A developmental perspective on the race-based size bias: Evidence from children, adolescents, and adults

a dissertation submitted by

Julia Wefferling

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Psychology
Tufts University
May 2025

Advisor:
Dr. Paul Muentener

Committee members:
Dr. Keith Maddox
Dr. Jessica Remedios
Dr. John Paul Wilson

Abstract

Black men are perceived as stronger, larger, and more capable of harm than White men by White perceivers (Wilson et al., 2017). However, the developmental origins of this race-based size bias remain unknown. Across three studies, we investigated the developmental onset of the race-based size bias and how contextual information changes the representation of this bias. In Study 1, children, adolescent, and adult participants heard negative vignettes paired with Black and White cartoon faces. Children used contextual, but not racial information to make physical strength judgements. Adolescents and adults used both contextual and racial information to make physical strength judgements. There was no consistent evidence for a race-based size bias for height or weight. Study 2 used a similar vignette method while manipulating the valence of events (e.g., positive, negative). Additional measures for Study 2 included forced-choice and open-response questions. Study 2 conceptually replicated the results from Study 1, and expanded our knowledge. The forced-choice measure paired real Black and White faces together and asked which person was stronger, taller, and heavier. Children, adolescents, and adults chose a Black character as stronger than a White character significantly more than chance, providing the first piece of evidence that children may be sensitive to a race-based size bias. In the open-response measure, participants reported using visual information (e.g., seeing their face, seeing their body) as well as contextual information (e.g., their behavior) to make their size judgements. Study 3 was conducted only with child participants and conceptually replicated Study 2 by using real faces rather than cartoon faces paired with the same vignettes. Children did not show evidence of a race-based size bias for strength, height, or weight when contextual information was included. However, in the absence of contextual information, children showed evidence of a race-based size bias for strength, replicating Study 2. Across three studies, we found evidence

that children as young as 6 years of age may be sensitive to the bias, and that it is consistent across measures starting in adolescence.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

White adults show a race-based size bias (Wilson et al., 2017; Freiburger et al., 2023) in which they perceive Black men as stronger, taller, heavier, and more capable of harm than White men of the same size (Wilson et al., 2017). There are significant consequences for these biased perceptions. Hester & Gray (2018), explained that Black men are associated with threat, and when they are also tall, they are stopped more frequently by police. Similarly, Milner, George & Alliston (2016) found through the NYPD database that tall and heavy Black and Latino men are at the highest risk for stop-and-frisk encounters, and they are also at an increased risk for experiencing police force. The connection between White adults perceiving Black men as threatening and large is so robust that it even impacts Black middle school adolescents. Freiburger et al. (2023) showed in a predominantly White adult sample that participants perceived Black adolescents as more threatening and stronger than White adolescents. Black adolescent boys are also more likely to get suspended from school than any other group, and previously suspended Black individuals are more likely to be incarcerated or involved in criminal activity as adults (Graves & Wang, 2022; Wolf & Kupchik, 2016). Taken together, there is clear evidence that Black people are being put at a disadvantage by White adults beginning in childhood and persisting throughout development.

This dissertation examines how a race-based size bias emerges over development. Although there may not be a simple solution for adults who already have the bias to perceive Black men as physically large and threatening, understanding the development of biases can help to create effective interventions to reduce the development and potential consequences of biased perceptions.

For this chapter, we begin by investigating the relation between the different components of the race-based size bias; the perception of height, weight, strength, threat, and race. First, we examine how White adults think about physical size, strength, threat, and race in the literature and analyze the possible directions of the race-based size bias. Then, we explore how the race-based size bias develops in children. Lastly, we introduce three studies that investigated the race-based size bias using a developmental perspective.

Adult Perspective from the Literature

White Adults and Anti-Black Biases

Research has shown that White adults think Black men are more threatening than White men. The association between Black men and violence is strong enough that it is easier for participants to pair weapons and assume non-dangerous objects are threatening when they are paired with Black men than White men (Payne, 2001; Correll et al., 2002; Todd et al., 2021) showing a clear association that White people have between Black people and threat in a laboratory setting. White people also react negatively to Black individuals in ambiguous real-world scenarios. In a seminal study by Duncan (1976), researchers found that White participants perceived an ambiguous interaction (e.g., pushing) as more violent when it was performed by a Black actor relative to a White one. This is evidence that when White participants are required to make a snap-judgment about a Black man, they are likely to associate Black men with threats.

White Adults' Perception of Black Men, Physical Size and Threat

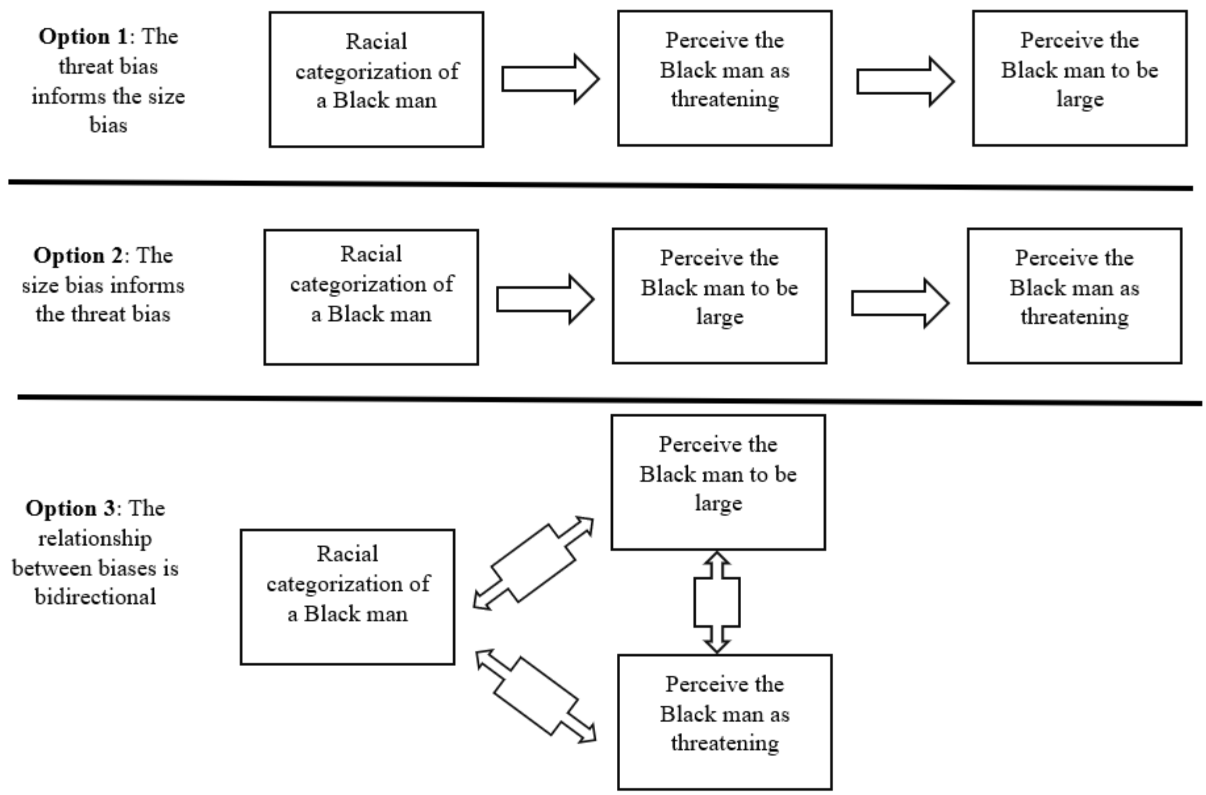
Black men as physically large, strong, and threatening

While there may be several explanations for Black men being perceived as more threatening than White men by White perceivers, some researchers propose that this bias is linked to biased representations of the size and strength of Black men (Waytz et al., 2015; Wilson et al., 2017). For example, Wilson et. al. (2017) showed that White participants rate photographs of Black men as larger than White men. They have also shown that this bias extends beyond a perceptual bias, given that even in the absence of visual information White participants expect people with stereotypically Black names to be larger than people with stereotypically White names (Wilson et al., 2017).

What is the relation between perceptions of race, physical size and threat? Are these independent and unrelated cognitive biases, or do they interact as we perceive and interact with other people? We consider three possible kinds of interactions in this paper (Table 1). First, racial biases related to threat and strength may influence size perception. White people may think Black men are large because they are perceived as threatening and strong. Second, racial biases related to physical size may influence representations of threat and strength. White people may think Black men are threatening because they represent them as large. Third, the relation between these biases could be bidirectional or context dependent. We outline current evidence for each possibility below.

Table 1

Proposed directions between the size and threat biases



Note. The table above depicts the different possibilities proposed in this paper surrounding the relation and direction between the size bias and the threat bias.

Follow-up research by Johnson and Wilson (2019) has shown that the race-based size bias may be linked specifically to biased representations of physical strength. In this study they found that race impacts only White participants' judgments of strength, but not height. Black men were perceived as stronger than White and Asian men; however, Black men were not perceived as taller than White men. Moreover, when looking at male targets and using individuating information on strength (e.g., bicep size), participants rated men with higher upper body strength and larger biceps as stronger than those with lower levels of strength, independent of race (Johnson & Wilson, 2019). Additionally, a replication of Wilson et al. (2017), using a top-down approach, did not find evidence of a race-based size bias for height and weight (Navon et al., 2025). These findings suggest that while there may be associations between representations of race, physical size, and threat, these associations may be driven primarily by a bias to represent Black men as stronger than White men.

Further support for this hypothesis comes from research showing that when White participants are more likely to show threat-related responses to Black men, but not other racial out-groups (e.g., Asians). This suggests that Black targets may be perceived as threatening by White participants which results in exaggerated size perceptions (Johnson & Wilson, 2019). Moreover, when threat is experimentally manipulated with a White man as the target, we see the same ‘superhumanization’ and threat perception that people typically associate with Black men (Payne & Correll, 2020; Waytz et al., 2015). Specifically, if participants are told that a White man is holding a weapon, participants perceive the White male targets to be larger in terms of height and muscularity (Fessler et al., 2012). Despite there being no prior associations between White people and threat, once the White target is placed in a threatening context, they become larger and more threatening to the perceiver.

On the other hand, some studies also suggest that perceptions of physical size cause participants to view targets as threatening. For example, in a sub-study within Wilson et al., (2017), when Black participants performed the same size and threat perception tasks that White participants did, Black participants rated Black men as larger than White men, but not more threatening (Wilson et al., 2017). Because Black individuals perceive Black men as physically larger but not more threatening than White men, we see support that Black men may be encoded as large first with the threat piece emerging in a later cognitive process. Additionally, when researchers manipulated the weight of Black men to be obese, participants were less likely to associate them with threat stereotypes that are typical of Black men, but still associated them with other stereotypes of Black individuals (e.g., unintelligent) (Sim et al., 2022). In this study, while participants still viewed Black men in-line with myriad negative stereotypes, when they were told that the target was obese, and thus, non-threatening, they no longer came to the conclusion that the person was threatening. Importantly, Sim and colleagues (2022) still found that Black men, who did not have reduced threat capacities (e.g., were weight unspecified) were still viewed as physically threatening. This research shows that physical size impacts threat perception, and that such threat perception can be reduced or inflated based on the size of the target.

Other studies suggest that race-based size biases may be context dependent. When participants are shown Black men in a non-threatening context (e.g., business owner), they are thought of as large, but not threatening (Hester & Gray, 2018; Holbrook et al., 2016). Because Hester & Gray (2018) do not specifically investigate the directions of these relations, these results should be interpreted with caution. It may be that the threat exacerbates an existing bias that Black men are physically large. Or, it could be that threatening contexts change how participants perceive physical size.

An alternative possibility is that the direction of this race-based size bias may differ depending on the age or developmental group (e.g., child, adolescent, adult) of the participant. Although existing literature may suggest that the bias is context dependent or bi-directional in adult participants, this may not hold true for children and adolescents. For example, it is possible that children learn anti-Black biases first, followed by Black-threat biases, which then result in race-based size biases; in adolescence and adulthood, these biases may be influenced by context and display bi-directional effects. Because researchers, to our knowledge, have not studied the race-based size bias developmentally, it is important to understand how children think about each component of the bias separately.

Children's Perspectives From the Literature

Despite having some information about how adults think about race, size, and threat, we do not know exactly how this relation forms. Do White children hold the same race-based size bias in relation to the perception of size, race, threat, and strength that we see in adults? Given children think about race, threat, and size, it is likely that they also think about the relation between them. Looking at children's understanding of this race-based size bias helps us to better understand the cognitive origins and underpinnings for the biases in adults. It is important to target these biases early on to reduce microaggressions, trauma, and other adverse consequences for Black children and adults (Liu et al., 2019; Metzger et al., 2021).

We can understand the relation between race, physical size, and threat by using a developmental approach. If we see the race-based size bias in one age range, but not a younger range, then we would have evidence that that specific age is where the bias begins. For instance, showing that the bias that Black men are perceived as threatening by White participants emerges earlier in development than the physical size bias would be convincing evidence for Black men being perceived as threatening informing their exaggerated size perception.

Through this section we discussed White children's anti-Black biases. Then, we looked to the literature to understand how children think about Black people in relation to threat and physical size, separately. As a result of little research existing on how White children think about Black people in relation to both physical size and threat, this piece is touched on briefly.

Children and Anti-Black Biases

To begin, it is important to understand how children's biases present and develop at a young, formative age. We see evidence of an implicit anti-Black bias in a study by Perszyk and colleagues (2019) where preschool-aged children feel more positively about neutral stimuli after being presented White faces relative to Black faces.

As White children get older, we see pro-White or in-group and anti-Black or out-group biases continue to develop. Positive in-group and negative out-group associations begin around age 3 and continue to develop throughout one's childhood, matching adult's attitudes around age 10 (Aboud, 2003; Baron & Banaji, 2006). At 5 years of age, White children who grow up in homogeneous environments display in-group favoritism and thus, feel more negatively about an out-group (Aboud, 2003; Rutland et al., 2005; McGlothlin & Killen, 2006; McGlothlin & Killen, 2010; Gedeon et al., 2021). The positive in-group and negative out-group biases that begin in early childhood not only provide evidence of racial categorization, but an important differentiation where children think negatively about group members that do not 'like' them.

Over development, racial biases form until they are at a level similar to adults. A study by Baron & Banaji (2006) found White children as young as 6 years of age show an implicit bias on a Child Implicit Association Task (IAT) where they have pro-White and anti-Black attitudes. While the researchers found that at 6 years of age, children are already showing a pro-White/anti-Black attitude, by 10 years of age, implicit pro-White/anti-Black attitudes were at the same magnitude as the adults in the study (Baron & Banaji, 2006).

These pro-White/anti-Black biases link to an awareness and endorsement of social stereotypes. Previous research has shown that by 6 years of age, children are aware of racial stereotypes (Ambady et al., 2001; Castelli et al., 2009; Devine, 1989; McKown & Strambler, 2009; McKown & Weinstein, 2003; Sierksma et al., 2022), and that by adolescence, children are more likely to explicitly endorse the stereotypes they have learned through society and their family's beliefs (Copping et. al., 2013; Okeke et al. 2009; Rowley et al., 2007). The endorsement of these stereotypes have negative consequences for young Black individuals and their academic self-concepts (Evans et al., 2011), which could theoretically link back to the disproportionate amount of Black children being suspended compared to their White counterparts (Caldera, 2018; Graves & Wang, 2022; Wolf & Kupchik, 2016).

With the present evidence about how children's racial biases develop, one may wonder if children map specific stereotypes to certain races. A study by Sierksma and colleagues (2022) found that children 4-8 years old do not apply many stereotypes to Asian, Black, and White children. However, they did see a reported preference for White children over Black and Asian children. The reported preference result is not surprising due to past research showing that White children show a racial in-group bias (Baron & Banaji, 2006). However, findings that child participants do not apply stereotypes to racial out-groups is, perhaps, surprising given adults often report knowledge of a stereotype and are able to map such stereotypes to specific groups. Children's inability to map stereotypes to specific groups provides evidence that the full mapping of stereotypes to specific groups may come later in development (Sierksma et al., 2022; Devine, 1989).

Taken together, research suggests that children categorize people according to race and have general implicit and explicit preferences and associations based on these categories, which appear to be stable across development.

Children's Perception of Black People, Physical Size and Threat

Not only do children have broad, negative implicit attitudes based on race, they also are biased when interpreting behaviors of people of different races. We have evidence that children perceive Black children as meaner or more threatening than White children (Sagar & Schofield, 1980; Lawrence, 1991; McGlothlin & Killen, 2010). In a seminal study by Sagar & Schofield (1980), 6th grade children rated ambiguously aggressive acts performed by Black children as meaner and more threatening than the same acts performed by White children. More recently, McGlothlin & Killen (2010) found that White children in racially homogeneous school environments interpret ambiguously aggressive acts by Black children as meaner than the same acts by White children.

The negative interpretations of ambiguous acts have consequences in real-world settings which could be related to the finding that cross-race friendships decline around 5th grade due to racial exclusion (Aboud et al., 2003). Further, Aboud and colleagues (2003) saw that children who were less prejudiced were more likely to be able to maintain cross-race friendships. In a laboratory setting, we see similar patterns where White children explicitly choose same-race White children as their friends rather than Black children (McGlothlin et al., 2005).

There is considerably less evidence on how children think about size and race together than topics related to racial discrimination. However, there is information about how children think about size in general. A study by Carvalho and colleagues (2021) found that children generally prefer people in smaller bodies. Additional research has shown that children are able to think about the cause of individuals' height and weight by four years of age (Carvalho et al., 2021; Cramer & Steinwert, 1998; Harrison et al, 2016; Musher-Eisenman et al., 2004; Penny & Haddock, 2007; Raman, 2014, 2017). Also, children, as young as 3 years of age, think about physical size and posture in order to understand power and who is in charge in a given situation (Lourenco et al., 2016; Terrizzi et al., 2019). Together, the literature tells us that children think about people's physical size in general while adults can think about physical size in relation to race. Given that adults' race-based size bias must develop from somewhere, it seems likely that as children develop, they may begin to think of size differently depending on race and the context presented to them.

It is apparent that preschool aged children begin to reason about race, threat, and physical size, but few studies have investigated the relation between them. One piece of evidence we have of children considering these factors together (e.g., race and threat; race and strength), is the above-mentioned Sagar & Schofield (1980) study. In this study, Black and White 6th grade children perceived Black characters as stronger than White characters showing that they are thinking about physical size and strength in relation to race.

However, there is no research to our knowledge that looks at how children think about race, threat, and physical size together. In order to understand the complicated relation between race, threat, and physical size, one must understand how each of these pieces develop. If researchers understand how the size, race, and threat biases develop and interact, it will be easier to see how such biases can be targeted and prevented from developing into adulthood. This dissertation begins to address the development of how size, race and threat develop and interact in relation to different contexts/vignettes including physical contact, social exclusion, stealing something, and destroying something by means of three experiments with age ranges from 6 years of age to adulthood.

Summary of Findings

Although we have some information about how White people think about size, race, and threat separately, there is a large gap on the combination of these three variables and how they develop over time. Thus far, we have reviewed research regarding three different possibilities for the relation between representations of size, race, and threat. Our first option is that Black men are perceived as threatening, leading to exaggerated size perception. We have evidence of this from Fessler et al. (2012) where threat was manipulated to make a White target be perceived as threatening, participants considered the target to be physically larger. Our second possibility is that the perception of exaggerated physical size leads to threat perception. Evidence for this comes from Wilson et al. (2017) where Black participants perceive Black men as physically larger, but not more threatening than White men, while White participants perceive Black men as both physically large and threatening. Finally, we considered the possibility that the relation is bidirectional or context dependent. There is support for this possibility in Hester & Gray (2018) where the experimenters manipulated the context a man was presented in and it influenced how participants perceived the target.

While each possibility has supporting evidence, we predict the most likely direction for the relation between size, race, and threat works is: you see a Black individual, perceive them as threatening, and then think that they are larger because of it. We see developmental evidence of children showing anti-Black biases early in childhood (Aboud, 2003; Fessler et al., 2012; McGlothlin & Killen, 2010), but scarce developmental evidence that young children possess a race-based size bias. This may be due to a bias in the literature since there is ample evidence about White children's negative biases about Black individuals, but no evidence that White children think Black people as physically larger than White people. However, based on the research available, it seems likely that anti-Black biases emerge first and strengthen over development.

Using the adult literature, when White men were manipulated to be threatening, they are perceived as larger (Payne & Correll, 2020; Fessler et al., 2012). Because we see evidence of White people being perceived as threatening or 'superhuman' only after manipulating threats, it is likely that the relationship works in this direction (Payne & Correll, 2020). Additionally, when threat is not manipulated, we see that the race-based size bias applies to Black individuals, not all out-groups (Johnson & Wilson, 2019; Navon et al., 2025) who are perceived as threatening. Together, this provides convincing evidence for seeing a Black individual, perceiving them as threatening, then thinking they are large.

While we have information on how children think of race, physical size, threat, and strength, having a larger body of research looking at how children think about these variables independently would also help to inform research on the relation between them. To test the development and relation between the variables of the race-based size bias, in Study 1 we investigated the race-based size bias across development (childhood, adolescence, adulthood) by having participants rate the perceived physical size of Black and White faces. If the race-based size bias is not present in children, this would be evidence that the race-based size bias has a later onset in development. In order to grasp a fuller picture of the developmental trajectory of the race-based size bias, investigating if adolescents perceive Black characters as physically larger and stronger than White characters, especially if children do not display a race-based size bias, is crucial. As an addition to our central question of interest, we include a measure investigating how specific vignettes, or contexts, impact how participants think about race and physical size. This measure is largely exploratory, but can provide information about how the bias to rate Black men as physically large and threatening emerges, and potentially, how the context around the person influences how participants think about their physical size.

In Study 2, we focused on whether contextual factors, such as the physicality (e.g., physical contact or social exclusion) and valence (e.g., positive or negative) of an event, changes how participants think about race and physical size across development. We also used Study 2 to replicate our findings from Study 1 to understand the emergence and development of the bias to rate Black men as physically large and threatening.

In Study 3, we investigated whether the cartoon faces used in the previous two studies were valid, or if results would change when looking at real Black and White faces. Study 3 was also used as a replication of the findings from Study 2 for the forced-choice, memory, and open-response questions.

Carrying out these studies gives us additional information on how size, race, and threat work together and aids in the understanding of the developmental origins and cognitive underpinnings of this race-based size bias. These studies aim to address if and when the race-based size bias appears developmentally. This paper is the first, to our knowledge, to explore the race-based size bias developmentally with children, adolescents, and adults. It also aims to address if the manipulation of context changes how the bias presents, or if the bias is static across contextual information. Understanding when the bias develops and under what conditions it presents gives crucial insight into the direction and development of the race-based size bias.

Chapter 2: Study 1

Method

Participants

Children

One hundred and seven (41 boys, 65 girls, 1 nonbinary) 6- to 11-year-old children were recruited from a local museum (n = 69) and through the online platform Children Helping Science (n = 38). Child participants were predominantly White (105/108 families voluntarily reported the child participant's race: 54.8% White, 1.9% Black, 23.1% Asian, 3.8% Latino, 14.4% Mixed-Race, and 1.9% Other). Participants were recruited in three age cohorts with at least 34 participants per age cohort (6-7 years, n= 34; 8-9 years, n=34; 10-11 years, n= 39) to ensure sufficient representation across the full age range. Two additional participants were dropped from the final sample; one because they rated all the events positively (see Preliminary Analysis section below) and one because they did not complete at least half of the trials.

Adolescents

Sixty eight (38 male, 30 female) 12- to 17-year-old predominantly White participants were recruited from a local museum. Participants were predominantly White (68/68 participants voluntarily reported race: 76.5% White, 2.9% Black, 7.4% Asian, 4.4% Latino, 4.4% Mixed-Race, 4.4% Other). Participants were recruited in two separate age cohorts (12-14 years, 15-17 years; n = 34/cohort) to ensure sufficient representation across the full age range. One additional participant was dropped from the final sample because they did not complete at least half of the trials.

Adults

Forty nine (10 male, 38 female, 1 non-binary) undergraduate students participated in this study. Participants were predominantly White (49/49 participants reported: 38.8% White, 12.2%

Black, 34.7% Asian, 4.1% Latino, 8.2% Mixed-Race, 2% Other). One additional participant was dropped from the final sample because they rated all the events positively (see Preliminary Analysis section below).

Power Analysis

Using G*Power a power-analysis was conducted for a dependent samples t-test with a power of 0.8 and a medium effect size of 0.5. Thirty-four participants were sufficient to detect a significant effect of character race with each age group. The total sample size across age groups was also sufficient to detect a significant interaction between Participant Age Group and Character Race.

Procedure

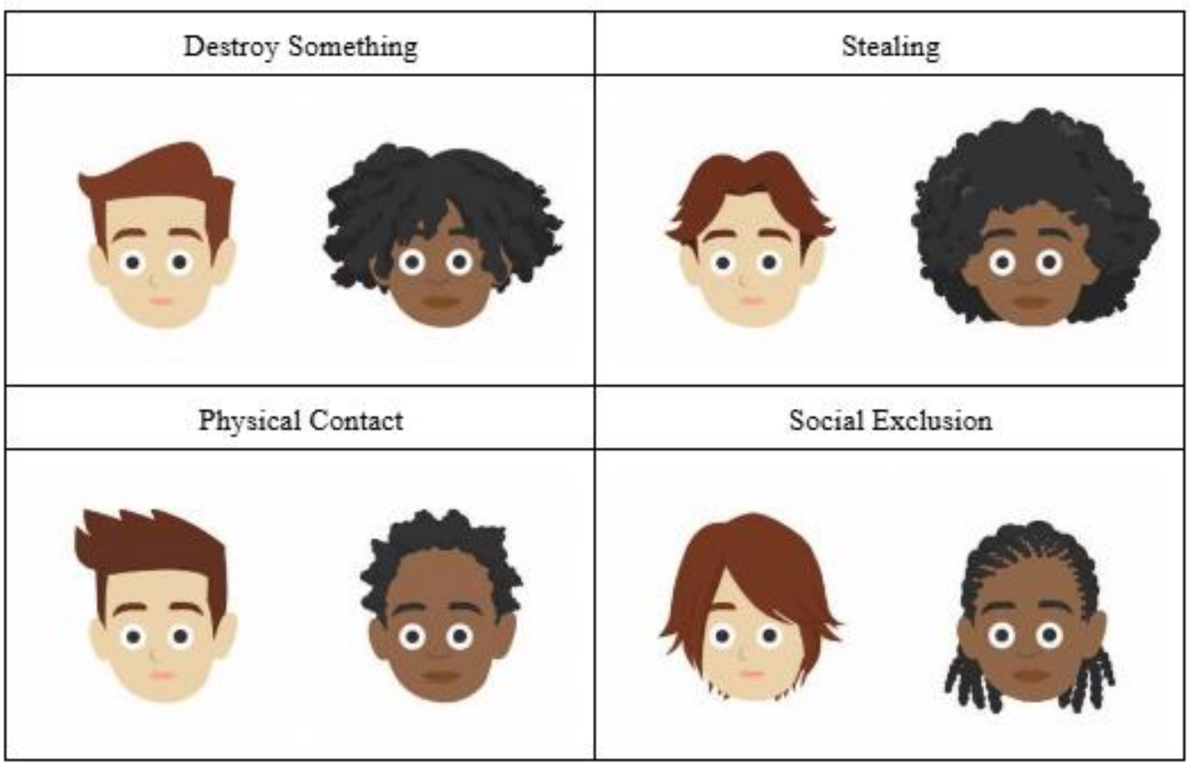
In-person child and adolescent participants were recruited at the Museum of Science after researchers identified them to be within the age range of 6-17 years old. Children under 18 years of age and their parents were told that they would be participating in a study about “race and physical size” or “the physical size and behaviors of Black and White children”. Once participants agreed to participate, their parents filled out a parental permission form and children provided verbal assent to participate. Children provided researchers with verbal affirmation that they wanted to participate and clicked a button that said ‘yes’ to continue the experiment. Participants were given the option to operate the computer themselves, but some preferred to point to the computer screen and have the experimenter press the buttons and operate the computer for them. Participants sat next to the experimenter so that they could help read out the scales or assist as necessary with the procedure. Museum participants were provided with a sticker for their participation.

Online child participants were recruited through Children Helping Science. On this site a description of the study was posted explaining the study and parents reached out if they wanted their child to participate. Children in the online study were run synchronously on Zoom with their screen shared so experimenters could see what the participants were seeing and could help walk them through the study as needed. Online participants and their parents followed the same permission and consent protocols as the in person participants did. Participants were provided a $5 Amazon gift card for their participation.

The procedure for the adult participants was identical to the procedure for the child and adolescent participants except the adults took the study asynchronously and had total control of the computer and clicked through the experiment themselves. Adults were given course credit for their participation in the study.

Figure 1

Cartoon faces and the Event Type category they were presented with in Study 1



All participants saw a total of eight faces one at a time in a randomized order; four faces were Black and the other four faces were White (Figure 1). Faces were chosen based on a pre-study based on the categorization of each of the faces as ‘Black’ and ‘White.’ The pre-study consisted of thirty-five additional undergraduate students categorizing 12 Black and 12 White cartoon faces. Participants were asked in an open-response format, ‘Which race is this character?’ and ‘How confident are you in your response? on a Likert scale from 1-5’. Four Black and four White faces were chosen based on which faces were categorized as ‘Black’ and ‘White’ most often, and with the highest confidence. The four selected Black cartoon faces as ‘Black’ with high confidence (M= 4.26, 85.1%), and the four selected White cartoon stimuli were categorized as ‘White’ with high confidence (M= 3.99, 79.8%).

Table 2

Descriptions of the vignettes broken down by category in Study 1

Event Type	Descriptions
Physical Contact	● Harry ran over to Gerald and bumped into him, making Gerald drop all of his books ● Landon walked over to Mike and pushed Mike off the swings
Destruction	● Jonas grabbed Ian's favorite bear he brought in for show and tell, and ripped off the bears arm ● Frank went over to Edmunds desk, picked up his homework, and ripped it in half
Theft	● David saw Casey walking through the hallway and saw his Ipad dropped out of his backpack. David picked up the Ipad and put it in his backpack to take home. ● Omar didn't like what he had for lunch, so he went over to Kieran, took his sandwich, and ate it in his seat.
Social Exclusion	● Walden said “There is no room for you” when his classmate Xavier asked if he could sit in an open seat at lunch ● Uriel was picking teams to play soccer at recess. When Vincenzo asked he could play, Uriel said “No” and told him to “do something else”

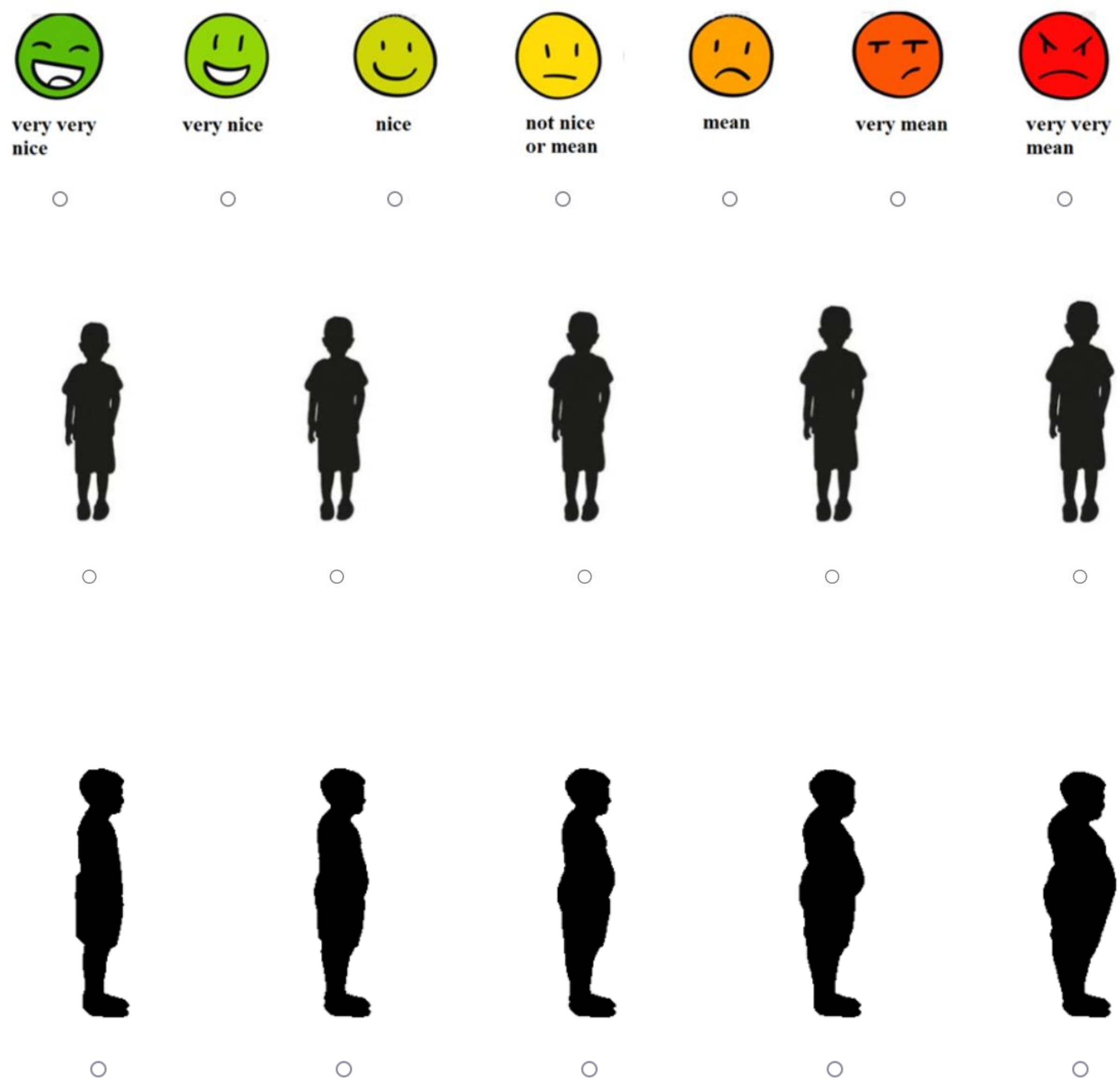
Each of the faces were presented with a verbal pre-recorded vignette under the categories of: Physical contact (e.g., pushed someone off the swing), destruction (e.g., ripping someone's homework), theft(e.g., taking someone's lunch), and social exclusion (e.g., won't let someone sit with them at lunch). A table including descriptions of each of the vignettes and their respective categories is listed in Table 1. To understand the developmental origins and development of the race-based size bias, we ask children, adolescents and adults to rate the physical size and behaviors of Black and White cartoon children. Following previous related vignette-based study designs (Sagar and Schofield, 1980; McGlothlin, Killen & Edmonds, 2005; McGlothlin & Killen, 2010), we used negative vignettes allowing for a level of 'threat' to be activated. While we do not want to elicit threat from participants using dangerous objects such as guns paired with children, instead, we aimed to provide child appropriate levels of threat through stories relating to bullying and stealing.

After each vignette-face pairing, participants answered a series of questions about the character's physical characteristics and traits. Participants were asked "How nice or mean was Frank" and answered on a scale of 1-7 ranging from "very very nice" to "very very mean" accompanied by happy and sad faces (Figure 2). Participants were then asked "How weak or strong was Frank?" and answered on a scale of 1-7 ranging from "very very weak" to "very very strong." There was no physical scale included as we wanted participants to interpret 'strength' on their own terms. This meant participants could be rating mental strength, physical strength, or some other form of strength the participant was considering.

Then, participants were presented with a memory check asking if they remembered what the character did. If they did not remember what the character did, they were reminded. In this study, we did not keep track of whether participants remembered the vignette on their own, or if they were reminded; we address this issue in Study 2. Following the memory check, participants were asked "How short or tall do you think Frank is?” and answered on a scale of 1-5 depicted by varying increasing heights (Figure 2). Finally, the participants were asked “How thin or heavy do you think Frank is?” and answered on a scale of 1-5 depicted by varying increasing weights (Figure 2).

Figure 2

Scales used for Study 1, 2 & 3: Niceness/meanness, height, weight



Presentation order was counterbalanced across participants in which half the time participants were asked about their trait characteristics first: how nice or mean was the character and how weak or strong was the character? Then, they were presented with a memory check asking if they remembered what the character did. Finally, they were asked: how short or tall was the character and how thin or heavy was the character? The other half of the participants were asked about the physical characteristics first (e.g., height, weight), then the memory check, followed by the trait characteristics (e.g., nice, strength).

Results

Preliminary analyses

We conducted a preliminary analysis on participant gender across ages and strength, height and weight ratings. There were no significant main effects of participant gender, so we did not include it in further analyses. Another preliminary analysis assessed effects of presentation order strength, height, and weight ratings across ages; this analysis did not reveal any significant main effects of presentation order, so it was not included in further analyses.

Although our intent, in line with prior research (Sagar and Schofield, 1980; McGlothlin, Killen & Edmonds, 2005; McGlothlin & Killen, 2006), was to present participants with mildly negative events, preliminary analysis of the meanness ratings revealed that participants viewed some events as positive interactions, choosing ‘very, very nice,’ ‘very nice’ or ‘nice’ (3 or below) on the meanness scale (child: N=60 trials including one full male participant who was run at the museum; adolescent: N=7 trials; adult: N=16 trials including one full female participant). To ensure we were assessing participant’s representations of the size of actors in negative events, we removed these individual trials from all subsequent analyses. This resulted in 107 (N=41 males) child participants, 68 (N=38 males) adolescent participants, and 47 (N= 10 males, 1 non-binary) adult participants analyzed in our final sample.

Finally, we conducted a preliminary analysis on specific Event Type across ratings. Although we recognized there would be variability in responses across these events, we did not predict that Event Type would have a significant impact on ratings. Contrary to that prediction, we found that Event Type did have a significant influence on size ratings across age. As such, we included event type as a fixed factor in all of the following analyses and report those findings below.

Overall analysis plan

We conducted linear mixed effects regression models for all the analyses reported below. Unless noted otherwise, all models included Character Race (Black, White) and Event Type (Destruction, Physical Contact, Social Exclusion, Theft) as fixed factors. Participant Age Group was entered as a fixed factor, either at the broad level (Child, Adolescent, Adult) for initial overall models or at the recruitment-based cohort level (Children: 6-7 years, 8-9 years, 10-11 years; Adolescents: 12-14 years, 15-17 years) for within age group follow-up analyses. All models included random intercepts for participants and random slopes for Event Type and Character Race. We also tested, and retained in the final models, any significant interactions between fixed factors.

Strength ratings

Strength ratings across development: children, adolescents, and adults

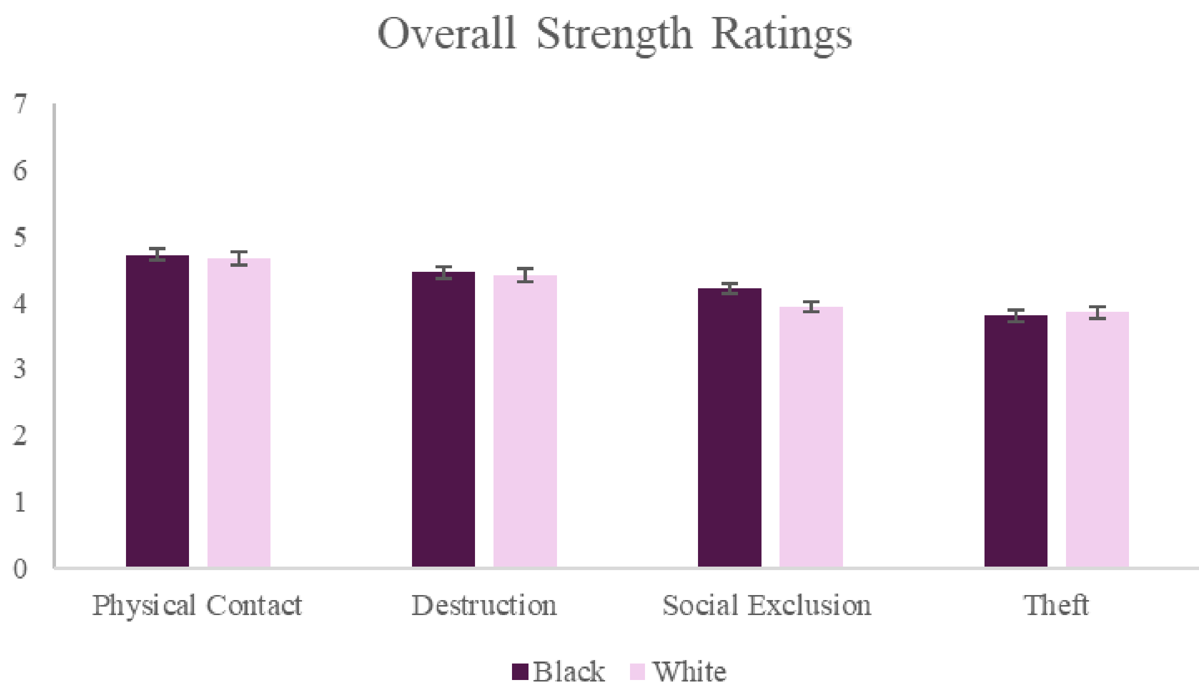
Character Race, $F(1, 243) = 5.05$, p= .025, Participant Age Group, $F(2, 213)= 10.52$, p< .001, and Event Type, $F(3, 337)= 43.89$, p <.001 all significantly predicted participants' strength ratings. Black characters (*M*= 4.30, *SD*= 1.29) were rated as significantly stronger than White characters, (*M*= 4.22, *SD*= 1.35; t(214)= 2.25, p=.025). Children (*M*= 4.55, *SD*= 1.32) rated characters as stronger overall than adolescents (*M*= 3.97, *SD*= 1.21; t(219)= 4.25, p <.001) and adults (*M*= 4.06, *SD*= 1.36; t(219)= 4.25, p=.002), but there was no difference between adolescents and adults' strength ratings, *t*(217)= .654, p= .51. Finally, characters who engaged in physical contact events were significantly stronger (*M*=4.69, *SD*= 1.28), compared to all other event types (all *p*s > .05).

These effects were qualified by a significant three-way interaction between Character Race, Participant Age Group, and Event Type, *F*(6, 1041)= 3.19, *p*= .004, as well as two-way interactions between Character Race and Participant Age Group, *F*(1, 243)= 3.63, *p*= .028, and Character Race and Event Type, *F*(3, 1041)= 3.54, *p*= .014. The Character Race and Event Type post-hoc tests revealed that this interaction was driven by the influence of Character Race on strength for social exclusion events. Specifically, Black characters who socially excluded others (*M*= 4.21, *SD*= 1.06) were perceived as stronger than White characters who socially excluded others (*M*= 3.93, *SD*= 1.03; *t*(848)= 3.84, *p*< .001), but there were no significant differences between Black and White characters perceived strength for any other event (all *p*s > .05). The findings are presented below in Figure 3.

To further explore the overall interaction between Character Race, Participant Age Group, and Event Type, we then fit separate models for each age group, assessing the influence of Character Race and Event Type, along with their interactions, on participants' strength ratings. We discuss significant findings from these models, in turn, below.

Figure 3

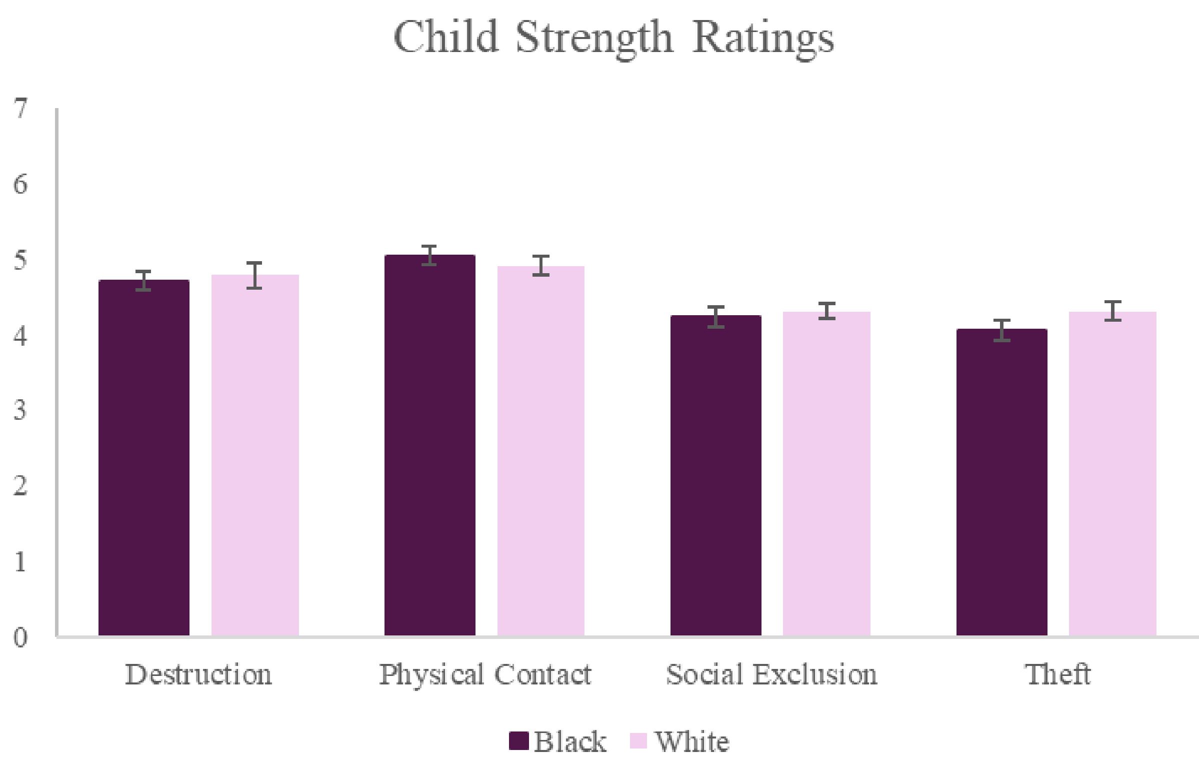
Strength ratings across children, adolescents and adults for Study 1



The model for child participants yielded a significant effect of Event Type, *F*(3, 145.8)= 23.22, *p* <.001. Children rated characters who engaged in physical events as significantly stronger (*M*= 4.98, *SD*= 1.23) compared to all other event types. Also, characters who destroyed something were perceived as significantly stronger than those who socially excluded, *t*(102)= 4.22, p< .001 and stole, *t*(101.5)= 5.33, p< .001. However, there were no significant effects of either Character Race, *F*(1, 112.2)= 0.539, *p*= .464 or Age Cohort, *F*(2, 97.9)= 2.07, *p*= .132 on participants' strength ratings. There were also no significant interactions between the factors, so they were removed from the model. These findings are displayed in Figure 4.

Figure 4

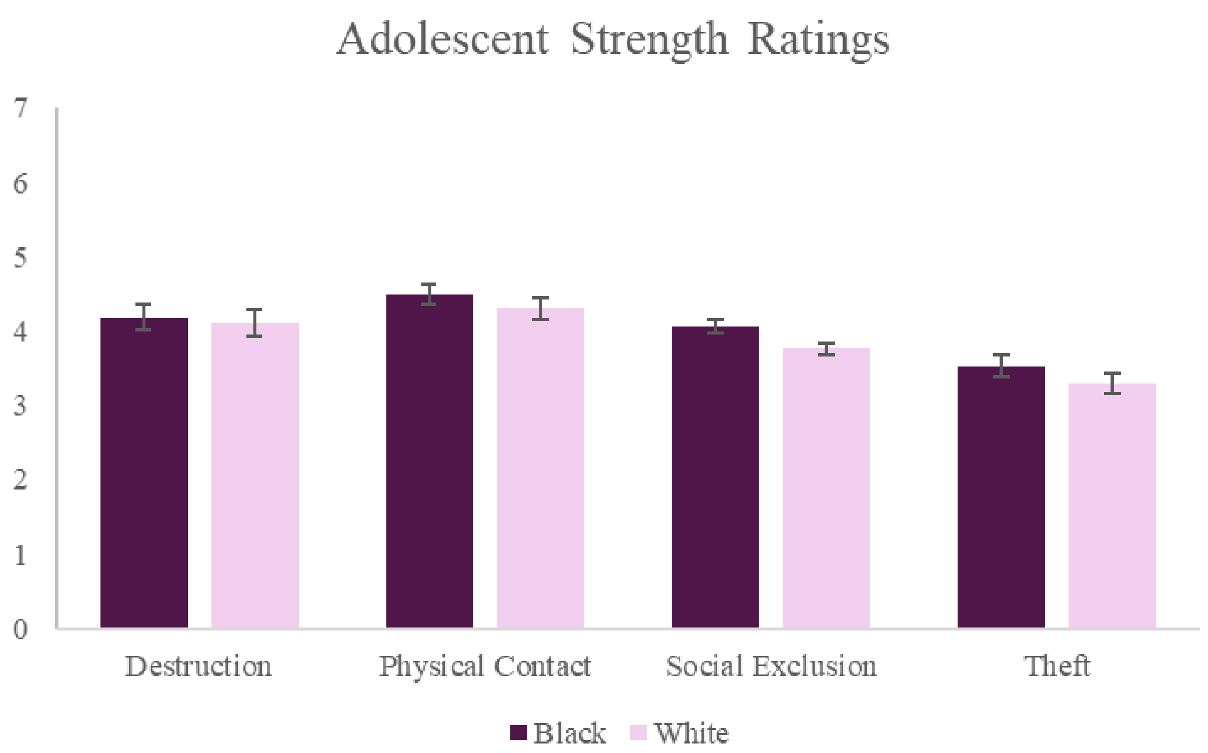
Child (6-11 years) strength ratings for Study 1



The model for adolescent participants yielded significant effects of Character Race, *F*(1, 74.7)= 6.85, *p*= .011, and Event Type, *F*(3, 100.2)= 22.81, *p*<.001. Participants rated Black characters as stronger (*M*=4.07, *SD*= 1.18) than White characters (*M*= 3.87, *SD*= 1.23). Adolescents' rated characters who engaged in physical contact events as stronger than all other event types. Those who engaged in destruction, *t*(66.9)= 6.16, p< .001 and social exclusion, *t*(66.9)= 4.27, p< .001 were stronger than characters who engaged in stealing. Similar to the child participants, this model yielded no significant effects of Age Cohort, *F*(1, 82.6)= 1.27, *p*= .264 on participants' strength ratings. There were no significant interactions between any factors in this model, so they were removed. These findings are displayed in Figure 5.

Figure 5

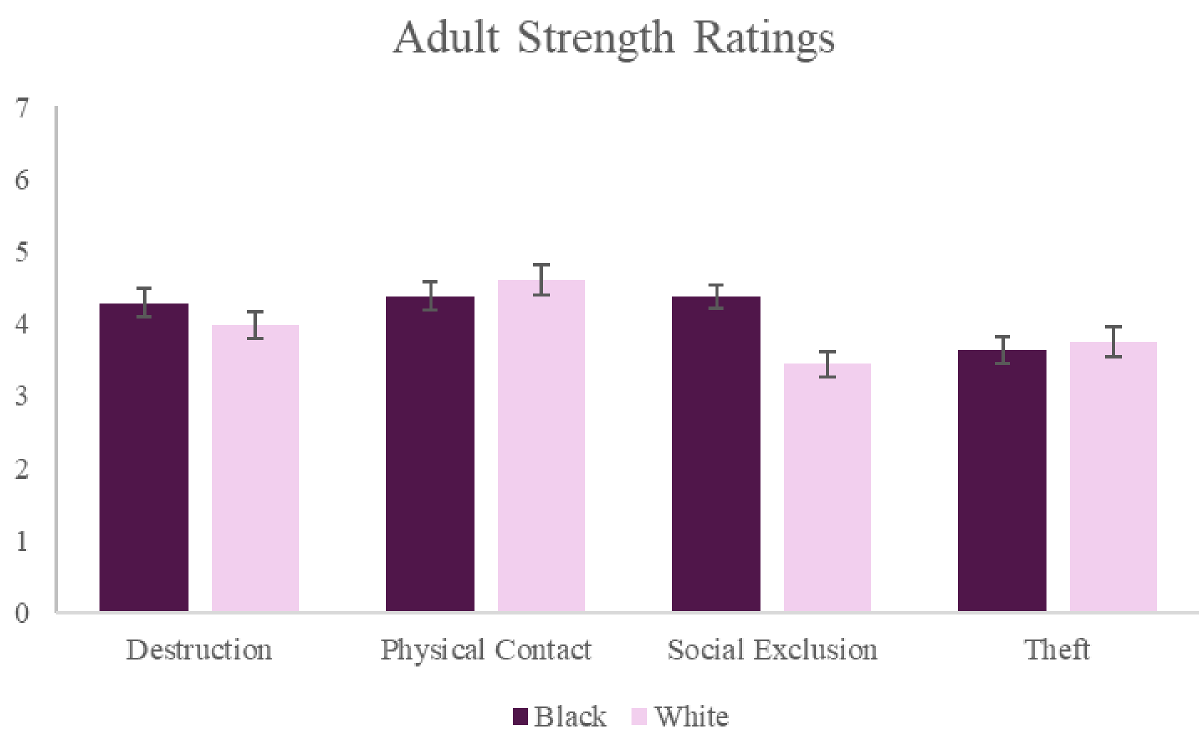
Adolescent (12-17 years) strength ratings for Study 1



For the adult participants, we again conducted a similar model described above, except Age was not entered as a fixed factor. We found that Character Race, *F*(1, 70.2)= 4.24, p= .043 and Event Type, *F*(3, 70.6)= 7.57, p < .001, both significantly predicted participants' strength ratings. Adults' rated Black characters as stronger (*M*= 4.17, *SD*= 1.29) than White characters (*M*= 3.94, *SD*= 1.41). Adults also rated characters who performed in physical contact events as stronger (*M*= 4.49, *SD*= 1.41) compared to destruction (*M*= 4.13, *SD*= 1.35); *t*(47.8)= 2.44, *p*= .019), theft (*M*=3.69, *SD*= 1.33; *t*(47.4)= 4.61, *p*< .001) and social exclusion (*M*= 3.91, *SD*= 1.23; t(47.7)= 3.55, p < .001) events. Finally, these effects were qualified by a significant interaction between Character Race and Event Type, *F*(3, 234.3)= 6.15, *p*= .001. Post-hoc tests revealed that this interaction was driven by the influence of Character Race on strength for social exclusion events. Specifically, Black characters who socially excluded others (*M*= 4.38, *SD*= 1.11) were perceived as stronger than White characters who socially excluded others (*M*= 3.44, *SD*= 1.17), $t(78.2) = 4.49$, $p < .001$, but there were no significant differences between Black and White characters perceived strength for any other event (all *p*s > .05). These findings are displayed in Figure 6.

Figure 6

Adult strength ratings for Study 1



Together, adolescents and adults perceived Black characters as stronger than White characters, but children did not. Instead, children used what a character did, rather than their race, to inform their judgments. Adolescents used both race and behavior to inform their judgments, while adults used race and behavior individually, as well as together, to inform their strength judgments. Specifically, children, adolescents, and adults all perceive characters who engage in physical events as stronger than those in other events. Adults perceived Black characters who socially excluded others as stronger than White characters who socially excluded others.

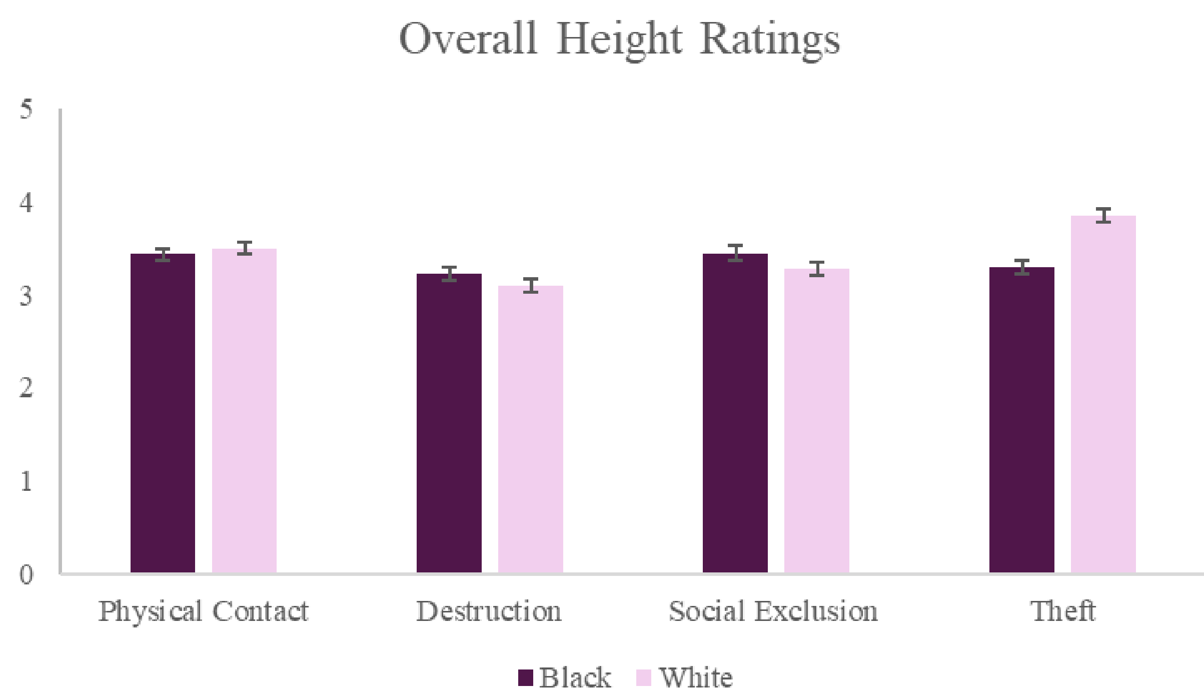
Height ratings

Height ratings across development: children, adolescents, and adults

Event Type, $F(3, 279)= 3.5$, $p= .016$ significantly predicted participants' height ratings, but Character Race, $F(1, 237)= .014$, $p= .906$, and Participant Age Group, $F(2, 215)= .184$, $p= .832$ did not. Participants' rated characters who engaged in physical contact events as physically taller (M=3.46, SD= 1.03), compared to destruction (M= 3.27, SD= 1.04; $t(220)= 2.92$, p=.004), and theft (M= 3.3, SD= .99; $t(217)= 2.2$, $p= .01$). There were no significant interactions, so they were removed from the model. The findings are presented below in Figure 7.

Figure 7

Height ratings across children, adolescents and adults for Study 1



Because we saw no evidence of a race-based size bias for height across age groups, the subsequent analyses broken down by age group are in the Supplemental Information, where there are some significant findings within the adolescent and adult age groups.

Weight ratings

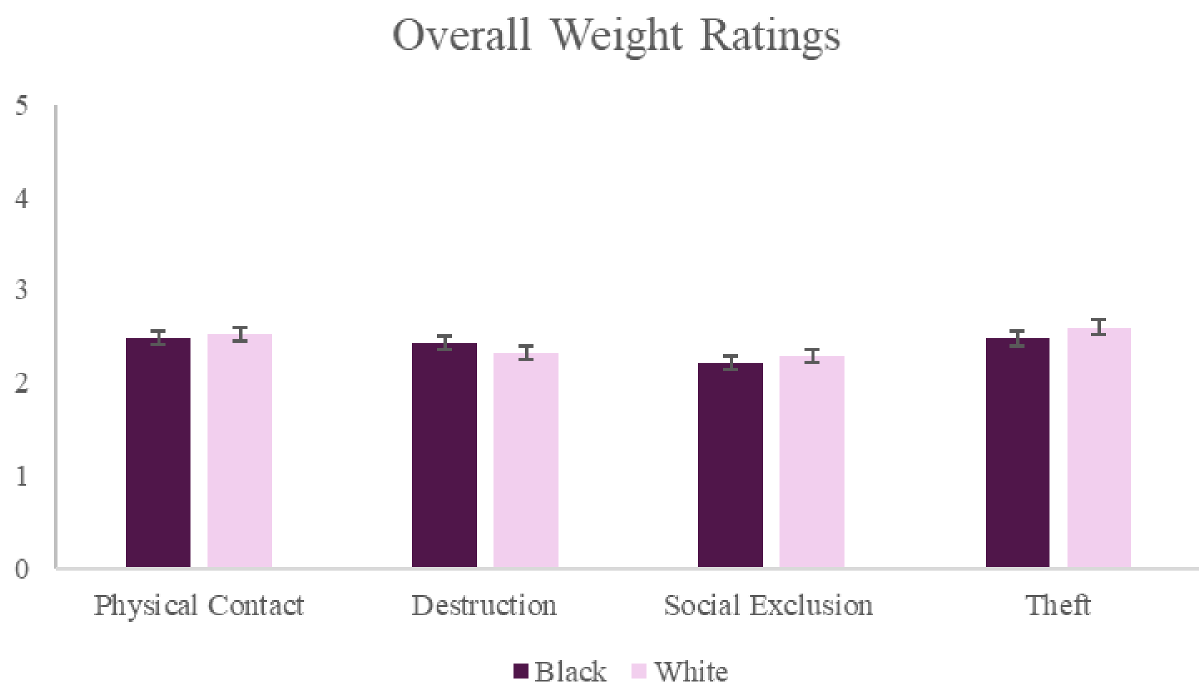
Weight ratings across development: children, adolescents, and adults

Event Type, $F(3, 415)= 13.37$, $p <.001$ significantly predicted participants' weight ratings, but Character Race, $F(1, 354)= 1.93$, $p= .17$ and Participant Age Group, $F(2, 216)= .656$, $p= .520$ did not. Using social exclusion as the reference category, post-hoc tests revealed that characters who engaged in social exclusion ($M= 2.25$, $SD= 1$) events were significantly *thinner* than those who engaged in physical contact (*M*= 2.5, *SD*= 1.04; *t*(213)= 4.66, *p*< .001), destruction (*M*= 2.38, *SD*= 1; *t*(213)=2.74, *p*= .007) and theft (*M*= 2.54, *SD*= 1.11; *t*(211)= 5.85, *p*< .001) events.

These effects were all qualified by a significant interaction between Participant Age Group and Event Type, *F*(6, 415)= 2.15, p= .047. Post-hoc tests revealed adolescents and adults had significant effects of Event Type on weight ratings using social exclusion as the reference category, but children did not. Adolescents perceived characters who engaged in social exclusion (*M*= 2.16, *SD*= 0.9) as *thinner* than those who engaged in all other event types (physical contact: *M*= 2.53, *SD*= .99, destruction: *M*= 2.44, *SD*= .90, theft: *M*= 2.64, *SD*= 1.03; all *p*s > .05). Adults perceived characters who engaged in social exclusion (*M*= 2.2, *SD*= 0.9) as *thinner* than those who engaged in physical contact (*M*= 2.6, *SD*= 1.01, *t*(212)= 3.07, *p*= .002) and theft (*M*= 2.69, *SD*= 1.19; *t*(212)= 3.99, *p*<.001) events, but not destruction events (*M*= 2.44, *SD*= 1.06; *t*(212)= 1.88, *p*= .06).

Figure 8

Weight ratings across children, adolescents and adults for Study 1



Because we see no evidence of a race-based size bias for height across age groups, the subsequent analyses broken down by age group are in the Supplemental Information, where there are some significant findings within age groups.

Correlations between Strength, Height and Weight

Finally, we assessed the relation between participants' ratings of strength, height and weight. Given height and weight are often correlated in the general population, we investigated whether participants' ratings of height and weight were correlated within our study; characters who were rated as taller and heavier may also be stronger.

The correlation matrix is presented in Table 3 below and displays Pearson correlation coefficients for strength, height and weight. Notably, we see a positive correlation between height and weight (r= .208, p=.002), but no other significant correlations. These findings appear to mimic real-world height and weight relationships where people who are taller tend to weigh more. Strength does not appear to be related to height or weight in a correlational manner, perhaps because race and context were considered more readily than the potential relations between strength, height, and weight.

Table 3

Correlation table between strength, height, and weight for Study 1

	Strength	Height	Weight
Strength	- -	- -	- -
Height	Pearson's r= .056 *p*= .408	- -	- -
Weight	Pearson's r= -.015 *p*= .827	Pearson's r= .208 *p*= .002	- -

Study 1: Summary of findings

Together, the results begin to explain the race-based size bias from a developmental perspective. Overall, participants perceived Black characters as stronger than White characters. Through investigating the interactions and relations between Character Race and Participant Age Group, we saw that children did not perceive Black characters as stronger than White characters, but adolescent and adult participants did. Children based their judgements on the way a character behaved rather than their race. Across ages, we also saw that Event Type mattered. Characters who engaged in physical contact events were perceived as stronger than many of the other event types. The contextual factors presented by the events were so robust that children used only the contextual information to make predictions about the targets' strength. Additionally, adolescents and adults used these contextual factors to determine their strength ratings, but they also took Character Race into consideration. Across height and weight ratings, Character Race did not influence participant ratings. Instead, the Event Type was consistently used to make predictions about characters height and weight. This provides us with important information about the environmental context someone is presented in and warrants further investigation of these contexts.

While Study 1 provided us with information about the potential onset of the race-based size bias, it left many open questions. Study 1 employed one single way to measure the race-based size bias. Likert scales were used to provide responses to cartoon faces and vignettes, but this one measure may have limited our understanding of the race-based size bias. It is possible that using a more sensitive measure, such as a forced-choice design, would show sensitivity to a race-based size bias in children for the first time. Additionally, Study 1 used only negative events. While this was intentional based on previous research (Sagar and Schofield, 1980; McGlothlin, Killen & Edmonds, 2005; McGlothlin & Killen, 2006), it is possible that positive events may operate in a different capacity. Perhaps White characters would be perceived as strong in positive situations, showing that participants think that strength is positive in some situations, while Black characters would be perceived as strong in negative situations. Determining when participants apply different heuristics could be informative about how they think about strength and size depending on the race of the character and the context they are presented in.

Another limitation of this study was that we did not track participant's memory for events. While we did ask participants if they remembered an event, and reminded them if they did not, we did not mark whether or not participants remembered the events. In the future, investigating how participants remember events could be informative. For example, will participants remember positive or negative events better? Will they remember positive events performed by White characters better than positive events performed by Black characters better?

Finally, all stimuli in Study 1 were cartoon faces. While this was beneficial for control and internal validity of the experiment, including ensuring that the Black and White faces were the same size, it limited the external validity of the experiment. In future experiments, real faces should be incorporated into different elements of the design to check the external validity and generalizability of the findings.

Chapter 3: Study 2

Study 2 continued to investigate the development of the race-based size bias that White adults have to perceive Black men as physically larger and stronger than White men of the same physical size (Wilson et al., 2017), while addressing limitations and future directions from Study 1. In Study 2, we aimed to answer: how the race-based size bias may change based on both positive and negative vignettes, how real faces may be perceived by children, adolescents and adults rather than cartoon faces, how people report they are making decisions about target's physical strength and size, and if participants remember events better under certain conditions (e.g., positive events by White targets, negative events by Black targets). Study 2 addresses the possibility that children may have a race-based size bias, but the measures of Study 1 were not sensitive enough to detect it.

To begin, in Study 1, all the vignettes were intended to be perceived as negative. Negatively valenced interactions would signal a potential level of 'threat' to participants, which we hypothesized would be an important variable to consider. Additionally, Study 1 was modeled conceptually off of previous research that investigated interactions between Black and White children that were made to be ambiguous, but were often perceived as negative when they included Black children (McGlothlin & Killen, 2010; McGlothlin & Killen, 2006; Sagar & Schofield, 1980). Creating novel negative interactions based on established work allowed for confidence in our predictions of how the events would be perceived. Because it was important to Study 1 that the events were perceived as negative, we dropped trials from participants that viewed negative events as positive. However, investigating only negatively valenced events tells only part of the story.

In Study 2, we included both positive and negative interactions. This could provide us with important theoretical information regarding differences in events by race. Specifically, starting at 3 years of age and persisting throughout adulthood, White people have positive in-group and negative out-group biases (Aboud, 2003; Baron & Banaji, 2006). It is possible that when participants see a White person doing something positive, they may think that person is stronger and physically larger in a favorable way. However, when White participants see a Black person doing something negative, they may think that person is stronger and physically larger because they are feeling threatened. Previous research shows that height is beneficial for White men and results in them seeming competent, while height for Black men often results in stop-and-frisk encounters by the police that may end in police force, and the biased perception that they are threatening (Hester & Gray, 2018; Milner et al., 2016). Given it appears that threat perception could be biasing individuals' judgments, dampening the threat perception could lead to a reduction in the race-based size bias. If we investigate the race-based size bias from both positive and negative interactions, it could better explain if the bias not only operates differently based on race, but also based on the contextual information provided and the amount of threat that is elicited. If Black people who perform negative events are perceived as physically larger and stronger than Black people who perform positive events, we would have support that 'threat' is leading towards Black people being perceived as physically larger and stronger than White people.

To conceptually replicate event type findings from Study 1, we included the use of social and physical events only. Because each of these event types significantly differed from one another, we felt they were important to consider looking into further. We decided to not include destruction and theft, for different reasons. First, our destruction vignettes may have provided more context than we would have liked on the strength measure. For example, ripping the arm off of a teddy bear must require some physical strength, which could bias our results. Anecdotally, child and adolescent participants often commented on this event saying that it would take significant strength to rip the arm off a teddy bear, so we felt the removal of this event type would be advisable. Additionally, we removed the theft event type for Study 2. This is because the theft events were ambiguous. While these events were meant to be perceived as negative, some participants did not think the children were actually stealing. Instead, they thought they may have been helping someone by picking up their item, or borrowing from them. Finally, there is not a comparable 'positive' version of theft or destroying something. Because it was important for us to investigate valence in Study 2, using social exclusion and physical contact were selected as the events. It was not expected that participants have difficulty understanding the positive and negative versions of each of these events.

In addition to the physical size and strength vignette ratings using the cartoon faces, we introduced a new forced-choice design in this experiment using real faces. This provided an opportunity for participants to decide if Black or White targets are stronger, taller, or heavier, without being given any additional information about the target. This measure is also much simpler and more sensitive than a Likert scale, and may be more likely to show evidence of children having a race-based size bias towards Black individuals in a way that a less sensitive measure does not detect. While we believe contextual information may be important for the race-based size bias in adults, the contextual information we provided may have dampened the effects of the race-based size bias in children. We predicted that adolescents and adults would continue to display a race-based size bias for strength using this forced-choice measure. To our knowledge, no research has yet investigated the race-based size bias using a forced-choice design developmentally, but it has been employed in studies relating to race research in the past with Black stereotypical and non-stereotypical faces and neutral or threatening expressions (Kleider-Offutt et al., 2018) and racial categorization for multiracial individuals with child and adult participants (Roberts & Gelman, 2015; Peery & Bodenhausen, 2008). We also predict we may see child participants begin to show race-based size biases for strength, as this measure is more explicit and is less open to additional contextual factors such as behavior.

Additionally, our forced-choice measure used photographs of real faces to potentially identify a race-based size bias in participants, rather than the cartoon faces used in the vignette ratings. This is designed to compliment our results from our vignette component of Study 2. While using cartoon faces provided experimental control by ensuring that all faces were the same size, it also limited the potential generalizability of the findings. Because of this, we included real faces as a component of the study design for Study 2. It's possible that the race-based size bias is reduced overall by cartoon faces, so the use of real faces should begin to determine if a race-based size bias isn't present in children due to a lack of the bias or due to a methodological issue. Previous research has shown evidence of the race-based size bias by simply eliciting the idea of a Black individual by using images of identical bodies labeled as Black or White or with stereotypically Black and White names (Wilson et al., 2017), so we predicted that both the Black and White cartoon faces, as well as the real faces, would be able to elicit the race-based size bias effectively.

We also included an additional open-response exploratory measure to assess participants' views on why they judge others to be stronger, taller, and heavier. Using qualitative data to compliment our quantitative responses can give us a more well-rounded view of the race-based size bias across ages. Rather than using measures that may not be clear to participants to try and understand their race-based size bias, participants can tell us explicitly how they are making their decisions. It also provides alternative explanations for our results by giving participants the freedom to provide any response, without the restraint put upon them by the researchers (Haddock & Zanna, 1998). This method is common in research dealing with attitudes and stereotypes and has been reliably used in research to test attitudes related to teacher's stereotypes of Black and White students (Chang & Demyan, 2007), and stigmatization of individuals with intellectual disabilities (Pelleboer-Gunnink et al., 2019). While we do not expect the majority of participants to mention race outright in their responses, we do still expect to see a variety of explanations for how people make physical size judgments. If participants do mention racial information outright, then it would provide evidence that people are aware of the race-based size bias and are aware they are using that bias to make their physical size judgements. Additionally, even if participants are not mentioning race as a determinant for their size judgments, we still hypothesized that the race-based size bias persists in both adolescents and adults. This discrepancy may be indicative of participants being unaware of their own race-based size biases or simply being uncomfortable mentioning such stereotypes.

Additionally, Study 2 used memory as a measure to determine if children, adolescents, and adults are tracking the information provided to them and to collect information on how participants are remembering positive and negative information by race. Human memory is susceptible to many kinds of errors, and investigating what is correctly remembered could give insight into participants' beliefs. Previous research provides mixed results on whether positive or negative events are better remembered. One study found that when participants were asked to recall either traumatic or positive events and write about them, there were no statistical differences for event valence recall controlling for emotional intensity and the amount of time that has passed (Waters et al., 2013). On the other hand, when researchers used social interactions as a model, they found that participants were more likely to forget negative events than positive ones, though it depended on the kindness of the behavior and who performed the behavior (Li, 2012). This contrasts previous studies suggesting in non-social events (e.g., photograph recall), negative emotions helped participants to better remember events (Kensinger et al., 2006).

Supporting the latter evidence, the leading belief in the field is that negative events are remembered better than positive ones (Proverbio, Mastra & Zani, 2016; Ochsner, 2000; Keightley et al., 2011). To begin, participants recall more negative, true events compared to positive, true events when asked to recall public events (Porter et al., 2008). Developmentally, children as young as 3 months of age privilege negative social information compared to positive information using a looking time paradigm (Hamlin et al., 2010). Similarly, 4 year old children are better at recalling negative social events compared to positive ones (Baltazar et al., 2012). Also, adolescents with depression and anxiety remembered negatively-valenced words compared to positive or neutral words (Ho et al., 2018), which is conceptually replicated in adult samples showing better recall for negative events in adults with depression (Dillon & Pizzagalli, 2018). Finally, both 5 year old children and adults detect threatening facial stimuli (e.g., angry) quicker than non-threatening facial stimuli (e.g., happy, sad) (LoBue, 2009). Taken together, we predicted that negative events would be remembered more overall than positive ones.

Research has shown that memory operates differently for stereotype-consistent and inconsistent behaviors. For example, after a delay, stereotype-inconsistent behaviors were misremembered as being performed by a stereotype-consistent actor, rather than the correct stereotype-inconsistent actor (Kledier et al., 2007). Additionally, Black individuals with stereotypically Black or Afrocentric features have been shown to be associated with violence by predominantly White adult perceivers (Kledier et al., 2012; Blair et al., 2004; Eberhardt et al., 2004). As a result of this biased association, when Afrocentric Black faces (stereotype-consistent) and non-Afrocentric Black faces (stereotype-inconsistent) were shown to participants and paired with labels such as artist and drug dealer, participants were more likely to remember stereotype-consistent faces in criminal roles and misremember stereotype-consistent faces as criminals (Kledier et al., 2012). Similarly, participants often remember events differently based on their own group membership. Developmentally, 7-10 year old children have heightened memory for positive behaviors of high status and ingroup members compared to low status and outgroup members (Corenblum, 2003; Rothbart et al., 1979).

Given the information provided by the literature, and the assumption that the majority of our participants are White, we predicted that participants would remember negative, physical events best when they are performed by Black characters. This is supported by the literature confirming that it is easier to recall events that are stereotype-consistent, and predominantly White participants associate Black people with threat, physicality, and criminality (Hester & Gray, 2018; Waytz et al., 2015; Kledier et al., 2012).

Investigating these additional components provided us with crucial information about the race-based size bias. While Study 1 provided us with important information regarding the developmental timeline for the race-based size bias, Study 2 builds on those findings by expanding our information on the importance of context for this bias in children, adolescents, and adults. It also provides insight into the generalizability of the findings from Study 1 by using real faces rather than only cartoon faces. Crucially, Study 2 dives deeper into the potential factors that could be influencing a race-based size bias developmentally. While it’s possible that children do not possess a race-based size bias, it's also possible that they do, but the measures of Study 1 were too limiting, and thus, not viable to detect a race-based size bias in children.

In sum, in Study 2 we continued the use of cartoon faces for the vignette physical size and strength ratings. We added in the use of positive events rather than just negative events under the categories of physical and social vignettes. Additionally, we kept track of participants' memory for events in Study 2. We also added several new measures to Study 2. First, we included a forced-choice design that used real faces of Black and White children. Second, we asked participants open-ended questions about how they make decisions about people's physical size and strength. These changes are detailed below.

Method

Participants

Children

Eighty six (39 boys, 47 girls) 6- to 11-year-old children were recruited from a local museum. Child participants were predominantly White (84/86 families voluntarily reported the child participant's race: 60.9% White, 2.4% Black, 15.4% Asian, 4.7% Latino, 13% Mixed-Race, and 2.4% Other). Participants were recruited in two age cohorts with at least 34 participants per age cohort (6-8 years; n= 44, 8-11 years; n= 42) to ensure sufficient representation across the full age range. Four additional participants were dropped from the final sample because they did not complete at least half of the trials.

Adolescents

Eighty one (45 male, 36 female) 12- to 17-year-old predominantly White participants were recruited from a local museum. Participants were predominantly White (80/81 families voluntarily reported the adolescent participant's race: 56.5% White, 0% Black, 15.8% Asian, 5.3% Latino, 17.1% Mixed-Race, 3.9% Other). Participants were recruited in two separate age cohorts (12-14 years; n=41, 15-17 years; n= 40) to ensure sufficient representation across the full age range. Four additional participants were dropped from the final sample because they did not complete at least half of the trials.

Adults

Eighty one (17 male, 61 female, 1 non-binary, 2 decline to answer) undergraduate students participated in and completed this study. Participants were predominantly White (81/81 participants reported: 63.8% White, 2.5% Black, 22.5% Asian, 3.8% Latino, 7.5% Mixed-Race).

Power Analysis

An a priori power analysis was conducted using the means and standard deviations collected from Study 1. Given our results, the analysis suggested that 79 participants were sufficient to detect, and thus replicate, a significant effect of Character Race in the adolescents with 0.8 power, in a matched pairs design. Although only 64 participants were required to detect a significant effect of Character Race for the adult participants, we aimed to collect 80 participants per age group: children, adolescents, and adults. Because we did not expect a significant effect of Character Race in the child age group, we did not power to detect an effect of Character Race. However, we were powered to detect a significant difference in Event Type, which required only 26 child participants. To ensure we had sufficient representation across all age ranges, we recruited participants into smaller cohorts within the child (6-8 years, n=44; 9-11 years, n=42) and adolescent (12-14 years, n=41; 15-17 years; n=40) samples; we included all 80 adults who completed the study for course credit.

Procedure

All child and adolescent participants were recruited at the Museum of Science after researchers identified them to be within the age range of 6-17 years old. Children under 18 years of age and their parents were told that they would be participating in a study about "race and physical size” or “the physical size and behaviors of Black and White children” if they wished to participate in the study. Parent permission and pre-study procedures were identical to those in Study 1.

Adults took the study asynchronously and had total control of the computer and clicked through the experiment themselves, but otherwise followed the same protocol as the children and adolescents, as they did in Study 1.

There were three main question blocks within this study: Physical size and strength ratings (including a memory check), forced-choice questions, and open-response questions. They are investigated separately below.

Physical size and strength ratings

All participants saw a total of sixteen cartoon faces one at a time in a randomized order; eight faces were Black and the other eight faces were White. The faces that were used in the experiment are presented below in Figure 9. Faces were selected using the same data as in Study 1, but sixteen, rather than eight, Black (n=8) and White (n=8) cartoon faces were selected based on which faces were categorized as ‘Black’ (M= 4.21, 84.3%) and ‘White’ (M= 3.95, 79%) most often, and with the highest confidence.

Every face was paired with each Event Type, between subjects. Each of the cartoon faces were presented with a verbal pre-recorded vignette under the categories of: Physical contact negative (e.g., shoved someone off the swings), physical contact positive (e.g., helped someone up off the ground), social exclusion negative (e.g., did not let someone play on their team), and social exclusion positive (e.g., asked someone to play). Similar to Study 1, the events were based on previous developmental research on children’s interpretations of cross-race interactions (Sagar and Schofield, 1980; McGlothlin, Killen & Edmonds, 2005; McGlothlin & Killen, 2006).

However, based on our results from Study 1, we chose to include only physical and social events as these two event types yielded the largest differences in Study 1. When creating these events, we controlled across physical and social events for the actual environment (e.g., cafeteria, PE class). That is, every environment appears in both the physical and the social events for additional control. Each of our positive and negative events each had an inverse of the other across conditions. For example, if in one condition Timmy tripped his classmate, in another condition, Timmy helped his classmate who tripped to get up. A table including descriptions of each of the vignettes and their respective categories is listed in Table 4.

Figure 9

Cartoon faces used in Study 2

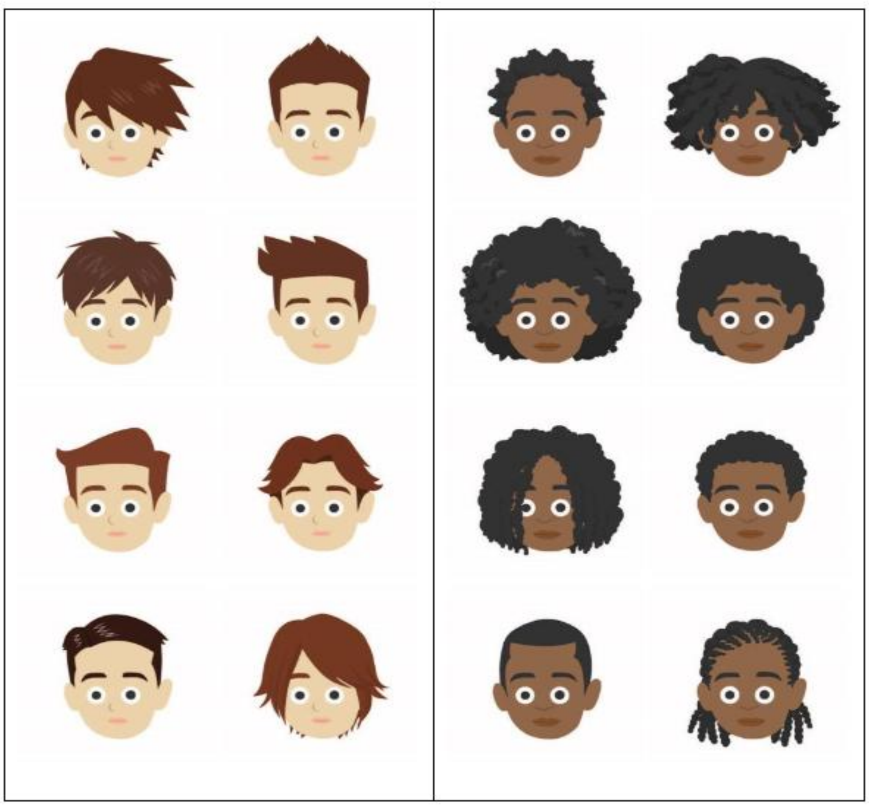


Table 4

Descriptions of the vignettes broken down by category in Study 2 and 3

Event Type	Descriptions
Physical - Positive	● When Timmy was at soccer practice, his classmate tripped over the ball and Timmy helped him back up. ● A classmate tripped during a class party and Sam helped his classmate back up. ● Mark helped his classmate stretch during a PE class. ● Jack helped a teammate off the ground during a game of tag.
Physical - Negative	● During lunch in the cafeteria, Alfred knocked his classmates' tray out of their arms. ● When James was at basketball practice, his teammate missed a basket, and he slapped his arm on purpose. ● During recess, David shoved someone off the swings. ● Ethan shoved his teammate at baseball practice instead of doing their secret handshake.
Social - Positive	● Alex's classmate was singing in the cafeteria. Alex clapped for his classmate's performance. ● When Aaron noticed his classmate standing by himself at basketball practice, he asked him to join the game he was playing. ● During recess, a new classmate approached Landon and asked to play soccer with him. Landon said yes and taught his new classmate the rules. ● Luke formed a baseball team, including all of his classmates, since they like the same things.
Social - Negative	● Kevin ignored his teammates' ideas and suggestions during their soccer practice. ● During a class party, a classmate who was sitting alone walked past Eric. Eric noticed they had nobody to hang out with, laughed at them and did not invite them to sit at his table with his friends. ● During PE class, Johnny did not let one of his classmates play on his team. ● Dominic told his classmate, who was sitting by himself, he could not join his game of tag.

Participants answered the same series of questions about the character's physical characteristics as they did in Study 1. The main difference between Study 1 and Study 2 in this block is that participants were provided with a visual scale for strength (Figure 10) to ensure that participants were rating physical strength. One additional difference is that Study 2 also counterbalanced whether the scales went from smaller to larger (e.g., very very weak to very very strong) and scales that went larger to smaller (e.g., very very strong to very very weak). All responses that went from larger to smaller were reverse-scored in Qualtrics prior to being exported.

Figure 10

Physical strength scale used in Study 2 and Study 3

Participants were presented with a memory check asking if they remembered what the character did. In Study 1, if participants did not remember what a character did, they were reminded, but their memory for the events was not recorded or measured. In Study 2, participants were presented with a memory check, and research assistants transcribed the children and adolescents' responses. They also recorded in the survey whether or not they reminded participants what the character did. Adults typed in their own responses. Following the memory check, participants were asked about the height and weight or strength and behavior of the character, as they had in Study 1.

Study 2 employed the same counterbalancing techniques as Study 1. Half of the participants were asked about trait characteristics first (e.g., mean, strength), followed by a memory check, and then physical characteristics, and the other half of the participants were about the physical characteristics first (e.g., height, weight), followed by a memory check, and trait characteristics. Half of the participants in Study 2 also saw the scales reversed, where rather than going from small and nice to large and mean, the scales were shown going from large and mean to small and nice.

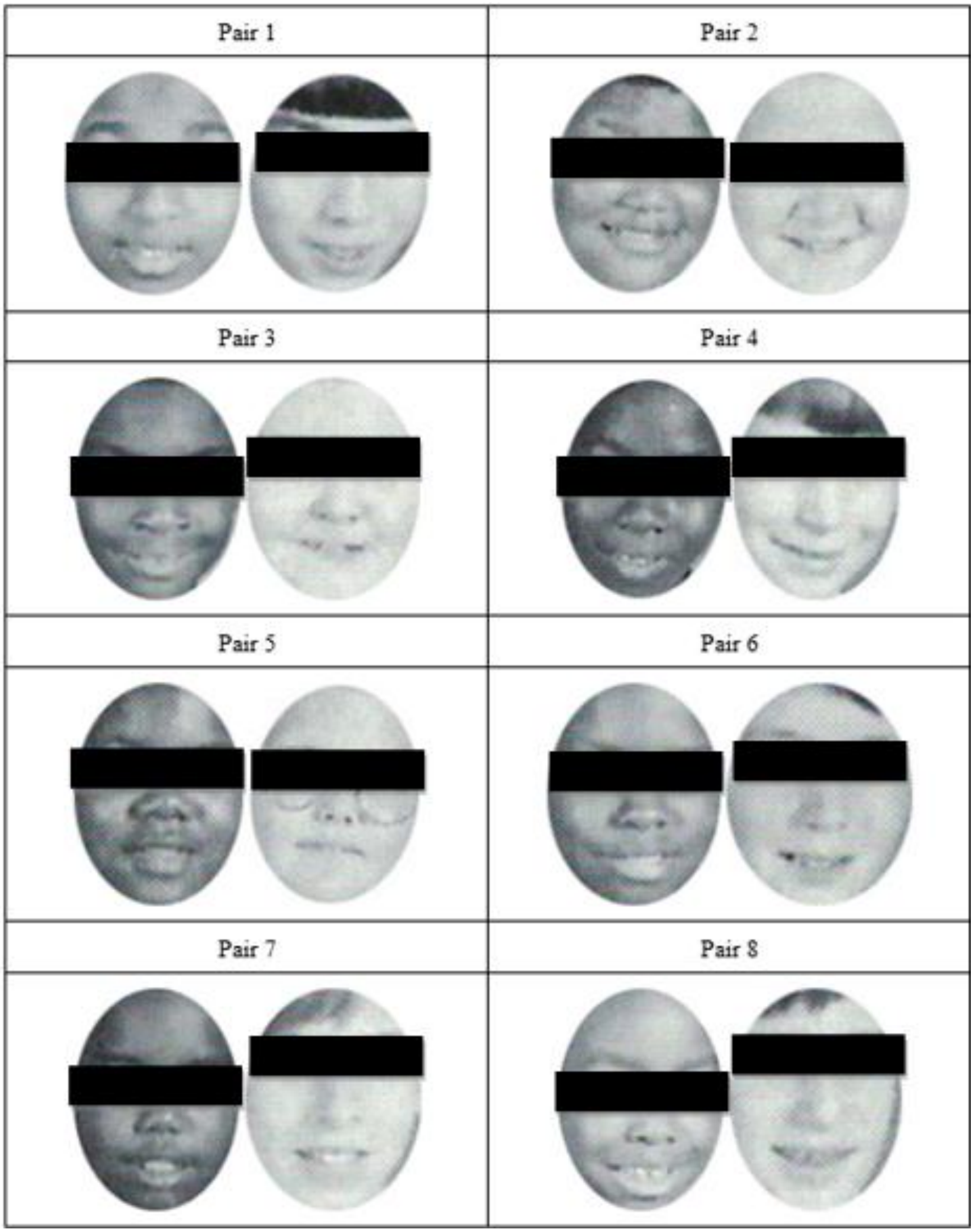
Forced-choice

Following the physical size and strength block, participants completed a forced-choice block. In this test block, participants were shown black and white images of real Black and White children in late childhood/early adolescence sourced from a public yearbook. All participants saw a total of eight real Black-White child face pairings. The face pairings that were used in the experiment are presented below in Figure 11. Faces were matched on the experimenter’s perceived size of the children as closely as possible without knowing the actual size of the children.

Participants were asked in a fixed order, ‘Who is stronger?,’ ‘Who is taller?,’ and ‘Who is heavier?’. Five children, two adolescents, and two adults did not complete the forced-choice trials.

Figure 11

Face pairings used in the forced-choice design of Study 2 and Study 3. The eyes are obscured to protect the individuals' identities whose pictures were used.



Open-response questions

Following the forced-choice block, participants were asked a series of open-response questions. The experimenter asked the participants, in a fixed order: “How can you tell someone is strong?”, “How can you tell someone is tall?”, and “How can you tell someone is heavy?”. The experimenter typed the participants’ responses in a textbook within the survey to be coded at a later time. Two children and four adults did not complete the open-response questions.

Results

Physical size and strength ratings

Preliminary analyses

We conducted a preliminary analysis on participant gender and presentation order across ages and strength, height and weight ratings. There were no significant effects of participant gender or presentation order, so we did not include them in further analyses.

As in Study 1, our goal was to present participants with mildly negative or positive events and ensure that participants understood each event to be negative or positive, respectively. Preliminary analysis of the meanness ratings revealed that participants viewed some negative events as positive interactions, choosing ‘very, very nice,’ ‘very nice’ or ‘nice’ (3 or below) on the meanness scale (child: N= 27 trials; adolescent: N= 14 trials; adult: N= 21 trials) or viewed some positive events as negative interactions (child: N= 26 trials; adolescent: N= 1 trial; adult:

N= 13 trials). To ensure we were assessing participant's representations of the size of actors in negative events, we removed these individual trials from all subsequent analyses.

Overall analysis plan

We conducted linear mixed effects regression models for all the analyses reported below, following the same analysis plan as Study 1 with one exception. We included Valence as a fixed factor and random slope in the model.

Strength Ratings

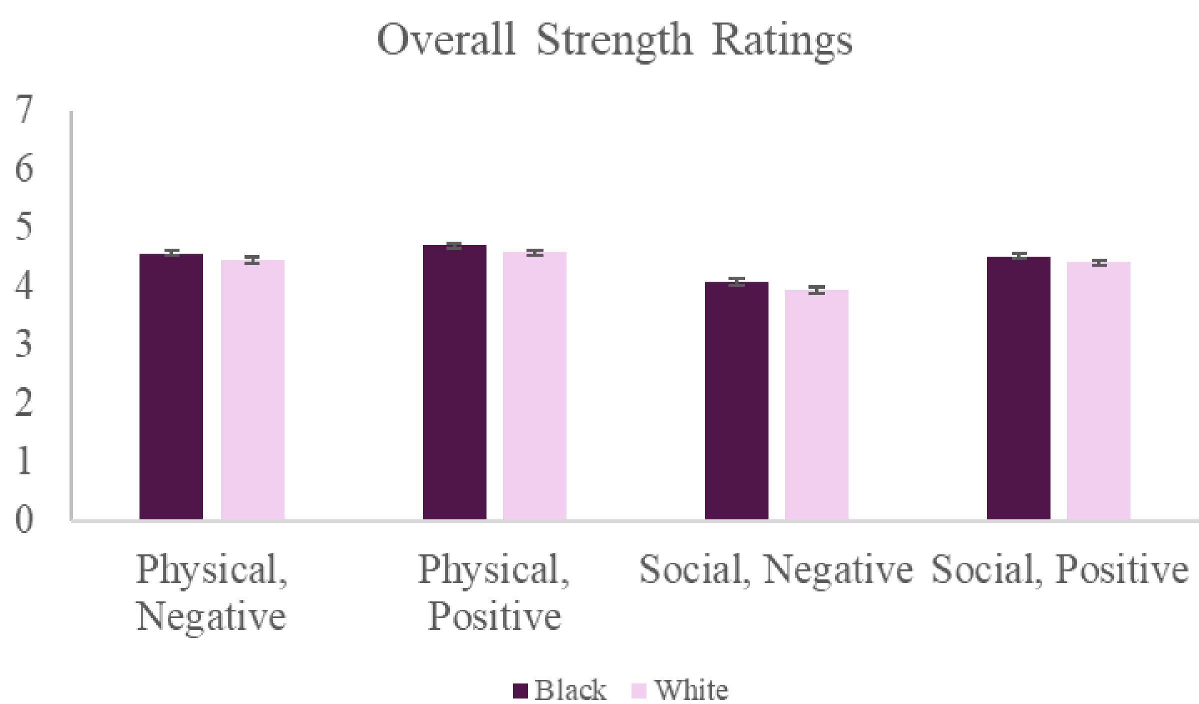
Strength ratings across development: children, adolescents, and adults.

Character Race, $F(1, 703) = 16.95$, $p < .001$, Event Type, $F(1, 228) = 93.97$, $p < .001$, Valence, $F(1, 236) = 27.02$, $p < .001$, and Participant Age Group, $F(2, 238) = 3.34$, $p = .037$ all significantly predicted participants' strength ratings. Participants rated Black characters as stronger ($M = 4.47$, $SD = 1.04$) than White characters ($M = 4.36$, $SD = 1.1$). Characters who engaged in physical events ($M = 4.58$, $SD = 1.08$) were rated as significantly stronger than those in social events ($M = 4.24$, $SD = 1.03$) and characters who engaged in positive events ($M = 4.56$, $SD = .093$) were perceived as significantly stronger than those who engaged in negative events ($M = 4.27$, $SD = 1.17$). Post-hoc tests using adults as the reference category revealed that adolescents ($M = 4.52$, $SD = .88$) rated characters as stronger overall than adult strength ratings, ($M = 4.33$, $SD = .98$; $t(236) = 2.56$, p= .01), but children's strength ratings did not differ from the adults ($M = 4.39$, $SD = 1.29$; $t(240) = 1.03$, p= .303).

These effects were all qualified by a significant interaction between Event Type and Valence, $F(1, 3012) = 32.98$, $p < .001$. Post-hoc tests revealed that Valence had an effect on both social and physical events, but the magnitude of the effect was stronger for social events than physical events. Targets who engaged in social - positive events (M= 4.47, SD= .93) were rated as stronger than those in social - negative events (M= 4.02, SD= 1.08; t(375)= 7.23, p< .001); targets who engaged in physical - positive events (M= 4.65, SD= .93) were also rated as stronger than those in physical - negative events, but to a lesser degree (M= 4.52, SD= 1.2; t(371)= 2.03, p= .043). There were no other significant interactions, so they were removed from the model. The findings are presented below in Figure 12.

Figure 12

Strength ratings across children, adolescents and adults for Study 2



Although there were no significant interactions between Participant Age Group and any of the factors, we still explored the effects by each age group individually for strength. Because it was hypothesized that strength, rather than height or weight, may be the more important factor in the race-based size bias, it was imperative to explore it further. In order to gain a stronger understanding of the development of the race-based size bias, we planned to explore the fixed factors within each of the age bins we recruited from.

The model for child participants yielded significant effects of Event Type, $F(1, 86.5)= 19.13$, $p <.001$, and Valence, $F(1, 81.4)= 8.09$, $p= .006$. Child participants rated characters who engaged in physical ($M= 4.56$, $SD= 1.3$) events as stronger than those who engaged in social ($M= 4.22$, $SD= 1.25$) events. Also, children rated characters who engaged in positive events ($M= 4.57$, $SD= 1.16$) as stronger than those who engaged in negative events ($M= 4.23$, $SD= 1.39$). However, there were no significant effects of either Character Race, $F(1, 127.5)= .61$, $p= .43$ or Age Cohort, $F(1, 82.4)= .01$, $p= .903$.

There were two significant two-way interactions. First, there was a significant interaction between Event Type and Valence, $F(1, 1010.6)= 7.03$, $p= .008$. Post-hoc tests revealed that valence had a significant effect on social, but not physical events. Specifically, characters who engaged in social - positive events ($M= 4.48$, $SD= 1.16$) were significantly stronger than those who engaged in social - negative events ($M= 3.97$, $SD= 1.29$; $t(123.4)=3.75$, $p< .001$), but there were no differences between strength ratings for physical - positive events ($M= 4.65$, $SD= 1.16$), and physical - negative events ($M= 4.47$, $SD= 1.43$; $t(119)= 1.49$, $p= .14$). There was also a significant interaction between Character Race and Age Cohort, $F(1, 127.5)= 9.8$, $p= .002$. Post-hoc tests revealed that younger children rated Black characters ($M= 4.51$, $SD= 1.53$) as significantly stronger than White characters ($M= 4.28$, $SD= 1.53$; $t(85.5)=2.72$, $p= .008$), but older children did not, t(78.8)= 1.72, p= .09.

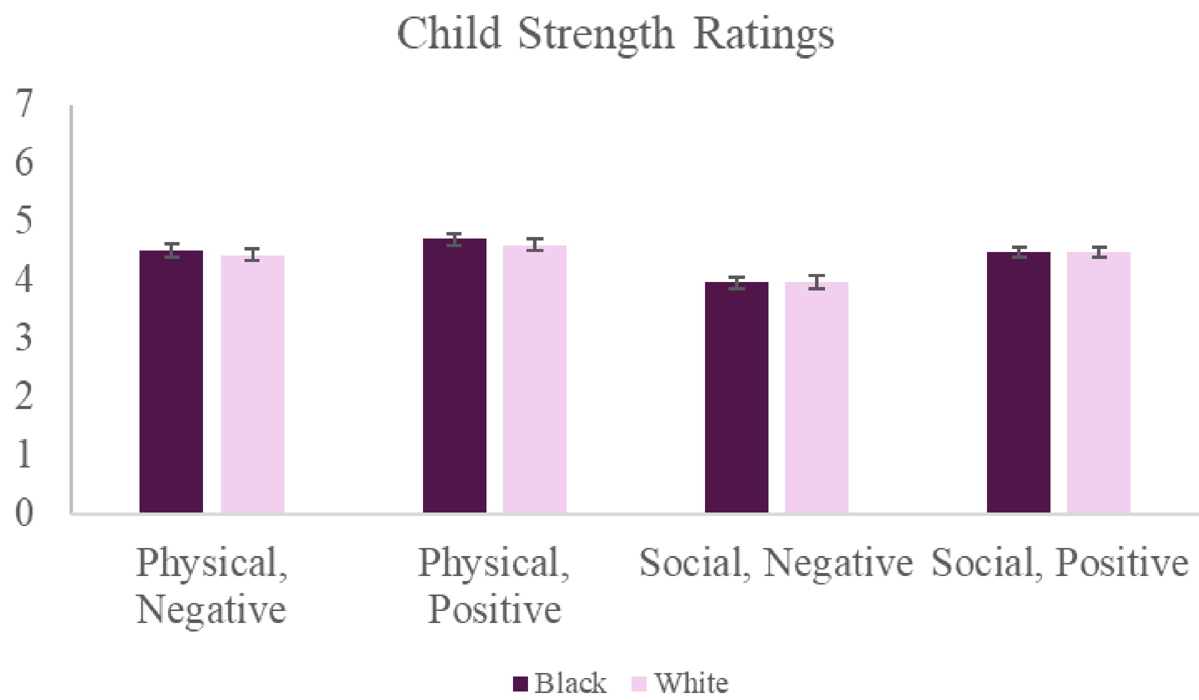
These effects were qualified by a significant three-way interaction between Event Type, Valence, and Age Cohort, $F(1, 1010.6) = 3.9$, $p = .049$. To explore this interaction further, we investigated the effect of Event Type and Valence within each Age Cohort separately.

The younger child model yielded significant main effects of Character Race, $F(1, 45) = 4.75$, $p = .035$, where Black characters (M= 4.51, SD= 1.53) were perceived as stronger than White characters (M= 4.28, SD= 1.53); Event Type, $F(1, 53.3) = 7.74$, $p = .007$, where characters who engaged in physical events (M= 4.57, SD= 1.56) were perceived as stronger than those who engaged in social events (M= 4.21, SD= 1.48) and Valence, $F(1, 39) = 6.05$, $p = .018$, where characters who engaged in positive events (M= 4.65, SD= 1.42) were perceived as stronger than those who engaged in negative (M= 4.15, SD= 1.6) events. There were no significant interaction effects between the factors, so they were removed from the model.

The older child model yielded significant main effects of Character Race, $F(1, 73.6) = 5.24$, $p = .025$, where White characters (M= 4.46, SD= 1.05) were perceived as stronger than Black characters (M= 4.32, SD= 1.01), and Event Type, $F(1, 40.6) = 12.83$, $p < .001$, where characters who engaged in physical events (M= 4.55, SD= 1.04) were perceived as stronger than those who engaged in social events (M= 4.23, SD= 1.01), but there was no main effect of Valence, $F(1, 43) = 1.95$, $p = .17$. There was also a significant interaction between Event Type and Valence, $F(1, 539.4) = 21.11$, $p < .001$, where Valence had an effect on social, $t(57.8) = -2.97$, p= .004, but not physical, $t(56.9) = .48$, p= .63, events. While both younger and older children use Event Type to inform their strength judgements, only younger children use Valence to inform their decisions, separate from Event Type. The separate models also each yielded effects of Character Race, but each in opposite directions, thus cancelling out any potential effects in the overall model.

Figure 13

Child (6-11 years) strength ratings for Study 2



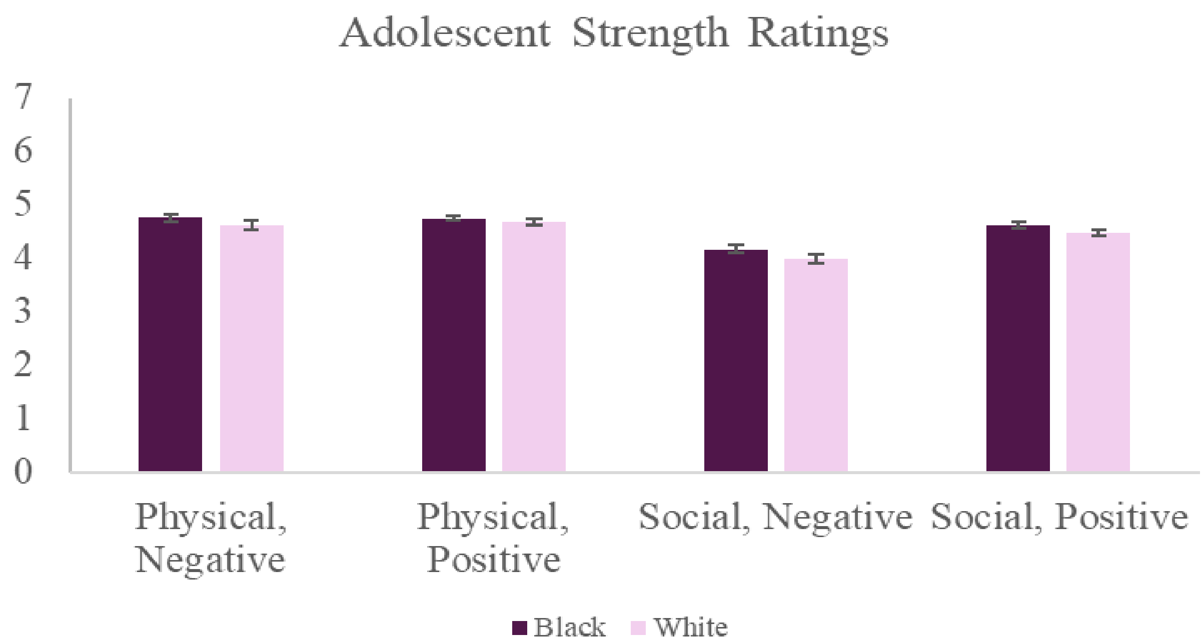
The model for the adolescent participants yielded significant effects of Character Race, $F(1, 616) = 10.12$, $p = .002$, Event Type, $F(1, 76.7) = 72.26$, $p < .001$, Valence, $F(1, 77.5) = 12.14$, $p < .001$, and Participant Age Cohort, $F(1, 77.3) = 7.18$, $p = .009$ on participants' strength ratings. Black characters were perceived as significantly stronger ($M = 4.59$, $SD = .85$) than White characters ($M = 4.46$, $SD = .91$). Characters who engaged in physical events ($M = 4.71$, $SD = .85$) were perceived as stronger than character who engaged in social events ($M = 4.33$, $SD = .87$). Additionally, characters who engaged in positive events ($M = 4.65$, $SD = .73$) were rated as significantly stronger than those who engaged in negative events ($M = 4.39$, $SD = 1.0$). Finally, younger adolescents rated characters as stronger (*M*= 4.63, *SD*= .82) than older adolescents (*M*= 4.41, *SD*= .93).

These effects were qualified by a significant interaction between Event Type and Valence, $F(1, 1005.8) = 30.97$, $p < .001$. Mirroring the child participants, post-hoc tests revealed that Valence had an effect on social, but not physical events. Specifically, characters who engaged in social - positive events (*M*= 4.56, *SD*= .76) were significantly stronger than those who engaged in social - negative events (*M*=4.08 , *SD*= .91; $t(127.1) = 5.67$, $p < .001$), but there were no differences between physical - positive (*M*= 4.72, *SD*= .69) and physical - negative (*M*=4.69 , *SD*= .98; $t(127.4) = 0.39$, $p = .69$) events.

There were no significant three-way interactions between the factors, so they were removed from the model. These findings are displayed in Figure 14.

Figure 14

Adolescent (12-17 years) strength ratings for Study 2



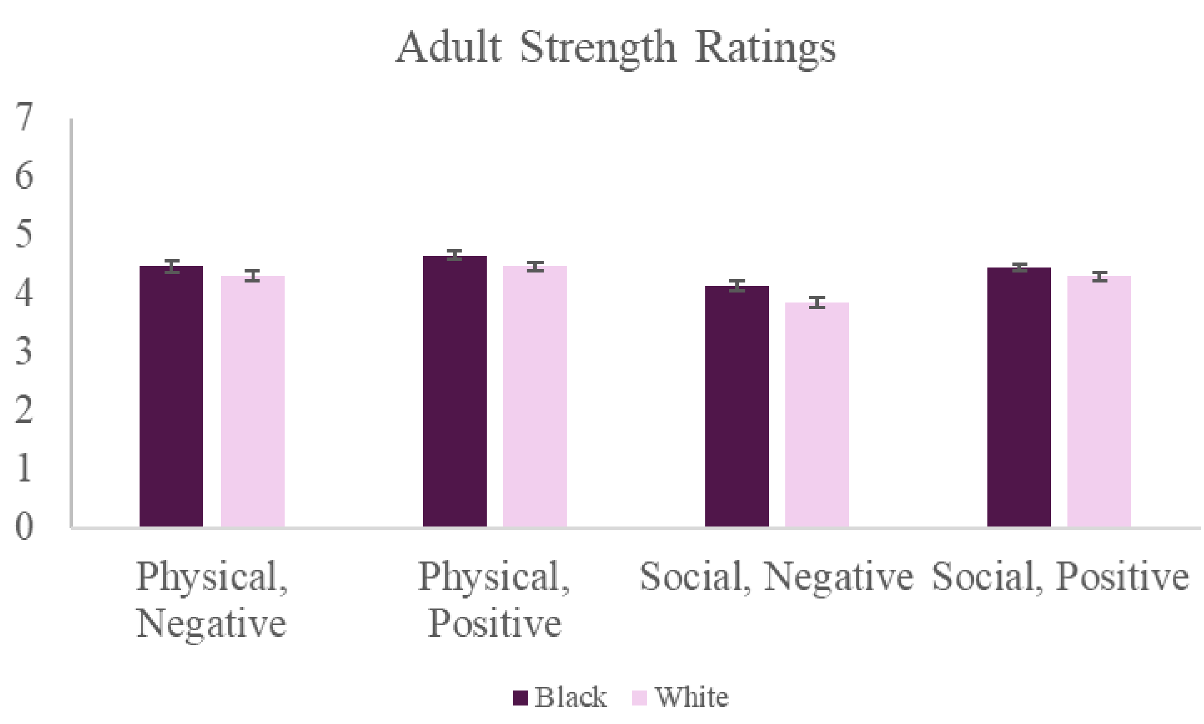
The adult participants model yielded significant effects of Character Race, $F(1, 741.6)= 15.54$, $p< .001$, Event Type, $F(1, 77)= 24.445$ $p< .001$, and Valence, $F(1, 77.8)= 11.16$, $p= .001$ on participant strength ratings. Adult participants perceived Black characters as stronger (*M*= 4.43, *SD*= .92) than White characters (*M*= 4.23, *SD*= 1.02); and characters who engaged in physical events (*M*= 4.48, *SD*= 1.0) were significantly stronger than those who engaged in social events (*M*= 4.19, *SD*= .94). Additionally, adults rated characters who engaged in positive events as stronger, overall (*M*= 4.47, *SD*= .84) than those who engaged in negative events (*M*= 4.19, *SD*= 1.08).

These effects were qualified by a significant interaction between Event Type and Valence, $F(1, 999.2)= 4.72$, $p= .03$. Replicating the pattern seen in child and adolescent participants, Valence had an effect on social, but not physical events. Specifically, characters who engaged in social - positive events (*M*= 4.38, *SD*= .81) were significantly stronger than those who engaged in social - negative events (*M*= 4.00, *SD*= 1.02; *t*(142.5)= 3.99, *p*< .001), but there were no differences between physical - positive (*M*= 4.56, *SD*= .86) and physical - negative (*M*= 4.39, *SD*= .86; *t*(144.1)= 7.5, *p*= .083) events.

There were no significant three-way interactions between the factors, so they were removed from the model. These findings are displayed in Figure 15.

Figure 15

Adult strength ratings for Study 2



Together, we understand that across children, adolescents and adults, Black characters are perceived as stronger than White characters. Also, characters who engage in positive events and physical events are perceived as stronger than those who engage in negative and social events, respectively. Individual Age Groups reveal that children do not factor race in when making strength judgments, but adolescents and adults do. Children, adolescents and adults showed that Valence had a significant effect on social, but not physical, events.

Adolescent and adult participants show evidence of a race-based size bias for strength where Black characters are perceived as significantly stronger than White characters, and they use behavior to inform their decisions. Children, however, do not use race as a predictor when rating Black and White character's strength. Instead, they only use behavior information to make their decisions. This conceptually replicates our findings from Study 1.

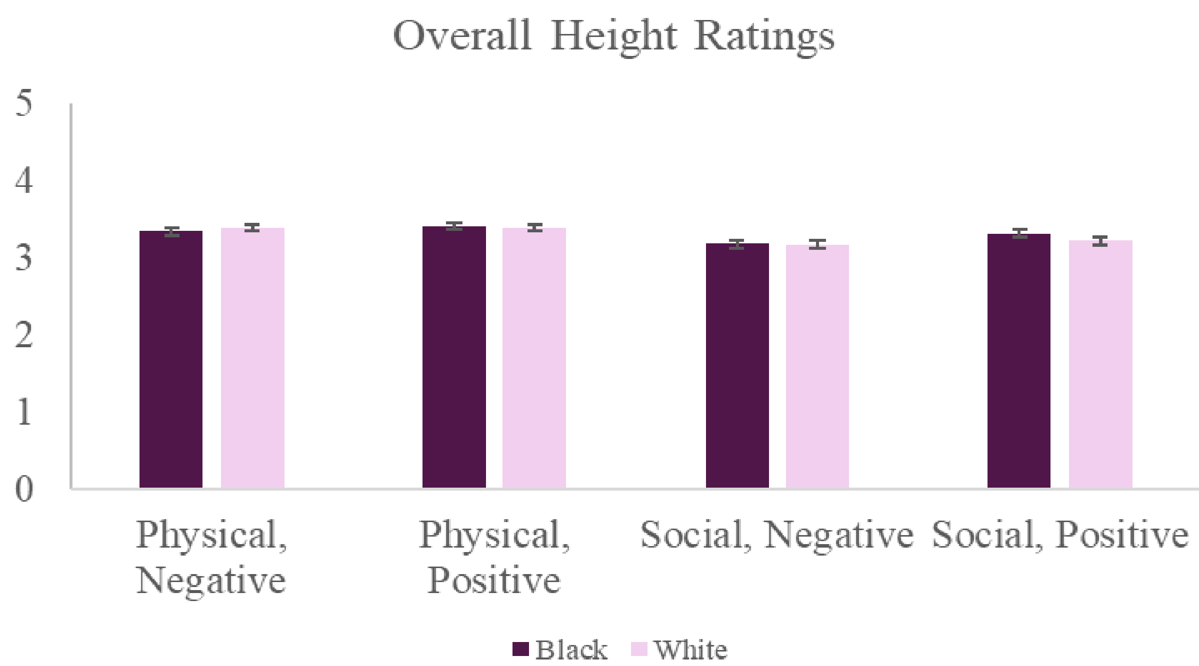
Height Ratings

Height ratings across development: children, adolescents, and adults.

Event Type, $F(1, 2483) = 27.8$, $p < .001$ and Participant Age Group, $F(2, 241) = 3.6$, $p = .028$ significantly predicted participants' height ratings. All participants rated characters who engaged in physical events as taller ($M = 3.38$, $SD = 1.02$) than those who engaged in social events ($M = 3.22$, $SD = 1.02$). Adolescent participants rated characters as significantly taller ($M = 3.41$, $SD = .91$) overall than children ($M = 3.25$, $SD = 1.13$; $t(241) = 2.42$, $p = .016$) and adults ($M = 3.24$, $SD = 1.01$; $t(238) = 2.77$, $p = .006$) did. There were no significant effects of Character Race, $F(1, 2780) = .19$, $p = .67$, or Valence, $F(1, 235) = 1.99$, $p = .16$. No interactions were significant, so they were removed from the model.

Figure 16

Height ratings across children, adolescents and adults for Study 2



Conceptually replicating Study 1, participants perceived characters as taller when they engaged in physical events, rather than social events, but they did not perceive characters differently based on their race. Adolescent participants also rated characters as taller overall than children and adults did. Because we see no evidence of a race-based size bias for height across age groups, the subsequent analyses broken down by age group are in the supplemental information.

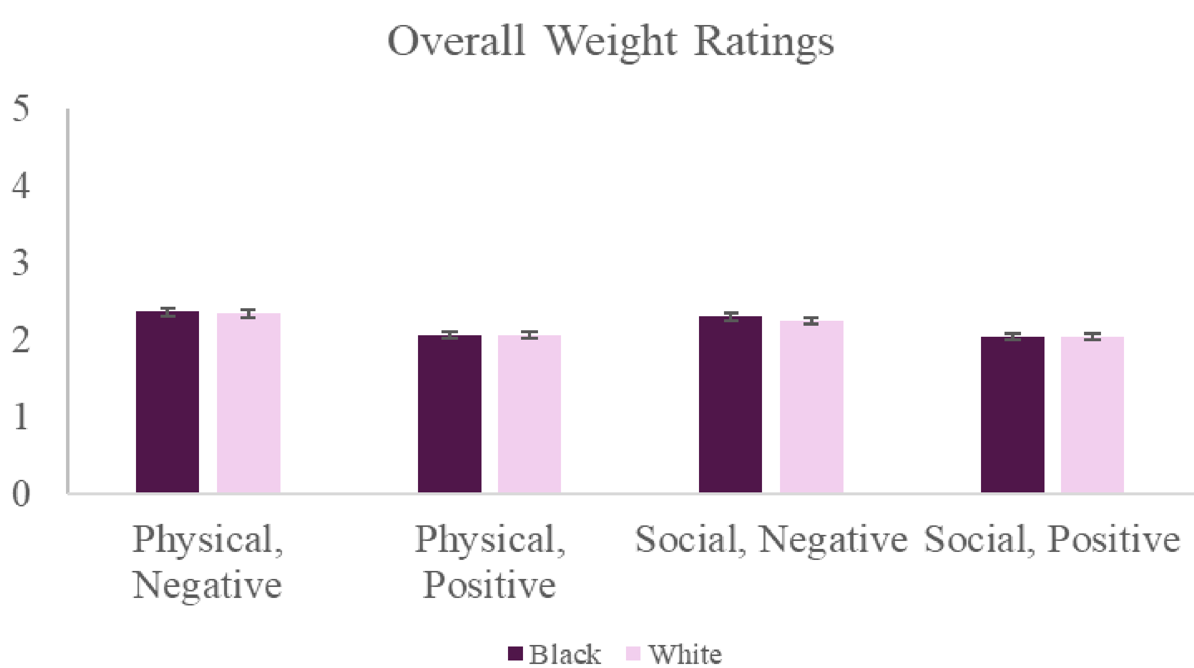
Weight Ratings

Weight ratings across development: children, adolescents, and adults.

Valence, *F*(1, 3268)= 102.55, *p*< .001 and Participant Age Group, *F*(2, 237)= 3.33, *p*= .037 significantly predicted participants' height ratings. All participants rated characters who engaged in negative events as heavier (*M*= 2.32, *SD*= 1.02) than those who engaged in positive events (*M*= 2.05, *SD*= .91). Adolescent participants rated characters as significantly heavier (*M*= 2.29, *SD*= .89) overall than adults (*M*= 2.07, *SD*= .97, *t*(235)= 2.53, *p*= .012), but not heavier than children (*M*= 2.2, *SD*= .89, *t*(236)= 0.78, *p*= .439). There were no significant effects of Character Race, *F*(1, 245)= .35, *p*= .55, or Event Type, *F*(1, 1650)= 2.65, *p*= .10. No interactions were significant, so they were removed from the model.

Figure 17

Weight ratings across children, adolescents and adults for Study 2



Together, these results suggest that participants in all age groups perceive characters who engaged in negative events as heavier than those who engaged in positive events. Additionally, adolescent participants rated characters as heavier overall than adult participants did. In Study 1, we did not explore Valence as a factor, but Study 1 did have significance of Event Type showing the importance of contextual information when considering weight suggesting a conceptual replication. Because we see no evidence of a race-based size bias for weight across age groups, the subsequent analyses broken down by age group are in the supplemental information.

Correlations between Strength, Height, Weight and Age

As in Study 1, we conducted a correlation analysis on the data including the factors: strength, height, and weight.

The matrix is presented in Table 5 below and displays Pearson correlation coefficients for all pairs of the variables. Notably, we see positive correlations between strength and height ($r = .507$, $p < .001$), and strength and weight ($r = .229$, $p < .001$), which we did not see in Study 1. Additionally, in contrast to Study 1, there was no significant correlation between height and weight ($r = .087$, $p = .177$). These findings suggest that strength is positively related to both height and weight, but height and weight did not have corresponding values.

Table 5

Correlation table between strength, height, and weight for Study 2

	Strength	Height	Weight
Strength	- -		
Height	Pearson's r= .507 *p*< .001	- -	
Weight	Pearson's r= .229 *p*< .001	Pearson's r= .087 *p*= .177	- -

Degree of Niceness and Meanness ratings

In order to investigate whether Character Race, Event Type, Valence and Age Group are influenced by the degree of niceness and meanness, responses to the question were changed from their 1-7 score to be based on degree. 'Nice' was recoded as 1, 'very nice' was recoded as 2, and 'very, very nice' was recoded as 3. Similarly, 'mean' was recoded as 1, 'very mean' was recoded as 2, and 'very, very mean' was recoded as 3. Any responses that were neutral are not included in the current analyses. As with the prior models, any responses that were rated with incorrect valence (e.g., the event was mean but was perceived as nice) were not included. The first model was the same as the model described above, but with the recoded degree of niceness as the dependent variable rather than strength, height, or weight. These results should be interpreted with caution, as they are exploratory measures.

The model yielded a significant effect of Age Group, $F(2, 237)= 9.5$, $p< .001$, where post-hoc tests revealed that children ($M= 1.86$, $SD= .79$) used more extreme scores on the scale than adolescents ($M= 1.6$, $SD= .66$; $t(239)= 4.31$, $p< .001$) and adults ($M= 1.7$, $SD= .7$; $t(239)=$

2.66, p= .008). Additionally, the model yielded three significant two-way interactions. To begin, there was a significant interaction between Character Race and Valence, $F(1, 2941) = 13.66$, $p < .001$ that showed participants rated Black characters as nicer than White characters overall. Specifically, Black characters who engaged in positive ($M = 1.75$, $SD = .74$) events as nicer than White characters who engaged in positive ($M = 1.67$, $SD = .71$; $t(2839) = 2.87$, $p = .004$) events, but rated Black characters who engaged in negative ($M = 1.69$, $SD = .74$) events as less mean than White characters who engaged in negative ($M = 1.76$, $SD = .72$; $t(2860) = 2.36$, $p = .018$) events. There was an additional significant interaction between Event Type and Valence, $F(1, 2935) = 46.27$, $p < .001$ showing that physical events are perceived as meaner than social events overall. Specifically, physical - negative ($M = 1.81$, $SD = .76$) events were meaner than social - negative ($M = 1.65$, $SD = .71$; $t(2904) = 5.36$, $p < .001$) events, but social - positive ($M = 1.77$, $SD = .74$) events were nicer than physical - positive ($M = 1.65$, $SD = .71$; $t(2894) = 4.25$, $p < .001$) events. The final two-way interaction was between Valence and Age Group, $F(2, 2862) = 4.72$, $p = .009$, which was explored with the breakdown by Age Group below. These effects were all qualified by a three-way interaction between Character Race, Valence, and Age Group, $F(2, 2941) = 3.94$, $p = .02$. To further explore this interaction, we broke down the model by Age Group below.

The child model revealed significant main effects of Event Type, $F(1, 744.5) = 7.78$, $p = .005$, and Age Cohort, $F(1, 82.4) = 10.19$, $p = .002$. Physical events ($M = 1.91$, $SD = .79$) resulted in more extreme values than social events ($M = 1.8$, $SD = .79$) did. Also, younger children ($M = 2.02$, $SD = .82$) use more extreme values on the Likert scale than older children ($M = 1.71$, $SD = .73$) do. These effects were all qualified by a significant interaction between Event Type and Valence, $F(1, 975.8) = 12.17$, $p < .001$, where Valence influenced degree of niceness and meanness for social events, but not physical events. Specifically, social - positive ($M = 1.89$, $SD =$

.8) events had more extreme scores than social - negative (*M*= 1.71, *SD*= .77; *t*(188.3)= 2.72, p= .007) did, but physical - positive (*M*= 1.87, *SD*= .8) and physical - negative (*M*= 1.95, *SD*= .78; *t*(183.2)= 1.39, p= .165) did not significantly differ.

The adolescent model revealed a significant interaction between Event Type and Valence, *F*(1, 974.2)= 22.69, p< .001, where physical - negative (*M*= 1.67, *SD*= .69) events were meaner than social - negative (*M*= 1.57, *SD*= .663; *t*(279.5)= 2.25, p= .025) events, but social - positive (*M*= 1.69, *SD*= .67=) events were nicer than physical - positive (*M*= 1.47, *SD*= .61; *t*(265.6)= 4.5, p< .001) events, mirroring the overall model. There was also a significant three-way interaction between Character Race, Valence, and Age Cohort, *F*(1, 976.6)= 4.49, p= .03. To further understand this interaction, we broke down the adolescent model by Age Cohort and re-ran the analyses.

The younger adolescents model did not yield any significant effects of Character Race, but it once again showed a significant interaction between Event Type and Valence, *F*(1, 497.4)= 5.86, p= .016. In this case, Event Type had an effect on positive events, but not negative ones. Specifically, social - positive (*M*= 1.65, *SD*= .67) events were rated with a higher Degree of niceness and meanness than physical- positive (*M*= 1.44, *SD*= .62; *t*(126.6)= 2.89, p= .004) events, but physical - negative (*M*= 1.54, *SD*= .62) and social - negative (*M*= 1.53, *SD*= .66; *t*(136.2)= .48, p= .64) events did not differ.

The older adolescent model yielded a significant interaction between Character Race and Valence, *F*(1, 478.4)= 7.99, p= .005, and Event Type and Valence, *F*(1, 478.4)= 17.69, p< .001. Valence had an effect on the Degree of niceness and meanness ratings for White, but not Black characters. Specifically, negative events performed by a White character (*M*= 1.77, *SD*= .68) were rated with more extreme values than positive events performed by a White character (*M*= 1.54, *SD*= .60; *t*(117.5)= 3.12, p= .002), but there were no differences between negative (*M*= 1.64, *SD*= .73) and positive events for Black characters (*M*= 1.7, *SD*= .68; *t*(124.9)= .77, p= .44). Valence had an effect on physical, but not social events for the Degree of niceness and meanness measure. Physical - negative events (*M*= 1.81, *SD*= .73) were rated with a higher Degree of niceness and meanness than physical - positive events, (*M*= 1.5, *SD*= .60; *t*(121.9)= 4.01, p< .001), but there were no differences between social - positive (*M*= 1.73, *SD*= .67) and social - negative events (*M*= 1.61, *SD*= .67; *t*(119)= 1.72, p= .09).

The adult model replicated the findings of the older adolescent model. The model yielded a significant interaction between Character Race and Valence, *F*(1, 980.7)= 18.62, p= .005, and Event Type and Valence, *F*(1, 977.6)= 14.36, p< .001. Valence had an effect on the Degree of niceness and meanness ratings for White, but not Black characters. Specifically, negative events performed by a White character (*M*= 1.79, *SD*= .7) were rated with more extreme values than positive events performed by a White character (*M*= 1.58, *SD*= .66; *t*(220)= 3.89, p< .001), but there were no differences between negative (*M*= 1.67, *SD*= .72) and positive events for Black characters (*M*= 1.76, *SD*= .7; *t*(224.8)= 1.76, p= .08). Valence had an effect on physical, but not social events for the Degree of niceness and meanness measure. Physical - negative events (*M*= 1.79, *SD*= .71) were rated with a higher Degree of niceness and meanness than physical - positive events, (*M*= 1.6, *SD*= .65; *t*(227.8)= 3.5, p< .001), but there were no differences between social - positive (*M*= 1.74, *SD*= .72) and social - negative events (*M*= 1.66, *SD*= .65; *t*(215.5)= 1.43, p= .16).

Together, we understand that overall, Black characters are rated as nicer than White characters. When broken down by age, older adolescents and adults differ in their Degree of niceness and meanness ratings for Black and White characters, but children and younger adolescents do not. These results build upon the finding that there are no consistent race effects in the child age group for the vignette ratings, but we do begin to see evidence of a race effect in adolescents and adults. Across ages, Event Type was an important predictor either as a main effect, or interacting with Valence. This is further evidence that the context surrounding an event matters, and participants make their decisions largely based on this information whether it is for a physical size measure or how extreme they rate events as nice or mean.

Memory check

Memory check responses were coded with either a 1, if participant verbal responses were accurate, or a 0, if participant verbal responses were not accurate. Accuracy was based on if participants remembered if the events were positive or negative and physical or social. The specific details of the event were not required to be considered accurate.

We conducted a generalized logistic mixed effects regression model for the memory check. Unless noted otherwise, all models included Character Race (Black, White), Event Type (Physical, Social) and Valence (Positive, Negative). Participant Age Group was also entered as a fixed factor, either at the broad level (Child, Adolescent, Adult) for initial overall models or at the recruitment-based cohort level (Children: 6-8 years, 9-11 years; Adolescents: 12-14 years, 15-17 years) for within age group follow-up analyses. All models included random intercepts for participants and random slopes for Event Type, Valence, and Character Race. We also tested, and retained in the final models, any significant levels of interactions between fixed factors to control for Type-I error. All other non-significant interactions were removed.

The overall model yielded significant effects of Event Type $\chi2(1)= 19.8$, $p< .001$, where participants remembered physical events (M= .95, SD= .23) better than social events (M= .90, SD= .23), and Participant Age Group $\chi2(2)= 55.19$, $p< .001$ where children remembered less events (M= .85, SD= .36) overall than adolescents (M= .97, SD= .18) and adults (M= .96, SD= .19). There were no significant differences between adolescent and adults' memory of the events. There were no significant main effects of Character Race, $\chi2(1)= 3.54$, $p= .06$ or Valence, $\chi2(1)= .18$, $p= .68$. We broke down the model further by age to gain a larger picture understanding of what may be happening developmentally. Because we saw effects of Participant Age Group, it was possible that race-based memory effects were being covered up by the Participant Age Group effects, but that we would see effects of race when it was broken down within age groups.

The child model reveals a significant effect of Event Type, $\chi2(1)= 19.54$, $p< .001$, where participants remembered physical events (M= .89, SD= .32) better than social events (M= .81, SD= .4). There were no significant effects of Character Race $\chi2(1)= 3.18$, $p= .07$ or Valence $\chi2(1)= 0.18$, $p= .674$. The adolescent model yielded no significant effects of Character Race $\chi2(1)= .12$, $p= .729$, Event Type $\chi2(1)= 2.92$, $p= .09$, or Valence $\chi2(1)= 1.02$, $p= .31$. The adult model yielded significant effects of Character Race $\chi2(1)= 4.18$, $p= .041$ where participants remembered what Black characters (M= .97, SD= .18) did more often than White characters (M= .96, SD= .2), and Event Type $\chi2(1)= 3.97$, $p= .046$, where participants remembered physical events more often than social ones. There was no significant effect of Valence $\chi2(1)= 1.08$, $p= .298$. Generally, children and adults remembered physical events better than social ones. Adults also remembered events performed by Black characters better than events performed by White characters. Adolescents' memory did not differ meaningfully with any of the factors.

Memory on physical size and strength models

To further investigate the effect of memory on strength, height, and weight, we re-ran our above physical size and strength models on strength, height, and weight, but only including trials where participants remembered events. Because we do not have enough data on children who did not remember, we only ran the subsequent analyses on those who do remember. These results should be interpreted with caution, as they are done on a subset of the larger dataset, and are largely exploratory.

The memory dependent strength model replicated all of the main effects and interactions from the original model. This model, however, yielded a significant interaction between Character Race and Age Group, $F(2, 201) = 3.99$, $p = .02$. Follow-up models revealed that all main effects and interactions across ages remained the same, with one exception - the child model yielded no significant interaction between Event Type and Valence, $F(1, 864.2) = 3.61$, $p = .058$. Both the height and weight memory dependent models fully replicated the original models. Because the majority of participants remembered the events, this measure did not significantly impact the results when accounted for in the memory dependent model.

Forced-choice ratings

Overall analysis plan

We began by calculating the total number of times participants chose a Black child rather than a White child as stronger, taller, and heavier, separately. Then, we created a percentage by dividing that number by the total number of trials they completed for each participant. Overall percentages for strength, height, and weight were then compared to chance (0.5) in 4 separate one sample t-tests: across ages, children, adolescents, adults. These results are reported in Table 6.

Strength ratings

Overall, participants chose a Black child as stronger than a White child 60.7% (*SD*= .18) of the time, which was significantly greater than chance t(234)= 8.92, p< .001, *d*= 0.58. This effect was found for children (55.4% (*SD*= .21); t(80)= 2.29, p= .024, *d*= 0.25), adolescents (63.2% (*SD*= .16); t(75)= 7.39, p< .001, *d*= 0.16), and adults (63.8% (*SD*= .17); t(77)= 7.3, p< .001, *d*= 0.83).

Height ratings

Overall, participants chose a Black child as taller than a White child 47.8% (*SD*= .19) of the time, which was *not* significantly less than chance t(234)= 1.8, p= .073, *d*= 0.12. This pattern was replicated in children (46.6% (*SD*= .02); t(80)= 1.38, p= .17, *d*= .15), adolescents (47.4% (*SD*= .17); t(75)= 1.35, p= .18, *d*= .16), and adults (49.4% (*SD*= .18); t(77)= .32, p= .75, *d*= .04).

Weight ratings

Overall, participants chose a Black child as heavier than a White child 44.8% (*SD*= .21) of the time, which was significantly *less* than chance t(234)= 3.81, p< .001, *d*= .25. This effect was not found in children (48.3% (*SD*= .22); t(80)= .69, p= .49, *d*= .08), but replicated in adolescents (41.1% (*SD*= .20); t(75)= 3.92, p< .001, *d*= .45) and adults (44.9% (*SD*= .20); t(77)= 2.28, p= .026, *d*= .06).

Together, participants reported that Black characters were stronger than White characters, but that White characters were heavier than Black characters. Black and White characters did not differ on height reporting.

Table 6

Percentage of participants choosing a Black character as stronger, taller and heavier in Study 2

	Overall % choosing a Black character	Child % choosing a Black character	Adolescent % choosing a Black character	Adult % choosing a Black character
Strength	60.7% p< .001	55.4% p< .001	63.2% p< .001	63.8% p< .001
Height	47.8% p= .073	46.6% p= .17	47.4% p= .18	49.4% p= .75
Weight	44.8% p< .001	48.3% p= .49	41.1% p< .001	44.9% p= .026

Open-response questions

Coding

Categories for the open-response questions were created by reviewing the responses from participants and creating categories that seemed relevant to both the questions and the responses. The categories that were decided on were: guessing, intuition, seeing their face, seeing the whole body, behavior, referencing another category, referencing the age of the character, referencing the race of the character, other, and no response. If participant responses fit into a category, coders put a 1 in the corresponding column. If participant responses did not fit into a category, coders put a 0 in the corresponding column. Given there was no limit placed on the participants' responses, the responses often included content from multiple categories.

Reliability

Responses were coded by two research assistants. After coding was completed, 30% of each research assistants' responses were re-coded by the other research assistant for reliability purposes. Reliability percentages were: 94.6% for strength, 93.4% for height, and 96.8% for weight. A breakdown of reliability percentages per category is included in the supplemental information.

Open-response category coding for strength, height, and weight

Overall participant open-responses were analyzed separately between strength, height, and weight. Percentages were collected by averaging each coding column, as each column had either a 1 or a 0. If participants referenced multiple coding categories in their responses (e.g., they looked tall and they played basketball), both responses were coded. Thus, percentages do not equal 100. After looking across ages, children, adolescents, and adults are analyzed separately following the same procedure.

Strength

Overall, participants reported the most likely reason for them to report someone as strong was based on seeing their face (55.1%), followed by their behavior (35%) and seeing their whole body (31.1%). This finding was replicated in children, and adolescents. Adults used the same three categories most often, but they used seeing their face, followed by seeing their whole body, and then their behavior last. One child, one adolescent, and two adults referenced race in their explanations. Over 90% of participants (97%) gave an answer to our open-response strength question. The percent of times each category was selected is presented below in Table 7.

Table 7

Explanations for how participants decide the strength of someone overall and by age for Study 2

Category	Overall Strength	Child Strength	Adolescent Strength	Adult Strength	Participant Example
Guessing	6%	9.8%	7.6%	0%	'Guessed'
Intuition	4.7%	2.4%	6.3%	6.8%	'Instinct'
Seeing their face	55.1%	58.5%	48.1%	58.1%	'Look at their face'
Seeing their whole body	31.1%	24.1%	23.7%	45.9%	'Their muscles'
Behavior	35%	28%	46.8%	29.7%	'They lift something really heavy'
Referencing another category	2.6%	3.7%	3.8%	0%	'If they're big then they're strong'
Referencing the age of the character	3.8%	0%	7.6%	4.1%	'Maybe if someone is older, you can presume that they're stronger'
Referencing the race of the character	1.7%	1.2%	1.3%	2.7%	'Maybe Black people are stronger'
Other	2.1%	4.9%	0%	1.4%	'Compared to people I know and see who looks stronger'

Note: Research assistants typed responses for children and teens, so response examples are likely paraphrased throughout

Height

Overall, participants reported they can tell someone was tall by: seeing their face (50%), followed by seeing their whole body (24,4%) and their behavior (18.6%). All age groups report seeing their face, seeing their whole body, and their behavior as how they determine the height of a target. Additionally, children report guessing and adults mention the age of the target being informative. No children or adolescents mentioned race in their responses, but three adults did. Over 90% of participants (92.4%) of participants answered the open-response question about height. The percent of times each category was selected is presented below in Table 8.

Table 8

Explanations for how participants decide the height of someone overall and by age for Study 2

Category	Overall Height	Child Height	Adolescent Height	Adult Height	Participant Example
Guessing	8.1%	14.6%	8.9%	0%	'Just a guess'
Intuition	4.7%	3.7%	2.5%	8%	'Intuition'
Seeing their face	50%	51.2%	50.6%	48%	'Their face, long or short'
Seeing their whole body	24.4%	27.2%	19.2%	26.7%	'They're just tall'
Behavior	18.6%	14.6%	31.6%	9.3%	'What they did'
Referencing another category	5.9%	4.9%	11.4%	1.3%	'Stronger people are usually taller'
Referencing the age of the character	5.5%	0%	7.6%	9.3%	'Their age, especially for kids'
Referencing the race of the character	1.3%	0%	0%	4%	'Black individuals tend to be taller'
Other	3.8%	6.1%	2.6%	1.3%	'You can tell if they're taller if you stand back to back'

Weight

Overall, participants reported they can tell someone was heavy by: seeing their face (72.8%), followed by seeing their whole body (22.6%) and their behavior (9.8%). This finding was consistent in children and adolescents. Adult participants reported they could tell how heavy someone was mostly by seeing their face, followed by seeing their whole body and referencing another size category. No children or adolescents referenced race in their responses, but one adult did. Over 90% of participants responded to the open-response question about weight (95.7%). The percent of times each category was selected is presented below in Table 9.

Table 9

Explanations for how participants decide the weight of someone overall and by age for Study 2

Category	Overall Weight	Child Weight	Adolescent Weight	Adult Weight	Participant Example
Guessing	6.4%	7.3%	10.3%	7.3%	'Guessed'
Intuition	1.7%	2.4%	2.6%	2.4%	'Intuition'
Seeing their face	72.8%	63.4%	76.9%	63.4%	'Their face'
Seeing their whole body	22.6%	25%	17.3%	25%	'Looking at their size'
Behavior	9.8%	9.8%	17.9%	9.8%	'Their actions'
Referencing another category	6.4%	6.1%	5.1%	6.1%	'If they are mean, they're heavy'
Referencing the age of the character	1.3%	0%	2.6%	0%	'Mostly just being older because it gives your body more weight'
Referencing the race of the character	0.9%	0%	1.3%	0%	'If you are a white kid and you're skinnier, those are the ones that are meaner… I try not to pay attention to their race…'
Other	3.4%	9.8%	0%	9.8%	'It doesn't really matter'

Study 2: Summary of findings

Study 2 conceptually replicated findings from Study 1, while providing more information about how the race-based size bias operates developmentally. To begin, overall, participants perceived Black characters as stronger than White characters. In the absence of contextual information (e.g., in the forced choice task), children use racial information to make judgements about characters' strength. However, when contextual information is included (e.g., vignette ratings), children use contextual information, rather than racial information to make their strength judgements, mirroring Study 1. Adolescents and adults, however, used both race and behaviors to determine their strength ratings across measures. Across ages, participants used contextual information (e.g., event type, valence) to make their strength judgements. Characters who engaged in physical and positive events were perceived as stronger than those who engaged in social or negative events.

For height and weight ratings, we did not see evidence of a race-based size bias overall. Instead, participants rated characters who engaged in physical events as taller than those who engaged in social events, and characters who engaged in negative events heavier than those who engaged in positive events. While this weight and valence interpretation goes beyond the scope of this paper, it is hypothesized that this is related to children and adults's preferences towards people in smaller bodies and the associations between thinness being positive, and heaviness being negative (Cramer & Steinwert, 1998; Carvalho et al., 2021). Across two studies, race appeared to be tied to physical strength more strongly than height and weight.

Additionally, the forced-choice measure in Study 2 was the first study to show evidence of children displaying a race-based size bias for strength. Children, adolescents, and adults all chose a Black child as stronger than a White child more than chance showing a link between race and strength. While we do not know the actual strength of the children in the photos used, we do not believe that their actual strength would explain the relation. This is because Wilson et al. (2017) has shown that adults have a bias about Black men being large and strong, independent of the truth. Additionally, the past two studies provide similar evidence. While we do not have evidence of how strong the cartoon faces used may be, they were matched in terms of their face shape and size, and there would be no reason to suspect that one is stronger than the other. Instead, we believe that the race-based size bias is controlling the result. Alternatively, our finding may have been the result of using real faces rather than cartoon faces in this measure. This possibility was addressed in Study 3.

As with our continuous measure above, height and weight did not result in Black children being chosen as taller and heavier than White children. This provides further evidence that strength is the more relevant measure in the race-based size bias and may be the most important piece of evidence to look towards when studying this bias.

In addition to the continuous and forced-choice measures, Study 2 employed two additional, exploratory measures: memory and open-response questions. To begin, overall, adolescent and adult participants remembered events better than child participants. Additionally, participants generally remembered physical events better than social events. While there were no effects of Character Race across ages, adult participants did remember what Black characters did more often than White characters. Despite previous literature (Li, 2012; Kensinger et al., 2006; Corenblum, 2003; Rothbart et al., 1979), this study provided no evidence in favor of negative events being remembered better than positive events. This is likely because the majority of participants remembered the events presented to them; the memory task itself was not taxing, even for 6-11 year old children. Participants were asked to recall what happened in a vignette less than a minute after it was presented to them.

For our exploratory open-response measure, we saw that across ages, participants report making their height judgments through seeing someone's face, seeing their body, and their behavior. While our results display a race-based size bias consistently across adolescents and adults, the participants do not appear to be aware of their race-based decision making. In fact, less than 2% of overall responses explicitly mentioned race as a way to determine someone's strength, height or weight. The few responses relating to race that were given, were quite explicit. For strength, one participant mentioned, 'maybe Black people are stronger.' For height, one participant stated, 'Black individuals tend to be taller,' and for weight, one participant stated, 'if you are a White kid and you're skinnier, those are the ones that are meaner. I know it sounds terrible, but I try not to pay attention to their race because that's not what I want to think about.' While extremely rare, participants that did mention race seemed to be alluding to the stereotype that Black people are larger than White people. Interestingly, regardless of how participants reported making their judgments, our other measures showed that participants are taking race into consideration as well as character's behavior.

Study 2 provided novel information about how Valence impacts strength, height, and weight ratings, as well as how it interacts with Event Type. As with Study 1, we found the importance of context and additional information when participants are making their judgments across age groups. Importantly, we also saw a replication of the race-based size bias findings in adolescents and adults - Black, child, cartoon characters were perceived as physically stronger than White characters. For the first time, using a forced-choice methodology, we saw child, adolescent, and adult participants choosing a photograph of Black child as stronger than a photograph of a White child with an absence of additional, contextual information. Because of the consistency of the strength rating, across methodologies, it appears that this measure may be the first marker or step in the perceptions of race-based size biases. The forced-choice finding leaves open an interesting question. Are child participants perceiving Black children as stronger than White children in this portion of the study because they are real faces rather than cartoon faces? Or, are they perceiving the Black children as stronger than White children because they are given no additional contextual information about the individuals to base their responses on? Lastly, will this finding replicate, or was it simply noise? These questions were addressed in Study 3.

Chapter 4: Study 3

Study 2 provided the first set of evidence that children may be susceptible to the race-based size bias for strength using the forced-choice design. In order to investigate this evidence further, a third study recruiting only child participants and real Black and White faces was employed. Because the original study used cartoon faces, provided contextual information, and used a Likert scale measure and did not see results, it was important to further investigate what was driving the significant forced-choice result. That is, were the cartoon faces the factor that was causing child participants to not display the race-based size bias with the vignettes? While many developmental studies use cartoon-like stimuli, there has also been some debate and criticism about the use of cartoon and computer-generated faces and their generalizability (Gaither et al., 2019; Packer & Moreno-Dulcey, 2022; Rakoczy, 2022; Revencu & Csibra, 2024; Wellman & Yu, 2022). Some researchers question the validity and generalizability of showing images that are not life-like for various reasons. Some researchers have found differences in how real and computer-generated faces are perceived by participants (Gaither et al., 2019), and others criticize the use of non-humans puppets that have no mental states and feel that the interpretations researchers make based on their non-human stimuli are unrealistic and are based on symbolic representations (Packer & Moreno-Dulcey, 2022). On the opposing side of the debate, researchers argue that the representation of non-human stimuli as human is a natural part of life and allows researchers to understand children (Rakoczy, 2022). Nevertheless, it is not entirely clear how representations of people (e.g., puppets, cartoons) are understood by children (Revencu & Csibra, 2024).

To address the potential concern relating to cartoon images, we put the real faces from the forced-choice portion of the study in place of the cartoon faces for the vignette portion of the study, but kept the rest of the study the same. If child participants show a race-based size bias using the real faces, it would indicate that cartoon faces may not be the most appropriate stimuli to show this bias in children, but that they do possess the bias. If child participants did not show a race-based size bias with the real faces, it would further validate the finding that children are not as sensitive to a race-based size bias, and that they are more likely to use contextual information to make their judgments, regardless of the type of stimuli used.

Another main motivation of Study 3 was to investigate if the race-based size bias for strength observed in children using the forced-choice measure would replicate. Because this was the first time there was evidence of children showing such a bias, it was important to ensure that this finding replicates in order to better understand the potential developmental foundations of the bias. Collecting additional data on memory and open-response questions was also of interest. This allowed us to follow the same study design provided in Study 2, and further investigate the impact of memory on the race-based size bias.

Method

Participants

Children

One hundred and nine (49 boys, 58 girls, 2 nonbinary) 6- to 11-year-old children were recruited from a local museum. Child participants were predominantly White (107/109 families voluntarily reported the child participant’s race: 70.1% White, 2.8% Black, 6.5% Asian, 8.4% Latino, 9.3% Mixed-Race, 0.9% Middle Eastern and 1.9% Other). Participants were recruited in two age cohorts with at least 50 participants per age cohort (6-8 years; n= 56 participants, 8-11 years; n= 53 participants) to ensure sufficient representation across the full age range. Nine additional participants were dropped from the final sample because they did not complete at least half of the trials, and one additional participant was dropped because they did not speak English. Due to experimenter error, some real faces did not appear in place of the cartoon faces. This was caught and corrected well before data collection completed, but resulted in dropping full trials (N= 7) as well as individual questions (N=9).

Power Analysis

An a priori power analysis was conducted using the means and standard deviations collected from the child participants of the forced-choice block in Study 2. Given our results, the analysis suggested that 97 child participants were sufficient to detect, and thus replicate, a significant effect of Character Race in the child forced-choice block with 0.8 power, in a one-sample design. Because we did not expect a significant effect of Character Race in the child age group, we did not power to detect an effect of Character Race. However, we were powered to detect a significant difference in Event Type (physical contact - negative compared to social exclusion - negative), which required only 36 participants in a matched-pairs design. To ensure we had sufficient representation across all age ranges, we recruited at least 50 participants into smaller cohorts within the child age group.

Procedure

The procedures were identical to Study 2 with one exception. For the physical size and strength ratings in Study 3, the real faces from the forced-choice task were used in place of the cartoon faces used in Studies 1 and 2 (see Figure 11).

Results

Preliminary Analyses

We conducted preliminary analyses on participant gender and presentation order across strength, height and weight ratings. There were no significant main effects of participant gender or presentation order, so we did not include them in further analyses.

As in Study 2, our goal was to present participants with mildly negative or positive events and ensure that participants understood each event to be negative or positive, respectively. Preliminary analysis of the meanness ratings revealed that participants viewed some negative events as positive interactions, choosing ‘very, very nice,’ ‘very nice’ or ‘nice’ (3 or below) on the meanness scale (N= 31 trials) or viewed some positive events as negative interactions, choosing ‘very, very mean,’ ‘very mean’ or ‘mean’ (N= 18 trials). These trials were spread across participants and no participant needed to be dropped as a result. To ensure we were assessing participant’s representations of the size of actors in negative events, we removed these individual trials from all subsequent analyses.

Physical size and strength ratings

Overall analysis plan

We conducted the same overall analysis plan as discussed in Study 2 with the caveat that we only had one Age Group, children. All analyses included Age Cohort (6-8 years; 9-11 years) rather than Age Group.

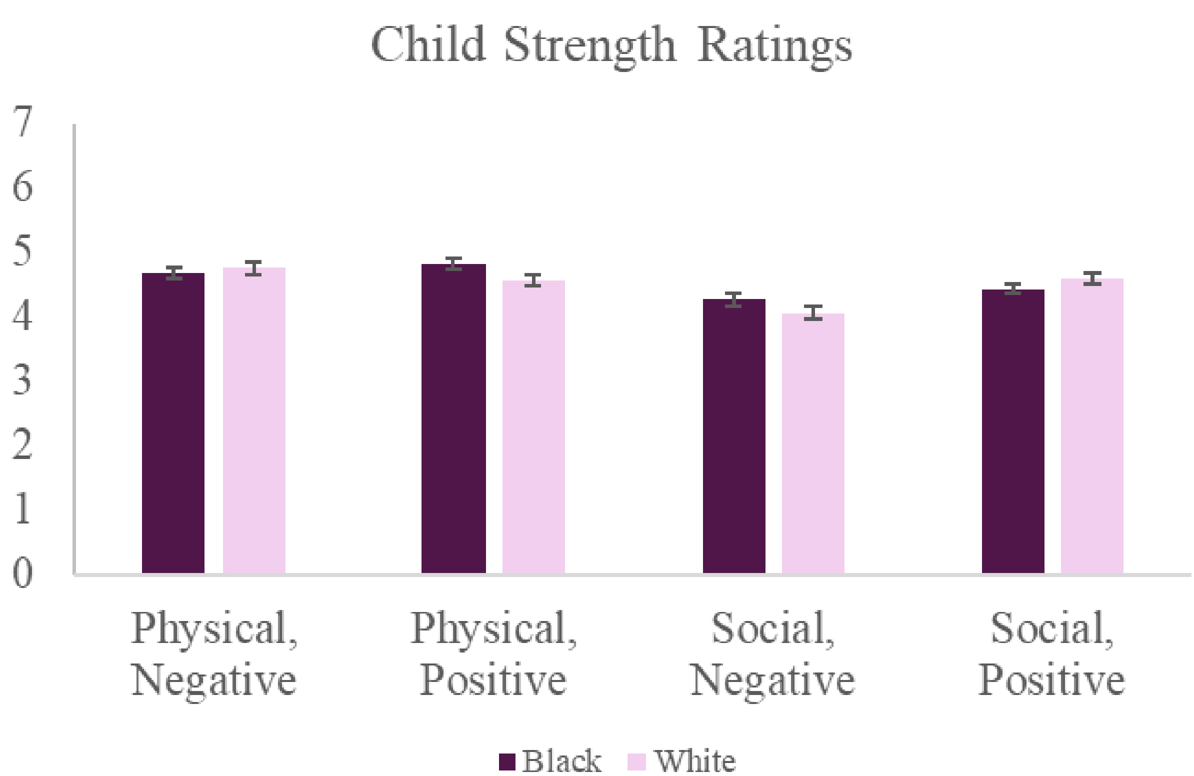
Strength ratings

Event Type significantly predicted children’s strength ratings, $F(1, 102) = 35.56$, $p < .001$. Characters who engaged in physical events (M= 4.71, SD= 1.24) were perceived as stronger than those who engaged in social events (M= 4.35, SD= 1.32,). However, there were no significant effects of Character Race, $F(1, 1228) = .78$, $p = .37$, Valence, $F(1, 111) = 3.19$, $p = .08$ or Age Cohort, $F(1, 106) = .85$, $p = .36$, and no significant interactions involving these factors, on participants’ strength ratings.

These effects were qualified by a significant interaction between Event Type, and Valence, *F*(1, 1292)= 14.09, p< .001. Post-hoc tests revealed that Valence had an effect on social, *t*(198)= 3.46, p< .001, but not physical, *t*(194)= .41, p= .69 events. Characters who participated in social - positive (M= 4.52, SD= 1.17) events were significantly stronger than those who participated in social - negative (M= 4.16, SD= 1.44) events, but physical - positive (M= 4.7, SD= 1.14) and physical negative (M= 4.72, SD= 1.34) events, did not differ. There were no other significant interactions. The results are displayed below in Figure 18.

Figure 18

Child (6-11 years) strength ratings for real faces for Study 3

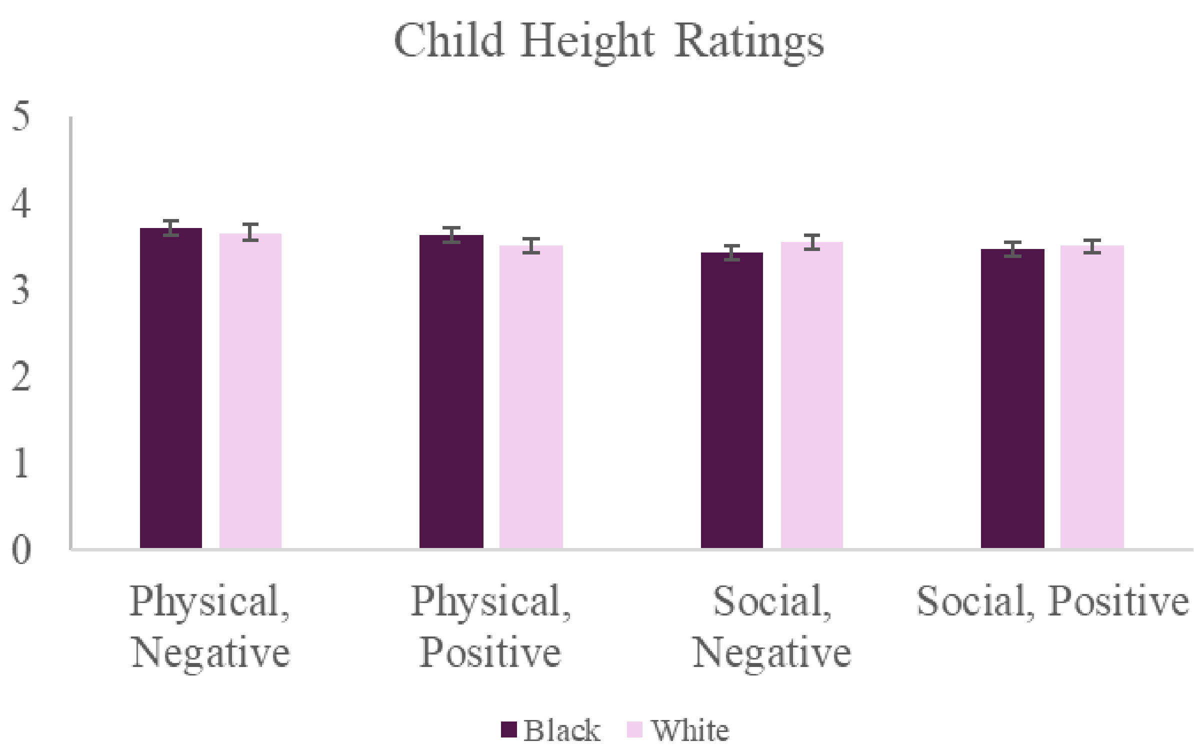


Height ratings

Event Type significantly predicted children's height ratings, *F*(1, 405)= 7.27, p= .007. Characters who engaged in physical events (M= 3.63, SD= 1.16) were perceived as taller than those who engaged in social events (M= 3.5, SD= 1.15). However, there were no significant effects of Character Race, *F*(1, 112)= .0009, *p*= .98, Valence, *F*(1, 105)= .85, p= .36 or Age Cohort, *F*(1, 105)= .69, *p*= .41, and there were no significant interactions. The results are displayed below in Figure 19.

Figure 19

Child (6-11 years) height ratings for real faces for Study 3



Weight ratings

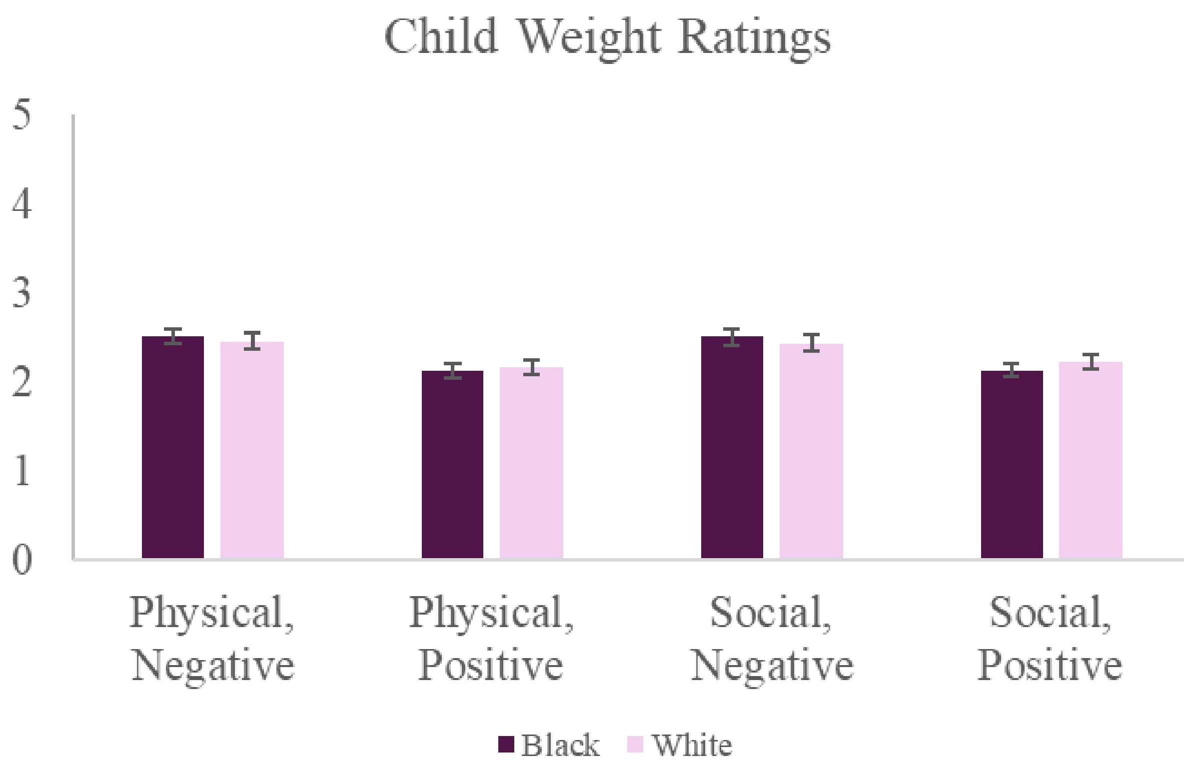
Valence significantly predicted children's weight ratings, *F*(1, 129)= 27.19, p< .001. Characters who engaged in negative events (M= 2.48, SD= 1.22) were perceived as heavier than those who engaged in positive events (M= 2.16, SD= 1.09). However, there were no significant effects of Character Race, *F*(1, 783)= .003, *p*= .96, Event Type, *F*(1, 266)= .035, p= .851 or Age Cohort, *F*(1, 107)= .37, *p*= .54.

These effects were qualified by a significant interaction between Character Race and Age Cohort, *F*(1, 786)= 3.91, p= .048. Post-hoc tests revealed no significant differences between any pairwise comparisons.

To follow up on this interaction, separate models were completed on the younger and older Age Cohorts separately. The model on younger children revealed an effect of Valence, *F*(1, 62.60)= 11.47, *p*< .001, where characters who engaged in negative events (M= 2.54, SD= 1.31) were perceived as heavier than those who engaged in positive events (M= 2.18, SD= 1.15). There were no significant effects of Character Race, *F*(1, 339.9)= 1.89, *p*= .17 or Event Type, *F*(1, 143.2)= .47, *p*= .49. The model on older children revealed an effect of Valence, *F*(1, 418.1)= 16.14, *p*< .001, where characters who engaged in negative events (M= 2.42, SD= 1.13) were perceived as heavier than those who engaged in positive events (M= 2.15, SD= 1.03). There were no significant effects of Character Race, *F*(1, 50.5)= 2.08, *p*= .16 or Event Type, *F*(1, 60.3)= .27, *p*= .61. There were no other significant interactions. The results are displayed below in Figure 20.

Figure 20

Child (6-11 years) weight ratings for real faces for Study 3



Correlations between Strength, Height, Weight and Age

We conducted a correlation analysis on the data including the factors: strength, height, and weight, as was done in Studies 1 and 2.

The matrix is presented in Table 10 below and displays Pearson correlation coefficients for all pairs of the variables. Notably, we see a positive correlation between strength and height measures ($r = .457$, $p < .001$), replicating Study 2. We do not see any correlations between strength and weight ratings ($r = .144$, $p = .134$) or height and weight ratings ($r = .035$, $p = .72$).

Table 10

Correlation table between strength, height, and weight for Study 3

	Strength	Height	Weight
Strength	- -		
Height	Pearson's r= .457 *p*< .001	- -	
Weight	Pearson's r= .144 *p*= .134	Pearson's r= .035 *p*= .720	- -

Degree of Niceness and Meanness ratings

The same coding scheme and analysis model from Study 2 was used for Study 3.

The model yielded a significant effect of Age Cohort, $F(1, 107) = 9.67$, $p = .002$ where post-hoc tests revealed younger children (M= 2.11, SD= .8) use more extreme scores than older children (M= 2.85, SD= .77; $t(108) = 3.11$, $p = .002$) when using a Likert scale. Additionally, the model yielded a significant interaction between Event Type and Valence, $F(1, 1206) = 32.27$, $p < .001$. Post-hoc tests revealed that Valence affected physical and social events inversely with physical events being perceived as meaner overall. Physical - negative events (M= 2.05, SD= .8) were rated as a higher degree of mean than social - negative events were (M= 1.87, SD= .81; $t(312) = 3.46$, $p < .001$), but physical - positive events (M= 1.89, SD= .78) were rated as a *lower* degree of nice than social - positive events were (M= 2.1, SD= .78; $t(305) = 4.08$, $p < .001$). As with all previous child models, there was no effect of Character Race, $F(1, 853) = .052$, $p = .82$.

Memory check

Memory check response coding and accuracy measures were the same as outlined Study 2. We conducted the same generalized logistic mixed effects regression models for the memory check as in Study 2.

The model yielded significant effects of only Age Cohort, $\chi2(1)= 18.22$, $p< .001$ where older children (M= .89, SD= .31) remembered more than younger children (M= .79, SD= .41). There were no significant effects of Character Race, $\chi2(1)= 2.21$, $p= .14$, Event Type, $\chi2(1)= 1.62$, $p= .2$, or Valence, $\chi2(1)= .05$, $p= .82$.

Memory on physical size and strength models

To further investigate the effect of memory on strength, height, and weight, we re-ran our above physical size and strength models on strength, height, and weight, but only including trials where participants remembered events. Because we do not have enough data on children who did not remember, we only ran the subsequent analyses on those who do remember. These results should be interpreted with caution, as they are done on a subset of the larger dataset, and are largely exploratory.

First, the memory dependent strength model replicated the main effects and interactions of the full model. Unlike the original model, this model includes an additional three-way interaction between Character Race, Event Type, and Valence, $F(1, 1050.6)= 9.11$, $p= .003$.

Follow-up models revealed that when physical events were isolated, positive events may be subject to a race-based size bias, with Black characters being perceived as stronger than White characters, $t(215)= 2.18$, $p= .03$, whereas this finding does not persist with negative events, *t*(195)= .72, p= .47. Specifically, Black (M= 4.86, SD= 1.05) characters in positive events were perceived as stronger than White (M= 4.63, SD= 1.14) characters in positive events, but Black (M= 4.72, SD= 1.23) and White (M= 4.81, SD= 1.34) characters in negative events were perceived to be equally strong. While this finding is exploratory in nature and should be interpreted with caution, it could be evidence of ambiguity being necessary for the race-based size bias to show up in child participants. The memory dependent height and weight models replicated the findings of the original models with one exception. The original weight model yielded an interaction between Character Race and Age, but this was not replicated in the memory dependent model. Because the majority of participants remembered the events provided, memory did not have a large impact on physical size and strength ratings.

Forced-choice

Analysis plan

The same analysis plan from Study 2 was followed for Study 3.

Strength ratings

Replicating the finding from Study 2, child participants chose a Black child as stronger than a White child 55.1% (*SD*= .218) of the time, which was significantly greater than chance t(93)= 2.28, p= .025, *d*= 0.24.

Height ratings

Child participants chose a Black child as taller than a White child 45.5% of the time, which was significantly *less* than chance, *t*(93)= 2.04, p= .045, *d*= 0.21. This did not replicate Study 2.

Weight ratings

Child participants chose a Black child as taller than a White child 48.7% of the time, which did not differ from chance, $t(92)= 0.48$, $p= .63$, $d= .05$, which did replicate Study 2.

Together, child participants reported that Black characters were stronger than White characters, but that White characters were taller than Black characters. There were no statistical differences between Black and White characters' perceived weight.

Table 11

Percentage of child participants choosing a Black character as stronger, taller and heavier in Study 3

	Child % choosing a Black character
Strength	55.1% p= .025
Height	45.5% p= .045
Weight	48.7% p= .63

Open-response questions

Coding

The same coding scheme as Study 2 was used for Study 3.

Reliability

Responses were coded by two research assistants. After coding was completed, 30% of each research assistants' responses were re-coded by the other research assistant for reliability purposes. Reliability percentages were: 96.67% for strength, 96.94% for height, and 97.2% for weight. A breakdown of reliability percentages per category is included in the supplemental information.

Open-response category coding for strength, height, and weight

As in Study 2, child open-responses were analyzed separately between strength, height, and weight. Percentages were collected by averaging each coding column, as each column had either a 1 or a 0. If participants referenced multiple coding categories in their responses (e.g., they looked tall and they played basketball), both responses were coded. Thus, percentages do not equal 100.

Strength

Children reported the most likely reason for them to rate someone as strong was based on their behavior, seeing their whole body, and seeing their face. These three categories were reported as the most likely reasons for rating someone as strong in Study 2 as well. Over 90% of participants (97.8%) gave an answer to our open-response strength question. The percentage of times each category was selected for strength, height, and weight is presented below in Table 12.

Height

Children reported the most likely reason for them to rate someone as tall was based on seeing their face, seeing their whole body, and seeing their behavior. This replicates the top three categories for rating someone as tall in Study 2. Over 90% of participants (91.3%) gave an answer to our open-response height question. The percentage of times each category was selected for strength, height, and weight is presented below in Table 12.

Weight

Children reported the most likely reason for them to rate someone as heavy was based on seeing their face (21.74%), seeing their whole body (32.61%), and their behavior (58.7%). This replicates the top three categories for rating someone as heavy in Study 2. Over 90% of participants (90.22%) gave an answer to our open-response weight question. The percentage of times each category was selected for strength, height, and weight is presented below in Table 12.

Table 12

Explanations for how child participants decide the strength, height, and weight of someone in Study 3

Category	Child Strength	Child Height	Child Weight	Participant Example
Guessing	2.2%	4.35%	5.43%	'Guessed'
Intuition	2.2%	6.52%	3.26%	'Intuition'
Seeing their face	21.74%	41.3%	52.17%	'By their face'
Seeing their whole body	32.61%	30.4%	34.78%	'Bigger build'
Behavior	58.7%	17.39%	13.04%	'By what they do, like if they play soccer a lot or basketball'
Referencing another category	4.35%	3.26%	5.43%	'Chubbier might be heavier than average'
Referencing the age of the character	1.09%	4.35%	0%	'Based on age'
Referencing the race of the character	0%	0%	1.1%	'People in Africa, people are skinny so they are smaller'
Other	1.09%	0%	9.78%	'By carrying them'

Note: Research assistants typed responses for children, so response examples are likely paraphrased

Study 3: Summary of findings

Study 3 exclusively investigated child participants' ratings using real faces rather than cartoon ones. Despite the change of stimuli, the child results from Study 3 largely replicated those of the first two studies. When contextual information was provided, children used context, rather than race, to make their strength judgements. However, in the absence of contextual information, and using a more simplified design, children once again chose a Black character as stronger than a White character.

Specifically, for the strength vignette ratings, child participants did not take race into consideration, but used Event Type and Valence to make their strength judgements. Characters who engaged in physical events were perceived as taller than those who engaged in social events, replicating previous studies. Valence impacted weight ratings where characters who engaged in negative events were perceived as heavier than those who engaged in social events, replicating Study 2. Together, we see that children consistently use contextual information (e.g., physical or social events) to determine the strength and height of characters, and use valence information to determine the weight of characters. However, race was not a predictor of physical size and strength ratings for children. Because the main results of Study 2 were replicated in Study 3 with real faces, we feel confident that the cartoon stimuli that were used in Study 1 and Study 2 were valid.

Interestingly, when only participants who remembered an event were included in our strength model, we begin to see a pattern where Black characters in positive events are perceived as stronger than White characters in positive events, but Black and White characters who engaged in negative events were perceived as equally strong. This may be evidence of potential ambiguity being necessary for children to show a race-based size bias. When an event is negative, it is possible that there's no 'room' for a bias to come through due to a potential link between something being negative and strong. It is possible that the link is threat itself, however, that is not directly tested in this study and we can only speculate about this potential relation. From this study alone, it is not entirely clear which direction the ratings are being pushed. For example, it is possible that the positive events are closer to baseline ratings for Black and White characters and that the negativity of the other events brings the White characters' strength ratings up.

Alternatively, it is possible that the negative events are closer to baseline ratings for Black and White characters, and the negative events only impact the Black characters strength ratings due to previous stereotypes about Black men being threatening. This general pattern of findings replicated the most closely related study to this one in the literature by Freiburger et al. (2024) where threatening facial stimuli attenuated a race-based size bias for height. In this case, the researchers explained that the perceived threat caused a size bias to appear in White targets and not change from baseline for Black targets. Based on my findings and Freiburger et al. (2024), I predict that it is the former: positive events would leave more 'gray area' to be filled in with stereotypes resulting in Black characters being perceived as stronger than White characters. Whereas, the negativity of an event may be more salient than the race of a character, which is what results in similar ratings across races. Future studies investigating perceptions of Black and White faces with no additional context may be influential in parsing out this distinction.

Replicating Study 2, when using a forced-choice design, children chose an image of a Black child as stronger than a White child. The forced-choice results from Study 2 and 3 are the first pieces of evidence that the race-based size bias begins in childhood, rather than adolescents. The forced-choice design is a very sensitive and direct measure, and may be more appropriate for testing the race-based size bias children. Alternatively, it may be the absence of contextual information resulting in evidence of a race-based size bias. It's possible that child participants are favoring the contextual information over the racial information in the vignettes. However, when they are given stimuli with no additional information, the racial information is being prioritized rather than eclipsed by the contextual information. This possibility will be explored as a future direction in the general discussion. The forced-choice method also resulted in White targets being perceived as taller than Black targets, but there were no differences in the perceived weights of Black and White children. Even using this more sensitive measure, it is apparent that height and weight are not the strongest predictors of the race-based size bias, but strength is.

Chapter 5: General Discussion

The goals of this dissertation were to investigate the developmental onset of the race-based size bias and to address how different contextual factors change the representation of the bias, if at all. To test these ideas, we employed three studies. In Study 1, experimenters tested children, adolescents, and adults for the presence of a race-based size bias towards Black and White children by pairing cartoon faces with mildly negative vignettes. Study 2 also tested children, adolescents, and adults, but the methods expanded significantly on Study 1. First, both positive and negative vignettes were introduced in Study 2. Additionally, real faces of young Black and White adolescents were introduced to participants in a forced-choice task asking participants which target was stronger, taller, and heavier. Memory for events were tracked by experiments in Study 2, unlike in Study 1. Finally, participants in Study 2 were asked open-response questions about how they can tell someone is strong, tall, and heavy. Study 3 only included child participants and was a replication of Study 2 with the main exception of using real faces instead of cartoon faces in the vignette physical size and strength rating portion of the experiment.

Across three studies, children used contextual information when it was available, rather than race, to make decisions about characters' strength. However, in the absence of contextual information and using a simplified measure, children do begin to show evidence of a race-based size bias for strength. Across two studies, adolescents and adults used both race and contextual information to make a decision about characters' strength, both in the absence and presence of contextual information.

While we had always predicted that Black characters would be perceived as stronger than White characters, and we did see this consistently in adolescents and adults, we did not begin this study with any specific predictions about how the events, or contexts, would affect the results. Now, we have a greater understanding of how contextual factors do and do not influence the race-based size bias. Specifically, characters who engaged in physical events, especially those that were negative, were more likely to be perceived as strong. Additionally, across studies and age groups, we gained evidence that Valence had an effect on social, but rarely on physical events. Characters who engaged in physical events were always perceived as strong, but only characters who engaged in social - positive events were perceived as strong. While this finding needs to be interpreted with caution, after breaking down the strength model to only include child participants who remembered events in Study 3, we saw that Black characters in positive events were stronger than White characters in positive events. However, there were no race differences between negative events. We interpreted this finding in relation to the importance of ambiguity in events highlighted in previous literature (Sagar & Schofield, 1980; McGlothlin & Killen, 2010). When something is perceived as negative, there's no ambiguity in the information, and it can be interpreted somewhat independently of race. However, when contextual information is more ambiguous, there is more 'room' for biases and heuristics to come into judgments. This finding must be explored in future studies where the prediction is made a priori the expectation is that only participants who remember events will be included in the final dataset.

Although we found evidence for the emergence of the race-based size bias related to strength judgements, we did not find consistent evidence for a race-based size bias for height or weight. For our height measures, we did not see any effects of race across ages. Instead, participants used contextual information of the type of event to make their height ratings. Across the three studies, characters who engaged in physical events were taller than participants who engaged in the other events. In Study 2 and Study 3, despite the manipulation of Valence, there was no effect of positive and negative events on height ratings. In Study 2, adolescents perceived characters as taller than children and adults, but we did not see evidence of this in Study 1. Despite our original hypotheses for Study 1, we found no evidence that children, adolescents, or adults perceived Black characters as taller than White characters.

This is not overly surprising given Navon et al. (2025) did not find a significant replication for a race-based size bias for Black men for height or weight, and Johnson & Wilson (2019) did not find evidence of this bias for height. However, they did replicate the finding that Black men are perceived as stronger than White men. Similarly, for our weight measures, we found no evidence that children, adolescents, or adults perceive Black characters as heavier than White characters. Instead, Valence seemed to be the strongest predictor of weight ratings. In Study 1, we did not manipulate Valence, but participants still used the contextual information from the event to make their ratings. In both Study 2 and Study 3, Valence was used to make their weight ratings. Specifically, when characters engaged in negative events, they were perceived as heavier than those who engaged in positive events. I hypothesize that this is related to people's generally negative opinions of people in larger bodies, and an association between negativity and heaviness (Cramer & Steinwert, 1998; Carvalho et al., 2021).

Because there were no strong pieces of evidence upholding a race-based size bias for height and weight across three studies, and mixed findings for the significance of height and weight in the literature (Wilson et al., 2017; Johnson & Wilson, 2019; Navon et al., 2025), I hypothesize that the race-based size bias is linked more strongly and consistently to strength than it is to height or weight. This is underscored by our forced-choice results where only strength showed evidence in the expected direction of the race-based size bias. Future research may be necessary to test this idea further, but the results suggest that since characters who engaged in physically threatening events (e.g., physical contact) were rated as stronger than other events, and Black characters are rated as stronger than White characters, there is a link between physical threat, strength, and Black characters. The exact mechanism behind the relation remains an open question, but this could be studied by measuring or manipulating threat specifically in children, adolescents, and adults.

From our forced-choice measure, we gained significant insight into the race-based size bias. Across age groups, and within each age cohort, participants chose a Black character as stronger than a White character. This was replicated a second time in child participants showing that when a measure is sensitive enough, children may possess the race-based size bias for strength. While originally it was an open question if the results were due to real faces being used instead of cartoon faces, this point was clarified with the employment of real faces used in place of cartoon faces in Study 3. Because children's ratings did not differ with the Likert scale vignette ratings for real faces compared to cartoon faces, we believe that the sensitivity and simplicity of the forced-choice measure was what elicited the race-based size bias for strength in children. What is still open for discussion is if the contextual information is covering up a potential race-based size bias in the first measures, or if it is the measure itself that allows for heightened sensitivity in detecting the bias. This idea is discussed in the 'Future directions' section below. Again, we do not see evidence of a race-based size bias in the predicted direction for height or weight. There were no significant findings in favor of Black characters being perceived as taller than White characters.

It is possible that the discrepancy between the physical size and strength ratings and the forced-choice ratings are because each of these measures are investigating different constructs, or that they are not valid for testing what we set out to measure. That said, we do not think this is the case. First, although children showed evidence of the race-based size bias for the forced-choice measure and not in the continuous measure, this discrepancy does not occur for adolescents and adults. Both adolescents and adults consistently, across measures, show evidence of a race-based size bias for strength. Because of this, we think it is more likely that the forced-choice measure was simpler for children, resulting in significant results in this measure, but not in the continuous measure. In the future (discussed below) it should be investigated whether children are not showing evidence of the race-based size bias due to the continuous measure including contextual factors, or if it was from difficulty with the measure itself.

Another possibility is that the representation of the race-based size bias changes across development. For example, it may be true that in adolescence and adulthood, the race-based size bias is robust and may not be tied to any particular measures. However, children may still be developing the race-based size bias resulting in the biased judgements only occurring in specific conditions (e.g., forced-choice measures, high levels of ambiguity) and favoring contextual information. While we cannot necessarily attest to what the developmental changes of the race-based size bias are or what is causing the change with our measures, we don’t believe this causes an issue with our results. Due to the consistent replication of our findings, we are confident in the measures we used to test for the race-based size bias. Future studies may investigate the specific developmental changes resulting in the consistent evidence of a race-based size bias occurring in adolescence.

A new open-response measure that was added to Study 2 and Study 3 assessed participants' self-reported reasoning for knowing if someone is strong, tall, or heavy. Results revealed that even when participants are using racial information to make their decisions a statistically significant amount, almost no participants report using racial information or stereotypes at all. Instead, participants report looking at the faces and judging, looking at the targets' whole body and judging, or looking at their behavior to make their judgements. Participants are correct in their assessments of the importance of characters' behaviors in their judgments, but adolescents and adults are not reporting a very important factor in their determination of someone's strength. This could be for a variety of reasons. First, participants may just be uncomfortable talking about race and racial information. Apfelbaum et al. (2008) have found that children, around age 10, do not mention racial information, even if it would be to their benefit in an identification game. White adults, which is the racial make-up of the majority of our adult participants, are also often uncomfortable speaking about race and try to cite ideas like 'equality' or a 'post-race' society rather than speaking about and confronting racism, biases, and stereotypes (Harries, 2014). If participants are uncomfortable mentioning racial information, they will not bring it up despite thinking it. One way to combat this is to ask participants outright if they are aware of the race-based size bias. This future direction is outlined below.

Alternatively, participants may not be aware that they are more likely to think Black characters and people are stronger than White characters, and their biased perception is more subconscious. This discrepancy between reported information and our actual size and strength measures pose an interesting question that should be followed up in subsequent research. Ultimately, adult participants consistently showed evidence of a race-based size bias for strength, but extremely rarely reported using racial information to make their judgements, despite being able to type their own responses anonymously and in private. This tells us that the open-response questions are not as much a limitation, as they are an opportunity to ask participants more questions about the bias in more explicit ways.

While not a perfect measure (discussed below) participants' memory for events were collected. In order to investigate participants' memory, we first looked to see if participants were remembering events, and the majority of participants were. Next, we investigated whether Character Race, Event Type, Valence, and/or Participant Age Group predicted memory as a dependent variable. In Study 2, we found that Participant Age Group and Event Type mattered in who remembered events, but only Age Cohort predicted memory in Study 3. Lastly, we removed all participants who did not remember events from the model and re-ran our original analyses with only the participants who remembered. Across the board, results did not change drastically when only participants who remembered events were included in the dataset, with one exception. Broken-down models including only participants who remembered the events and isolating physical events showed that Black characters in positive events were stronger than White characters in positive events, but there were no significant differences between Black and White characters' strength in negative events. As previously stated, this is an interesting finding as it could be indicative of a potentially interesting effect, but must be interpreted with caution. If this effect was to replicate, it would provide information on some of the conditions required for a race-based size bias to emerge. For example, when situations are clearly negative, but still controlled, participants may be unlikely to activate their stereotypes on race and physical size. When situations are more positive, they could be interpreted more ambiguously and result in more biased size and strength judgments. Now, we do not see this effect throughout the study despite the possibility of Valence and Character Race interacting in multiple models, so this needs to be interpreted cautiously. However, it may warrant a further investigation into this relation by powering the study to an interaction between Valence and Character Race in the future.

Another aspect of memory we hoped to investigate was a relation between memory and valence. Despite a wealth of literature debating whether positive and negative events help or hinder memory, we did not see evidence of valence predicting memory. Because this was not one of our main measures, it is possible that we were not powered to detect this effect. Additionally, the majority of our participants did remember the events we gave them, so there was little room to detect differences between groups. Despite the intention to gain additional information about memory and valence, we were unable to detect a relation between these two measures using this methodology.

As an exploratory measure, we looked at the Degree of niceness and meanness used as a dependent variable. Across children, adolescents, and adults, we saw that Event Type was consistently a significant factor when participants rated how nice and mean events were either as a main effect, or as a factor interacting with Valence. Across ages, participants perceived Black characters as nicer overall, but this does not replicate in individual age groups. Interestingly, this measure did not align with what would be expected based on previous literature where Black characters are perceived more negatively than White characters (McGlothlin & Killen, 2010; Sagar & Schofield, 1980). Due to this finding being based on an interaction in an exploratory measure that does not replicate within age groups, we did not interpret this finding much further. Because we used our ‘nice or mean’ measure as more of a filtration device than its own variable, all results should be interpreted with caution. However, this additional variable underscores the importance of Event Type and Valence in not only our Degree of niceness and meanness measure, but in the data overall.

Conclusion

These series of studies did not set out to answer the ‘why’ or the ‘how’ of the race-based size bias, nor did it intend to serve as a finalizing conclusion to the directionality piece of this bias. Instead, these studies aimed to address when the race-based size bias appears developmentally. Based on the results described above, I believe that we succeeded in gaining information on when the race-based size bias develops. Children show evidence of a race-based size bias *only* using an incredibly sensitive measure. Otherwise, children use information about the target’s behavior to make their judgments of size and strength. However, in adolescence, participants began to consistently show evidence of a race-based size bias for strength while also using information about the target’s behavior to inform their judgments. The racial and behavioral information was used consistently though adulthood, and in more ‘ambiguous’ events, race and contextual information interacted. These results showed that the race-based size bias emerges sometime between the ages of 12 and 17. This is incredibly informative because interventions can be made targeting this age group and trying to understand what external and internal factors may be contributing to the emergence of this bias. While it was not necessarily the goal of the projects, I also believe that we began to uncover the ‘how’ of the bias.

We began to discover the ‘how’ of the bias by looking at the contextual factors (e.g., different types of events, valence) and determining how they impacted ratings. While we likely did not have enough power to detect strong interactions between Character Race and Event Type and/or Valence, we do see the importance of these contextual factors when participants across ages are making their judgments. Children only use contextual information to make their judgments, adolescents use both racial and contextual information, and adults use racial and contextual information both separately and in interaction to make their conclusions. This tells us that it is not just the race of a person that matters, but the additional information surrounding the person that matters.

This project alone does not provide enough evidence to make a decision about the potential direction of the bias. However, if we bring in other literature, it does begin to point towards the race-based size bias building off of a negativity bias towards Black individuals. Many previous studies have shown that Black children are perceived as meaner than White children (McGlothlin & Killen, 2010; Sagar & Schofield, 1980), but no studies to date have shown that child participants perceive Black characters as stronger and physically larger using similar methodologies. While the aforementioned studies are not necessarily measuring threat, they are measuring the bias to think that Black children are meaner, which to a child, could also be perceived as threatening. Due to a wealth of prior research stating that White children have negative opinions of Black children early in development, but significantly less evidence that White children perceive Black children as larger and stronger, with the race-based size bias appearing later in development (e.g., adolescence). Additionally, the contextual information provided by the above studies shows that the race-based size bias is influenced by types of events as well as valence. Thus, I predict that (1) the threat bias informs the size bias and (2) contextual information matters significantly when making physical strength and size judgements. While it is not necessarily a limitation that this study does not fully answer the directionality piece, this study is not without limitations.

Limitations

One potential limitation to Study 1 and Study 2 was the use of cartoon faces. While the cartoon faces had very strong internal validity, they lacked the external validity that real faces of children or young adolescents could provide. In addition to the potential issue of external validity and generalizability, we also saw results where children began to show evidence of a race-based size bias using a forced-choice methodology. In order to further test this, we replaced the cartoon faces from Study 1 and Study 2 with the faces that were used in the forced-choice section of Study 2. With this, we found that our Study 2 results replicated using real faces or cartoon faces, validating the use of cartoon faces and the results from Study 1 and Study 2. Study 3 added significantly to the value of the overall project as we now have evidence that children only begin to show evidence of the race-based size bias for strength using very sensitive measures (e.g., forced-choice methodology) and that the cartoon faces do not negatively impact the results or main takeaways of the project.

While we investigated whether children, adolescents, and adults remembered events, this measure may not be the most accurate way to test for participant memory for a few reasons. First, children and adolescents had to verbally describe what happened in the study to the research assistant, who then had to type their responses as quickly as possible. If children or adolescents did not remember the event, they were reminded before moving forward in the experiment. Due to counterbalancing, some participants would have given strength and niceness ratings first and others would have given height or weight ratings first, complicating the impact of their memory on results. In our original model, we did not account for participants memory, but follow-up analyses that did control for participant memory yielded largely the same results as the original model, so it is unlikely memory significantly impacted results. On the other hand, adult participants had to type their own responses, signaling a discrepancy in methods. Because the majority of participants remembered the events, and when we removed participants who did not we sustained largely the same results, we do not think this memory measure invalidates our experiments or results.

Finally, another potential limitation revolves around the inability for our study to test specifically for threat. While we did manipulate Valence and Event Type, which could be correlated or related to threat, we did not ask specifically about threat. We also did not operationalize what threat may have looked like in this study for children, adolescents, and adults. Threat perception may have been a measure that went beyond the scope of this project, but having this link could have assisted in understanding which direction(s) the race-based size bias operates in.

Future directions

One future research direction would be continuing to investigate how contextual factors do and do not influence the race-based size bias. While the three studies provide substantial evidence that children are using contextual information rather than racial information to make their size and strength judgements, it is still an open question what would happen in the absence of contextual information. It's possible that we would see a replication of the pattern that young children simply do not have a race-based size bias. Alternatively, it is possible that with the removal of contextual information, the race-based size bias presents in children. This could be due to contextual information being more salient than racial information for children, thus 'over-powering' the racial information. Because there was consistent evidence of children perceiving Black children as stronger than White children in a sensitive forced-choice task, it's possible that similar results would exist in a continuous measure that is easier for children to track.

Another potential future direction would be to investigate the race-based size bias using Black and White girls instead of just boys. Because much of the race-based size bias has been validated with men, starting with boys was well motivated. However, there is some existing literature that motivates investigating the race-based size bias in girls as well. As explained in Johnson & Wilson (2019), Black women have many stereotypes and consequences associated with them that are similar to Black men. For example, Black women are stereotyped as confrontational and assertive (Smith-Evans et al., 2014). Also, Black women are imprisoned more often than their White counterparts (Gross, 2015). These findings relating to stereotyped behavior and incarceration also apply to young, Black girls. Black girls are perceived by White perceivers as more culpable for their crimes and are thought of as in need of less nurturing and protection because they are thought of as more adultlike than their non-Black peers (Epstein et al., 2017). Black girls are also more likely to be suspended from school than White girls (Caldera, 2018). For size-based judgements, Johnson & Wilson (2019) found that when they controlled for strength, Black women were rated as stronger than White women showing evidence that there may be a race-based size bias for Black women. Slightly more removed, previous research has shown that women who take part in risky, life-threatening activities are perceived as larger than women who do not (Fessler et al., 2014). While this finding does not relate to race specifically, it does show the importance of contextual information for size related ratings impacting women and not just men. The literature clearly outlines several ways in which Black girls and women are discriminated against relating to threat or size, and it seems likely that the race-based size bias may apply to them as well. It is additionally important to study Black girls in this case because there is limited research on Black women and girls in general leading to invisibility in a variety of settings, research, and experiences (Fryberg & Townsend, 2008; Purdie-Vaughns & Eibach, 2008; Sesko & Biernat, 2010). As researchers, we should strive to be more inclusive in our work and expand our findings to groups and intersectional identities that are often overlooked.

Asking participants if they are aware of the race-based size bias outright is another piece of information that could be useful when investigating the race-based size bias. Of course, this would need to be asked at the very end of the study to not influence the rest of the results, but it could be an interesting component to a future study. This question could provide further insight into how and when the bias develops. For example, if child participants report not knowing about this bias, but adolescents and adults both report knowledge of the bias, then it would become clear that knowledge of the stereotype is a prerequisite for developing the race-based size bias. It is likely that even if child participants are not consciously aware of the race-based size bias, they are likely still getting inundated with race-based size messaging from the media. Previous research has shown that elementary school children are watching the news (Drew & Reeves, 1980; York & Scholl, 2015) or getting news-related information through social media (Duvekot et al., 2024). With the exposure to media, there are consistent themes of Black people being shown negatively as threatening and dangerous in the news (Iheme, 2020; Smiley & Fakunle, 2016) or as ‘superhuman’ and ‘angry’ when describing Black athletes (Cramer & Donofrio, 2021; Deeb & Love, 2017). Because children learn from the media as young as age two (Kirkorian et al., 2008; Dinleyici et al., 2016), it is likely that they have certain exposures to stereotypes, but it is not until the knowledge of the stereotype is formed that they’ll be able to apply it to new situations.

A different direction for this project could revolve around investigating the contextual information around the participants themselves, rather than just contextual information’s influence on the stimuli. Looking at parent responses and how those responses may correlate with their children's responses could give some insight into the formation of their child's biases. In doing this study, it would also be beneficial to understand if and how parents are talking to their children about race, and if children are involved in conversations about race at school. When parents and their children identify strongly with each other, racial prejudice in parents often shows in their children as well (Sinclair et al., 2005). If this assumption held true for the race-based size bias, it would be another factor that could be targeted in future interventions. Rather than focusing solely on the children, interventions could be targeting both parent and children to reduce the overall impact of the race-based size bias.

All in all, across several studies and measures, we saw support for a race-based size bias for strength, but not for height and weight, in adolescents and adults. We also provided evidence for the importance of contextual information when children, adolescents, and adults make judgments on the physical size and strength of Black and White individuals. In two studies, we found evidence that children may be sensitive to a race-based size bias, but only when they are given a sensitive measure with no additional context. It should be investigated whether it was the lack of context or the measure itself that began to elicit the race-based size bias in children. While these studies provided a solid, evidence-based jumping off point for the development of the race-based size bias, more research should continue to be conducted furthering this line of research.

References

Aboud, F. E. (2003). The formation of in-group favoritism and out-group prejudice in young children: Are they distinct attitudes? *Developmental Psychology*, 39(1), 48–60. https://doi.org/10.1037/0012-1649.39.1.48

Aboud, F. E., Mendelson, M. J., & Purdy, K. T. (2003). Cross-race peer relations and friendship quality. *International Journal of Behavioral Development*, *27*(2), 165–173. https://doi.org/10.1080/01650250244000164

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, *12*(5), 385–390. https://doi.org/10.1111/1467-9280.00371

Apfelbaum, E. P., Pauker, K., Ambady, N., Sommers, S. R., & Norton, M. I. (2008). Learning (not) to talk about race: When older children underperform in social categorization. *Developmental Psychology*, *44*(5), 1513–1518. https://doi.org/10.1037/a0012835

Baltazar, N. C., Shutts, K., & Kinzler, K. D. (2012). Children show heightened memory for threatening social actions. *Journal of Experimental Child Psychology*, *112*(1), 102–110. https://doi.org/10.1016/j.jecp.2011.11.003

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological science*, 17(1), 53–58. https://doi.org/10.1111/j.1467-9280.2005.01664.x

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*(10), 674–679. https://doi.org/10.1111/j.0956-7976.2004.00739.x

Caldera, A. (2018). Suspendable and expendable, kicking out and throwing away Black girls: An analysis of a school district's policies and practices. Texas Education Review, 7(1), 30-47. http://dx.doi.org/10.26153/tsw/10

Carvalho, K., Peretz-Lange, R., & Muentener, P. (2021). Causal explanations for weight influence children's social preferences: Biological-essentialist explanations reduce, and behavioral explanations promote, preferences for Thin Friends. *Child Development*, 92(2), 682–690. https://doi.org/10.1111/cdev.13519

Castelli, L., Zogmaister, C., & Tomelleri, S. (2009). The transmission of racial attitudes within the family. *Developmental Psychology*, *45*(2), 586–591. https://doi.org/10.1037/a0014619

Chang, D. F., & Demyan, A. L. (2007). Teachers' stereotypes of Asian, black, and white students. *School Psychology Quarterly*, *22*(2), 91–114. https://doi.org/10.1037/1045-3830.22.2.91

Copping, K. E., Kurtz-Costes, B., Rowley, S. J., & Wood, D. (2013). Age and race differences in racial stereotype awareness and endorsement. *Journal of Applied Social Psychology*, *43*(5), 971–980. https://doi.org/10.1111/jasp.12061

Corenblum, B. (2003). What children remember about ingroup and outgroup peers: Effects of stereotypes on children's processing of information about group members. *Journal of*

Experimental Child Psychology, *86*(1), 32–66. https://doi.org/10.1016/s0022-0965(03)00105-x

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(6), 1314–1329. https://doi.org/10.1037/0022-3514.83.6.1314

Cramer, L. M., & Donofrio, A. R. (2021). Threatening Whiteness: “Angry Russell” and the Rhetoricity of Race. Rhetoric Society Quarterly, 51(2), 152–166. https://doi.org/10.1080/02773945.2021.1877802

Cramer, P., & Steinwert, T. (1998). Thin is good, fat is bad: How early does it begin? *Journal of Applied Developmental Psychology*, *19*(3), 429–451. https://doi.org/10.1016/s0193-3973(99)80049-5

Deeb, A., & Love, A. (2017). Media representations of multiracial athletes. Journal of Sport and Social Issues, 42(2), 95–114. https://doi.org/10.1177/0193723517749598

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Dillon, D. G., & Pizzagalli, D. A. (2018). Mechanisms of memory disruption in depression. *Trends in Neurosciences, 41*(3), 137–149. https://doi.org/10.1016/j.tins.2017.12.006

Dinleyici, M., Carman, K. B., Ozturk, E., & Sahin-Dagli, F. (2016). Media use by children, and parents' views on children's media usage. Interactive Journal of Medical Research, 5(2). https://doi.org/10.2196/ijmr.5668

Drew, D. G., & Reeves, B. B. (1980). Children and television news. Journalism Quarterly, 57(1), 45–114. https://doi.org/10.1177/107769908005700107

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(4), 590–598. https://doi.org/10.1037/0022-3514.34.4.590

Duvekot, S., Valgas, C. M., de Haan, Y., & de Jong, W. (2024). How youth define, consume, and evaluate news: Reviewing two decades of research. New Media & Society, 14614448241262809.

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*(6), 876–893. https://doi.org/10.1037/0022-3514.87.6.876

Epstein, R., Blake, J., & Gonzzlez, T. (2017). Girlhood interrupted: The erasure of Black Girls Childhood. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3000695

Evans, A. B., Copping, K. E., Rowley, S. J., & Kurtz-Costes, B. (2011). Academic self-concept in black adolescents: Do race and gender stereotypes matter? *Self and Identity*, *10*(2), 263–277. https://doi.org/10.1080/15298868.2010.485358

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(4). https://doi.org/10.1371/journal.pone.0032751

Fessler, D., Holbrook, C., Tiokhin, L., & Snyder, J. (2014). Sizing up Helen. *Journal of Evolutionary Psychology*, *12*(2–4), 67–80. https://doi.org/10.1556/jep-d-14-00009

Freiburger, E., Sim, M., Halberstadt, A. G., & Hugenberg, K. (2023). A race-based size bias for black adolescent boys: Size, innocence, and threat. *Personality and Social Psychology Bulletin*, 014616722311679. https://doi.org/10.1177/01461672231167978

Fryberg, S. A., & Townsend, S. S. (2008). The psychology of invisibility. *Commemorating Brown: The Social Psychology of Racism and Discrimination.*, 173–193. https://doi.org/10.1037/11681-010

Gaither, S. E., Chen, J. M., Pauker, K., & Sommers, S. R. (2018). At face value: Psychological outcomes differ for real vs. computer-generated multiracial faces. *The Journal of Social Psychology*, *159*(5), 592–610. https://doi.org/10.1080/00224545.2018.1538929

Gedeon, C., Badea, C., & Esseily, R. (2021). Racial categorization and intergroup relations in children: The role of social status and numerical group size. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.719121

Graves, S. L., & Wang, Y. (2022). It’s not that they are big, it’s just that they are black: The impact of body mass index, school belonging, and self esteem on black boys’ school suspension. *School Psychology Review*, *52*(3), 279–291. https://doi.org/10.1080/2372966x.2022.2072693

Gross, K. N. (2015). African American women, mass incarceration, and the politics of protection. *Journal of American History*, *102*(1), 25–33. https://doi.org/10.1093/jahist/jav226

Haddock, G., & Zanna, M. P. (1998). On the use of open-ended measures to assess attitudinal components. *British Journal of Social Psychology*, *37*(2), 129–149. https://doi.org/10.1111/j.2044-8309.1998.tb01161.x

Hamlin, J.K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science, 13*(6), 923–929. https://doi.org/10.1111/j.1467-7687.2010.00951.x

Harries, B. (2014). We need to talk about race. *Sociology*, *48*(6), 1107–1122. https://doi.org/10.1177/0038038514521714

Harrison, S., Rowlinson, M., & Hill, A. J. (2016). "No fat friend of mine": Young Children's responses to overweight and disability. *Body Image*, *18*, 65–73. https://doi.org/10.1016/j.bodyim.2016.05.002

Hester, N., & Gray, K. (2018). For Black men, being tall increases threat stereotyping and police stops. PNAS Proceedings of the National Academy of Sciences of the United States of America, 115(11), 2711–2715. https://doi.org/10.1073/pnas.1714454115

Ho, S. M. Y., Cheng, J., Dai, D. W., Tam, T., & Hui, O. (2018). The effect of positive and negative memory bias on anxiety and depression symptoms among adolescents. *Journal of Clinical Psychology*, *74*(9), 1509–1525. https://doi.org/10.1002/jclp.22597

Holbrook, C., Fessler, D. M. T., & Navarrete, C. D. (2016). Looming large in others' eyes: Racial stereotypes illuminate dual adaptations for representing threat versus prestige as physical size. *Evolution and Human Behavior*, *37*(1), 67–78. https://doi.org/10.1016/j.evolhumbehav.2015.08.004

Iheme, W. C. (2020). Blackness in America and the presumption of innocence: How the American police and mass media poisoned everything. Hum. Rts., 15, 153.

Johnson, D. J., & Wilson, J. P. (2019). Racial Bias in Perceptions of Size and Strength: The Impact of Stereotypes and Group Differences. Psychological Science, 30(4), 553–562. https://doi.org/10.1177/0956797619827529

Keightley, M. L., Chiew, K. S., Anderson, J. A., & Grady, C. L. (2011). Neural correlates of recognition memory for emotional faces and scenes. *Social Cognitive and Affective Neuroscience*, *6*(1), 24–37. https://doi.org/10.1093/scan/nsq003

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language*, *54*(1), 99–112. https://doi.org/10.1016/j.jml.2005.05.005

Kirkorian, H. L., Wartella, E. A., & Anderson, D. R. (2008). Media and young children's learning. *The Future of Children*, *18*(1), 39–61. https://doi.org/10.1353/foc.0.0002

Kleider, H. M., Pezdek, K., Goldinger, S. D., & Kirk, A. (2007). Schema-driven source misattribution errors: Remembering the expected from a witnessed event. *Applied Cognitive Psychology*, *22*(1), 1–20. https://doi.org/10.1002/acp.1361

Kleider, H. M., Cavrak, S. E., & Knuycky, L. R. (2012). Looking like a criminal: Stereotypical black facial features promote face source memory error. *Memory & Cognition*, *40*(8), 1200–1213. https://doi.org/10.3758/s13421-012-0229-x

Lawrence, V. W. (1991). Effect of Socially Ambiguous Information on White and Black Children's Behavioral and Trait Perceptions. Merrill-Palmer Quarterly, 37(4), 619–630. Retrieved from: http://www.jstor.org/stable/23087442

Li, K. K. (2012). Asymmetric memory recall of positive and negative events in social interactions. *Experimental Economics*, *16*(3), 248–262. https://doi.org/10.1007/s10683-012-9325-9

Liu, W. M., Liu, R. Z., Garrison, Y. L., Kim, J. Y. C., Chan, L., Ho, Y. C. S., & Yeung, C. W. (2019). Racial trauma, microaggressions, and becoming racially innocuous: The role of acculturation and White supremacist ideology. American Psychologist, 74(1), 143–155. https://doi.org/10.1037/amp0000368

LoBue, V. (2009). More than just another face in the crowd: Superior detection of threatening facial expressions in children and adults. *Developmental Science, 12*(2), 305–313. https://doi.org/10.1111/j.1467-7687.2008.00767.x

Lourenco, S. F., Bonny, J. W., & Schwartz, B. L. (2016). Children and adults use physical size and numerical alliances in third-party judgments of dominance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.02050

McGlothlin, H., Killen, M., & Edmonds, C. (2005). European-american children's Intergroup Attitudes About Peer relationships. British Journal of Developmental Psychology, 23(2), 227–250. https://doi.org/10.1348/026151005x26101

McGlothlin, H., & Killen, M. (2006). Intergroup attitudes of European American children attending ethnically homogeneous schools. Child Development, 77(5), 1375–1386. https://doi.org/10.1111/j.1467-8624.2006.00941.x

McGlothlin, H., & Killen, M. (2010). How social experience is related to children's intergroup attitudes. European Journal of Social Psychology, 40(4), 625–634. https://doi.org/10.1002/ejsp.733

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development*, *74*(2), 498–515. https://doi.org/10.1111/1467-8624.7402012

McKown, C., & Strambler, M. J. (2009). Developmental antecedents and social and academic consequences of stereotype-consciousness in middle childhood. *Child Development*, *80*(6), 1643–1659. https://doi.org/10.1111/j.1467-8624.2009.01359.x

Metzger, I. W., Anderson, R. E., Are, F., & Ritchwood, T. (2021). Healing Interpersonal and Racial Trauma: Integrating Racial Socialization Into Trauma-Focused Cognitive Behavioral Therapy for African American Youth. Child Maltreatment, 26(1), 17–27. https://doi.org/10.1177/1077559520921457

Milner, A. N., George, B. J., & Allison, D. B. (2016). Black and Hispanic men perceived to be large are at increased risk for police frisk, search, and Force. *PLOS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0147158

Musher-Eizenman, D. R., Holub, S. C., Miller, A. B., Goldstein, S. E., & Edwards-Leeper, L. (2004). Body size stigmatization in preschool children: The role of control attributions. *Journal of Pediatric Psychology*, *29*(8), 613–620. https://doi.org/10.1093/jpepsy/jsh063

Navon, M., Reggev, N., & Moran, T. (2025). Top-down racial biases in size perception: A registered replication and extension of. *Journal of Experimental Social Psychology*, *116*, 104690. https://doi.org/10.1016/j.jesp.2024.104690

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? the experience and process of recognizing feelings past. *Journal of Experimental Psychology: General*, *129*(2), 242–261. https://doi.org/10.1037//0096-3445.129.2.242

Okeke, N. A., Howard, L. C., Kurtz-Costes, B., & Rowley, S. J. (2009). Academic race stereotypes, academic self-concept, and racial centrality in African American youth. *Journal of Black Psychology*, *35*(3), 366–387. https://doi.org/10.1177/0095798409333615

Packer, M. J., & Moreno-Dulcey, F. A. (2022). Theory of puppets?: A critique of the use of puppets as stimulus materials in psychological research with young children. *Cognitive Development*, *61*, 101146. https://doi.org/10.1016/j.cogdev.2021.101146

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(2), 181–192. https://doi.org/10.1037/0022-3514.81.2.181

Payne, B. K., & Correll, J. (2020). Race, weapons, and the perception of threat. *Advances in Experimental Social Psychology*, 1–50. https://doi.org/10.1016/bs.aesp.2020.04.001

Pelleboer-Gunnink, H. A., van Weeghel, J., & Embregts, P. J. (2019). Public stigmatisation of people with intellectual disabilities: A mixed-method population survey into stereotypes and their relationship with familiarity and discrimination. *Disability and Rehabilitation*, *43*(4), 489–497. https://doi.org/10.1080/09638288.2019.1630678

Penny, H., & Haddock, G. (2007). Children’s stereotypes of overweight children. *British Journal of Developmental Psychology*, *25*(3), 409–418. https://doi.org/10.1348/026151006x158807

Perszyk, D. R., Lei, R. F., Bodenhausen, G. V., Richeson, J. A., & Waxman, S. R. (2019). Bias at the intersection of race and gender: Evidence from preschool-aged children. Developmental science, 22(3), e12788. https://doi.org/10.1111/desc.12788

Porter, S., Taylor, K., & ten Brinke, L. (2008). Memory for media: Investigation of false memories for negatively and positively charged public events. *Memory*, *16*(6), 658–666. https://doi.org/10.1080/09658210802154626

Proverbio, A. M., La Mastra, F., & Zani, A. (2016). How negative social bias affects memory for faces: An electrical neuroimaging study. *PLOS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0162671

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, *59*(5–6), 377–391. https://doi.org/10.1007/s11199-008-9424-4

Rakoczy, H. (2022). Puppet Studies present clear and distinct windows into the child’s mind. *Cognitive Development*, *61*, 101147. https://doi.org/10.1016/j.cogdev.2021.101147

Raman, L. (2014). Children’s and adults’ understanding of the impact of nutrition on biological and Psychological Processes. *British Journal of Developmental Psychology*, *32*(1), 78–93. https://doi.org/10.1111/bjdp.12024

Raman, L. (2017). Do children think that inheritance determines height and weight? *Infant and Child Development*, *27*(1). https://doi.org/10.1002/icd.2041

Revencu, B., & Csibra, G. (2024). Puppets as symbols in early development: From whether to how in the theory of puppets debate. *Cognitive Development*, *71*, 101487. https://doi.org/10.1016/j.cogdev.2024.101487

Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirming events: Memory processes and the maintenance of social stereotypes. *Journal of Experimental Social Psychology*, *15*(4), 343–355. https://doi.org/10.1016/0022-1031(79)90043-x

Rowley, S. J., Kurtz-Costes, B., Mistry, R., & Feagans, L. (2007). Social status as a predictor of race and gender stereotypes in late childhood and early adolescence. *Social Development*, *16*(1), 150–168. https://doi.org/10.1111/j.1467-9507.2007.00376.x

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children’s implicit and explicit intergroup attitudes. Child Development, 76(2), 451–466. https://doi.org/10.1111/j.1467-8624.2005.00856.x

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(4), 590–598. https://doi.org/10.1037//0022-3514.39.4.590

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black Women. *Journal of Experimental Social Psychology*, *46*(2), 356–360. https://doi.org/10.1016/j.jesp.2009.10.016

Sierksma, J., Brey, E., & Shutts, K. (2022). Racial Stereotype Application in 4-to-8-Year-Old White American Children: Emergence and Specificity. Journal of cognition and development : official journal of the Cognitive Development Society, 23(5), 660–685. https://doi.org/10.1080/15248372.2022.2090945

Sim, M., Almaraz, S. M., & Hugenberg, K. (2022). Stereotyping at the intersection of race and weight: Diluted threat stereotyping of obese black men. *Journal of Experimental Social Psychology*, *99*, 104274. https://doi.org/10.1016/j.jesp.2021.104274

Sinclair, S., Dunn, E., & Lowery, B. (2005). The relationship between parental racial attitudes and children's implicit prejudice. *Journal of Experimental Social Psychology*, *41*(3), 283–289. https://doi.org/10.1016/j.jesp.2004.06.003

Smiley, C., & Fakunle, D. (2016). From "Brute" to "Thug:" the demonization and criminalization of unarmed black male victims in America. Journal of Human Behavior in the Social Environment, 26(3–4), 350–366. https://doi.org/10.1080/10911359.2015.1129256

Smith-Evans, L., George, J., Goss Graves, F., Kaufmann, L. S., & Frohlich, L. (2014). Unlocking opportunity for African American girls: A call for action for educational equity. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4705316

Terrizzi, B. F., Brey, E., Shutts, K., & Beier, J. S. (2019). Children's developing judgments about the physical manifestations of power. Developmental psychology, 55(4), 793–808. https://doi.org/10.1037/dev0000657

Todd, A. R., Johnson, D. J., Lassetter, B., Neel, R., Simpson, A. J., & Cesario, J. (2021). Category salience and racial bias in weapon identification: A diffusion modeling approach. Journal of Personality and Social Psychology, 120(3), 672–693. https://doi.org/10.1037/pspi0000279

Waters, T. E., Bohanek, J. G., Marin, K., & Fivush, R. (2013). Null's the word: A comparison of memory quality for intensely negative and positive events. *Memory*, *21*(6), 633–645. https://doi.org/10.1080/09658211.2012.745877

Waytz, A., Hoffman, K. M., & Trawalter, S. (2015). A Superhumanization Bias in Whites' Perceptions of Blacks. Social Psychological and Personality Science, 6(3), 352–359. https://doi.org/10.1177/1948550614553642

Wellman, H. M., & Yu, C.-L. (2022). Theory of puppets or theory of mind? misunderstanding how children construe puppets in psychological research: A commentary on packer and Moreno-Dulcey (2022). *Cognitive Development*, *63*, 101218. https://doi.org/10.1016/j.cogdev.2022.101218

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. Journal of Personality and Social Psychology, 113(1), 59–80. https://doi.org/10.1037/pspi0000092

Wolf, K. C., & Kupchik, A. (2016). School suspensions and adverse experiences in adulthood. *Justice Quarterly*, *34*(3), 407–430. https://doi.org/10.1080/07418825.2016.1168475

York, C., & Scholl, R. M. (2015). Youth antecedents to news media consumption. Journalism & Mass Communication Quarterly, 92(3), 681–699. https://doi.org/10.1177/1077699015588191

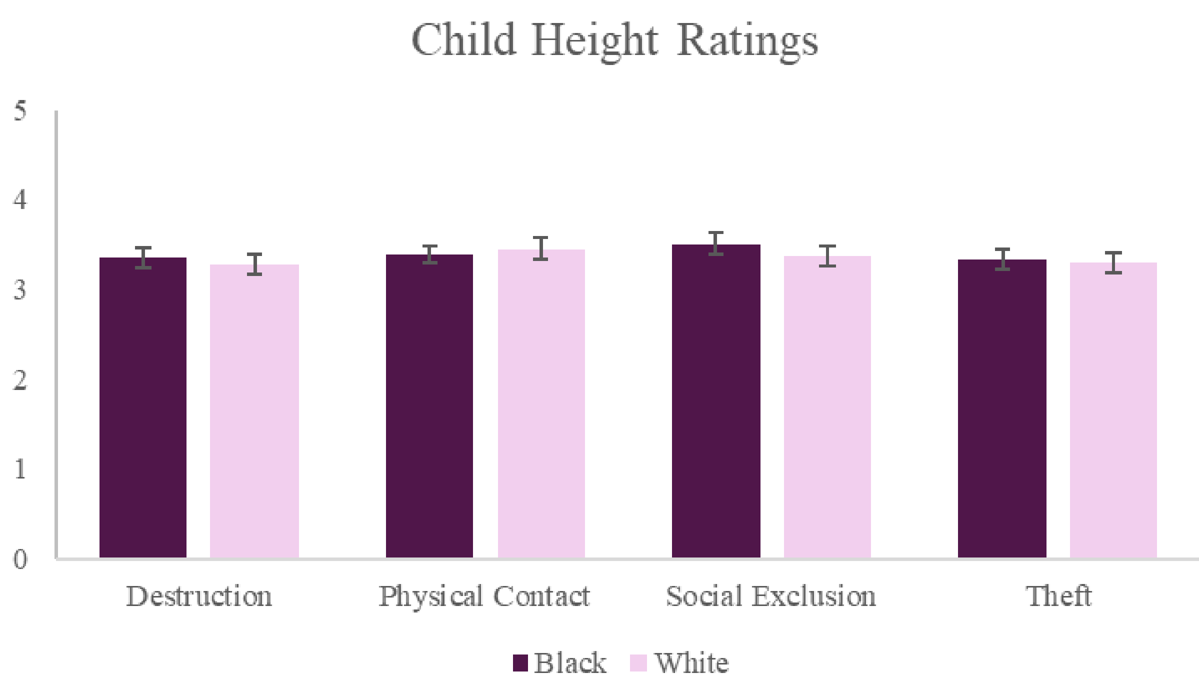
Supplemental Information
Study 1 Analyses

Height ratings separately by age group

The child participants model yielded no significant effects of Character Race, *F*(1, 105)= 0.294, *p*= .580, Event Type, *F*(3, 124)= .726, *p*= .538, or Age Cohort, *F*(1, 100)= 1.12, *p*= .33, on participants' weight ratings. There were no significant interactions, so they were removed from the model. These findings are displayed in Figure 8.

Figure 21

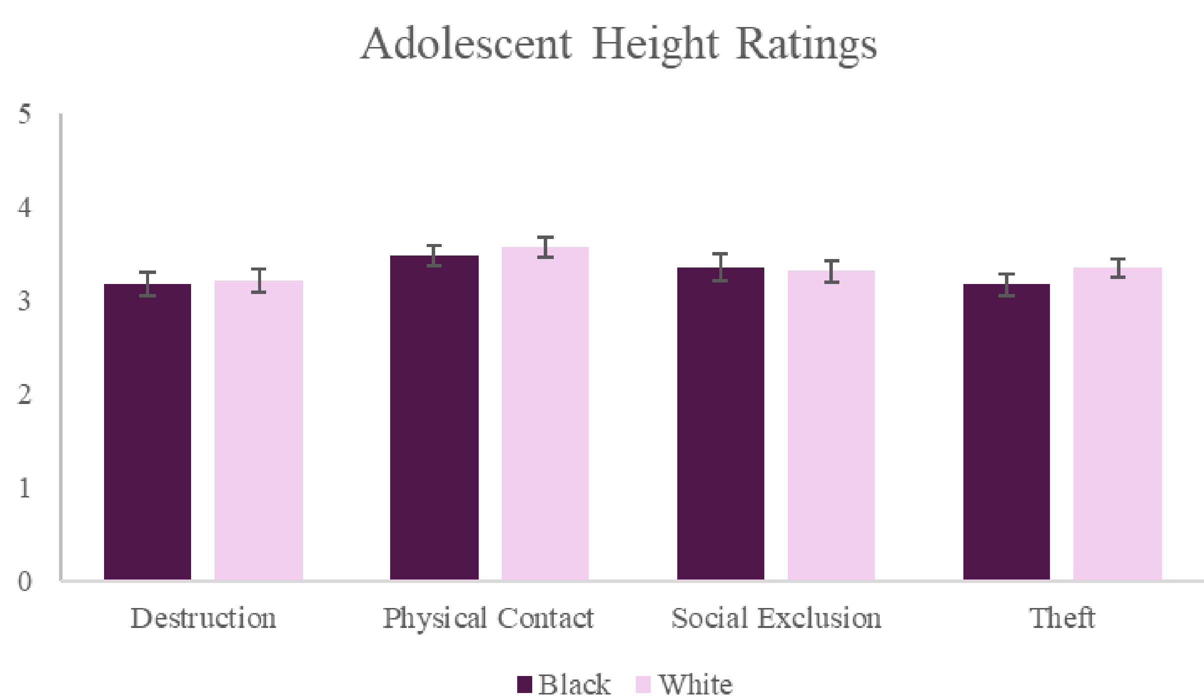
Child (6-11 years) height ratings in Study 1



The adolescent participant model resulted in significant effect of Event Type, *F*(3, 87.7)= 2.91, *p*= .039 where participants' rated characters as taller when they participated in physical contact events (*M*= 3.53, *SD*= .896), compared to destruction (*M*=3.19, *SD*= .97; *t*(66.9)= 2.59, *p*= .012) and theft (*M*= 3.26, *SD*= .89; *t*(66.8)= 2.45, *p*= .017). Similar to the child model, the adolescent model yielded no significant effects of either Character Race *F*(1, 225.3)= .783, *p*= .377 or Age Cohort, *F*(1, 66)= 0.617, *p*= .435 on participants' height ratings. There were no significant two- or three-way interactions between any factors in this model, so they were removed. These findings are displayed in Figure 9.

Figure 22

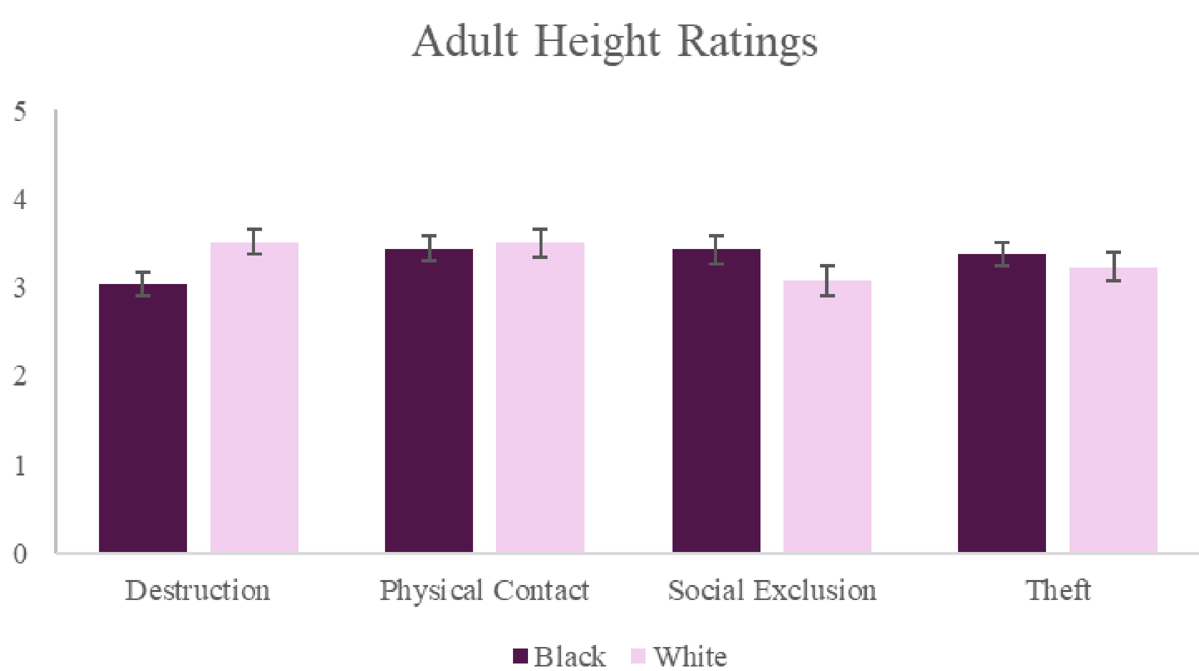
Adolescent (12-17 years) height ratings for Study 1



For the adult participants, Character Race, *F*(1, 62.9)= .03, *p*= .87 and Event Type, *F*(3, 59.7)= .91, *p*= .44, did not significantly predict participants' strength ratings. However, there was a significant interaction between Event Type and Character Race, *F*(3, 188.2)= 3.55, *p*= .015 where Black characters who engaged in destruction (*M*= 3.04, *SD*= .922) were rated as *shorter* than White characters who engaged in destruction (*M*= 3.51, *SD*= .96, *SE* = .137), *t*(183.3)= -2.51, *p*= .013, which goes against our predictions. There were no other significant differences between Character Race and Event Type. These findings are displayed in Figure 10.

Figure 23

Adult height ratings for Study 1

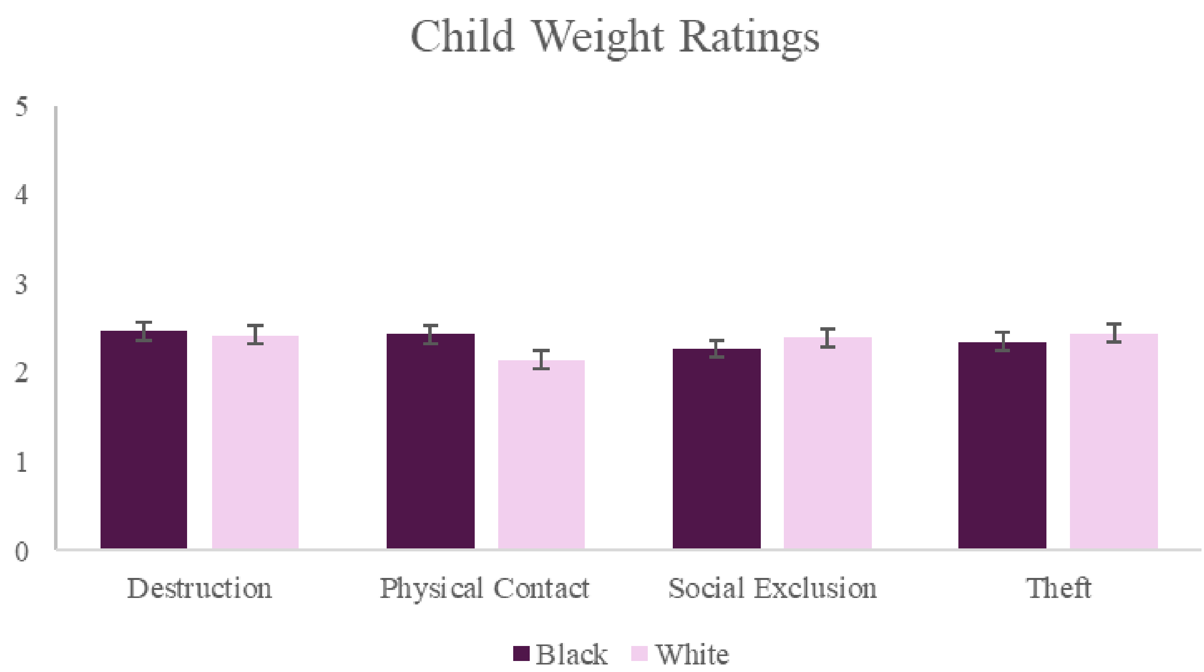


Weight Ratings

The child participant model yielded no significant effects of Character Race, *F*(1, 94.9)= .109, *p*= .74, Event Type, *F*(3, 134)= 1.21, *p*= .31, or Age Cohort *F*(2, 100.1)= 1.66, *p*= .195 on participants' weight ratings. There was a significant interaction between Character Race and Event Type, *F*(3, 394.5)= 3.13, *p*= .025. Post-hoc tests revealed Black characters who engaged in destruction events (*M*= 2.47, *SD*= 1.07) were perceived as heavier than White characters who engaged in destruction events (*M*= 2.14, *SD*= .98; *t*(379)= 2.71, *p*=.007, but there were no other significant differences across Character Race and Event Type. These findings are displayed in Figure 11.

Figure 24

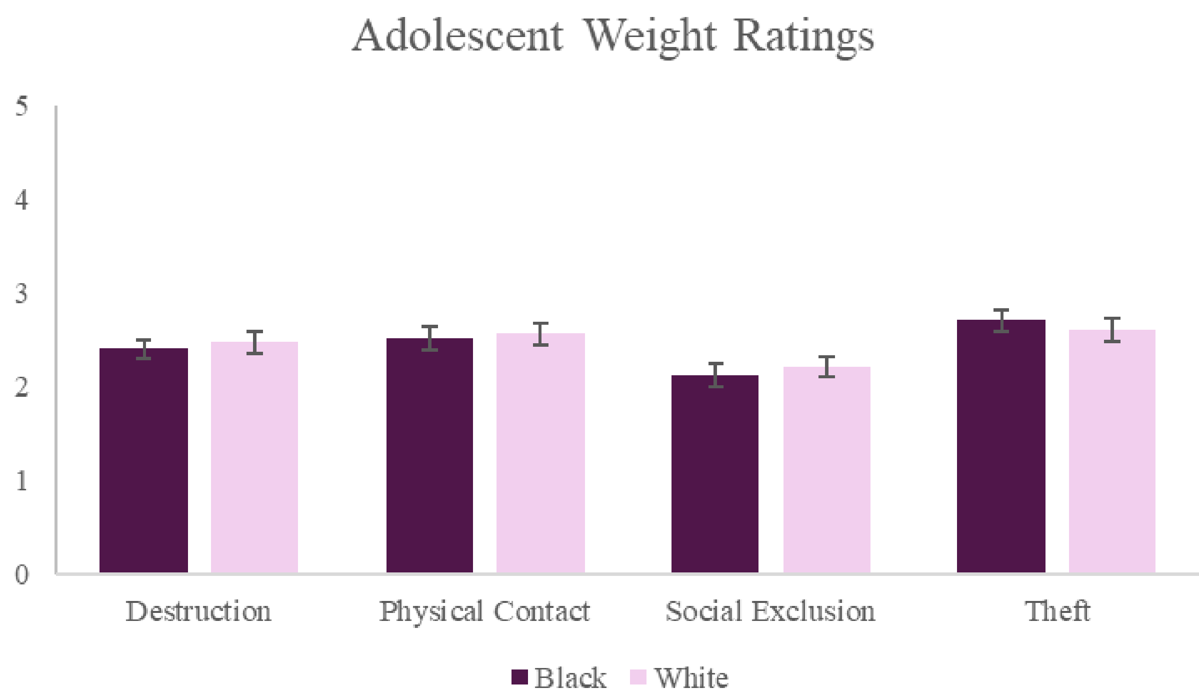
Child (6-11 years) weight ratings for Study 1



A follow-up model for the adolescent participants yielded a significant effect of Event Type, *F*(3, 130.8)= 8.75, *p*<.001. Here, we found participants' rated characters as thinner when they performed social exclusion actions (*M*= 2.16, *SD*= .932) compared to all other event types. Participants who engaged in theft were also perceived as heavier than those who engaged in destruction, *t*(66.8)= 2.20, *p*= .031. However, this model yielded no significant effects of either Age Cohort, *F*(1, 66)= 0.3, *p*=0.586 or Character Race *F*(1, 169.1)= 0.274, *p*= .601 on participants' weight ratings. There were no significant interactions between any factors in this model, so they were removed. These findings are displayed in Figure 12.

Figure 25

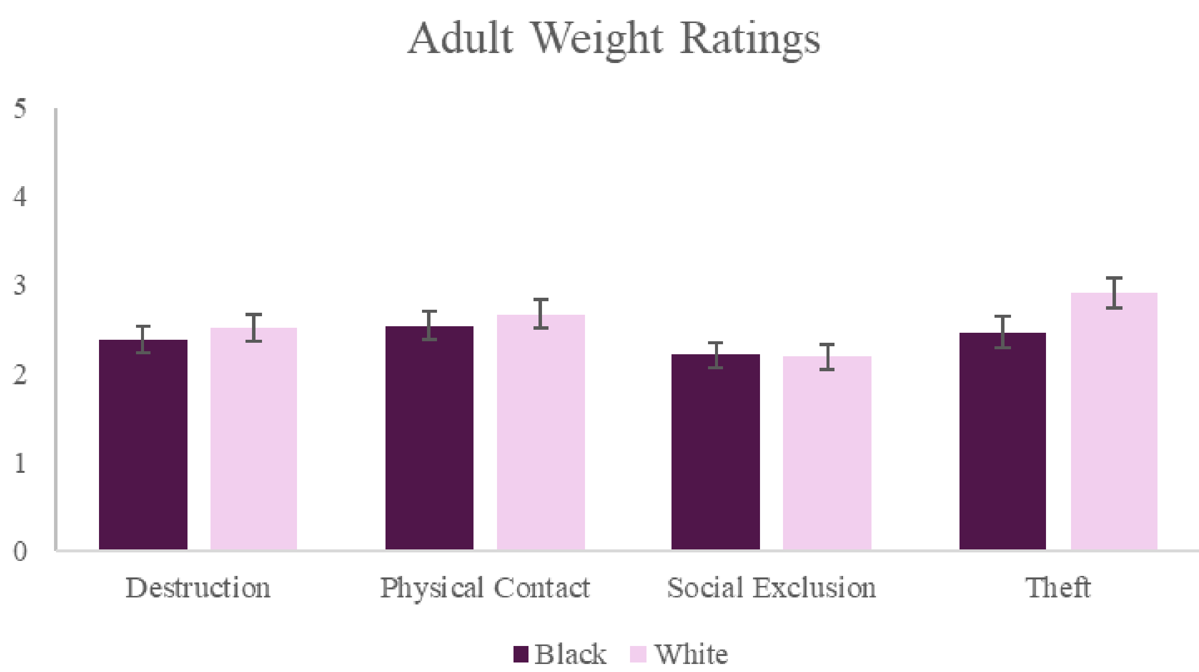
Adolescent (12-17 years) weight ratings for Study 1



The adult participants had a significant effect of Event Type, *F*(3, 67.9)= 4.53, *p*= .006, on weight ratings, but Character Race, *F*(1, 87.3)= 2.6 , *p*= .11 did not. Characters who socially excluded (*M*= 2.2, *SD*= .941) were rated as significantly *thinner* than those who used physical contact (*M*= 2.6, *SD*= 1.01; *t*(47.7)= 2.65, *p*= .011) and engaged in theft (*M*= 2.69, *SD*= 1.19; *t*(47.1)= -3.35, *p*= .002). There were no significant two-way interactions between any factors in this model, so they were removed. These findings are displayed in Figure 13.

Figure 26

Adult weight ratings for Study 1



Study 2 Analyses

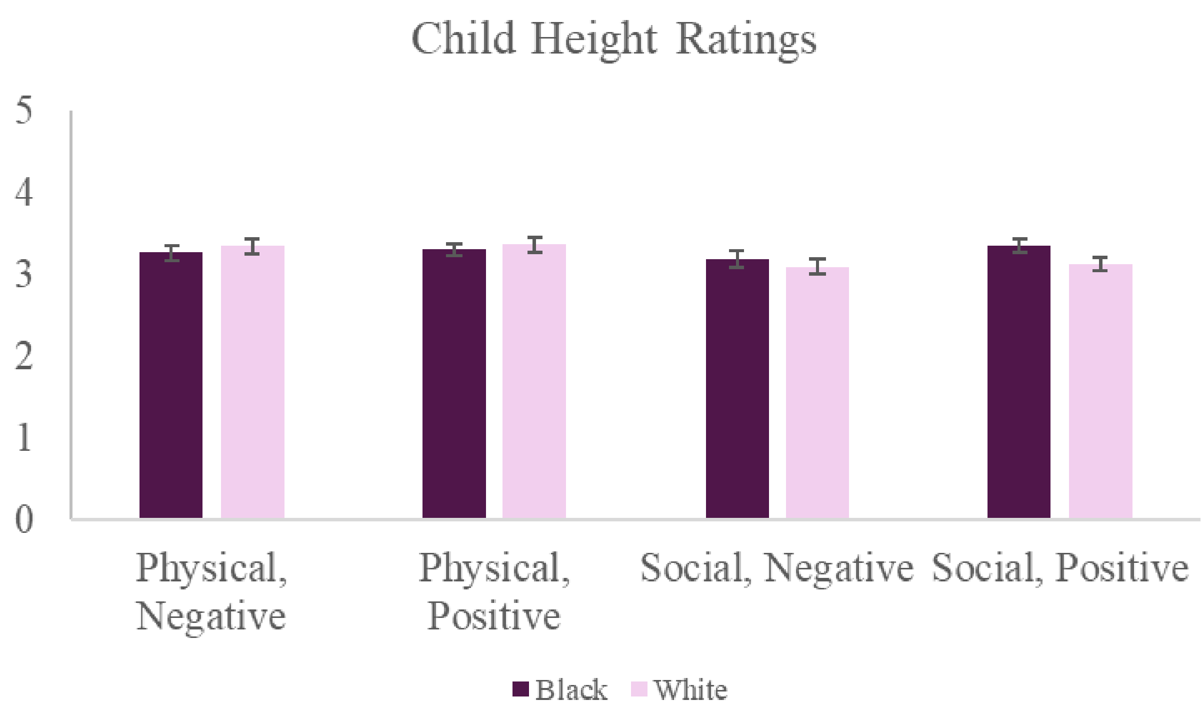
Physical size and strength ratings

Child height ratings

A follow-up model for child participants yielded a significant effect of Event Type, $F(1, 462.9)= 4.6$, $p= .032$ where characters who engaged in physical events were perceived as significantly taller ($M= 3.31$, $SD= 1.13$) than those who engaged in social events ($M= 3.19$, $SD= 1.13$). The model also yielded a significant interaction between Character Race and Event Type, $F(1, 1{,}097.2)= 4.6$, $p= .033$. Post-hoc tests revealed that White characters who performed physical events ($M= 3.35$, $SD= 1.17$) were significantly taller than White characters who performed social events ($M= 3.11$, $SD= 1.09$; $t(81.8)= 2.14$, $p= .035$), but there were no significant differences within the Black characters or between the Black and White characters. No other interactions were significant.

Figure 28

Child (6-11 years) height ratings for Study 2



Adolescent height ratings

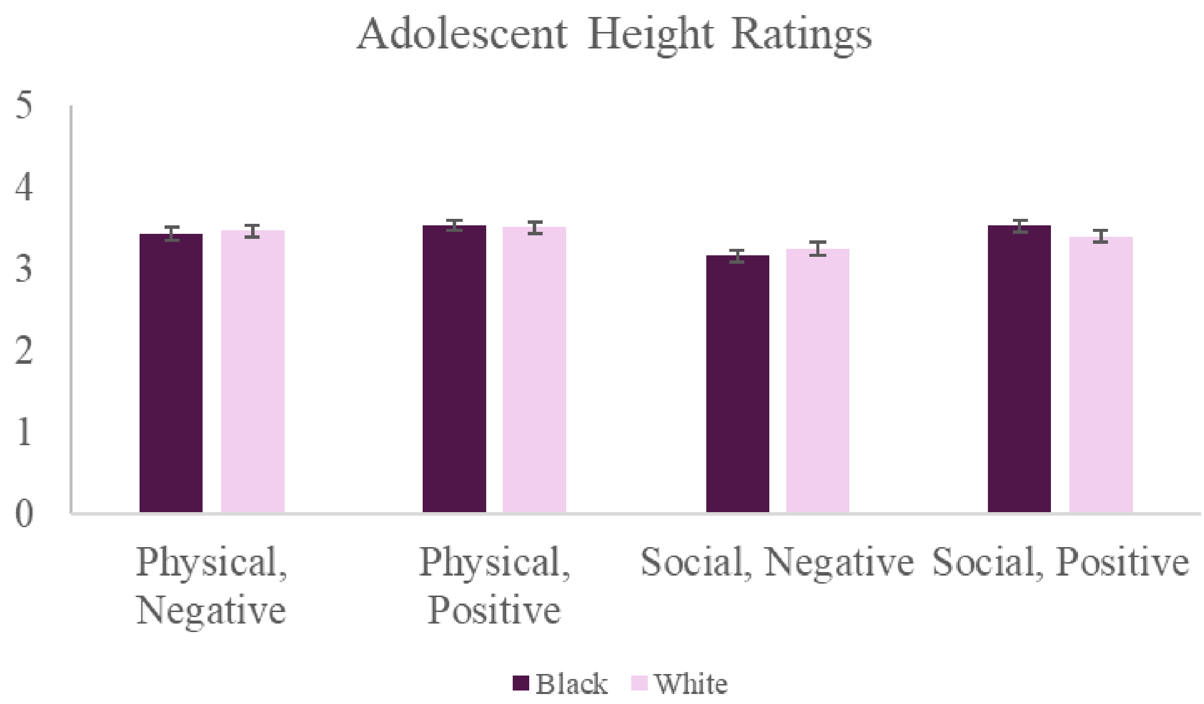
The follow-up model for adolescent participants yielded a significant main effect of Event Type $F(1, 117.8) = 10.07$, $p = .002$, where characters who engaged in physical events were perceived as taller ($M = 3.49$, $SD = .91$) than those who engaged in social events ($M = 3.33$, $SD = .91$); and Valence, $F(1, 77.9) = 7.68$, $p = .007$, where characters who engaged in positive events ($M = 3.49$, $SD = .85$) were perceived as taller than those who engaged in negative events ($M = 3.33$, $SD = .96$). There also was a significant main effect of Age Cohort, $F(1, 775) = 11.0$, $p = .001$, where younger participants perceived characters as taller ($M = 3.55$, $SD = .93$) than older participants ($M = 3.26$, $SD = .87$).

These effects were qualified by significant interactions between Event Type and Valence, *F*(1, 1005.4)= 2.2, *p*= .039, **Post-hoc tests revealed that valence had a significant effect on social, but not physical events.** Specifically, characters who engaged in social - negative events (*M*= 3.2, *SD*= .92) were significantly shorter than those who engaged in physical - negative events (*M*= 3.45, *SD*= .98; *t*(264)= 3.71, *p*< .001), physical - positive events (*M*= 3.52, *SD*= .83; *t*(76.9)= 3.84, *p*< .001), and social - positive events (*M*= 3.46, *SD*= .87; *t*(186.2)=3.45, *p*< .001), but there were no other significant interactions

Additionally, there was a significant interaction between Character Race and Age Cohort, *F*(1, 215.5)= 4.7, *p*= .031. Younger adolescents perceive Black (*M*= 3.61, *SD*= 0.9; *t*(76.9)= 3.72, *p*< .001) and White (*M*= 3.5, *SD*= 0.95; *t*(116.1)= 2.81, *p*= .006) characters as taller than older adolescents perceive Black characters (*M*= 3.22, *SD*= 0.88). Younger adolescents also perceive Black characters as taller than older adolescents perceive White characters (*M*= 3.31, *SD*= 0.88; *t*(77)= 2.99, *p*= .003). However, within younger adolescents and older adolescents separately, there is no difference in their ratings of Black and White characters. Finally, there's no difference in how younger and older adolescents perceive White characters' height.

Figure 29

Adolescent height ratings for Study 2

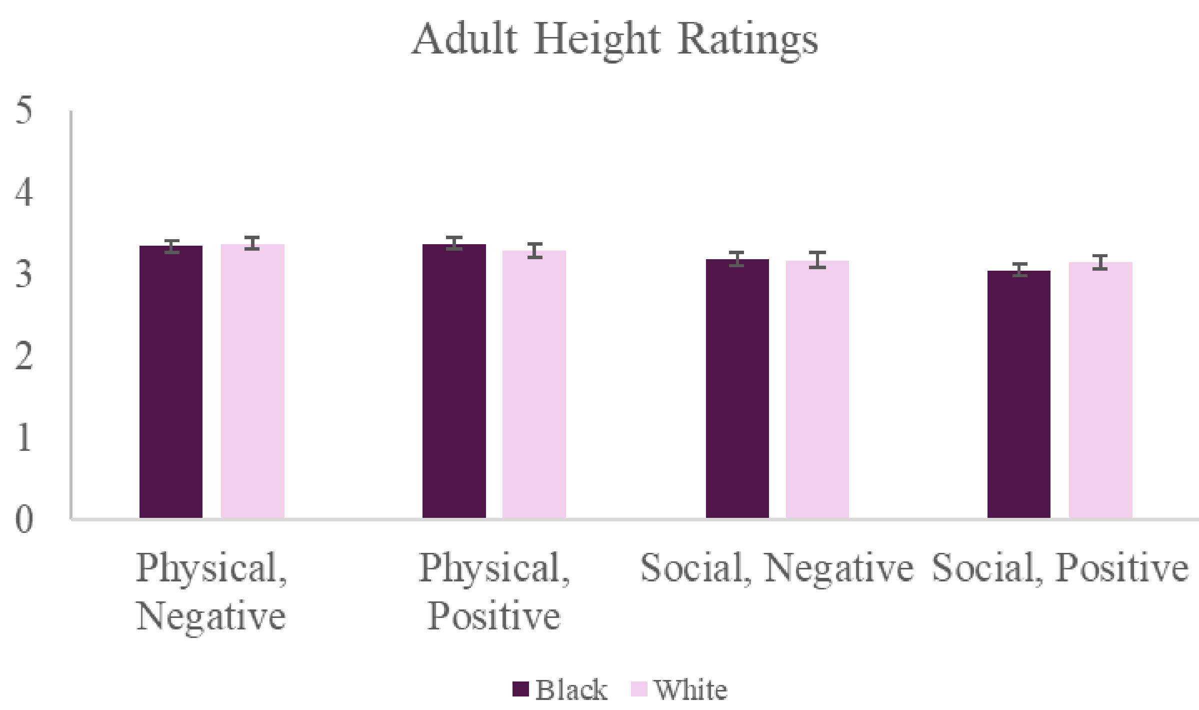


Adult height ratings

The follow-up model for adult participants yielded a significant main effect of Event Type, $F(1, 290.7) = 15.19$, $p < .001$, where characters who engaged in physical events were perceived as taller ($M = 3.35$, $SD = .1$) than those who engaged in social events ($M = 3.14$, $SD = 1.01$). There were no other significant main effects or any significant interactions.

Figure 30

Adult height ratings for Study 2

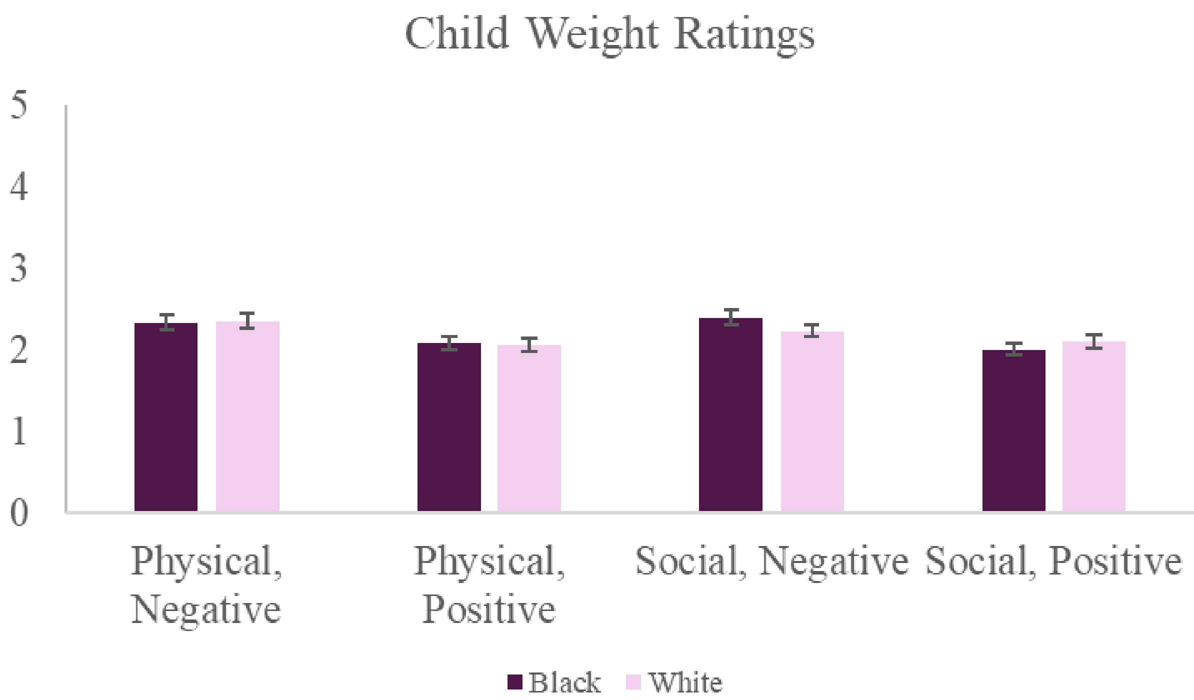


Child weight ratings

The follow-up child model yielded a significance of Valence, $F(1, 1181.5) = 28.88$, $p < .001$ where characters who engaged in negative events ($M = 2.33$, $SD = 1.08$) were perceived as significantly heavier than those who engaged in positive events ($M = 2.07$, $SD = .98$). No other main effects or interactions were significant.

Figure 31

Child (6-11 years) weight ratings for Study 2

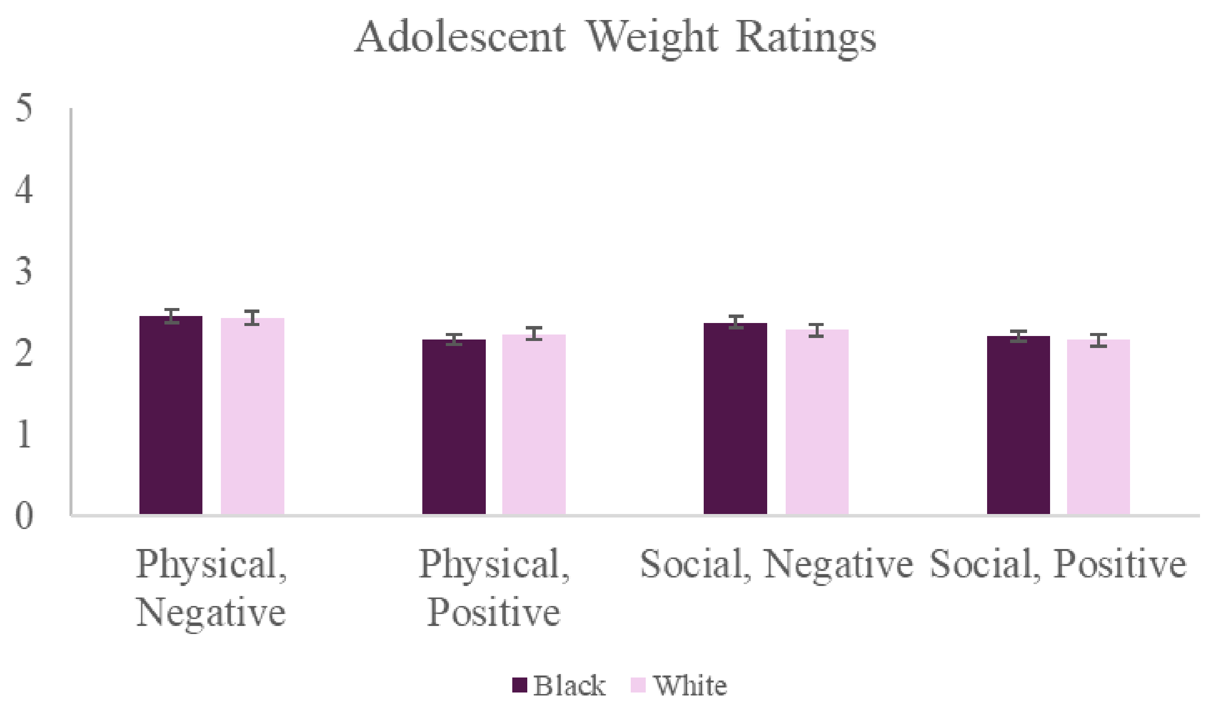


Adolescent weight ratings

The follow-up adolescent model yielded a significance of Valence, $F(1, 1085.9)= 23.2$, $p< .001$ where characters who engaged in negative events ($M= 2.39$, $SD= .92$) were perceived as significantly heavier than those who engaged in positive events ($M= 2.20$, $SD= .85$). No other main effects or interactions were significant.

Figure 32

Adolescent weight ratings for Study 2



Adult weight ratings

The adult model yielded a significance of Valence, *F*(1, 1078.4)= 56.2, $p < .001$ where characters who engaged in negative events (*M*= 2.24, *SD*= 1.03) were perceived as significantly heavier than those who engaged in positive events (*M*= 1.9, *SD*= .87). No other main effects or interactions were significant.

Figure 33

Adult weight ratings for Study 2

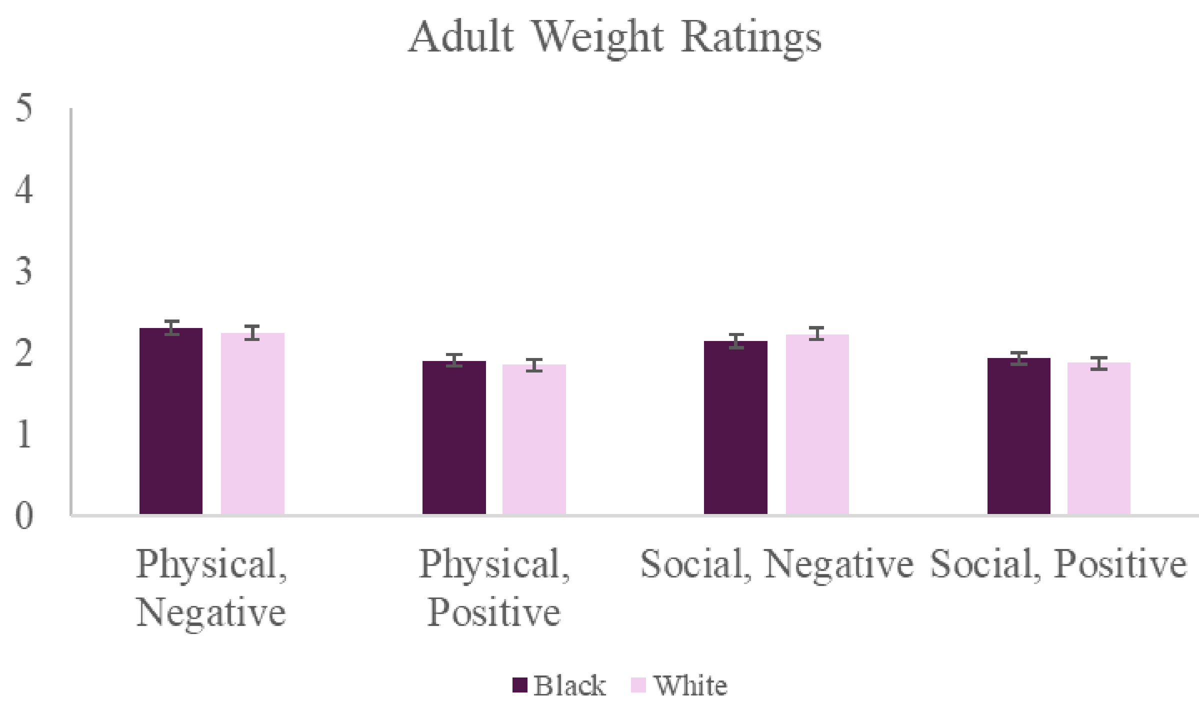


Table 13

Reliability percentages between coders for open-response questions for Study 2

Category	**Strength**	**Height**	**Weight**
Guessing	98.6%	98.6%	98.6%
Intuition	98.6%	97.2%	100%
Seeing their face	83.1%	83.3%	93.1%
Seeing their whole body	85.5%	80%	91.2%
Behavior	90.1%	93.1%	100%
Referencing another category	100%	94.4%	95.8%
Referencing the age of the character	100%	100%	100%
Referencing the race of the character	98.6%	100%	100%
Other	91.5%	90.3%	90.3%

Study 3 Analyses

Table 14

Reliability percentages between coders for open-response questions for Study 3

Category	Strength	Height	Weight
Guessing	100%	100%	100%
Intuition	97.2%	94.4%	94.4%
Seeing their face	97.2%	86.1%	91.67%
Seeing their whole body	88.9%	97.2%	97.2%
Behavior	88.9%	97.2%	94.4%
Referencing another category	97.2%	100%	97.2%
Referencing the age of the character	100%	100%	100%
Referencing the race of the character	97.2%	94.4%	97.2%
Other	100%	100%	100%