1 and 7 on a slider scale below the array.

Results and discussion. We entered targets' race, actual height, and actual weight as fixed factors in a mixed model that crossed participants and stimuli as random factors. The results again confirmed our hypothesis: Participants selected more muscular bodies for the Black targets (M = 3.57, SD = 0.75) than for the White targets (M = 3.24, SD = 0.66), B = 0.36, SE = 0.15, 95% CI [.05, .66], $t(96.21) = 2.34$, $p = .02$, $d = 0.48$. Thus, not only do perceivers overestimate young Black men as taller and heavier than young White men in numerical terms (Studies 1A and 1B), they also couple them with larger, more muscular, and thus more formidable physiques when attempting to match their faces to bodies.

Study 1D

The results of Studies 1A-1C established that perceivers estimate young Black men to be taller, heavier, and more muscular than young White men. In Study 1D, we measured physical formidability more directly by asking participants to estimate the targets' physical strength.

Method.

Participants. We recruited 60 US residents from MTurk for a study on person perception. We excluded one Black participant and one participant who did not report his race, leaving 58 participants (35 male, 23 female; M_{age} = 36.4 years, *SD* = 11.8).

Stimuli and procedure. We used the same 90 athlete faces as in Studies 1B and 1C, asking the participants to indicate how strong the person appeared from 1 (*Not at all strong*) to 7 (*Very strong*) for each face presented in random order.

Results and discussion. The results of a mixed model analysis again confirmed our hypothesis: Participants rated Black targets (*M* = 4.52, *SD* = 0.71) as stronger than White targets (*M* = 4.07, *SD* = 0.73), *B* = 0.50, *SE* = 0.13, 95% CI [0.24, 0.76], *t*(130.32) = 3.81, *p* < .001, *d* = 0.67. To assure that these perceptions did not simply reflect actual strength differences, we repeated this study with a new set of targets in Study 1E.

Study 1E

Upper-body strength reliably signals fighting ability and can be accurately estimated from photos of individuals' bodies and faces (Archer & Thanzami, 2007; Sell et al., 2009; Zilioli et al., 2015). However, our previous studies did not match Black and White targets on actual upper body strength. We therefore downloaded photos of participants in the 2015 NFL Draft Combine for whom measures of upper-body strength were publicly available. These targets consist of men who have completed their college football careers and been invited to participate in a series of physical challenges to measure their athletic ability, including performing as many repetitions of a 225-lb bench press as possible. Although this specific task requires notable physical endurance, previous research has found an extremely high correlation between this multiple-repetition task and a 1-repetition bench press maximum, indicating that it is a valid measure of upper body strength (Mayhew et al., 1999). Consistent with our hypothesis, we expected that participants

would judge Black targets as more formidable (i.e., stronger) than White targets, even when controlling for their actual physical strength.

Method

Participants. Although we recruited 60 US residents from MTurk, 65 actually completed the survey. We excluded one Black participant, leaving 64 participants (35 male, 27 female, 2 unreported; M_{age} = 33.6 years, *SD* = 10.0).

Stimuli and procedure. We downloaded the bench press performance data for every player who participated in the 2015 NFL Draft combine from a sports journalism website (2015 NFL Combine Results). Because Black players (*n* = 172) outnumbered White players (*n* = 32), we selected every White player with a listed bench press total and matched him with a Black player with a similar bench press total. In the resulting target sample, the Black targets (*M* = 22.94, *SD* = 6.07) bench pressed no more than the White targets (*M* = 22.86, *SD* = 6.05), *t*(62) = 0.04, *p* = .97, *d* = 0.01, 95% CI of difference [-2.97, 3.10]. The Black targets (*M* = 264.03 lbs, *SD* = 38.37) also weighed about the same as the White targets (*M* = 270.53 lbs, *SD* = 35.62), *t*(62) = -0.70, *p* = .49, *d* = -0.17, 95% CI of difference [-25.00, 12.00], though the Black targets (*M* = 74.25 in., *SD* = 2.36) were slightly shorter than the White targets (*M* = 75.47 in., *SD* = 2.36), *t*(62) = -2.07, *p* = .04, *d* = -0.51, 95% CI of difference [-2.40, -0.04]. This difference works against our directional hypothesis. The participants indicated how strong each of the 64 targets appeared to be from 1 (*Not at all strong*) to 7 (*Very strong*) in random order.

Results and Discussion

As above, we subjected the strength ratings to a mixed model cross-classified analysis. We first computed a model with only race as a fixed factor. As predicted, participants perceived the Black targets (*M* = 4.92, *SD* = 0.69) as stronger than the White targets (*M* = 4.41, *SD* = 0.90), *B*

= 0.46, *SE* = 0.13, 95% CI [0.21, 0.80], *t*(95.44) = 3.42, *p* < .001, *d* = 0.70. Of central importance to our bias hypothesis, this racial difference remained when we controlled for bench press totals, target height, and target weight, *B* = 0.48, *SE* = 0.15, 95% CI [0.18, 0.78], *t*(90.21) = 3.18, *p* = .002, *d* = 0.67. Although actual bench press totals marginally predicted strength judgments in a model that did not include the targets' race, height, or weight, *B* = 0.02, *SE* = 0.01, 95% CI [0.00, 0.05], *t*(62.51) = 1.75, *p* = .09, *d* = 0.44, they did not predict strength judgments in the full model, *B* = 0.01, *SE* = 0.01, 95% CI [-0.01, 0.04], *t*(59.44) = 0.89, *p* = .38, *d* = 0.23. Thus, we found evidence for substantial racial bias in judgments of strength. As the data from Studies 1A-1E strongly suggest some amount of racial bias in perceptions of formidability from the face across a variety of measures, we proceeded to investigate downstream consequences of these biased perceptions.

Study 2

In Study 1, we found that a man's race reliably biased perceivers' judgments of his physical size and formidability. In Study 2, we sought to explore the impact of these misperceptions on evaluations of the physical capacity for harm. In terms of our theoretical position, this study critically contributes to understanding how race can bias both basic physical judgments and more subjective downstream judgments—a relationship that we examine further in Studies 3 and 4. For example, although people have robust accuracy in judging fighting ability from the face (e.g., Sell et al., 2009), this appears to occur alongside reliable biases when judging body size. In our case, to the extent that perceivers use the face to estimate body size when judging formidability, people may thereby misperceive others' physical harm capability.

This study also addresses applied aspects of our research, as an adversary's perceived harm capability guides real-world decisions to engage in force and may also influence observers'

judgments about the appropriateness of force post hoc. Thus, we investigated whether individuals might estimate Black men's ability to harm others in step with misperceptions about their physical size.

Method

Participants. We recruited 185 US residents from MTurk. We excluded 17 Black participants, leaving 168 participants eligible for analysis (92 male, 76 female; M_{age} = 33.9 years, *SD* = 11.0). We recruited a larger sample in this study so that we would have the power to explore the possibility of a gender difference in harm capability bias. This exploratory analysis is reported in the SOM, along with a report of the potential moderating role of gender across all studies.

Stimuli and procedure. We used the same stimuli and procedure as in Study 1B except that, instead of assessing perceived body size, we asked the participants to imagine being in a physical altercation with each target. Specifically, we asked participants to "imagine that you are arguing with this person and he becomes physically threatening. If you were to be in a fight with this person, how capable would he be of physically harming you?". Participants selected a response on a scale from 1 (*Not at all capable*) to 8 (*Very capable*).

Results and Discussion

We again analyzed the data using cross-classified models, regressing harm capability on race with the targets' heights and weights as covariates. Supporting our hypotheses, participants judged the Black targets (*M* = 4.42, *SD* = 1.19) as more capable of harm than the White targets

(*M* = 3.91, *SD* = 1.23), *B* = 0.54, *SE* = 0.12, 95% CI [0.30, 0.78], *t*(146.73) = 4.50, *p* < .001, *d* = 0.74.[2]

When asked to consider Black and White men's potential capacity for harm, participants again showed a robust race-based bias, perceiving Black men as more capable of harm than White men. Because we equalized the two groups according to physical size and also controlled for any actual size differences in our regression model, the results strongly suggest that this race-based bias in perceived formidability results from perceivers' *beliefs* about race (i.e., stereotypes), rather than an accurate inference of physical size based on facial cues. This serves as initial evidence that biased judgments of size may influence downstream inferences of physical capabilities. Below, we investigated this link further.

Study 3

The results of Study 1 indicated a systematic bias in perceptions of Black versus White men's size, muscularity, and strength. In Study 2, we found that Black men were perceived as more capable than White men of doing harm overall. These accumulating data support our broader theoretical argument that people judge Black men as more physically formidable than White men (see Study 1) because they consider them more capable of causing physical harm

[2] To maintain consistency with the other studies, we did not include gender in the primary model. There was a significant main effect of gender, however: Female participants judged the targets to be substantially more capable of causing them harm than male participants. Gender also interacted with target race, such that male participants showed more race bias than female participants did. This interaction does not appear to be due to a ceiling effect among female participants. These findings are reported in more detail in the SOM.

than White men (see Study 2). In Study 3, we tested this hypothesis by asking the same participants to judge both the physical muscularity (as an index of physical formidability; Fessler & Holbrook, 2013) and harm capacity of Black and White men. This allowed us to investigate whether individual differences in racially biased estimates of muscularity predict individual differences in racially biased estimates of harm. We also included a simple measure of explicit prejudice to explore whether the relationship between race-based misperceptions of size and harm occurs independently of the more basic tendency to evaluate Black people more negatively than White people (e.g., Hugenberg & Bodenhausen, 2003).

Method

Participants. We recruited 120 US residents from MTurk for a study on person perception, excluding 10 Black participants' data for 110 participants in total (61 male, 49 female; M_{age} = 34.6 years, *SD* = 11.2). This sample size would provide approximately 90% power to detect a correlation of *r* = .30 between the two bias scores. We excluded one participant from analysis whose harm and muscularity bias difference scores fell 5 *SD*s above the sample mean.

Stimuli and procedure. We asked participants to rate each of the 90 athlete faces used in Studies 1B-1D and 2 on both muscularity (using the body arrays from Study 1C) and harm potential (as in Study 2). We organized the trials for each judgment into separate blocks counterbalanced in order between participants with target order randomized within each block. Finally, we asked participants to complete separate feeling thermometers for Black people and for White people. Previous research has shown that feeling thermometers reliably correlate with other measures of explicit prejudice (e.g., McConnell & Leibold, 2001).

Results

Replication of Study 1C and Study 2 results. We first used cross-classified models to analyze participants' scores for the Black and White targets on each of the muscularity and harm judgments to confirm that the race differences that we found in Study 1C and Study 2, respectively, replicated here. Indeed, participants rated the Black targets (M = 3.70, SD = 0.78) as more muscular than the White targets (M = 3.29, SD = 0.73), B = 0.45, SE = 0.15, 95% CI [0.15, 0.74], $t(112.74) = 2.95$, $p = .004$, $d = 0.56$, and rated the Black targets (M = 4.52, SD = 1.19) as more capable of harm than the White targets (M = 3.91, SD = 1.25), B = 0.65, SE = 0.13, 95% CI [0.39, 0.90], $t(144.74) = 4.92$, $p < .001$, $d = 0.82$.

Relationship between perceptions and prejudice. To analyze the relationship between participants' perceptions and measures of their prejudice, we subtracted each individual's mean score for the White targets from his or her mean score for the Black targets to create Black-White difference scores in which positive values indicated greater muscularity (harm) for Black over White targets. We similarly calculated a difference score on the feeling thermometer items such that positive scores represented warmer feelings toward Whites than Blacks.

We first analyzed these data with zero-order correlations. The muscularity and harm bias difference scores significantly correlated across participants, $r(107) = .49$, $p < .001$ (see Figure 3). The feeling thermometer difference scores modestly correlated with the harm capability difference scores, $r(107) = .22$, $p = .02$, but not with muscularity perception difference scores, $r(107) = .05$, $p = .61$. Furthermore, the correlation between muscularity differences and harm differences remained relatively unchanged when controlling for the feeling thermometer differences in a partial correlation, $r(106) = .50$, $p < .001$. Participants' biases to perceive Black men as more muscular therefore related to their biases to also perceive them as more capable of harm, but neither difference strongly related to anti-Black prejudice.

Next, we conducted a hierarchical regression analysis to examine the potential unique contributions of muscularity differences and prejudice differences to harm capability differences. We regressed the harm capability difference scores onto the muscularity perception difference scores in Step 1 of the model. Consistent with the zero-order correlations, muscularity differences positively predicted harm differences, $B = 0.63$, $SE = 0.11$, 95% CI [0.41, 0.84], $t(107) = 5.87$, $p < .001$, $R^2 = .24$. Adding prejudice differences to the model in Step 2 significantly increased the variance explained, $R^2 = .28$, $\Delta F(1, 106) = 5.76$, $p = .02$, $\Delta R^2 = .04$. Furthermore, although muscularity differences still powerfully predicted harm differences in this model, $B = 0.62$, $SE = 0.11$, 95% CI [0.41, 0.83], $t(106) = 5.87$, $p < .001$, prejudice differences significantly predicted harm differences as well, $B = 0.006$, $SE = 0.002$, 95% CI [0.001, 0.011], $t(106) = 2.40$, $p = .02$.

Discussion

In Study 3, we simultaneously replicated the results of Study 1C (in which participants rated Black men as more muscular than White men) and the results of Study 2 (in which participants rated Black men as more capable of harming them than White men). More important, because the same participants provided each judgment, we were able to test whether the degree to which individuals judged Black men as more capable of harm than White men significantly correlated with the analogous racial gap in their judgments of Black and White men's size (i.e., muscularity). Although size and harm correlated independently of the perceivers' explicit racial prejudice, preferences for Whites over Blacks did weakly associate with the participants' harm capability judgments. Race-based differences in perceptions of physical size are therefore not easily explained by general anti-Black prejudice but, instead, likely result from specific stereotypes associating Blacks with size and threat. To investigate this

further, we conducted Study 4 with a sample of both White and Black participants to determine the extent to which these perceptions may be based in more broadly held cultural stereotypes.

Study 4

Thus far, we have consistently found that non-Black perceivers rate Black men as physically larger, more muscular, stronger, and more capable of causing physical harm than White men, even when controlling for Black and White targets' objective size. Although the results of Study 3 showed that preferences for White over Black people relate to differences in perceptions of how capable they are of causing physical harm, we wanted to further explore the contributing role of intergroup perception. Specifically, non-Black perceivers may be particularly biased against Black targets because they are members of a disliked outgroup (e.g., Stephan & Renfro, 2002), and because threat perceptions are especially attuned to outgroups (Miller, Maner, & Becker, 2010). Alternatively, these effects may partly occur due to widely shared cultural stereotypes that Black men are dangerous. If the latter is true, both Black and White perceivers may express similar racial biases when judging Black and White men's size and harm capability.

We conducted Study 4 as an exploratory investigation of whether the relationship between misperceptions of size and harm capability among Black and White targets would differ according to participants' race. Looking at both types of perceptions among Black and White targets provide a number of useful tools in investigating the cause of the bias that we have observed in the first three studies. Specifically, we were able to test not just whether there are mean differences in race bias in size and harm perceptions between Black and White perceivers, but we were also able to test whether the relationship between harm and size bias is consistent for Black and White perceivers. Finding that size and harm bias are strongly correlated for both groups, for example, might suggest that stereotypes lead to threat, which looms in perception,

which leads to elevated judgments of harm capability, across perceiver groups. However, if participant race moderates the link between size and harm bias, this would suggest that both stereotypes and outgroup threat can contribute to size bias, but that outgroup threat is necessary for size bias to translate to harm bias. Given that muscularity perceptions are less evaluatively laden than inferences of harm capability, stereotypes about race and size may be consistent across perceiver race. However, harm capability judgments may be subject to a more complex process. Harm capability judgments, like size perceptions, may have a root basis in group-specific stereotypes (e.g., Black men = larger) but are likely also affected by broader cognitions linking the outgroup to potential harm (e.g., Miller et al., 2010). Thus, for White participants, this combination of stereotyping Black men as larger and believing the outgroup to be antagonistic could lead to a strong race difference in harm capability judgments. Black participants, however, might also stereotype Black men as larger than White men but not expect them to be more capable of harm because they share a social group identity. We tested this in Study 4.

Method

We requested 60 participants from each race (Black, White) and gender (male, female) combination without any other demographic restrictions through Qualtrics panels (a service that recruits participants according to desired demographic characteristics) for a final sample of 240 US residents (60 Black male, M_{age} = 47.0 years, *SD* = 14.3; 60 White male, M_{age} = 51.1 years, *SD* = 12.3; 60 Black female, M_{age} = 44.3 years, *SD* = 14.3; 60 White female, M_{age} = 45.4 years, *SD* = 13.1). We used the same stimuli and procedures as in Study 3, except that we did not assess explicit prejudice.

Results

Muscularity perceptions. We first entered target race, participant race, and their interaction term as fixed factors in a cross-classified model predicting muscularity estimates that included the targets' heights and weights as covariates. We included random error components for the intercept, target race, participant race, and the interaction between target and participant race for participants, and a random error component for the intercept for targets. As above, participants judged Black targets (M = 3.69, SD = 0.84) as more muscular than White targets (M = 3.42, SD = 0.80), B = 0.32, SE = 0.12, 95% CI [0.07, 0.56], t(75.75) = 2.60, p = .01, d = 0.60, but the participant's race qualified this difference, B = -0.17, SE = 0.07, 95% CI [-.30, -.04], t(196.40) = -2.62, p = .01, d = 0.37 (see Figure 4).

We decomposed this interaction by conducting the analyses separately for White and Black participants. White participants rated the Black targets (M = 3.71, SD = 0.81) as much more muscular than the White targets (M = 3.35, SD = 0.76), B = 0.40, SE = 0.12, 95% CI [0.15, 0.65], t(112.25) = 3.23, p = .002, d = 0.61. Black participants also rated the Black targets (M = 3.67, SD = 0.88) as more muscular than the White targets (M = 3.49, SD= 0.83) but to a considerably lesser extent, B = 0.23, SE = 0.11, 95% CI [0.00, 0.46], t(103.85) = 2.02, p = .046, d = 0.40. Although Black participants appear to have judged White targets as more muscular than White participants did in Figure 4 because of our use of within-subjects error bars, their ratings did not significantly differ, B = 0.14, SE = .10, 95% CI [-.06, .34], t(235.92) = 1.35, p = .18, d = 0.19. Similarly, Black participants did not judge Black targets to be less muscular than White participants did, B = -0.03, SE = .11, 95% CI [-.25, .18], t(236.39) = -0.31, p = .76, d = -0.04. Thus, the significant interaction seems to stem from differences in the Black versus White participants' relative ratings of the Black and White targets' muscularity, rather than differences in how they rated either of the two target groups independent of the other.

Harm perceptions. We next tested whether participant race would qualify the estimates of the targets' perceived ability to cause physical harm using a cross-classified model analogous to that computed for the muscularity perceptions above. Here, Black targets (M = 4.10, SD = 1.39) were rated as only marginally more capable of harm than White targets (M = 3.94, SD = 1.32), B = 0.18, SE = 0.09, 95 % CI [0.00, 0.36], t(133.03) = 1.97, p = .051, d = 0.34. Participant race qualified this marginally significant difference, however, B = -0.24, SE = 0.09, 95 % CI [-0.41, -0.07], t(238.00) = -2.75, p = .006, d = 0.36: White participants showed a clear tendency to see Black targets (M = 4.27, SD = 1.25) as more capable of harm than White targets (M = 3.99, SD = 1.23), B = 0.30, SE = 0.10, 95% CI [0.10, 0.51], t(154.32) = 2.93, p = .004, d = 0.47, whereas Black participants saw both Black (M = 3.92, SD = 1.50) and White (M = 3.88, SD = 1.41) targets as similarly capable of harm, B = 0.05, SE = 0.10, 95% CI [-0.14, 0.25], t(163.13) = .55, p = .58, d = 0.09 (see Figure 5). Upon examining their judgments of the two target groups individually, we saw that Black participants judged Black targets as less capable of harm than White participants did, B = -0.35, SE = .18, 95% CI [-.70, -.00], t(230.43) = -1.98, p = .049, d = -0.26, but did not significantly differ in their assessment of White targets' harm capability, B = -0.11, SE = .17, 95% CI [-.45, .22], t(233.30) = -0.67, p = .50, d = -0.09.

Relationship between muscularity and harm capability. We next examined the relationship between perceptions of targets' muscularity and their ability to harm. In Study 3, we found that race-based differences in individuals' estimates of muscularity correlated with race-based differences in their estimates of harm potential. Here, we conducted a similar analysis using multiple regression so that we could include an interaction term for participant race. As in Study 3, we again calculated a Black-White difference score for muscularity estimates to use as

the predictor variable and a Black-White difference score for harm estimates to use as the outcome variable.

First, we replicated the zero-order correlation between muscularity perception differences and harm perception differences, $r(238) = .23$, $p < .001$. Next, to test the possibility that different relationships exist within the different participant groups, we regressed the harm-capability difference scores onto the muscularity-perception difference scores with participant race as a moderator in Model 1 of Hayes's (2013) PROCESS script with 5,000 bootstrapped resamples. The overall model reached significance, $F(3, 236) = 13.45$, $p < .001$, $R^2 = .14$. As predicted, muscularity perception differences positively predicted harm capability differences, $B = 1.37$, $SE = 0.25$, 95% CI [0.87, 1.87], $t(236) = 5.42$, $p < .001$, $d = 0.71$. Furthermore, participant race moderated this link, $B = -0.78$, $SE = 0.17$, 95% CI [-1.12, -0.45], $t(236) = -4.59$, $p < .001$, $d = 0.60$. Decomposing this interaction and plotting the data by participant race showed a significant relationship between the muscularity and harm capability differences for White, $B = 0.59$, $SE = 0.11$, 95% CI [0.38, 0.80], $t(118) = 5.44$, $p < .001$, $d = 1.00$, but not Black participants, $B = -0.20$, $SE = 0.13$, 95% CI [-0.46, 0.06], $t(118) = -1.48$, $p = .14$, $d = 0.27$ (see Figure 6 for the slope of each group).

Readers might be curious about whether Black and White participants differed on their baseline level of harm bias (e.g., how much harm bias would be shown by participants who show no muscularity bias?). For the sake of illustration, we separately calculated predicted values for such participants based on the regression equations for each participant group, finding that we would expect Black and White participants not showing a muscularity bias to display very little harm bias ($\hat{y} = .07$ for White participants and $\hat{y} = .08$ for Black participants). Moreover, restricted range in the Black participants' difference scores does not easily explain these different

relationships between muscularity and harm capability. Rather, Black participants' muscularity bias scores (*SD* = 0.68) varied almost just as much as White participants' muscularity bias scores (*SD* = 0.66) and Black participants' harm bias scores (*SD* = 0.44) varied less than White participants' harm bias scores (*SD* = 0.54) but not significantly, Levene's $F = 3.75$, $p = .054$.

Analyses by target. To further elaborate on why White participants associated muscularity bias with harm bias more than Black participants did, we conducted additional analyses assessing the correlation between judgments of a target's perceived muscularity and harm capability within each participant. In other words, we assessed the extent to which each participant's judgments of muscularity correlated with his or her judgments of harm capability, testing whether this relationship was stronger among White participants than among Black participants.

To do this, we calculated sensitivity correlations that estimated the within-subject relationship between each participant's muscularity judgments of the target faces with their harm capability judgments for the same targets separately for the White and Black targets and converted each resulting correlation to a Fisher's *z*. We then subjected these sensitivity correlations to a 2 (participant race) × 2 (target race) ANOVA with repeated measures on the second factor.

As predicted based on the previous regression interaction, we observed a main effect of participant race, $F(1, 216) = 6.14$, $p = .01$, $\eta^2_{partial} = .03$, 95% CI [.02, .15], such that the overall sensitivity correlation between muscularity and harm judgments was stronger for White participants ($M = .28$, $SE = .02$) than for Black participants ($M = .20$, $SE = .02$; see Figure 7). Furthermore, a main effect of target race also emerged, $F(1, 216) = 11.46$, $p = .001$, $\eta^2_{partial} = .05$, 95% CI [.02, .08], showing stronger correlations between muscularity and harm judgments for

Black targets (M = .26, SE = .02) than for White targets (M = .21, SE = .02). Participant and target race did not interact, $F(1, 216) = 0.37$, $p = .54$, $\eta^2_{partial} < .01$. Both Black and White participants therefore seemed to be more sensitive to the perceived muscularity of Black targets when making harm capability judgments, as both groups associated muscularity with harm more for Black targets than for White targets. Critically, White participants used perceived muscularity to inform their harm capability judgments significantly more than Black participants did. This may partly explain why we observed a nonsignificant relationship between muscularity bias and harm bias among Black participants in the regression analyses.

Discussion

In Study 4, we extended the results of the preceding studies in a number of important ways. Foremost, we recruited Black participants to determine whether the racial biases that we observed earlier reflected an outgroup bias or a more general effect of internalized cultural stereotypes. The answer to this question was nuanced, however. The mean analyses showed that participants' race predicted the amount of bias in their perceptions of physical formidability: Although both Black and White participants perceived Black targets as more physically muscular than White targets, the difference was significantly smaller for Black participants. Furthermore, Black participants did not show the race-based difference in harm perceptions that White participants did.

Among White participants, we replicated the findings of Study 3: White perceivers' tendency to judge Black targets as more muscular than White targets translated into judging Black targets as more capable of harm. Black participants, however, did not show this same pattern. Black participants still showed an observable tendency to rate Black targets as larger than White targets, but this was much weaker than the effect observed for Whites. Further,

Black participants did not translate this larger size for Black men into an increased capacity for harm. These results suggest that multiple processes may contribute to racially biased size and harm judgments, and that they do so differentially for different perceiver groups. For White perceivers, group-specific stereotypes and outgroup threat cognitions could act in concert to produce strong biases in both types of judgments. However, Black perceivers may subscribe to size stereotypes without the associated group-based threat. For these perceivers, Black targets presumably are judged as larger merely as a result of stereotypes, and not because of a threat-looming mechanism. In turn, any potential for these small size differences to translate to harm bias is likely negated by ingroup-favoring biases in judgments involving conflict. An outgroup threat account may also account for the seeming tendency for Black perceivers to judge White targets as larger than the White perceivers did, although as we pointed out above, this comparison did not yield a significant difference. Relatedly, although we made efforts to clearly instruct participants that they should assess each target specifically with regard to his *physical* capability for causing harm, we cannot rule out that participants may have construed this question with targets' *intentions* in mind. Future work will therefore be needed to more carefully disentangle ability from intention.

Study 5

In Studies 1-4, we consistently found biases in perceivers' judgments of the size, strength, and overall physical formidability of Black relative to White men. Furthermore, these biases in size perceptions predicted corresponding biases in harm perceptions, at least among non-Black perceivers. Although these effects appear to be quite robust, we have yet to establish whether they might explain justifications for the use of force against a person. We thus investigated whether participants' perceptions of the size and harm capability of a target would predict their

opinions about how appropriate it would be for police officers to use force against the person. Such judgments are of critical importance. Although they do not simulate the judgments of a "reasonable officer" who actually witnessed a shooting incident (Klinger & Brunson, 2009), they may better simulate inferences made by members of the general public and by jurors evaluating whether force would have been necessary to subdue a suspect.

We predicted that participants would judge force as more appropriate against Black targets than against White targets. Furthermore, we expected that estimates of the targets' physical size and formidability would predict this racial bias in force justifications, with perceptions of harm capability mediating the role of perceived size on the use of force. In other words, we thought that racially biased size judgments would influence harm perceptions, and that harm perceptions would influence force justifications. We tested this path from race to force justifications (through size and then harm perceptions) in a serial mediation model at the target level, using estimates averaged across perceivers for each measure. Here, it is important to note that we differentiate between the broad threat stereotype that is commonly linked to race (and is the basis for biased formidability judgments) and specific judgments of harm capability, which are most directly a function of target formidability (e.g., Sell et al., 2009). Thus, our predicted model begins with the impact of race on formidability judgments, which then impact harm perceptions and, subsequently, force justifications.

Method

Participants. We recruited 80 US residents from MTurk for a study on person perception, excluding three Black participants for 77 participants total (39 male, 38 female; M_{age} = 35.6 years, SD = 10.8).

Stimuli and procedure. We asked participants to imagine that each of the 90 men whose faces we used above (e.g., Study 1b) had behaved aggressively toward a police officer but was not wielding a weapon. They then viewed each target face in random order and judged the extent to which it would have been appropriate for police to use force to subdue him, from 1 (*Not at all appropriate*) to 7 (*Very appropriate*).

Analytic strategy. We first tested whether participants' force justification ratings differed by race in a cross-classified model. Next, we investigated the relationship between force justification and the consensus measures of size and formidability borrowed from the studies above. Because we intended to show that these target characteristics mediated the relationship between force justifications and target race, we focused on the target as the unit of analysis. We calculated the mean level of force justification for each target by averaging across participants. We then averaged the scores for each size from the above studies, averaging across multiple studies when applicable (estimated height and weight, Study 1B; estimated muscularity, Study 1C, Study 3, and Study 4; and estimated strength, Study 1D), and converted these means to *Z*-scores for each type of judgment. Because the scores on the different judgments strongly correlated (Cronbach's $\alpha = .93$), we combined them into one Formidability composite. We similarly calculated target averages for perceived physical harm capability from the data of non-Black participants in Studies 2-4 and combined these into a Harm composite (Cronbach's $\alpha = .99$).

Finally, we tested the primary mediation model of interest by investigating the link from target race to force justifications through a serially mediated path involving both the Formidability and Harm composites. We hypothesized a path such that target race would predict Formidability perceptions, which would then predict Harm perceptions, which would then

predict force justifications. We tested this path by building a serial mediation model using Model 6 of Hayes's (2013) PROCESS script with 5000 bootstrapped resamples. In this model, we dummy-coded target race to serve as the predictor variable (Black = 1, White = 0), entered the Formidability composite scores as the first mediator, entered the Harm composite scores as the second mediator, and included the mean force justification ratings for each target as the outcome variable.

Results and Discussion

As expected, participants rated the use of force against Black men (M = 3.99, SD = 1.38) as more justified than the use of force against White men (M = 3.53, SD = 1.23), B = 0.48, SE = 0.14, 95% CI [0.20, 0.76], $t(118.75) = 3.38$, $p = .001$, $d = 0.62$. Accordingly, the total effect of race on force justification was significant in the serial mediation model, B = 0.23, SE < 0.01, 95% CI [0.16, 0.30], $t(87) = 6.48$, $p < .001$, which fit the data well overall, $F(3, 86) = 403.00$, $p < .001$, $R^2 = .93$. Consistent with our hypothesis, we observed a significant serial indirect effect through the Formidability and Harm composites, B = 0.11, SE = 0.03, 95% CI [0.05, 0.17], $p = .002$. The direct effect of race on force justification remained significant, B = 0.07, SE = 0.01, 95% CI [0.04, 0.09], $t(87) = 5.20$, $p < .001$, $d = 1.12$.

Race alone seems to have accounted for some of the variance in force justification, but Formidability and Harm accounted for a significant portion of the variance. This confirms that target race and its serial influence on perceptions of physical formidability and harm capability biased perceivers' justifications to use force against hypothetical criminal suspects. In other words, the racial bias in size and harm perceptions that we observed above extended to justifications of the hypothetical use of force against unarmed suspects of crime. Participants were more likely to indicate that police force was justified to detain Black versus White targets.

Critically, perceptions of Black men as larger and more capable of harm partially accounted for this, providing crucial evidence for our central research question. Increased justification of force (a downstream judgment resulting from judgments of a person's ability to cause physical harm), was strongly associated with racially biased judgments of overall size and formidability. People judged Black men as larger and more harmful than White men, thus rendering them more suitable recipients of physical force.

Although it cannot be concluded from these data that the process that we have observed necessarily unfolds in real-world force decisions, we have provided evidence that the basic judgments perceivers make about the harm potential of others are subject to this racial bias in size judgments. Although much more research would be needed to understand the extent to which such judgments manifest in the high-arousal situations faced by police *in vivo*, information about the physical and social-category cues that contribute to this bias might help to move towards that understanding. We therefore investigated the cues and processes underlying these judgments in Studies 6 and 7.

Feature- and Category-Based Influence

In Studies 1-4, we repeatedly found racial biases in perceptions of the size and overall physical formidability of young Black versus young White men, and then found that these biases feed into decisions about the use of force in Study 5. These biased perceptions seem to relate to stereotypes associated with the targets' social categories (rather than overt anti-Black prejudice), as even Black participants overestimated the size of Black targets relative to White targets, despite not expecting them to be more capable of physical harm.

Exactly what cues trigger these stereotypes is unclear, however. Feature-based biases in facial phenotypicality could distort person perception such that targets with darker skin tone or

more Black-prototypic features may be judged to be more formidable. On the other hand, perhaps category-based influences could trigger these differences in a top-down manner, such that beliefs regarding the social group as a whole result in biased judgments. Indeed, stereotypic responding can result from both feature-based perceptual associations and from top-down categorical associations simultaneously (see Blair et al., 2002; Livingston & Brewer, 2002). We thus hypothesized that both feature-based cues to racial phenotypicality and top-down effects of racial categorization likely contributed to the effects that we observed above. To learn more about these processes, we investigated whether feature-based cues promote the race-formidability link in Study 6, and tested whether merely believing a body to be Black versus White (holding the physical body itself constant) would distort judgments of size, strength, and physical formidability in Study 7. We conducted these final two studies with the goal of providing insight into the multiple sources of information that may feed into the phenomena that we observed in the first five studies. If biases in size and formidability judgments vary between Black and White targets, but do not vary between individuals within those groups who look more prototypical of the two races, then a pure social categorization account may be most appropriate. If, on the other hand, category labels of racially ambiguous targets do not yield different size judgments, perhaps the bias uncovered in the previous studies results from appearance-based correlates of race rather than from the racial group itself. However, based on existing literature showing that both types of information can cause bias in person perception, we predicted that both bottom-up racial prototypicality and top-down semantic information would be associated with race-based differences in size perception.

Study 6

In Study 6, we examined whether targets' racial prototypicality influences perceptions of their physical formidability. Past research has reliably found that racial prototypicality can influence person perception and judgment independent of top-down stereotype activation (Blair et al., 2002; Blair, Judd, & Chapleau, 2004; Dixon & Maddox, 2005; Eberhardt et al., 2006; Livingston & Brewer, 2002). That is, people judge faces that are more prototypical of their racial group (e.g., "Afrocentric" Black faces) as more stereotypic of their category. This prototypicality bias is so powerful that it operates across racial boundaries, such that even White targets high in Afrocentricity (i.e., who have facial structures reminiscent of Black people) trigger Black stereotypes (Blair et al., 2002). Moreover, racial prototypicality biases can affect perception early and automatically (e.g., influencing rapid decisions to shoot another person in hypothetical laboratory tasks; Ma & Correll, 2011).

Here, we examined whether perceivers' racial bias in formidability judgments would be more pronounced for Black and White targets high in Afrocentricity by asking participants to rate their faces for how Afrocentric they look. We then calculated the extent to which the mean Afrocentricity ratings for each face corresponded to each target's average score on the formidability measures we computed in Study 5. In short, we expected to find that Afrocentric Black and White faces would be seen as larger, more muscular, and more physically formidable than Eurocentric faces. In addition to these global Afrocentricity ratings, we also collected ratings of the faces' skin tone and structural features (e.g., lip and nose width) from two separate samples, as past work has shown that individual components of Afrocentricity can independently influence target judgments (Hagiwara, Kashy, & Cesario, 2012).

Finally, we wanted to test the extent to which Afrocentricity might relate to other physical characteristics that signal aggression, testosterone, or masculinity. We therefore measured each

target's facial width-to-height ratio (fWHR), a structural characteristic of the face linked to testosterone reactions (r = .18; Lefevre, Lewis, Perrett, & Penke, 2013) that predicts perceived and actual aggression (Carré, McCormick, & Mondloch, 2009). We investigated whether fWHR differs by race and the extent to which it correlates with ratings of targets' formidability. If Black targets have a higher fWHR than White targets, and fWHR positively correlates with formidability perceptions, then it would suggest that the racial differences reported above could be attributed to low-level physical differences that predict perceptions of threat. As such, there might not be anything race-specific about the differences in formidability judgments that we have observed but, rather, that such race-based differences represent an incidental byproduct of a related underlying difference. If, however, Black and White targets do not differ in fWHR, it would suggest that any relationship between fWHR and formidability judgments occurs independent of race.

Method

We recruited 60 US residents (35 male, 25 female; M_{age} = 35.3 years, SD = 12.5) from MTurk to rate the faces for global Afrocentricity and an additional 120 US residents (68 male, 51 female, 1 unspecified; M_{age} = 33.5 years, SD = 11.2) randomly assigned to rate the faces' skin tone or specific features. Two research assistants measured the fWHR of each face using ImageJ software (Abrramoff, Magalhaes, & Ram, 2004). Their ratings were highly consistent (r = .76), so we averaged them to create our fWHR measure.

Using the 90 high school athlete faces from the studies above, we asked participants to rate the targets' global Afrocentricity from 1 (*Not at all Afrocentric*) to 7 (*Very Afrocentric*) after informing them that people who are highly Afrocentric tend to have darker skin tone, a wider nose, and thicker, fuller lips. We blocked the faces by race so that participants saw all faces of

one race before seeing faces of the other race; we counterbalanced the block order and randomized the presentation of the faces within the blocks.[3]

For the skin tone ratings, we used the Gaussian blur function in Adobe Photoshop CS3 to obscure the faces' internal features while retaining skin tone. We then asked participants to rate the Afrocentricity of each blurred face based on its skin tone using the same scale and presentation format described above. For the feature-specific ratings, we used Adobe Photoshop CS3 to convert each image to a grayscale line drawing, similar to the procedures used by Hagiwara et al. (2012). We then asked the participants to rate each face's Afrocentricity following the same instructions as in the global rating condition, except that we did not mention skin tone.

To investigate whether facial Afrocentricity makes targets seem larger and more formidable, we calculated the zero-order correlations between each target's mean global Afrocentricity rating, skin tone rating, feature-specific rating, fWHR, Formidability and Harm scores computed in Study 5, and the targets' actual height and weight. We also calculated parallel partial correlations controlling for height, weight, and fWHR. Because the Afrocentricity scores were bimodally distributed according to race (i.e., Black targets were unsurprisingly judged to be much more Afrocentric than White targets), we normalized them within race for each rating type to simplify data interpretation. In other words, each individual target's Afrocentricity score was a function of his appearance relative to other targets of his own race.

[3] We did not analyze the data for race order effects because Afrocentricity ratings were analyzed with the target at the unit of analysis, precluding the possibility participant-level differences influenced the analyses.

We thus tested the extent to which Afrocentricity relative to other own-race targets was associated with size and formidability judgments. Target race did not qualify the relationship between Afrocentricity and the Formidability and Harm measures, so we collapsed the two races together when correlating those variables. Collapsing across race provided the benefit of increasing power to a more acceptable level for these target-based analyses.

Results

First, we verified the consistency of the Afrocentricity measures by examining the extent to which ratings of the skin tone and features predicted the global Afrocentricity ratings. Global Afrocentricity correlated strongly with skin tone, $r(88) = .52$, $p < .001$; however it correlated with the feature-specific ratings significantly stronger, $r(88) = .83$, $p < .001$; Stieger's $z = 5.15$, $p < .001$. A multiple regression model confirmed the stronger weight of targets' features than of targets' skin tone: the feature-specific ratings contributed unique variance to the global Afrocentricity ratings, $B = 0.79$, $SE = 0.07$, 95% CI [0.64, 0.94], $t(87) = 10.71$, $p < .001$, $d = 2.30$, but skin tone did not, $B = 0.06$, $SE = 0.06$, 95% CI [-0.09, 0.21], $t(87) = 0.81$, $p = .42$, $d = 0.17$.

Second, the Black ($M = 1.97$, $SD = 0.20$) and White ($M = 2.00$, $SD = 0.13$) targets did not differ in fWHR, $t(88) = -0.68$, $p = .50$, 95% CI of difference [-0.10, 0.05], $d = -0.15$. Moreover, although fWHR weakly correlated with the feature-based measure of Afrocentricity, its correlations with the global Afrocentricity ratings and skin-tone ratings were both very small (see Table 1). Similarly, some but not all of Formidability's component measures positively correlated with fWHR. Overall, then, fWHR does not seem to explain the race-based biases that we observed throughout this set of studies, at least within the range of targets that we tested here.

This suggests that race differences in formidability judgments are not due to actual differences in a testosterone-linked facial structural characteristic.

Finally, we examined how the three Afrocentricity measures related to the Formidability composite from Study 5 and to the measures that it comprised (i.e., weight, muscularity, strength, and harm capability, but not height). Despite the unique contribution of features to global Afrocentricity ratings, each Afrocentricity measure positively correlated with the Formidability composite and its components. Thus, the more that targets looked prototypically Black (regardless of their race), the more formidable they seemed (but see Tables S4 and S5 in the SOM for correlations decomposed by target race). Justification of force also positively correlated with Afrocentricity. These relationships were independent of the targets' actual size and fWHR (see partial correlations, below the diagonal, in Table 1); in fact, actual height and weight did not correlate with any measure of Afrocentricity. Moreover, both Formidability and each of its constituent measures correlated similarly with global Afrocentricity, skin tone, and the feature-specific ratings (all Stieger's zs < 1.30, all ps > .19).

Discussion

Here, we found that face-based perceptions of racial prototypicality predicted judgments of physical formidability. Judgments of physical formidability therefore appear to be subject not just to social categorical information related to race, but also may relate to perceptions of specific race-related appearance cues. This relationship comports with past work on the role of race-typical features (e.g., skin tone) in producing bias. For example, exposing participants to a darker-skinned perpetrator of a violent crime aroused greater emotional discomfort than when the perpetrator was lighter-skinned (Dixon & Maddox, 2005), consistent with our data on harm judgments above. Such perceptions can have dangerous consequences. For example, participants

in a first-person shooter task mistakenly shot unarmed White targets who looked less prototypically White (Ma & Correll, 2011) and more Afrocentric-looking individuals may be punished more severely in court even to the point of execution (Blair, Judd, & Chapleau, 2004; Eberhardt et al., 2006).

Racial prototypicality therefore associates with perceptions that may lead to biased outcomes. That is, despite the null relationship between targets' racial prototypicality and their *actual* size, participants showed a strong tendency to judge prototypically Black targets as larger, stronger, and more dangerous than targets who looked more prototypically White. Beyond contributing to the comprehensiveness of the current investigation, such information also suggests that the race biases we have observed thus far may unfortunately be more difficult to control than biases rooted only in top-down social category effects (see Blair, Judd, & Fallman, 2004). These data, in combination with the findings from Studies 1-5, paint a picture of race-based threat perception that is multiply caused, strikingly robust, and partly based on low-level perceptual elements of racial phenotypicality.

Study 7

In Study 6, we observed that feature-based cues to race may bias judgments of men's size. This result is quite sensible given the influence of racial prototypicality on other outcomes in person perception (e.g., Blair et al., 2004; Dixon & Maddox, 2005; Eberhardt et al., 2006). Yet it seems unlikely that the effects described throughout this research result only from feature-based biases. To the contrary, beliefs about targets' groups can often powerfully influence social cognition. For instance, receiving semantic information about race can result in a wide variety of biased consequences, from hiring decisions and interpersonal distancing behavior to differences in face memory (Amodio & Devine, 2006; Dovidio & Gaertner, 2000; Pauker et al., 2009). We

therefore investigated whether imposing beliefs about the race of racially ambiguous targets would similarly affect judgments of physical size in Study 7 by examining how top-down racial category information might influence perceivers' size judgments even when holding the objective stimulus constant. This allowed us to further test our theoretical perspective that one's beliefs can systematically bias perceptual judgments.

To achieve this, we presented participants with images of male bodies that we color-inverted to conceal their race, leading the participants to believe that the bodies were either Black or White by presenting them with a Black or White face (or stereotypically Black or White name) that supposedly identified the body's identity. We then measured their estimates of the target's height and weight, predicting that they would rate the bodies as larger when they believed the targets were Black versus White.

Method

Participants. Although we recruited 120 US residents (60 per between-subjects condition) from MTurk, 127 completed the survey. We excluded six Black participants, leaving 121 participants in the final sample (63 male, 56 female, 2 uncategorized; M_{age} = 32.2 years, *SD* = 10.0).

Stimuli and procedure. Target stimuli consisted of 16 photographs of White men photographed in the lab wearing a skintight body suit. We cropped the photographs below the head and above the ankles. To obscure the targets' actual race, we used Adobe Photoshop to color-invert each photo, resulting in images lacking natural skin-tone or other racial cues (see Figure 8 for an example).

We instructed the participants that would be asked to estimate the height and weight of bodies from photographs that had been visually altered. For each measure, each participant

viewed eight Web browser pages displaying four bodies each. We grouped the pages by target race (e.g., four pages of ostensibly White targets followed by four pages of ostensibly Black targets), counterbalanced between participants. Due to the lack of identifying information in the stimuli, we showed participants each individual target both as a putatively Black target and as a putatively White target. Participants provided height and weight estimates for each target in random order within randomly ordered blocks separated by judgment type. As such, there were 16 total pages consisting of four judgments each: four pages with Black targets and four pages with White targets (for eight total pages) for each of the two judgments (height, weight).

We assigned the participants to one of two between-subjects conditions: Faces or Names. Participants in the Faces condition viewed an array of four full-color photographs depicting Black or White men's faces at the top of their screen and read that each body corresponded to one of the four faces but that the order of the bodies was random, and so there was no way to match individual faces to bodies. Hence, the purpose of the face array was to subtly communicate the bodies' ostensible race. We specifically instructed the participants to provide their height and weight estimates on the basis of the bodies and not the faces, which were not visible onscreen while the participants scrolled down the page to make their estimates. Participants in the Names condition completed the same trials without any faces on the slides. Rather, we displayed a forename underneath each body, taken from a list of the most distinctively Black and White names compiled from births in California since 1961 (Levitt & Dubner, 2005). We used the top 16 names of each race. In both conditions, participants made their judgments using the same scales described in Study 1A.

Results

We again analyzed the data using cross-classified models in which we effect-coded target race (Black = .5, White = -.5) and display condition (Faces = .5, Names = -.5) and computed an interaction term from their product. For height, we observed the predicted main effect of target race, $B = 0.40$, $SE = 0.08$, 95% CI [0.24, 0.55], $t(119) = 4.94$, $p < .001$, $d = 0.92$, such that participants estimated the targets as taller when alleged to be Black ($M = 71.88$ in., $SD = 1.83$) than when alleged to be White ($M = 71.44$ in., $SD = 1.78$). This was qualified by a marginally significant interaction between target race and display condition, $B = 0.31$, $SE = 0.16$, 95% CI [-0.01, 0.63], $t(119) = 1.92$, $p = .06$, $d = 0.35$, such that participants in the Faces condition showed a marginally stronger bias to overestimate the putative Black men's height than participants in the Names condition did.

Conducting the same analysis with weight as the dependent variable, we again observed the predicted main effect of target race, $B = 1.96$, $SE = 0.62$, 95% CI [0.73, 3.19], $t(119.02) = 3.17$, $p = .002$, $d = 0.58$, such that participants estimated the targets as heavier when alleged to be Black ($M = 192.55$ lbs, $SD = 20.86$) than when alleged to be White ($M = 190.58$ lbs, $SD = 19.78$). Here, display condition did not qualify this difference, $B = 0.51$, $SE = 1.24$, 95% CI [-1.94, 2.96], $t(119.02) = 0.41$, $p = .68$, $d = 0.08$.

Discussion

In Study 7, we found that perceivers' associations between race and physical size were strong enough to bias their judgments of the size of *identical* bodies simply because they were led to believe that they were Black or White. Specifically, participants perceived racially ambiguous bodies as both taller and heavier when labeled as Black than when labeled as White, regardless of whether we conveyed this information visually (through a face) or semantically

(through a name). Thus, race appears to bias individuals' perceptions of men's size whether their race is communicated through either category- or feature-based cues.

General Discussion

In this research, we found that Americans demonstrated a systematic bias in their perceptions of the physical formidability imposed by Black men. Non-Black perceivers overestimated young Black men as taller, heavier, stronger, more muscular, and more capable of causing physical harm than young White men. Critically, these size and harm perceptions predicted the extent to which perceivers saw force as justified against hypothetical suspects of crime. Specifically, judgments of size fed into biased perceptions of harm capability, and these harm perceptions mediated the link between size perception and force justification. Finally, we showed that raced-based biases in perceptions of formidability manifest through both feature-based visual cues related to racial prototypicality, as well as to top-down semantic information about targets' social category membership.

Some aspects of the racial bias in formidability perceptions appeared to be more generalizable than others. When assessing overall size and muscularity, for instance, Black perceivers also tended to overestimate the size of Black targets, albeit to a much smaller degree than did White participants. This accords with work showing that even members of minority groups may show automatic outgroup favoritism (Ashburn-Nardo, Knowles, & Monteith, 2003; Dasgupta, 2004; Nosek, Banaji, & Greenwald, 2002; Richeson, Trawalter, & Shelton, 2005), internalize culturally held stereotypes (e.g., Allport, 1954; Clark & Clark, 1947), and use that information in social judgment (e.g., Correll et al., 2002). Unlike non-Black perceivers, however, Black participants did not show a bias in harm perceptions, nor did bias in their size perceptions correlate with bias in their perceptions of harm. Thus, although Black individuals may have

learned the same cultural stereotypes about the size of Black men, they do not seem to apply these misperceptions the same way that non-Black people do. These results also suggest that multiple mechanisms may be involved in creating racial bias in judgments of size and overall formidability. Among White participants, both stereotypes and associations between young Black men and threat may work in tandem, leading to a strong bias in perceptual judgment and encompassing size as well as downstream harm-related judgments. Among Black participants, however, race-size stereotypes alone, absent the experience of threat from young Black male targets, may create biased size judgments that are weaker than those seen in White participants. Corroborating this lack of experienced threat, Black participants size judgments do not seem to lead to broader inferences regarding greater harm capability of Blacks than Whites.

This participant race difference also raises questions regarding whether size judgments hold distinct values for ingroup versus outgroup members. Across these studies, we found a clear racial bias in perceptions of individuals' size and strength. However, it would be overly simplistic to assume that these perceptions are necessarily always negatively valenced. It stands to reason that the physical formidability of an opponent would be experienced negatively by a perceiver who is in a threatening situation. Physical formidability might be neutral or even positive in non-threatening situations—for instance, research on body image clearly shows that men value a muscular physique (Cohane & Pope, 2001; Pope, Olivardia, Gruber, & Borowiecki, 1999). Even in a threatening context, it may seem quite positive to feel that one is supported by physically formidable allies. In several studies here, perceivers much more strongly attributed muscular builds to Black men relative to White men. However, the perception of such a physique seems to have been imbued with some degree of negativity, as it corresponded to elevated perceptions of harm capability among White perceivers. In contrast, Black participants who

judged Black targets as larger more weakly judged these targets as more capable of causing harm, and muscularity judgments predicted White participants' harm judgments better than Black participants' harm judgments. One explanation may be that Black participants also used social group information to make their harm judgments. Specifically, Black participants may not have seen ingroup members as more capable of harm due to greater familiarity or perceived affiliation, despite the group-specific stereotype that likely contributed to elevated muscularity perceptions. Indeed, although people may ascribe potentially negative traits (e.g., dominance, aggression) to ingroup members, they may value the presence of such traits when used in the service of their group's interest (see Rule et al., 2010; Stirrat & Perrett, 2012). Much more research is needed to provide more a more definitive disentangling of the multiple possible causes of the observed biases.

In Studies 6 and 7, we also found that formidability perceptions were biased both by feature-based perceptions and by category-based beliefs. Men whose faces appeared more prototypically Black were more likely to be judged as heavier, stronger, more muscular, and more capable of harm than men who looked more prototypically White. This suggests that perceivers are quite sensitive to features associated with race, despite the absence of a relationship between these features and the men's actual size. Importantly, even in the absence of such phenotypic facial features, top-down beliefs about targets' race also led to biases in the perception of targets' physical size when participants drew inferences of height and weight from photos of bodies. Simply believing a target is a Black man (even with a Black name) was sufficient to elicit the effect. Thus, we have evidence that the racial bias in estimates of targets' formidability emerges both from feature-based and category-based racial cues.

Critically, we also showed that biased formidability perceptions manifested in hypothetical force justification. Much more research is necessary to determine whether a similar process would play out in actual force decisions. To the extent that police officers are subject to the same biased perceptions that the current participants exhibited (see Eberhardt et al., 2004), it is easy to see how they might be more likely to decide to shoot dangerous-looking yet unarmed Black men compared to similarly threatening White men, but the present data cannot speak directly to this possibility.

Threat Distorts Perceptions of Size and Proximity

This work is consistent with research showing the influence of threat-based factors on judgments of the social world. In some of this past work, motivationally relevant factors, such as threat, have been shown to affect perceptions and representations of the physical environment. For example, spiders seem closer than they actually are (Cole et al., 2013) and racially prejudiced White participants estimated a city with a high Black population as closer than less prejudiced participants did (Cesario & Navarette, 2014). In similar work, threatening the identity of New York University students led them to estimate rival Columbia University to be closer, Americans threatened by Mexican immigration perceived a shorter distance to Mexico City, and New York Yankees baseball fans estimated Fenway Park (the stadium of a disliked outgroup: the Boston Red Sox) to be closer than New Yorkers who were not Yankees fans did (Xiao & Van Bavel, 2012). Thus, threats loom large. However, in this past research, there was no attempt to link broad threat stereotypes about social groups to judgments about their basic physical characteristics and possible downstream consequences. We attempted to help fill that gap here.

Indeed, the current research suggests that threat may influence another important perception of the physical environment—others' size and physical formidability. Some of the

research on motivated perception has demonstrated that such misperceptions can be adaptive; it may be sensible to see a hill as steeper when wearing a heavy backpack (Proffitt, 2006) or to judge a drink as closer when thirsty (Balcetis & Dunning, 2010), as this may help guide behavior. Here, we observed effects that are likely harmful to others: Unarmed people may be more likely to be judged as physically formidable, and thus subjected to force, if they belong to groups stereotyped as threatening. Our work clearly indicates that, holding objective size constant, young Black men are more likely to be perceived and treated as a formidable threat than are young White men. Although one limitation of the current work is that we have not provided strong evidence for effects on visual perception (see Firestone & Scholl, 2014, for a critical perspective), it is at least quite clear in the present studies that a target's race biases perceivers' *judgments*. These results largely comport with an error management perspective on social perception (Haselton & Nettle, 2006), such that overperceiving the capacity for physical harm (size, muscularity, capability of harm) in outgroup members may have conferred and continue to confer adaptive advantages.

One remaining question concerns the possibility that gender may moderate the racial biases that we observed in each study. Although we did not conduct this research with any hypotheses about gender differences, there is some reason to suspect that a difference might emerge. Past work has shown that men may be most attuned to intergroup conflict, reliably displaying more xenophobic and ethnocentric attitudes than women (Ekehammar, 1985; Furnham, 1985; Qualls, Cox, & Schehr, 1992; Sidanius, Cling, & Pratto, 1991; Sidanius & Ekehammar, 1980). Some have posited a "male warrior hypothesis" in which men are more competitive with outgroup members and more cooperative with ingroup members when the intergroup context is salient (McDonald, Navarrete, & Van Vugt, 2012; Van Vugt, De Cremer,

& Janssen, 2007). If men are indeed more attuned than women to signals of intergroup conflict, they may also be more likely to perceive outgroup members (here, Black men) as more capable of causing physical harm. This is especially relevant for perceptions that may lead to the use of physical force, as men commit the majority of documented acts of violence (Lauritsen, Heimer, & Lynch, 2009) and constitute the majority of police officers in the US by a wide margin (Langton, 2010). Men may therefore be much more likely to find themselves in situations in which they must consider using force when assessing an adversary. As we report in detail in the supplementary materials, meta-analysis of the present effects showed a small but significant tendency for men to exhibit more racial bias than women (mean weighted r = .05, 95% CI [.003, .11], p = .04), but this tendency was much more pronounced for the measures of harm capability and force justification (r = .21, 95% CI [.13, .30] than for the size judgments (r = -.05, 95% CI [-.12, .02]. Though exploratory, this finding tentatively supports the male warrior hypothesis. Some questions remain, however. It is possible that women's reduced race bias is at least partially reflective of a general tendency to see both target groups as highly capable of causing them harm, although we did not observe ceiling effects in female judgments. We plan to conduct confirmatory tests in future work.

Moreover, we conducted a number of studies not included in this main manuscript detailed in the SOM. These include one study (Study S1) similar to Study 1a that suffered from low target power and did not reach significance, one study (Study S2) largely redundant with Study 3, one study (Study S3) somewhat similar to Study 7 that used a different method to manipulate top-down perceptions of race, and one study (Study S4) similar to Study 5 that measured hypothetical force decisions rather than force justifications. The results of two of these studies strongly supported our hypotheses, and the results of two others did not (but did show

results in the same direction as those reported in the main text). We have also conducted a mini meta-analysis of every study conducted (see Goh, Hall, & Rosenthal, 2016), showing strong support for the overall size bias across all of our studies, yielding a mean weighted effect size of $r = .28$, $Z = 10.12$, $p < .001$.

Limitations

It would be valuable for future research to investigate whether the biases that we have observed here manifest in face-to-face interactions outside of the laboratory. Despite this limitation, we believe that the consistency of the effects that we have observed from multiple sets of face and body photographs is quite striking on its own. We used a sample of targets matched for height and weight (and in which the White men were actually descriptively taller and heavier than the Black men), finding clear racial differences in judgment for every measure that we tested. Furthermore, the theoretically informed moderating effects of participant race helped to demonstrate that the present findings did not simply arise due to stimulus artifacts. Furthermore, people's beliefs about targets' race extended from faces to direct perceptions of body size in Studies 6 and 7. Thus, although there is great value to conducting future iterations of this work with methods that would amplify its ecological validity, the present data provide an intriguing step towards a meaningful perceptual bias with real-world implications.

It is also possible that the restricted range of actual target size limits the generalizability of our findings. In most of the studies, we used images depicting talented athletes in a sport that rewards size and strength. This likely restricts the range of physical size of our stimuli to larger than average men. This might account for some surprising patterns in the present data—such as the absence of a significant relationship between strength estimates and actual strength in Study 1E, as has been found in previous research (e.g., Sell et al., 2009). However, there is reason to

expect that bias would occur among a more representative sample of targets. For instance, we observed race-based differences in judgments of height and weight among a non-athlete sample in Study 1A (stimuli used in past stereotyping research) and therefore believe that there is little reason to expect that the biases we observed in the other studies presented here would be limited to athletes. We nevertheless strongly endorse the sentiment that that future work with a wider range of targets would provide a better understanding of bias and accuracy in interpersonal formidability judgments. Such work should also consider female targets while considering that true size differences exist between the average Black and White woman in the US population (Fryar, Gu, & Ogden, 2012), and any work finding a supposed bias in size judgments must be considered in light of true group differences.

Further, the current data cannot fully exclude the possibility that kernels of truth underlie race-based formidability judgments; thus, the question of accuracy must be taken seriously. Our data do indicate that perceivers are biased in their judgments of physical size—participants reliably overestimated the height and weight of Black targets relative to the height and weight of White targets. However, it remains possible that, holding height and weight constant, Black men are actually stronger and more capable of inflicting physical harm than White men are. Some researchers have argued that stereotypes are, on the whole, highly accurate (Jussim et al., 2015), and such a perspective might suggest that perceivers in the present research simply employed accurate stereotypes. The present findings may not comport well with such a perspective, however. First, national health data indicate that American White and Black men are of similar weight and height (with Whites actually being slightly taller than Blacks; Fryar et al., 2012). Second, we believe that the specific patterns of our data are not fully consistent with this interpretation either. For instance, the race-based size perception biases of participants in Study 3

closely tethered to the same individuals' harm capability judgments. The fact that individuals who were most likely to judge Black targets as larger than White targets also judged Black targets as more capable of harm suggests that broader inferences about the physical harm a person is capable of inflicting may be borne out of bias. Moreover, the Black participants in Study 4 also showed a race difference in size perceptions, but not in judgments of harm capability. If our participants had accurately extrapolated the enhanced physical ability of Black men over White men, we would expect the opposite pattern of results (i.e., no muscularity difference, but a significant harm capability difference). Taken together, these findings do not completely rule out an accuracy account, but do suggest that people are inaccurate at least some of the time.

Another open question concerns how much the effects we observed stem from race versus a constellation of race-related characteristics, such as perceived socioeconomic status. Race has been argued to serve as a powerful cue to one's overall social ecology (Neuberg & Sng, 2011). Black people in America tend to experience lower socioeconomic status (SES) than Whites, even to the extent that the racial stereotypicality of one's name serves as an accurate cue of SES (Fryer & Levitt, 2004). Skin tone similarly correlates with SES, as darker-skinned Black people suffer worse SES outcomes than lighter-skinned Black people (Keith & Herring, 1991). To the extent that race cues perceivers to targets' social ecologies, perceivers may assume that Black and dark-skinned targets are from more disadvantaged backgrounds and perhaps have a history of witnessing and participating in physical conflict (e.g., Singer, Anglin, Song, & Lunghofer, 1995). Furthermore, recent work has shown that ecology-related stereotypes may trump race-specific stereotypes in assumptions about others' life histories (Williams, Sng, & Neuberg,

2016). We have yet to conduct work that attempts to disentangle race from perceived ecology but hope that future work will focus on resolving this question.

An additional limitation of the current work is that we did not directly measure implicit black-violence stereotypes. It is reasonably clear from our data that explicit prejudice cannot provide a satisfactory explanation for the biases that we observed, but we have not affirmatively demonstrated the role of this group-based stereotype in producing biased size and formidability judgments. We see this as an important next step in this work and hope that scholars will provide more mechanistic evidence in future work.

Finally, we also hope that future research will be able to directly link these perceptions of physical formidability to use-of-force decisions in a more ecologically valid context. Although we showed that the bias in harm perception mediated the link between perceptions of physical formidability and use-of-force judgments for target faces, the present data do not allow us to conclude whether similar perceptual judgments play out in high-arousal decisions in the field. There is some evidence that Black and Hispanic men who were *perceived as* large were more likely to be stopped and frisked by the New York Police Department (Milner, George, & Allison, 2016), but this study did not include targets' actual size. It is furthermore quite important to establish whether similar processes unfold with real shooting decisions, and in more realistic contexts. For example, Correll et al. (2007) found that police officers did not show the same decision criteria for shooting Black targets that lab participants did, and James et al. (2012) found that police, civilian, and military participants took longer to shoot Black targets and were more likely to fail to shoot armed Black targets when tested in high-fidelity training simulators. Thus, there is good reason to believe that police officers may show different response patterns than typical research participants, highlighting the need for more studies using realistic and

ecologically valid situations. This significant limitation notwithstanding, the present work may provide an important first step towards understanding the cognitive processes relevant to these real-world phenomena, thereby potentially helping to shed light on recent incidents of racial bias unfolding in contemporary society, such as the tragic shootings mentioned in the introduction.

Conclusion

Across a range of different stimuli and dependent variables, perceivers showed a consistent and strong bias to perceive young Black men as larger and more capable of harm than young White men (at least among non-Black participants). Such perceptions may have disturbing consequences for how both civilians and law enforcement personnel perceive and behave towards Black individuals. The studies reported here serve as a clear demonstration of this important phenomenon and provide theoretically meaningful knowledge about both feature-based and category-based influences on the bias to misperceive Black men as larger and more threatening. We hope that stakeholders are able to apply this information to formulate interventions that can meaningfully reduce these biases in the future.

References

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International*, *11*, 36-42.

Allport, G. W. (1954). *The Nature of Prejudice*. Cambridge, Mass: Addison-Wesley.

Allport, G. W., & Postman, L. J. (1947). *The Psychology of Rumor*. New York: Holt, Rinehart & Winston.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*, 652-61.

Archer, J., & Thanzami, V. (2007). The relation between physical aggression, size and strength, among a sample of young Indian men. *Personality and Individual Differences*, *43*, 627-633.

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition*, *21*, 61-87.

Balcetis, E., & Dunning, D. (2010). Wishful seeing: More desired objects are seen as closer. *Psychological Science*, *21*, 147-152.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*, 674-679.

Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgment. *Journal of Personality and Social Psychology*, *87*, 763-778.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, *83*, 5-25.

Brunson, R. K., & Miller, J. (2006). Young black men and urban policing in the United States. *British Journal of Criminology, 46,* 613-640.Burton, C. M., & Rule, N. O. (2013). Judgments of height from faces are informed by dominance and facial maturity. *Social Cognition, 31,* 672-685.

Burton, C. M., & Rule, N. O. (2013). Judgments of height from faces are informed by dominance and facial maturity. *Social Cognition, 31,* 672-685.

Capehart, J. (2013). Playing "games" with Trayvon Martin's image. Retrieved from http://www.washingtonpost.com/blogs/post-partisan/wp/2013/02/06/playing-games-with-trayvon-martins-image/ on May 26, 2015.

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20,* 1194-1198.Cesario, J., & Navarrete, C. D. (2014). Perceptual bias in threat distance: The critical roles of in-group support and target evaluations in defensive threat regulation. *Social Psychological and Personality Science, 5,* 12-17.

Cesario, J. & Navarrete, C.D. (2014). Perceptual bias in threat distance: The critical roles of ingroup support and target evaluations in defensive threat regulation. *Social Psychological and Personality Science, 5,* 12-17.

Chiricos, T., & Eschholz, S. (2002). The racial and ethnic typification of crime and the criminal typification of race and ethnicity in local television news. *Journal of Research in Crime and Delinquency, 39,* 400-420.

Clark, K. B., & Clark, M. P. (1947). Racial identification and preference in Negro children. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (pp. 602-611). New York: Holt, Rinehart, & Winston.

Coetzee, V., Chen, J., Perrett, D. I., & Stephen, I. D. (2010). Deciphering faces: Quantifiable visual cues to weight. *Perception, 39*, 51-61.

Cohane, G. H., & Pope, H. G. (2001). Body image in boys: A review of the literature. *International Journal of Eating Disorders, 29*, 373-379.

Cole, S., Balcetis, E., & Dunning, D. (2013). Affective signals of threat increase perceived proximity. *Psychological Science, 24*, 34-40.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83*, 1314-1329.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology, 37*, 1102-1117.

Correll, J., Wittenbrink, B., Park, B., Judd, C. M., & Goyle, A. (2011). Dangerous enough: Moderating racial bias with contextual threat cues. *Journal of Experimental Social Psychology, 47*, 184-189.

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.

Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research, 17*, 143-169.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.

Dixon, T. L., & Linz, D. (2000). Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. *Journal of Communication, 50*, 131-154.

Dixon, T. L., & Maddox, K. B. (2005). Skin tone, crime news, and social reality judgments: Priming the stereotype of the dark and dangerous Black criminal. *Journal of Applied Social Psychology, 35*, 1555-1570.

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.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87*, 876-893.

Ekehammar, B. (1985). Sex differences in socio-political attitudes revisited. *Educational Studies, 11*, 3-9.

Fessler, D. M., & Holbrook, C. (2013). Bound to lose: Physical incapacitation increases the conceptualized size of an antagonist in men. *PLoS ONE, 8*, 1-7.

Fessler, D. M. T., Holbrook, C., & Gervais, M. (2014). Men's physical strength moderates conceptualizations of prospective foes in two disparate human societies. *Human Nature, 25*, 393-409.

Fessler, D. M., Holbrook, C., & Snyder, J. K. (2012). Weapons make the man (larger): Formidability is represented as size and strength in humans. *PLoS ONE, 7*, 1-9.

Firestone, C., & Scholl, B. J. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science, 25*, 38-46.

Frederick, D. A., & Peplau, L. A. (2007). The UCLA Body Matrices II: Computer-generated images of men and women varying in body fat and muscularity/breast size to assess body satisfaction and preferences. In *8th Annual Meeting of the Society for Personality and Social Psychology*.

Fryar, C. D., Gu, Q., & Ogden, C. L. (2012). Anthropometric reference data for children and adults: United States, 2007-2010. *Vital and health statistics. Series 11, Data from the national health survey*, (252), 1-48.

Fryer Jr, R. G., & Levitt, S. D. (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, 767-805.

Furnham, A. (1985). Adolescents' sociopolitical attitudes: A study of sex and national differences. *Political Psychology*, 621-636.

Geller, W. A. (1982). Deadly force: What we know. *Journal of Police Science and Administration, 10*, 151-177.

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: an egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology, 78*, 211-222.

Goff, P. A., Jackson, M. C., Di Leone, B. A. L., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing Black children. *Journal of Personality and Social Psychology, 106*, 526-545.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on Why and a primer on How. *Social and Personality Psychology Compass*, *10*, 535-549.

Hagiwara, N., Kashy, D. A., & Cesario, J. (2012). The independent effects of skin tone and facial features on Whites' affective reactions to Blacks. *Journal of Experimental Social Psychology*, *48*, 892-898.

Harris-Lacewell, M. (2001). No place to rest: African American political attitudes and the myth of Black women's strength. *Women & Politics*, *23*, 1-33.

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, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Hayes, M. (2014). Dontre Hamilton's autopsy shows police officer shot him from behind. *Buzzfeed*. Retrieved from http://www.buzzfeed.com/mikehayes/dontre-hamilton-autopsy#.ckLlbLqP2 on December 1, 2015.

Holbrook, C., & Fessler, D. M. (2013). Sizing up the threat: The envisioned physical formidability of terrorists tracks their leaders' failures and successes. *Cognition*, *127*, 46-56.

Holbrook, C., Fessler, D. M., & 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.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14,* 640-643.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science, 15,* 342-345.

James, L., Vila, B., & Daratha, K. (2013). Results from experimental trials testing participant responses to White, Hispanic and Black suspects in high-fidelity deadly force judgment and decision-making simulations. *Journal of Experimental Criminology, 9,* 189-212.

Johnson, J. D., Trawalter, S., & Dovidio, J. F. (2000). Converging interracial consequences of exposure to violent rap music on stereotypical attributions of Blacks. *Journal of Experimental Social Psychology, 36,* 233-251.

Jones, J. M. (2015). In U. S., confidence in police lowest in 22 years. *Gallup Poll Social Series*. Retrieved from http://www.gallup.com/poll/183704/confidence-police-lowest-years.aspx on December 1, 2015.

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., 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., Crawford, J. T., Anglin, S. M., Chambers, J., Stevens, S. T., & Cohen, F. Stereotype accuracy: One of the largest relationships in all of social psychology. To appear in T. Nelson (ed.), *Handbook of prejudice, stereotyping, and discrimination* (*2nd ed*). Hillsdale,

Keith, V. M., & Herring, C. (1991). Skin tone and stratification in the Black community. *American Journal of Sociology*, 760-778.

Klein, W. M., & Kunda, Z. (1993). Maintaining self-serving social comparisons: Biased reconstruction of one's past behaviors. *Personality and Social Psychology Bulletin, 19*, 732-739.

Klinger, D. A., & Brunson, R. K. (2009). Police officers' perceptual distortions during lethal force situations: Informing the reasonableness standard. *Criminology & Public Policy, 8*, 117-140.

Langton, L. (2010). Women in law enforcement, 1987-2010. *Bureau of Justice Statistics*. Retrieved from http://www.bjs.gov/content/pub/pdf/wle8708.pdf

Lauritsen, J. T., Heimer, K., & Lynch, J. P. (2009). Trends in the gender gap in violent offending: New evidence from the national crime victimization survey. *Criminology, 47*, 361-399.

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: Facial width is associated with testosterone levels in men. *Evolution and Human Behavior, 34*, 273-279.

Levitt, S. D., & Dubner, S. J. (2005). *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: William Morrow.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, *82*, 5-18.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology, 39*, 821-31.

Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology*, *47*, 391-396.

Maner, J. K., Kenrick, D. T., Becker, D. V., Robertson, T. E., Hofer, B., Neuberg, S. L., ... & Schaller, M. (2005). Functional projection: how fundamental social motives can bias interpersonal perception. *Journal of Personality and Social Psychology*, *88*, 63-78.

Mayhew, J. L., Ware, J. S., Bemben, M. G., Wilt, B., Ward, T. E., Farris, B., ... & Slovak, J. P. (1999). The NFL-225 Test as a Measure of Bench Press Strength in College Football Players. *The Journal of Strength & Conditioning Research*, *13*, 130-134.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology*, *37*, 435-442.

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

McDowell, M. A., Fryar, C. D., Ogden, C. L., & Flegal, K. M. (2008). Anthropometric reference data for children and adults: United States, 2003-2006. *National Health Statistics Reports*, *10*. Retrieved from http://www.cdc.gov/nchs/data/nhsr/nhsr010.pdf.

Miller, S. L., Maner, J. K., & Becker, D. V. (2010). Self-protective biases in group categorization: Threat cues shape the psychological boundary between "us" and "them". *Journal of Personality and Social Psychology*, *99*, 62-77.

National Football League (NFL) Combine 2015. (February 23, 2015). Retrieved October 1, 2015 from http://www.sbnation.com/nfl-combine.

Navarrete, C. D., McDonald, M. M., Molina, L. E., & Sidanius, J. (2010). Prejudice at the nexus of race and gender: An outgroup male target hypothesis. *Journal of Personality and Social Psychology*, *98*, 933-45.

Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle. *Evolution and Human Behavior, 26*, 88-105.

Neuberg, Steven L., Kenrick, D. T., & Schaller, M. (2011). Human threat management systems: Self-protection and disease avoidance. *Neuroscience & Biobehavioral Reviews, 35*, 1042-1051.

Neuberg, S. L., & Schaller, M. (2016). An evolutionary threat-management approach to prejudices. *Current Opinion in Psychology, 7*, 1-5.

Neuberg, S. L., & Sng, O. (2013). A life history theory of social perception: Stereotyping at the intersections of age, sex, ecology (and race). *Social Cognition, 31*, 696-711.

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*, 101-115.

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science, 309*, 785-787.

Pauker, K., Weisbuch, M., Ambady, N., Sommers, S. R., Adams, R. B., & Ivcevic, Z. (2009). Not so black and white: Memory for ambiguous group members. *Journal of Personality and Social Psychology, 96*, 795-810.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81*, 181-192.

Plant, E. A., Goplen, J., & Kunstman, J. W. (2011). Selective responses to threat: the roles of race and gender in decisions to shoot. *Personality and Social Psychology Bulletin, 37*, 1274-81.

Pope, H. G., Olivardia, R., Gruber, A., & Borowiecki, J. (1999). Evolving ideals of male body image as seen through action toys. *International Journal of Eating Disorders, 26,* 65-72.

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science*, *1*, 110-122.Re, D. E., & Rule, N. O. (2016). About face: New directions for the physician's general survey. *Current Directions in Psychological Science*. In press.

Qualls, R. C., Cox, M. B., & Schehr, T. L. (1992). Racial attitudes on campus: Are there gender differences? *Journal of College Student Development*, *33*, 524-530.

Re, D. E., & Rule, N. O. (2016). Heavy matters: The relationship between just noticeable differences in perceptions of facial adiposity and facial attractiveness. *Social Psychological and Personality Science, 7*, 69-76

Richeson, J. A., Trawalter, S., & Shelton, J. N. (2005). African Americans' implicit racial attitudes and the depletion of executive function after interracial interactions. *Social Cognition*, *23*, 336-352.

Rule, N. O., Ambady, N., Adams, R. B., Jr., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: Prediction and consensus across cultures. *Journal of Personality and Social Psychology*, *98*, 1-15.

Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, *39*, 590-598.

Schultz, C. (2015). A city of two tales. Retrieved from http://www.politico.com/magazine/story/2015/02/tamir-rice-cleveland-police-115401.html#.VWS2iPlVhBc on May 26, 2015.

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 of London B: Biological Sciences*, *276*, 575-584.Stephan, W. G., Boniecki, K. A., Ybarra, O., Bettencourt, A., Ervin, K. S., Jackson, L. A., ... & Renfro, C. L. (2002). The role of threats in the racial attitudes of Blacks and Whites. *Personality and Social Psychology Bulletin*,*28*, 1242-1254.

Sidanius, J., Cling, B. J., & Pratto, F. (1991). Ranking and linking as a function of sex and gender role attitudes. *Journal of Social Issues, 47*, 131-149.

Sidanius, J., & Ekehammar, B. (1980). Sex-related differences in socio-political ideology. *Scandinavian Journal of Psychology, 21*, 17-26.

Singer, M. I., Anglin, T. M., yu Song, L., & Lunghofer, L. (1995). Adolescents' exposure to violence and associated symptoms of psychological trauma. *Journal of the American Medical Association*, *273*(6), 477-482.

Stephan, W. G., & Renfro, C. L. (2002). The role of threat in intergroup relations. *From prejudice to intergroup emotions: Differentiated reactions to social groups*, 191-207.

Stirrat, M., & Perrett, D. I. (2012). Face structure predicts cooperation: Men with wider faces are more generous to their in-group when out-group competition is salient. *Psychological Science, 23,* 718-22.

Trawalter, S., Todd, A. R., Baird, A. A., & Richeson, J. A. (2008). Attending to threat: Race-based patterns of selective attention. *Journal of Experimental Social Psychology*, *44*, 1322-1327.

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

Waytz, A., Hoffman, K. M., & Trawalter, S. (2014). A superhumanization bias in Whites' perceptions of Blacks. *Social Psychological and Personality Science*, 1-8.

Williams, K. E., Sng, O., & Neuberg, S. L. (2016). Ecology-driven stereotypes override race stereotypes. *Proceedings of the National Academy of Sciences, 113*, 310-315.

Xiao, Y. J., & Van Bavel, J. J. (2012). See your friends close and your enemies closer: Social identity and identity threat shape the representation of physical distance. *Personality & Social Psychology Bulletin, 38*, 959-72.

Zilioli, S., Sell, A. N., Stirrat, M., Jagore, J., Vickerman, W., & Watson, N. V. (2015). Face of a fighter: Bizygomatic width as a cue of formidability. *Aggressive Behavior, 41*, 322-330.

Figure 1. Sample athlete target photographs used in Studies 1B-1D, and Studies 2-5.

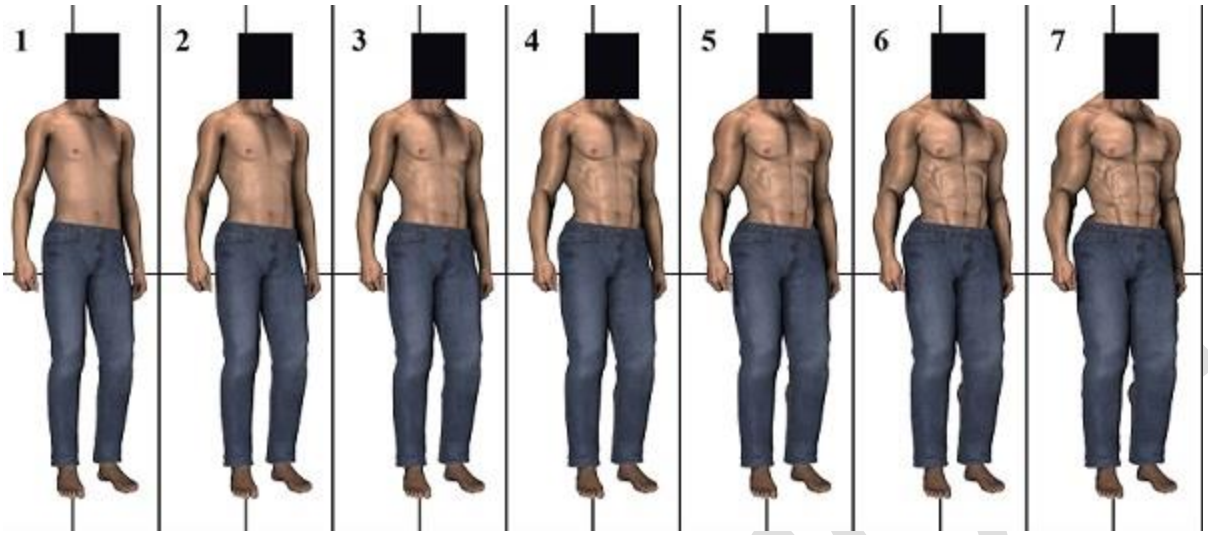

Figure 2. Body muscularity array used in Study 1C, Study 3, and Study 4.

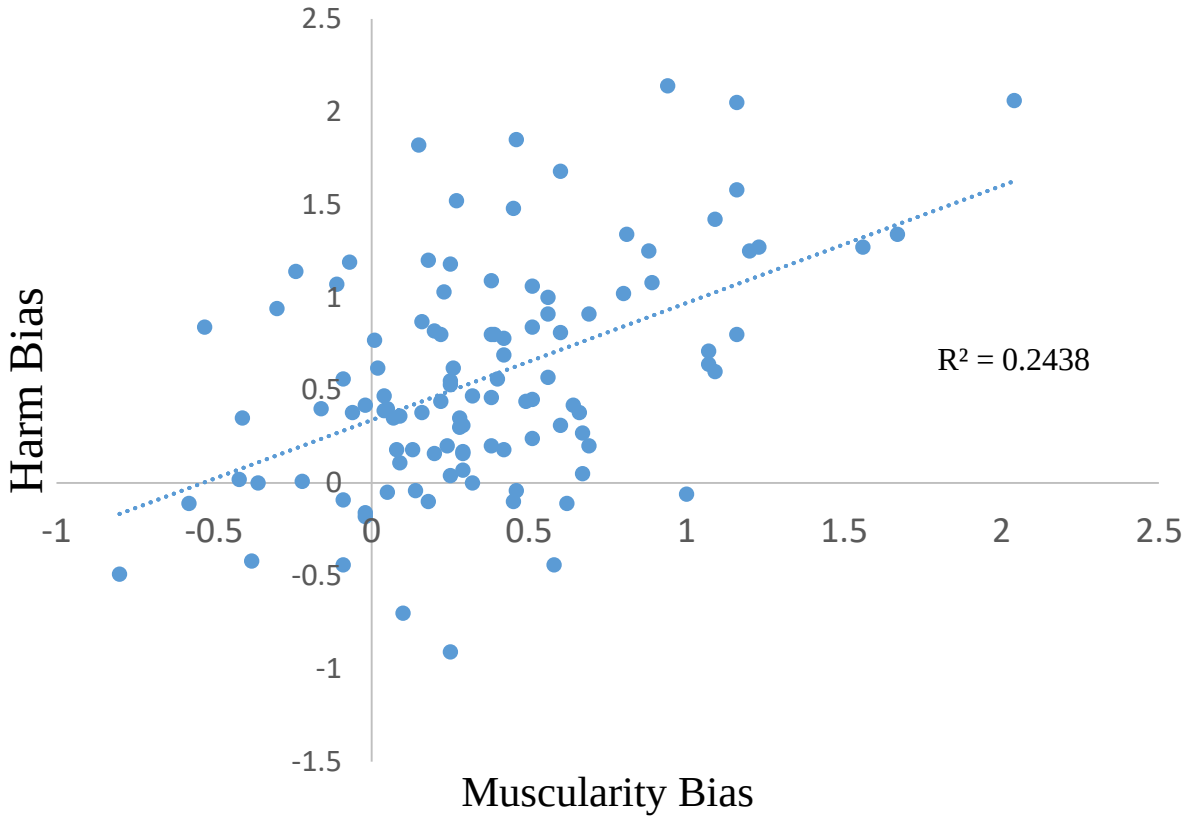

Figure 3. Scatterplot of the relationship between muscularity bias and harm bias. Positive scores reflect increased perceptions of muscularity and harm capability for Black over White targets.

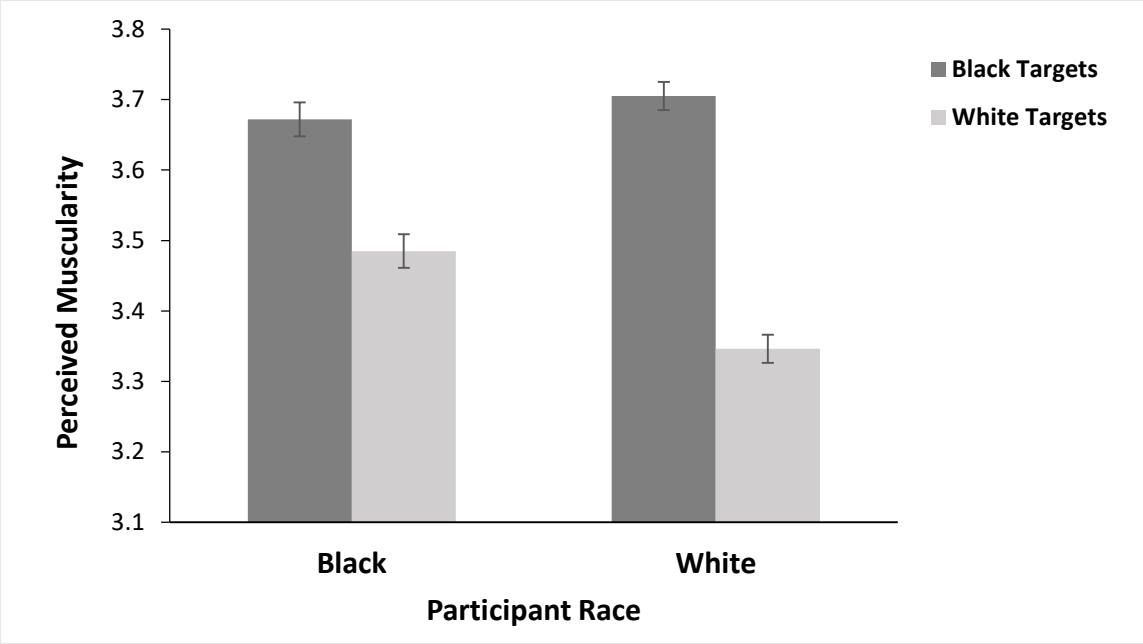

Figure 4. Means and within-subject standard errors for Black and White participants' perceptions of Black and White targets' muscularity in Study 4.

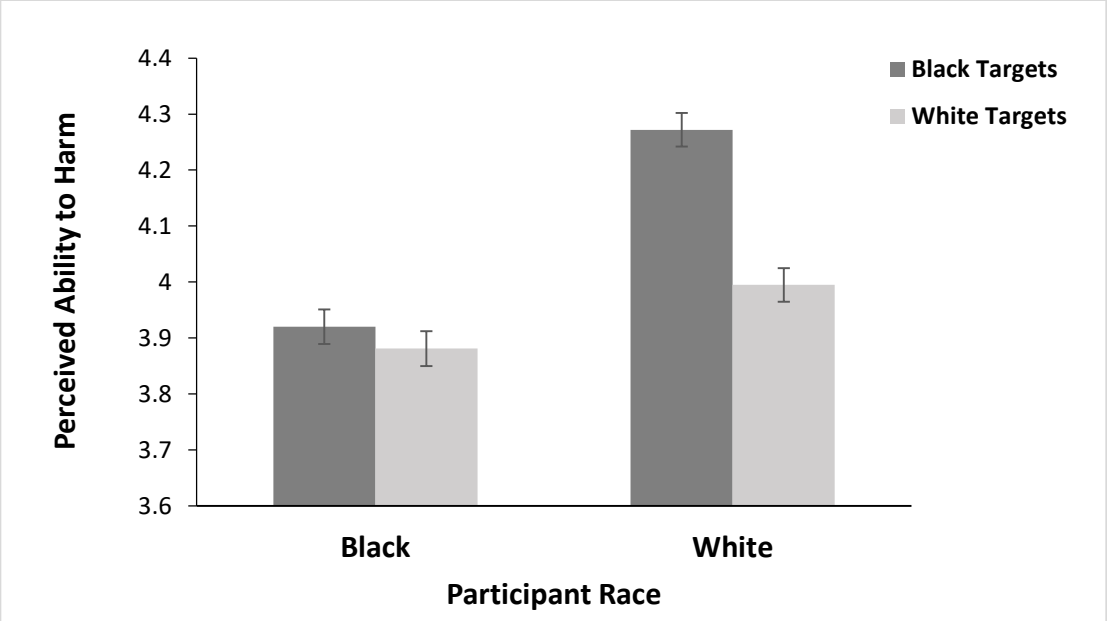

Figure 5. Means and within-subject standard errors for Black and White participants' perceptions of Black and White targets' ability to harm them in Study 4.

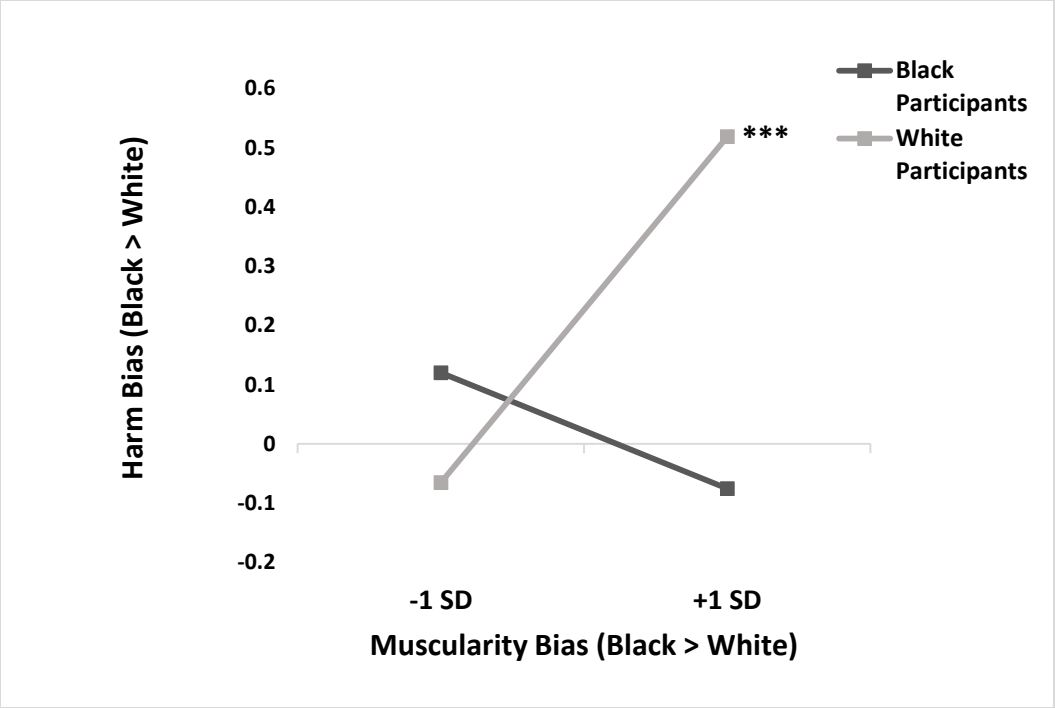

Figure 6. The relationship between Black-White difference scores for Black and White participants' perceptions of targets' muscularity and capacity for harm.

Note. *** $p < .001$. Points plotted at ± 1 *SD* for the Black-White muscularity difference scores.

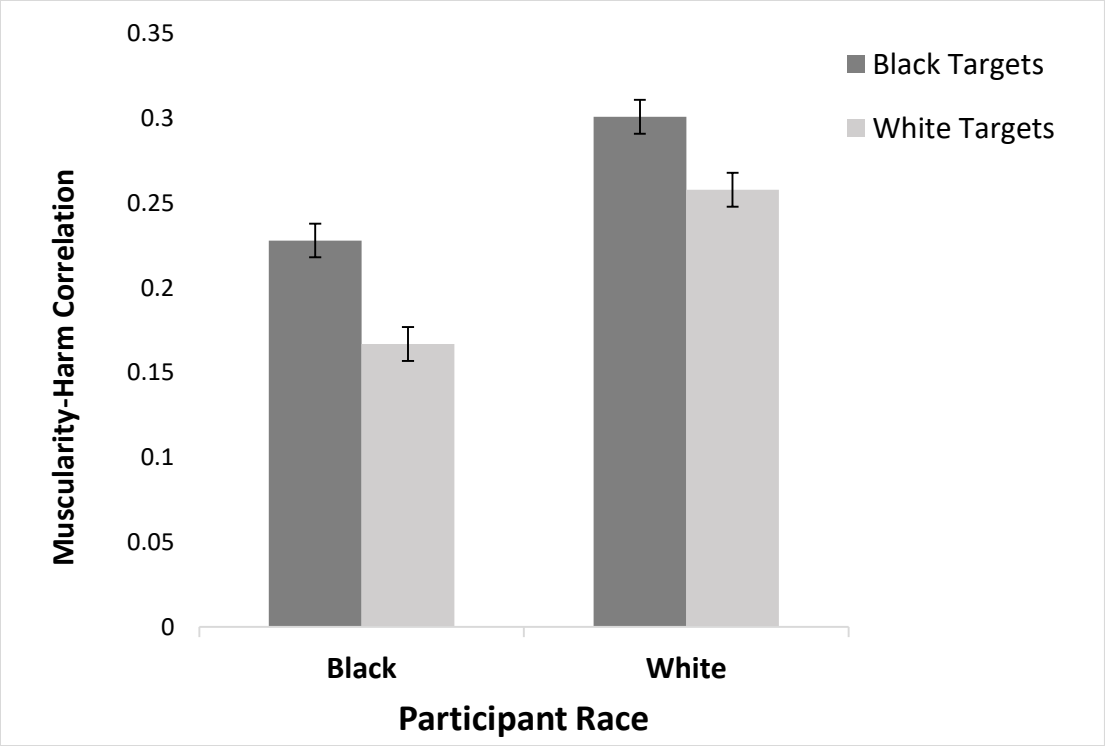

Figure 7. Means and within-subject standard errors for the sensitivity correlation between muscularity and harm capability judgments according to participant and target race.

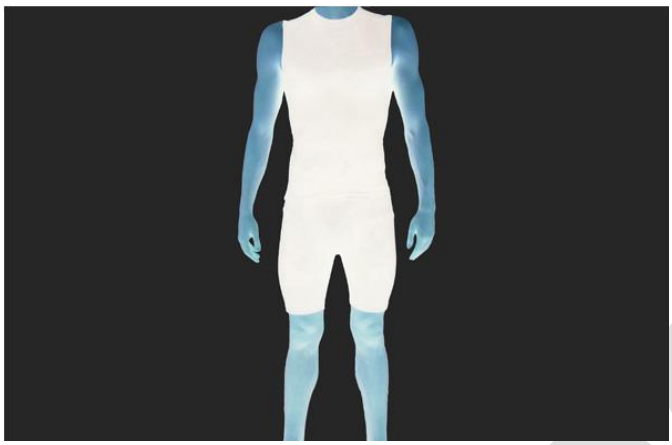

Figure 8. Example of a target body used in Study 7.

Table 1

Relationships between the Mean Ratings of the Extent to Which Targets Look Afrocentric and Various Perceptions of Their Formidability

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Afrocentricity [a]		.83***	.52***	.06	.26*	.28**	.30**	.25*	.28**	.25*	-.13	-.03	.12
2. Feature-Specific [a]	.83**		.58***	.09	.31**	.35**	.33**	.30**	.28**	.22*	-.08	-.03	.21
3. Skin Tone [a]	.52**	.58***		.10	.24*	.23*	.29**	.24*	.30**	.26*	.01	-.01	.01
4. Estimated Height [b]	.08	.12	.11		.66***	.53***	.63***	.78***	.54***	.50***	.12	.16	-.09
5. Estimated Weight [b]	.22*	.27*	.25*	.77***		.95***	.89***	.96***	.75***	.64***	-.02	.29**	.37***
6. Muscularity [c, d, e]	.23*	.30**	.24*	.66***	.93***		.92***	.94***	.75***	.64***	-.10	.21*	.44***
7. Strength [f]	.26*	.29**	.29**	.71***	.88***	.91***		.95***	.86***	.77***	-.07	.14	.34**
8. Formidability Composite [h]	.21*	.26*	.24*	.86***	.97***	.94***	.94***		.80***	.71***	-.02	.22*	.29**
9. Harm Capability [d, e, g]	.25*	.24*	.29**	.58***	.75***	.75***	.85***	.78***		.95***	-.07	.10	.22*
10. Force Justification [h]	.22*	.18[†]	.26*	.53***	.66***	.66***	.78***	.71***	.95***		-.04	.14	.14
11. Actual Height												.66***	-.03
12. Actual Weight													.05
13. fWHR [a]													

Note. Values above the diagonal (df = 88) represent bivariate correlations; values below the diagonal (df = 86) represent partial correlations controlling for the targets' actual height and weight and fWHR.

[a] Based on data from Study 6, [b] Based on data from Study 1B, [c] Based on data from Study 1C, [d] Based on data from Study 3, [e] Based on data from Study 4, [f] Based on data from Study 1D, [g] Based on data from Study 2, [h] Based on data from Study 5.

[†]$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$.

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

Supplemental Online Materials

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

Meta-Analysis of Gender Differences Across Studies

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

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

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

This analysis showed that gender strongly moderated the (Black > White) race bias for conflict-related variables, $r = .24$, 95% CI [.13, .30], $p < .001$, but not for size-related variables, $r = -.05$, 95% CI [- 12, .02], $p = .14$. We acknowledge that this assessment is post hoc and merely exploratory, and that there are other potential explanations for these findings. For example, women may be inferring that virtually all male targets are capable of physically harming them. Somewhat consistent with this possibility, inspection of the descriptive statistics (Table S2) shows that female participants tended to rate both groups of targets as more capable of harm than male participants, but that this relative difference was smaller for Black targets than White targets. Thus, reduced race bias tended to be driven more by elevated judgments of White targets' harm potential than by reduced judgments of Black targets' harm potential (e.g., Studies 2, 3, and 4). This view of the data casts some doubt on the male warrior hypothesis as the sole mechanism of the gender difference. However, these descriptive statistics also show that we did not observe ceiling effects in our data. For example, female participants' mean harm capability ratings never exceeded 4.83 (on a 7-point scale) for any target group. We also point out that we observed a strong gender difference in Study 5, in which participants assessed force justification

and did not implicitly consider their own physical capabilities (as in Study 2, for example). In this study, women were only slightly less likely than men to see force as justified for White targets, but they were much less likely than men to see force as justified for Black targets. As such, we see mixed evidence for the male warrior hypothesis and grant the possibility that the gender difference may not be theoretically meaningful. We ultimately see this as an open question and, accordingly, will seek to confirm this pattern in future work.

Additional Correlations

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

Additional Studies

We conducted four additional studies not reported in the main text. We excluded these studies from the manuscript for various reasons and include them here for completeness and transparency. We excluded Study S1 in part because the target sample size was small and appears to have been underpowered to detect effects using the cross-classified analyses employed elsewhere in the main text. Similarly, we excluded Study S3 because we had used only

a single race-ambiguous target. We excluded Study S2 because it was redundant with Study 3 but did not include the measures of prejudice. We excluded Study S4 because we decided that it was more appropriate to focus on force justification than force decisions, given that a decision to use force is quite far removed from the experimental context that we created here. In all four excluded studies, the results either directionally (Study S1, S3), marginally (Study S1), or significantly (Study S2, Study S4) supported our hypotheses.

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

Method.

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

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

Procedure. The participants viewed the images in blocks, such that they estimated all 42 targets' height or weight (order counterbalanced between participants) individually in random order before moving on to the next judgment. The target image appeared above a slider scale for

each height (weight) rating. The scale for the weight ratings ranged from 120 to 300 lbs, with the possible responses in increments of 1 lb. The scale for the height ratings ranged from 60 in. (5 ft 0 in.) to 78 in. (6 ft 6 in.), with the possible responses in increments of 1 in. We selected these ranges to include plausible values for men of average size (McDowell, Fryar, Ogden, & Flegal, 2008).

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

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

Method.

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

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

Results and discussion.

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

Relationship between size and threat. To analyze the relationship between muscularity and harm perceptions, we subtracted each participant's mean score for the White targets from his or

her mean score for the Black targets to create Black-White difference scores in which positive values indicated greater muscularity (harm) for Black over White targets, whereas negative values indicated greater muscularity (harm) for White over Black targets. The targets' mean muscularity and harm bias difference scores significantly correlated: $r(88) = .37$, $p < .001$. This finding replicated the pattern that we reported in Study 3, increasing our confidence that the tendency to show racial bias in these two measures is strongly related.

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

Method.

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

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

Participants read one of eight individual vignettes, each of which subtly indicated the race of the suspect by referring to him using one of four stereotypically White (Neil, Brett, Brendan, Todd) or Black (Jamal, Rasheed, Tremayne, Kareem) names (Bertrand & Mullainathan, 2003).

The participants read only one vignette, after which we asked them to rate the target on all of height, weight, muscularity, strength, and harm capability using the scales employed in Studies 1 and 2 of the main manuscript.

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

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

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

Method.

Participants. We recruited 120 US residents from MTurk for a study on person perception, but an additional three participants completed the study without collecting compensation for a total of 123 participants. We excluded nine Black participants and an additional 20 participants

who gave the same response on every trial, however, leaving 96 participants in the final sample (52 male, 44 female; M_{age} = 34.6 years, *SD* = 12.0).

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

Results and discussion.

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

Figure S1. Target body used in Study S3

Table S1

Parameter Estimates and Effect Size Estimates for Gender Interaction Among Studies in Main Manuscript

Study (DV)	*B*	*SE*	*df*	*p*	*r*	Conflict
1a (Height)	-0.30	0.37	53	.42	-.11	0
1a (Weight)	-0.49	2.14	44	.82	-.03	0
1b (Height)	0.03	0.15	28	.85	.04	0
1b (Weight)	-1.67	1.03	28	.12	-.29	0
1c (Muscularity)	-0.11	0.09	53	.25	-.16	0
1d (Strength)	-0.31	0.16	56	.05	-.26	0
1e (Strength)	-0.07	0.16	60	.70	-.06	0
2 (Harm Cap.)	0.32	0.13	166	.01	.19	1
3 (Muscularity)	0.08	0.12	107	.49	.07	0
3 (Harm Cap.)	0.33	0.14	107	.02	.22	1
4 (Muscularity)	-0.07	0.06	237	.28	-.07	0
4 (Harm Cap)	0.23	0.09	237	.01	.18	1
5 (Force Justify)	0.76	0.24	74	<.01	.35	1
7 (Height)	0.14	0.17	117	.41	.08	0
7 (Weight)	-0.42	1.26	117	.74	-.03	0

Note. B = estimate from cross-classified linear mixed model, *SE* = standard error of estimate, *df* = degrees of freedom from cross-classified linear model, *r* = effect size of interaction estimate (positive values indicate larger race bias for male participants than female participants), Conflict: 1 = Yes, 2 = No.

Table S2

Means and Standard Deviations (in Parentheses) by Participant Gender and Target Race.

Study (DV)	Male Ps		Female Ps	
	Black	White	Black	White
1a (Height)	70.28 (1.34)	69.94 (1.16)	70.14 (1.61)	69.82 (1.56)
1a (Weight)	188.36 (15.95)	178.07 (15.27)	187.58 (17.41)	175.72 (14.20)
1b (Height)	71.73 (1.57)	70.79 (1.51)	72.24 (1.98)	71.36 (2.07)
1b (Weight)	180.47 (18.98)	177.41 (18.91)	182.42 (19.92)	176.01 (19.76)
1c (Muscularity)	3.54 (0.75)	3.28 (0.71)	3.58 (0.74)	3.21 (0.59)
1d (Strength)	4.45 (0.61)	4.11 (0.67)	4.65 (0.85)	4.00 (0.82)
1e (Strength)	4.91 (0.62)	4.43 (0.81)	4.96 (0.76)	4.41 (1.04)
2 (Harm Cap.)	4.24 (1.08)	3.58 (1.02)	4.65 (1.29)	4.31 (1.34)
3 (Muscularity)	3.71 (0.80)	3.27 (0.77)	3.69 (0.76)	3.34 (0.69)
3 (Harm Cap.)	4.30 (1.10)	3.54 (1.05)	4.83 (1.25)	4.41 (1.33)
4 (Muscularity)	3.85 (0.82)	3.61 (0.78)	3.53 (0.84)	3.22 (0.77)
4 (Harm Cap)	4.04 (1.46)	3.76 (1.34)	4.16 (1.32)	4.11 (1.28)
5 (Force Justify)	4.48 (1.21)	3.64 (1.10)	3.48 (1.40)	3.40 (1.37)
7 (Height)	71.78 (2.00)	71.33 (1.82)	71.72 (1.28)	71.41 (1.50)
7 (Weight)	190.52 (20.08)	188.76 (18.48)	193.00 (18.37)	190.82 (17.71)

Table S3

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

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Afrocentricity [a]		.86***	.26†	.02	.20	.21†	.27†	.20	.14	.13	-.05	-.05	.21
2. Feature-Specific [a]	.85**		.43**	.02	.23	.28†	.29†	.22	.17	.14	-.05	-.01	.31*
3. Skin Tone [a]	.20	.35*		.12	.31*	.30†	.38*	.30*	.34*	.33*	.04	.06	.36
4. Estimated Height [b]	.09	.08	.20		.76***	.69***	.68***	.83***	.44**	.33*	.24	.46**	-.17
5. Estimated Weight [b]	.19	.16	.24	.81***		.96***	.88***	.97***	.69***	.57***	.05	.37*	.31*
6. Muscularity [c, d, e]	.17	.20	.19	.78***	.95***		.91***	.96***	.69***	.56***	.02	.29†	.38*
7. Strength [f]	.25	.23	.31*	.74***	.87***	.90***		.94***	.80***	.71***	.05	.25†	.28†
8. Formidability Composite [h]	.19	.18	.25	.88***	.97***	.97***	.94***		.71***	.60***	.09	.37*	.22
9. Harm Capability [d, e, g]	.12	.14	.30†	.46***	.70***	.69***	.80***	.71***		.94***	.05	.14	.15
10. Force Justification [h]	.11	.11	.32*	.38*	.61***	.58***	.72***	.62***	.95***		-.05	.02	.13
11. Actual Height												.71***	-.35*
12. Actual Weight													-.13
13. fWHR [a]													

Note. Values above the diagonal (*df* = 43) represent bivariate correlations; values below the diagonal (*df* = 40) represent partial correlations controlling for the targets' actual height and weight and fWHR.

Table S4

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

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Afrocentricity [a]		.79***	.77***	.10	.34*	.37*	.40**	.34*	.51***	.46**	-.19	-.01	.08
2. Feature-Specific [a]	.80***		.72***	.17	.41**	.44**	.46**	.42**	.46**	.38*	-.10	-.05	.15
3. Skin Tone [a]	.81***	.75***		.10	.18	.19	.25†	.20	.32*	.31*	-.01	-.07	-.07
4. Estimated Height [b]	.14	.20	.10		.54***	.31*	.45**	.68***	.45**	.40**	.16	.10	-.01
5. Estimated Weight [b]	.31*	.43**	.30†	.81***		.93***	.92***	.97***	.84***	.73***	-.02	.32*	.48**
6. Muscularity [c, d, e]	.34*	.47**	.35*	.78***	.90***		.94***	.91***	.79***	.66***	-.12	.25	.56***
7. Strength [f]	.40**	.47**	.39*	.74***	.88***	.91***		.94***	.85***	.72***	-.05	.23	.54***
8. Formidability Composite [h]	.31*	.42**	.30*	.88***	.97***	.89***	.93***		.83***	.72***	-.00	.26†	.44**
9. Harm Capability [d, e, g]	.51**	.48**	.44**	.46***	.76***	.70***	.80***	.77***		.94***	.05	.14	.15
10. Force Justification [h]	.45**	.38*	.39*	.38*	.67***	.60***	.67***	.65***	.95***		.02	.28†	.28†
11. Actual Height												.63***	.09
12. Actual Weight													.14
13. fWHR [a]													

Note. Values above the diagonal (*df* = 43) represent bivariate correlations; values below the diagonal (*df* = 40) represent partial correlations controlling for the targets' actual height and weight and fWHR.

Table S5

Means, Standard Deviations, and Test Statistics Comparing Participants' Ratings Along Multiple Measures of Physical Formidability (Height, Weight, Muscularity, Strength) and Harm Capability for a Single Target Purported to be Black or White

Measure	Race	*M*	*SD*	*t*	*p*	Cohen's *d*
Height (in.)	Black	72.19	2.30	1.71	.09	0.28
	White	71.59	2.03			
Weight (lbs)	Black	206.11	24.01	1.55	.12	0.25
	White	199.91	25.71			
Muscularity	Black	3.46	1.30	1.18	.24	0.19
	White	3.20	1.43			
Strength	Black	4.57	1.37	1.51	.13	0.24
	White	4.26	1.14			
Harm Capability	Black	5.16	1.40	0.90	.37	0.15
	White	4.97	1.22			

Note. *df* = 154 for each comparison; muscularity, strength, and harm rated on a 1-7 Likert scale.

References

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

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

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