



Racial Stereotype Application in 4-to-8-Year-Old White American Children: Emergence and Specificity

Jellie Sierksma, Elizabeth Brey & Kristin Shutts

To cite this article: Jellie Sierksma, Elizabeth Brey & Kristin Shutts (2022) Racial Stereotype Application in 4-to-8-Year-Old White American Children: Emergence and Specificity, Journal of Cognition and Development, 23:5, 660-685, DOI: [10.1080/15248372.2022.2090945](https://doi.org/10.1080/15248372.2022.2090945)

To link to this article: <https://doi.org/10.1080/15248372.2022.2090945>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 18 Jul 2022.



[Submit your article to this journal](#)



Article views: 538



[View related articles](#)



[View Crossmark data](#)



This article has been awarded the Centre for Open Science 'Open Data' badge.






This article has been awarded the Centre for Open Science 'Open Materials' badge.



This article has been awarded the Centre for Open Science 'Preregistered' badge.

Racial Stereotype Application in 4-to-8-Year-Old White American Children: Emergence and Specificity

Jellie Sierksma ^{a,b}, Elizabeth Brey ^c, and Kristin Shutts ^b

^aUtrecht University, Netherlands; ^bUniversity of Wisconsin – Madison; ^cUniversity of Hawaii – Manoa

ABSTRACT

Young children’s racial stereotyping is poorly understood even though stereotyping can influence individuals’ attitudes and behavior toward others. Here we present two preregistered studies (Total N = 257) examining White American children’s (4–8 years) application of six stereotypes (about being American, smart, wealthy, sporty, honest, and nice) when considering Asian, Black, and White children. We observed clear and consistent evidence for one cultural stereotype across the two studies: participants indicated that Asian and Black children were less American than White children. In a measure of racial attitudes, participants also preferred White children over Black and Asian children. Taken together, this research suggests that, in contrast to findings from previous work, White American children only consistently applied stereotypes about being American. Moreover, this research suggests that children’s cultural stereotypes might diverge from children’s attitudes early in development. These studies raise new questions about the emergence of racial stereotype application early in childhood – including how best to study it.

KEYWORDS


Stereotype application; racial attitudes; discrimination; child development

Understanding how and when racial stereotypes (i.e., thoughts about groups’ attributes) and racial attitudes (i.e., general feelings about racial groups) emerge is paramount to addressing social group biases from an early age onward. Although much is known about the development of racial attitudes, far less is known about the emergence and application of racial stereotypes in childhood. However, stereotypes can serve as the basis for discrimination and exclusion (Hilton & Von Hippel, 1996). As such, understanding the developmental trajectory of racial stereotyping is important for addressing differential treatment of members of negatively stereotyped racial groups early in life.

Attitudes versus stereotypes about racial groups

In social psychology, attitudes refer to an “overall evaluation of a group” whereas stereotypes are generally defined as “associations, and attributions of specific characteristics to a group” (Dovidio, Hewstone, Glick, & Esses, 2010). In addition to differing in specificity (i.e., attitudes are more general, stereotypes are more specific), research reveals that adults’ application of racial stereotypes cannot be perfectly predicted from adults’ racial attitudes. For example, in the U.S., White adults generally hold more positive attitudes about White

CONTACT Jellie Sierksma  j.sierksma1@uu.nl  Department of Developmental Psychology, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, Netherlands

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15248372.2022.2090945>

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

people (their in-group) compared to Black and Asian people, but they do not always rate their in-group higher than out-groups on positive traits in measures of their stereotyping. Indeed, research reveals that White U.S. adults think White people are wealthier and *more* intelligent than Black people, but *less* athletic and *less* musical than Black people; further White U.S. adults rate White people as *more* athletic but *less* intelligent than Asian people (Zou & Cheryan, 2017).

The few studies that have examined stereotype application in young children (i.e., up to 8 years) typically do not clearly delineate whether children's responses are guided by specific cultural stereotypes (i.e., widely held stereotypes about groups that are situated within a particular culture) or by general attitudes toward racial groups (i.e., feeling more positively or negatively toward different groups). For example, one study tested White American children's (4–9 years) memory for stories in which Black and White characters were both portrayed as mean, dirty, and lazy. Participants were more likely to remember that Black characters were mean, dirty, and lazy compared to White characters. The authors' main interpretation of these results was that children remembered stereotype-consistent information better than counter-stereotypic information (Bigler & Liben, 1993). However, another interpretation of Bigler & Liben's (1993) results and those from similar studies (e.g., Augoustinos & Rosewarne, 2001; Chiesi & Primi, 2006; De la Peña, Ortiz, Berrocal, & Barrett, 2003) is that young children assume groups they feel positively toward will have more positive characteristics (or more of a given positive characteristic) and that groups they feel negatively toward will have more negative characteristics (or more of a given negative characteristic). In other words, in these studies, children may not be applying cultural stereotypes they have gleaned from the world about people from specific groups; rather, when asked, children may simply rely on their racial attitudes to guide their responses.

To understand whether children apply specific, cultural stereotypes to groups, or apply attributes to groups in line with their racial attitudes, it is crucial to test whether and when children's general attitudes dissociate from the manner in which they apply specific cultural stereotypes. As noted earlier, adults sometimes rate their most liked group less positively on a given characteristic than a group they like the least. By assessing whether children's attitudes ever pattern differently from their stereotype ratings, it is possible to identify whether and when children differentiate in their stereotype application in an adult-like manner¹ (vs. based on their racial attitudes). To this end, we can ask whether White children, for example, evaluate Asian people as smarter than White people (i.e., as White adults do: Fiske, Cuddy, Glick, & Xu, 2002; Zou & Cheryan, 2017) even though White children generally evaluate White people more or equally positively compared to Asian people on measures of racial attitudes (Dunham, Baron, & Banaji, 2006; Persson & Musher-Eizenman, 2003). And we can ask whether White children, like White adults (Zou & Cheryan, 2017), evaluate Black people as more athletic than White people even though they generally evaluate White people more positively than Black people on measures of racial attitudes (i.e., Raabe & Beelmann, 2011).

¹Note that by "adult-like" we do not mean "correct" or "acceptable", but rather matching adults' responses.

Emergence of racial attitudes and stereotypes

In general, there is consistent and robust evidence that explicit racial in-group favoritism emerges around the age of 4 to 5 years, at least among children from high-status groups (Raabe & Beelmann, 2011). Because stereotypes are more complex, requiring multifaceted representations of groups, it seems likely that they would emerge later in development compared to racial attitudes (Brown, Ali, Stone, & Jewell, 2017; Mackie, Hamilton, Susskind, & Rosselli, 1996). This might mean that at first, young children would not engage with the content of specific cultural stereotypes in an adult-like manner but instead would apply traits for racial groups in line with their racial attitudes (reflecting “halo” or “pitchfork effects”; Koenig & Jaswal, 2011).

Studies that examine stereotype application indeed suggest that adult-like stereotyping begins to emerge in middle- to late-childhood (Brigham, 1974; McKown & Strambler, 2009; McKown & Weinstein, 2003). With a few exceptions (e.g., Bigler, Averhart, & Liben, 2003), the majority of studies that assess stereotype application in children (i.e., up to 13 years) have focused on White or majority-White samples. These studies reveal that around 8 to 9 years of age, White children in the U.S. infer that White people are more American than Asian people (tested 5–10 year-olds, DeJesus, Hwang, Dautel, & Kinzler, 2018); that White people are smarter than Black people but that Black people are better at sports (tested 9–13 year-olds, Copping, Kurtz-Costes, Rowley, & Wood, 2013; Rowley, Kurtz-Costes, Mistry, & Feagans, 2007); and that Asian people are better at math than White people (tested 9–13 year-olds, Cvencek, Meltzoff, & Greenwald, 2011; tested 4th and 5th graders; Nasir, McKinney de Royston, O’Connor, & Wischnia, 2017). In sum, there is evidence that children as young as 4 years of age exhibit racially biased attitudes, and that at around 8 to 9 years of age, White children apply specific adult-like cultural stereotypes to racial groups even when such stereotypes contradict their in-group favoring racial attitudes.

There is a lack of empirical evidence on younger children’s (i.e., < 9 years) cultural stereotype application because studies with this age group are scarce. Moreover, the few studies conducted on young children provide mixed evidence as to the developmental trajectory of stereotyping. As reviewed earlier, in some studies (e.g., Bigler & Liben, 1993), assessments of children’s stereotypes are conflated with their racial attitudes. Other studies have focused on just one specific cultural stereotype: For example, research on children’s stereotypes about wealth indicate that White U.S. children between 3 and 6 years expect Black individuals to be less wealthy than White individuals (Mandalaywala, Tai, & Butler, 2020; Olson, Shutts, Kinzler, & Weisman, 2012; Shutts, Brey, Dornbusch, Slywotzky, & Olson, 2016). Moreover, research on intelligence suggests that in the U.S. 5 and 6-year-old children of diverse backgrounds evaluate White men as smarter than Black men (Jaxon, Lei, Shachnai, Chestnut, & Cimpian, 2019), but that 5-to-13-year-old Asian children do not explicitly endorse the stereotype that Asian people are good at math (Ambady, Shih, Kim, & Pittinsky, 2001). In addition, research that focused on one particular group suggests that 6-year-old children are familiar with stereotypes about Arab Muslims in the U.S (Brown et al., 2017; a majority of the participants were White).

To our knowledge, just one study has examined stereotype application in young children across a broad spectrum of traits and groups: Pauker, Ambady, and Apfelbaum (2010) investigated American children’s (3–10 years, 89.9% European American) application of various stereotypes to three racial groups (i.e., Asian, Black, and White children). They

showed participants (who were members of all three groups) pairs of children from two racial groups and asked them who had demonstrated positive and negative stereotypic behaviors. For example, participants might be asked whether a Black or White child stole from someone, acted aggressively, or was a good dancer. Children were asked about a range of positive and negative cultural stereotypes for each group and the researchers examined participants' assignment of positive and negative stereotypes (collapsed across stereotypes) to in-group and out-group targets (using each participant's race to determine the in-group and combining scores for their two out-groups). Results showed that older children (over 6 years) applied both positive and negative cultural stereotypes to out-groups (at similar levels; 6 to 10 years) and in-groups (7 to 10 years), but younger children did not.

The results of Pauker and colleagues' (2010) study suggest that children might begin to apply out-group stereotypes as early as 6 years. However, because the researchers collapsed children's stereotypes across the two out-groups and across the various stereotypes for each group, it remains unclear which, if any, specific stereotypes children applied to Asian, Black, and White groups. For example, the results for younger children might reflect that they did not apply any stereotypes (in line with the interpretation from Pauker et al., 2010) or might indicate that young children applied only a subset of stereotypes (e.g., that Asian people are smarter than White people, but not that Black people are more untrustworthy or better at dancing than White people). Similarly, because the results were collapsed across multiple racial out-groups, it is unclear if older children's stereotype application varied across the three racial groups featured in the study. This lack of specificity in the reported analyses raises the possibility that extant work has underestimated or overestimated the age at which children begin to apply specific, cultural stereotypes about Asian, Black, and White people.

The present research

The current research aims to delve deeper into the question of *when* the application of cultural stereotypes about racial groups emerges in childhood (distinct from racial attitudes) and *which* stereotypes young children apply to various racial groups. We examined White American children's ratings for three racial groups: Asian, Black, and White. As a first step, we focused on White children's stereotype application because White children generally show clear and robust racial attitudes favoring their own group (e.g., Raabe & Beelmann, 2011), which is a prerequisite for testing our hypotheses about the role of racial attitudes vs. adult-like differentiation in stereotype application. Moreover, the scarce literature on children's application of stereotypes predominantly focuses on White children and we aim to shed more light on these findings. Because White children start to express in-group favoritism at age 4 (Raabe & Beelmann, 2011) and previous research suggests that White children show consistent application of cultural racial stereotypes by 8–9 years (e.g., Brown et al., 2017; Cvencek et al., 2011; DeJesus et al., 2018; Rowley et al., 2007), we focused on 4-to-8-year-old children to study the emergence and development of racial stereotyping in this group.

The way previous research has assessed stereotype application in adults and older children differs greatly: In some studies, participants are asked to rate groups on various traits (e.g., Zou & Cheryan, 2017); other studies have used examples of behavior that align with traits (i.e., sharing cookies to represent being nice, e.g., Pauker et al., 2010). Studies differ also in whether they present positive *and* negative traits to participants (e.g., Pauker et al., 2010) or solely positively framed traits with assessments of intensity (e.g., how

intelligent is this group on a scale of 1 to 7; Fiske et al., 2002). Here we decided to focus on six positive trait domains: intelligence (how “smart” people are), athleticism (how “sporty” people are), nationality (in particular, how “American” people are), kindness (how “nice” people are), wealth (how “rich” people are), and honesty.

We focused on traits rather than examples of behavior because research shows that young children have a deeper understanding of traits compared with behaviors highlighting certain traits (e.g., Boseovski & Lee, 2006). Moreover, young children use traits to make inductive inferences about others (Heyman & Gelman, 2000). In addition, we decided to only include positive trait domains as young children are likely aware of the social norms about saying bad things about other people. As such, we expected that children would be more likely to express stereotypes in our study if we presented them with traits that were positively framed rather than negatively framed. There are also ethical concerns about asking children to apply negative words to individuals. Moreover, previous research found evidence of stereotyping using only positive trait domains with adults and older children (e.g., Brown, 2011; Devos & Banaji, 2005; Fiske et al., 2002; Mandalaywala et al., 2020; Rowley et al., 2007; Zou & Cheryan, 2017).

For the present study, we focused on positive traits young children are familiar with (see Lane, Wellman, & Gelman, 2013 on nice, smart, honest; see Brown, 2011, on being American; see Horwitz, Shutts, & Olson, 2014 on being rich; see Sierksma & Shutts, 2021, on being sporty). Importantly, our collection of traits included at least one trait for each racial out-group on which White American adults evaluate that group more positively than their in-group: Adults generally view Asian people as smarter than White people and Black people as more athletic than White people (Zou & Cheryan, 2017). In our main analyses we thus focus on the relative differences between groups on the traits (i.e., comparing White vs. Black people, and White vs. Asian people) and whether this differed from children’s general tendency to evaluate White people more positively than racial out-groups.

We preregistered two contrasting hypotheses. One hypothesis (1a) was that White children would apply traits in alignment with their racial attitudes. This would mean that for all traits, participants would evaluate White children more positively than Asian children and Black children. The alternative hypothesis (1b) was that children would apply traits in line with prevailing cultural stereotypes applied by adults (i.e., Asian people are smarter than White people, Black people are more athletic than White people; White people are nicer, wealthier, and more American than Asian and Black people; White people are more honest than Black people, see Table 1 for an overview of adults’ stereotypes). We also considered whether the basis for children’s responses might change over development such that younger children would apply the traits according to their racial attitudes, whereas older children’s application of the traits will be more aligned with cultural stereotypes applied by adults.

Study 1

Method

Participants

Previous research on stereotype application in children resulted in a medium effect size (i.e., interaction of age and stereotyping, $\eta^2 = .10$; Pauker et al., 2010). We therefore followed recommendations by Cohen (1992) for a medium effect size, power of 0.80, and error

Table 1. Adults application of stereotypes.

Stereotype	Order of the groups for adults
Athletic	Black > White > Asian ¹
Rich	White > Asian > Black ¹
Honesty	White = Asian, White > Black, Asian > Black ¹
Intelligent	Asian > White > Black ¹
Nice	White > Black > Asian ²
American	White > Black > Asian ³

Note. ¹based on Zou & Cheryan (2017; Study 3), participants answered the following question: “How much are <racial group> stereotyped as <trait>?” ²Based on how warm/cold participants think these groups are viewed by others in Fiske et al. (2002), ³based on Devos & Banaji (study 1, 2005) and (study 1, Cheryan & Monin, 2005), participants were asked how American groups were.

probability of 0.05 (i.e., 64 children). However, because we were also interested in exploring age effects, we doubled this number in order to have a reasonable number of children across all ages, allowing us to acquire sufficient power to run analyses for age. Therefore, we aimed for a final sample of 128 White participants. Summary-statistics-based sensitivity analysis for mixed-effects modeling of nested data (Murayama, Usami, & Sakaki, 2022) showed that with our sample ($n = 128$), power of 80% and alpha set at .05, we were able to detect a small effect (i.e., $d = .25$) in Study 1 and 2.

We did not know children’s race or ethnicity in advance of the testing sessions and thus tested any children ($N = 178$) whose parents were interested in the study. Parents reported their children’s race or ethnicity in a questionnaire: 137 children were White and non-Hispanic/Latinx, 12 children were Black, 11 children were Hispanic or Latinx, 3 children were Asian, 9 children were identified as “other,” and information was missing for 5 children. Only White non-Hispanic/Latinx children were included in the main analyses; 9 children were excluded for failing the attention check ($n = 2$) or experimenter error ($n = 7$). This resulted in a final sample of 128 children (48.5% girls, 51.5% boys) between 4 and 8 years ($M = 5.98$, $SD = 1.41$; 25 4-year-olds, 28 5-year-olds, 25 6-year-olds, 25 7-year-olds, 25 8-year-olds). Of parents who completed the questionnaire ($N = 127$), 84.4% reported obtaining a college education or higher, and 58.3% reported that last year’s total family income was \$100,000 or more. We asked parents to estimate the racial diversity in their neighborhoods and their children’s schools; these data are available in the supplemental materials. Participants were tested in a mid-sized city in the Midwestern region of the US. The institutional review board of the university where the research was conducted approved all studies (ID 2017–0817). Data for nonwhite children can be found on OSF (see here). Preregistrations and data can be found at OSF (Study 1, see here; Study 2, see here, data see here). Note that some children were tested before the studies were preregistered (7 children for Study 1, 5 children for Study 2).

Design

Children’s stereotype application was measured for six traits (“smart,” “sporty,” “American,” “nice,” “rich,” “honest”) for three racial groups (Asian, Black, and White children). Similar to previous research on stereotyping (e.g., DeJesus et al., 2018; Pauker et al., 2010) participants rated pictures of individual children, belonging to different groups, without including group labels. Stereotype application was measured on a 6-point scale where a score of 6 corresponds to a positive evaluation (e.g., “really smart”) and a score of 1 corresponds to a less positive

evaluation (e.g., “not so smart”). All 3 racial groups and all 6 traits were measured within participants. Children completed 6 blocks of trials, one for each trait. On each trial, children rated one target child on a trait and this was repeated for two different target children within each racial group for a total of 6 trials in each block. Across the study, each participant provided a total of 36 ratings for 6 targets with 6 traits. The order of the 6 traits and the order in which children saw each racial group were counterbalanced across participants, and children never rated two targets from the same group consecutively. The gender of the target children was matched to the participant’s gender. The pilot we conducted beforehand to examine whether the procedure worked well for children across a wide age range, is described in the Supplemental Materials.

Materials

Participants indicated their stereotype application by applying traits to children from different racial groups using a “ladder”-scale. The scale consisted of six yellow vertical bars with a hook in the middle of each bar where children could hang pictures (see [Figure 1](#)). The vertical bars referred to the level of each trait such that if a participant placed a picture at the lowest bar, it indicated that they thought the child had low levels of the trait (e.g., “not so smart”), and if they placed a picture at the highest bar, it indicated that they thought the child had high levels of the trait (e.g., “very smart”). The ladder was mounted to a wall at a height that all children could reach. Laminated cards with words that marked the endpoints for each trait (i.e., “very smart,” “not so smart”) were clipped to each end of the ladder.

The target pictures consisted of 12 photographs taken from The Child Affective Facial Expression (CAFE) set (LoBue & Thrasher, 2015). Six pictures depicted smiling girls and six depicted smiling boys, and within each gender there were two pictures each of Asian, Black, and White children. The same six photographs were used in every block of the task. We did not include racial group labels as we were interested in children’s spontaneous application of stereotypes and previous research demonstrates that even young children are able to classify photographs of children by race (Aboud et al., 2003).



Figure 1. The ladder used in study 1 to measure children’s application of traits.

Procedure

The session took approximately 20 minutes and was conducted by a White female experimenter. When children entered the laboratory, they were seated in a chair next to the ladder and were told that they would be using it to think about other people.

Practice phase. Children had the opportunity to practice first. The experimenter clipped cards saying “thirsty” and “not thirsty” at the ends of a ladder. Then, she presented the child with a stick figure and told them that it was “very very thirsty” and showed children how to place the stick figure at the top of the ladder. Then, she explained that if the stick figure was not thirsty at all, it should go on the bottom of the ladder, and when the stick figure was somewhat thirsty, she said it would go somewhere in the middle and put the stick figure at the third bar. Then the child went through the three placement options and were corrected when they made mistakes.

Test phase. Children were then told, “now we are going to do it with real people and some other words.” The experimenter unclipped the cards that said “thirsty” and “not thirsty” and replaced them with the first trait (depending on the counterbalancing order) and defined the word (see a list of definitions in Supplemental Materials). Subsequently, she walked over to a table (about 7 ft from the ladder) with the child and placed all six pictures (gender matched to the child) on the table, grouped according to race in order to make racial groups salient (i.e., two pictures of Asian children next to each other, two pictures of Black children next to each other and two pictures of White children next to each other – with order of racial groups counterbalanced across children). Then participants were told: “Now of course you don’t know these children, but sometimes you can guess how people are just by looking at them. Now what we will do is, for each picture you get to hang them on the ladder where you think they should go. And after you have done that, you bring the picture back and we’ll do the next one. I am going to sit over here and I will not watch your answers, so you can just be honest, ok?” She then gave the first picture to the child and said: “So how about this kid, do you think he or she is [trait word], not so [trait word] or in between?” The child then walked over to the ladder and hung the picture on the hook to indicate their selection. Although children this age generally do not hide racial biases (e.g., Apfelbaum, Pauker, Ambady, Sommers, & Norton, 2008), the experimenter sat with her back to the ladder to reduce social desirability. Then, the child removed the picture from the hook and returned the picture to the experimenter at which point the experimenter would give the child the next picture and repeat the question.

Attention check. Because the experimenter did not watch children’s responses, we administered an attention check after children completed the fourth block (out of six). Children were given the stick figure again and asked to hang it on the ladder to indicate that it was “very hungry” and that it was “not hungry at all.” If children failed to hang the stick figure at the top and then at the bottom, they were excluded from analyses ($n = 2$).

Scoring

Each child’s responses were video-recorded and trained research assistants scored children’s responses according to where they hung the pictures on the ladder (scores for each trial ranged from 1 to 6). Children gave a total of 36 responses from which we calculated 18 separate scores (averaging across ratings for the 2 same-gender photos in each of the three

racial group and six trait combinations); paired t-tests for each trait showed that children's ratings for the two same-gender pictures that belonged to the same racial group (e.g., the two White boys) did not differ (i.e., all $ps > .05$). Higher scores indicate that children rated targets more positively, whereas lower scores indicate a more negative evaluation. We created an overall evaluation score for each racial group by averaging participants' scores across the six traits, and also calculated 6 scores for each of the individual traits.

Analyses

Linear mixed models were specified in R using the package lme4 (Bates, Maechler, Bolker, & Walker, 2015) including a random intercept.² To compare children's stereotype application for the three racial groups, we used centered contrasts (i.e., White vs. Black, White vs. Asian, Black vs. Asian). Children's age in months was included in the models as a continuous z-score.

First, for each trait we specified a model comparing children's ratings of White vs. Black and White vs. Asian targets and one for Black vs. Asian targets (i.e., models were estimated for each trait separately to avoid multicollinearity). Second, we examined children's overall trait application by collapsing across all traits to test whether children's responses aligned with previous studies. Therefore, we computed children's overall evaluation score and specified one model including contrasts for White vs. Black targets and White vs. Asian targets and one model for Black vs. Asian targets. All models additionally included a main effect for age as well as interactions with each contrast. Significant interactions with age were followed up using simple slope analyses (i.e., the slope between the contrasts comparing the racial groups and stereotype application is examined at different values of age, in our case older (1 *SD* above the mean) and younger (1 *SD* below the mean) children, see Aiken, West & Reno, 1991) as well as simple effects analyses for each racial group (i.e., the extent to which changes in age are related to how children view a particular racial group). Simple slopes are the preferred method for analyzing continuous predictors compared to dichotomizing predictors (i.e., examining separate age group groups) as dichotomizing predictors results in a loss of information and power (MacCallum, Zhang, Preacher, & Rucker, 2002). Results are reported in Table 2, Figures 2 and 3. For a schematic overview of all significant and non-significant results, see supplemental materials.

Results

Trait ratings

Smart. Children indicated that White and Black targets were equally smart ($p = .99$) but rated White targets as smarter than Asian targets ($p < .001$). Black and Asian targets were rated as equally smart ($p = .06$). There was no influence of age.

Sporty. Children rated White and Black targets as equally sporty ($p = .30$) but rated Asian targets as less sporty than White targets ($p < .001$). The latter effect was moderated by age ($p = .005$). Simple slope analyses showed that older children rated Asian targets as less sporty than White targets ($\beta = .89$, $p < .001$), whereas younger children did not

²Note that we preregistered to use linear mixed models with random slopes for the contrasts that compared racial groups. Doing so, however, resulted in singular fit suggesting that the models were too complex for the data and suggestions in the field to solve this issue did not work (see Bates et al., 2015). We therefore specified models including only random intercepts (See Barr, Levy, Scheepers, & Tily, 2013).

differentiate between the two groups ($\beta = .17, p = .35$). Children’s age was not related to their ratings of how sporty White targets were ($\beta = .03, p = .75$), but was negatively related to their ratings of how sporty Asian targets were ($\beta = -.29, p = .001$). Children rated Black targets as more sporty than Asian targets ($p = .004$), and this was not moderated by age.

American. Children indicated that White targets were more American than both Black targets ($p < .001$) and Asian targets ($p < .001$). A significant interaction emerged between children’s age and scores for White vs. Asian targets ($p < .001$). Simple slope analyses showed that older children indicated Asian targets were less American compared to White targets ($\beta = 1.13, p < .001$), whereas younger children did not differentiate ($\beta = .00, p = .99$). Furthermore, age was positively associated with children’s ratings of how American the White targets were ($\beta = .26, p = .001$) and was negatively related to how American children thought Asian targets were ($b = -.38, p < .001$). Children rated Asian and Black targets as equally American ($p = .81$). There was no significant interaction for children’s age and scores for White vs. Black targets or Asian vs. Black targets.

Table 2. Results for study 1 and study 2.

Traits	White vs. Black		White vs. Asian		Black vs. Asian	
	Study 1 β (SE)	Study 2 β (SE)	Study 1 β (SE)	Study 2 β (SE)	Study 1 β (SE)	Study 2 β (SE)
Overall Score	.13 (.11)	.53*** (.15)	.72*** (.11)	.82*** (.15)	.30** (.11)	.14* (.13)
* age	-.03 (.11)	.22 (.15)	.34** (.11)	-.04 (.15)	.18 (.11)	-.13 (.13)
Smart	-.00 (.12)	.57*** (.17)	.42*** (.12)	.25 (.17)	.21 (.11)	-.16 (.15)
* age	-.13 (.13)	.22 (.17)	.23 (.13)	-.19 (.17)	.18 (.11)	-.21 (.15)
Sporty	-.13 (.13)	-.04 (.20)	.53*** (.13)	.39* (.16)	.33** (.11)	.22 (.17)
* age	-.09 (.13)	-.06 (.20)	.36** (.13)	.12 (.16)	.23 (.11)	.09 (.16)
American	.50*** (.13)	.50** (.16)	.56*** (.13)	.76*** (.14)	.03 (.13)	.13 ¹ (.13)
* age	.17 (.13)	.26 (.16)	.56*** (.23)	.53*** (.14)	.20 (.13)	.13 (.13)
Nice	-.10 (.12)	.27 (.17)	.58*** (.12)	.28 (.17)	.34** (.11)	.00 (.15)
* age	-.09 (.12)	-.02 (.17)	-.06 (.12)	-.24 (.17)	.01 (.11)	-.11 (.15)
Rich	.03 (.13)	-.09 (.17)	.31* (.13)	.64*** (.17)	.14 (.11)	.36* (.15)
* age	-.11 (.13)	-.02 (.17)	.09 (.13)	-.02 (.17)	.10 (.11)	-.10 (.15)
Honest	.10 (.12)	.16 (.17)	.25 (.12)	.22 (.17)	.07 (.11)	.03 (.15)
* age	.04 (.13)	.35* (.17)	-.01 (.13)	-.50** (.17)	-.02 (.11)	-.42** (.15)

Note. ¹ Mixed linear model only including random intercept due to singular fit *** $p < .001$ ** $p < .01$ * $p < .05$ + $p < .10$.

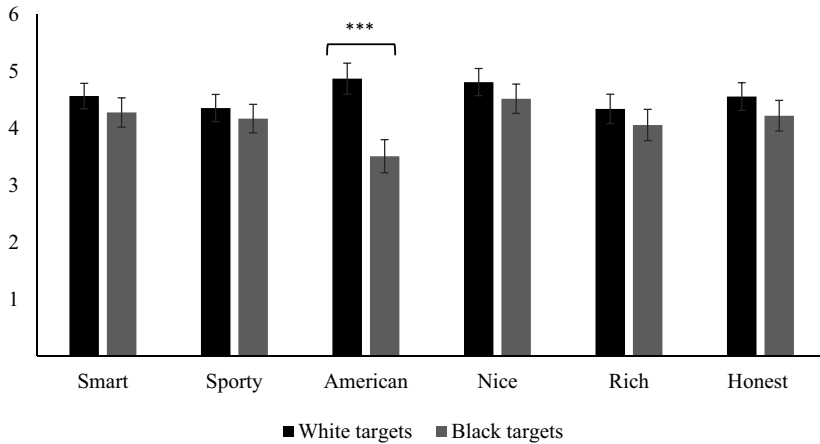


Figure 2. Children's application of all six traits for white vs. black targets, study 1. Note. Error bars represent 95% CI. *** $p < .001$.

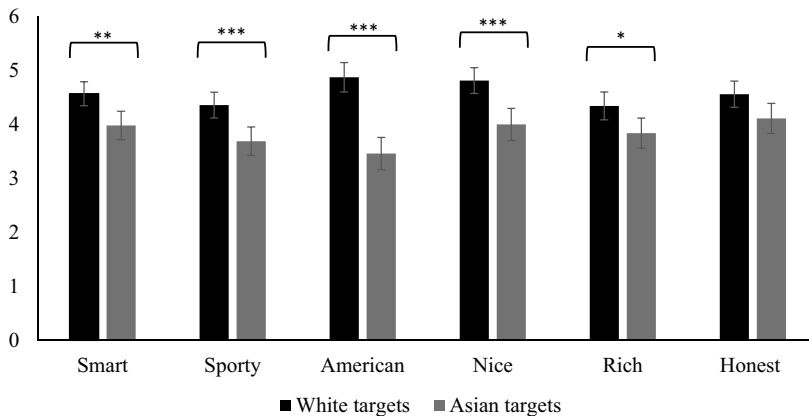


Figure 3. Children's application of all six traits for white vs. asian targets, study 1. Note. Error bars represent 95% CI. *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$

Nice. Children indicated that White and Black targets were equally nice ($p = .42$) but rated White targets as nicer than Asian targets ($p < .001$). Children also rated Black targets as nicer than Asian targets ($p = .002$). There was no influence of age.

Rich. Children indicated that parents of White and Black targets were equally rich ($p = .83$) but indicated that Asian parents were less rich than parents of White targets ($p = .02$). Asian and Black parents were rated as equally rich ($p = .22$). There was no effect of age.

Honest. Children indicated that White and Black targets were equally honest ($p = .43$) and also rated White and Asian targets as equally honest ($p = .05$). Black and Asian targets were rated as equally honest ($p = .50$). There was no influence of age.

Overall rating score

Overall, children evaluated White and Black targets similarly ($p = .28$). However, they were more positive about White compared to Asian targets ($p < .001$) and rated Black people more positively than Asian people ($p = .01$). There was a significant interaction between children's age and their overall evaluation of White compared to Asian targets ($p = .003$). Simple slope analyses showed that younger and older children were more positive about White compared to Asian targets, although this preference was stronger for older children (younger children: $\beta = .38$, $p = .01$; older children: $\beta = 1.06$, $p < .001$). Age was not related to children's evaluation of White targets ($\beta = .09$, $p = .29$), but with age, children evaluated Asian targets more negatively ($\beta = -.22$, $p = .008$). No other interactions emerged with age.

Discussion

The results of Study 1 provide mixed evidence for cultural stereotyping in young children: Children's ratings of traits for the three racial groups did not resemble adults' stereotype application in most cases. First, children rated Black individuals similarly to White individuals for five out of six traits. Only with regard to the trait American, did children give Black individuals lower ratings than White individuals. As such, children's evaluation of being American was the only stereotype for the comparison of White and Black people that aligned with adult stereotypes (Cheryan & Monin, 2005; Devos & Banaji, 2005), and some previous research with older children (Brown, 2011; Brown et al., 2017; we discuss this finding more fully in the general discussion). Second, Asian people were evaluated more negatively than White individuals on five out of the six traits (i.e., only for honesty results revealed no difference). Here the results are in line with adults' stereotypes for being athletic (but only for older children), being rich, honest, nice and American (but only for older children). Third, the comparison between Asian and Black individuals revealed that children evaluated them similarly for four out of six traits (i.e., intelligence, American, honesty, wealth); this pattern is in line with adults' stereotypes for only two traits (being athletic and nice). In general, the results did not reveal consistent age patterns: Older children's responses were no more adult-like than those of younger children.

Thus, while not providing consistent evidence for cultural stereotype application, the results of Study 1 also do not provide consistent evidence in line with the idea that young children's application of traits is mostly driven by their racial attitudes (hypothesis 1a). Specifically, although previous studies show that White children feel generally more positively toward White people compared with Black people (Raabe & Beelmann, 2011), children in Study 1 evaluated White and Black children similarly for most traits. The racial attitudes literature suggests that White children sometimes prefer Asian and White people equally and sometimes are slightly less positive about Asian people compared with White people (Dunham et al., 2006; Persson & Musher-Eizenman, 2003), yet in Study 1 children evaluated Asian children less positively than White children across most traits.

The results from Study 1 might mean that children this age truly do not apply racial stereotypes about these three racial groups in an adult-like manner and that their application of traits is also not driven by their racial attitudes. However, it could also be that our method was not appropriate to capture stereotyping in this age group. For example, young

children might be less inclined to use traits as a basis for person inference when these are presented to them as dimensional (i.e., Likert scales, see Gonzalez, Zosuls, & Ruble, 2010). Another possibility is that children in this sample did not hold the racial attitudes that have typically been found in previous studies. Finally, it is possible that children viewed the traits we presented less positively than intended. To address these issues, we conducted a second study using a different method to capture children's stereotype application, and included a measure of children's racial attitudes as well as a measure of how they viewed the different traits.

First, in Study 2, we used a forced choice measure to determine whether children might show more evidence of stereotype application using a measure that arguably is more likely to push children to rely on their knowledge of cultural stereotypes, if such knowledge is present (Mandalaywala et al., 2020; Sierksma & Shutts, 2020; Signorella, Bigler, & Liben, 1993). Second, we included a measure of children's attitudes toward each of the three racial groups in order to assess whether children's attitudes and trait inferences diverge, as the results from Study 1 seem to suggest. Third, to examine how children perceived the traits in general, we also asked them to evaluate the valence of each trait.

Study 2

Method

Participants

We aimed for a similar sample size as in Study 1 (128 White children). A total of 154 children participated in Study 2 (136 White non-Hispanic/Latinx, 11 Hispanic/Latinx, 2 Black, and 5 did not report). For the final sample, only White non-Hispanic/Latinx children were included and an additional 7 children were excluded because an error occurred in recording ($n = 5$) or because they had previously participated ($n = 2$). The final sample consisted of 129 children (we collected data for 1 additional child because they were already scheduled to participate; 48.1% boys, 51.9% girls) between 4 and 8 years ($M = 5.96$, $SD = 1.45$; 29 4-year-olds, 24 5-year-olds, 25 6-year-olds, 25 7-year-olds, and 26 8-year-olds). Data for nonwhite children can be found on OSF ([here](#)). Parents' estimates of the racial diversity of their children's neighborhoods and schools can be seen in supplemental materials. For parents who filled out the questionnaire ($n = 122$ of 129), 90.3% reported a college education or higher, and 60% had a family income of \$100,000 or more last year. Similar to Study 1, participants again came from a mid-sized city in the Midwestern region of the US. Participants in this study did not participate in Study 1.

Design

Children's stereotype application for the same six traits and the same three racial groups as in Study 1 was assessed using a forced choice method. Specifically, for each of the six traits children saw three pairs of pictures of target children who differed by race and were asked to select which child was best described by the traits, resulting in a total of 18 trials. The pairs were: White–Black, White–Asian, and Asian–Black. The order of traits and pairs of targets were counterbalanced across participants. The pictures were randomized across participants.

Materials

We used 36 pretested pictures, 12 for each racial group, and 6 per gender within each racial group (12 pictures that were used in Study 1 and 24 new pictures; see supplemental materials for details about pretesting). We added additional pictures in this study so that children saw new pictures in each trial. This was done to keep children engaged throughout the task, and so that judgments on one trial did not influence their judgments on another trial. In addition, we used 7 boxes in which children were asked to put the pictures to indicate their answers. These boxes (sized 8inch x 8inch x 4inch) had a small opening on the top (where children put the pictures). Six white boxes had a card with the trait attached on the top (“the answer boxes”), and there was also one yellow box that was for the picture children did not pick (“the box for the extras”).

Procedure

The study took approximately 15 minutes and was conducted by one of two White female experimenters. When children entered the laboratory room they were seated in a chair at a table. They were shown the boxes and told they would use them to think about people.

Practice phase. The experimenter then said they would practice first. She showed the participant two pictures of dogs and asked them: “Who has spots?” (Only one of the dogs had spots). Children then picked the dog with spots and the experimenter explained that their answer would always go in the “answer box” and pointed out which one it was, then participants walked over and put the picture in the box. The experimenter then explained that the picture that was left over, should be placed in the “box for the extras, the yellow one” and participants placed the dog without spots in the yellow box. This practice was repeated for a fluffy and not fluffy cat (“Who is fluffy?”) and again for dogs that were smelly and not smelly (to teach children that not all questions would have obvious correct answers). After children were able to correctly place the pictures in each box, the experimenter repeated the instructions and said: “Now we are going to do it with real people and some other words.”

Test phase. The experimenter then picked up the first box (depending on the counterbalancing order) and said, “Ah, this one is about [trait].” She would then define the trait as in Study 1. She then placed a pair of pictures of the targets on the table and sat facing away from the boxes (so that she could not see the child’s response) and said, “Who is [trait]?” (Depending on the counterbalancing order). For each trait, children were asked about three pairs of pictures. For ethical reasons, children were never forced to pick. Thus, when children asked whether they could pick both, the experimenter said this was ok, but also emphasized it was fine to pick one if children thought they were just a little more [trait].

Evaluation of traits. To assess how positively children perceived each trait, we asked participants to evaluate each trait after rating all of the targets. The experimenter said: “So, we talked about a lot of different things and now I just want to know what you think about it.” She would then pull out a 3-point smiley face and said: “So, if you think something is good it is this one (happy smiley), if you think something is just ok it is this one (neutral)

and when you think something is bad it is this one (sad).” She then asked children to rate each (positive) trait. To ensure that children used the scale appropriately, the experimenter also asked participants to rate the trait “mean.”

Racial group attitudes. In the last phase, the experimenter told children that the last thing they would do is look at pictures of children again and tell her whom they liked. For a total of 18 cross-race pairs (identical pictures and identical to order of stereotype assessment) children were asked “Who do you like?”

Scoring

Videos were coded by one of three trained research assistants. In a few cases, the video camera was not functioning so the experimenter live-coded children’s answers (while making sure children did not see this). For each trial, the racial group that was chosen received a score of “1” and the racial group that was not chosen received a score of “0.” Across all trials, children saw two pictures from each racial group for each trait. Thus, each racial group could receive a score ranging from 0 to 12 in total, and 0 to 2 for each trait. In 0.02% of the cases children put both pictures in the answer box and in 0.009% of the cases children said they liked both targets. Because these percentages were so low it is unlikely that they influenced the results in any way and we thus included these answers in our analyses. For children’s evaluation of the traits, when they picked the smiling face it was coded as “3,” the neutral face was coded as “2” and the frowning face was coded as “1.”

Analyses

We preregistered an analysis plan that involved analyzing the data using linear mixed models in R using the package lme4 (Bates et al., 2015). To compare children’s stereotype application for each racial group we used centered contrasts (i.e., White vs. Black, White vs. Asian, Black vs. Asian) with a random slope. Children’s age was included in the models as well (continuous z-score, fixed effect). Note that the models were estimated without random intercepts due to singularity problems (see Brauer & Curtin, 2018). Specified models were the same as in Study 1. Results can be seen in Table 2, Figures 4 and 5. Exploratory analyses for the association between stereotype application and racial group attitudes are reported in the supplemental materials.

Results

Trait ratings

Smart. Children indicated that White targets were smarter than Black targets ($p < .001$) but indicated that White and Asian targets were similarly smart ($p = .16$). Black and Asian targets were rated as equally smart ($p = .28$). None of these main effects was moderated by age.

Sporty. Children indicated that White and Black targets were equally sporty ($p = .83$) and indicated that White targets were more sporty than Asian targets ($p = .02$). Children rated Black and Asian targets as equally sporty ($p = .17$). None of these main effects was moderated by age.

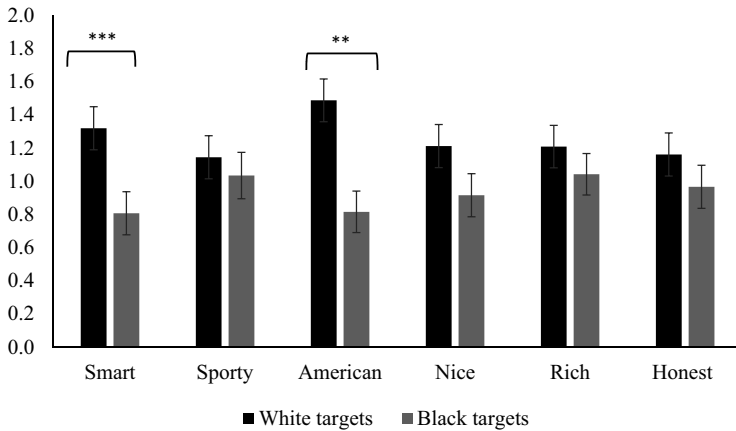


Figure 4. Children's application of all six traits for White vs. Black targets, study 2. Note. Error bars represent 95% CI. *** $p < .001$, ** $p < .01$

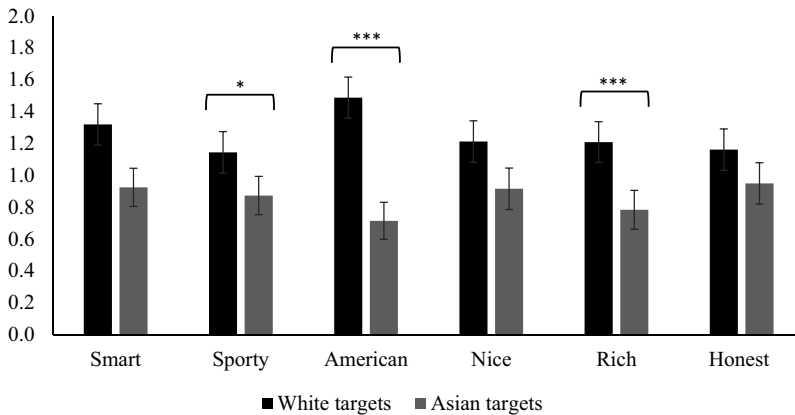


Figure 5. Children's application of all six traits for White vs. Asian targets, study 2. Note. Error bars represent 95% CI. *** $p < .001$, * $p < .05$, + $p < .10$

American. Children indicated that White targets were more American than Black targets ($p = .002$) and Asian targets ($p < .001$). However, children's evaluation of how American White versus Asian targets were was moderated by age. Simple slope analyses showed that younger children did not differ in their ratings of White and Asian targets ($\beta = .22, p = .26$), however, older children indicated that White targets were more American than Asian targets ($\beta = 1.29, p < .001$). Moreover, age was positively related to children's evaluation of how American White targets were ($\beta = .39, p < .001$) and negatively related to how American they thought Asian targets were ($\beta = -.27, p < .001$). Children rated Asian and Black targets as equally American ($p = .32$).

Nice. White and Black targets were rated as equally nice ($p = .11$), White and Asian targets were also rated as equally nice ($p = .097$), and Black targets and Asian targets were rated as equally nice ($p = .98$). There was no influence of age.

Rich. Children indicated that parents of White and Black targets were equally rich ($p = .61$) but rated parents of White targets as richer than parents of Asian targets ($p < .001$). Parents of Black targets were rated as richer than parents of Asian targets ($p = .017$). None of these main effects was moderated by age.

Honest. There were no main effects for either contrast (White vs. Black, $p = .92$; White vs. Asian, $p = .21$; Black vs. Asian, $p = .85$). However, a significant interaction emerged between age and children's evaluation of White versus Black targets ($p = .04$). Follow-up analyses showed that younger children that indicated White and Black targets were equally honest ($\beta = -.18, p = .17$), whereas older children indicated that White targets were more honest than Black targets ($\beta = .51, p = .033$).

A significant interaction also emerged between age and children's evaluation of White versus Asian targets ($p = .004$). Simple slope analyses showed that younger children indicated that White targets were more honest than Asian targets ($\beta = .72, p = .004$) but older children indicated that White and Asian targets were equally honest ($\beta = -.28, p = .25$). Children's age was not related to their ratings of White targets' honesty ($\beta = .01, p = .25$), but age was positively related to children's ratings of honesty in Asian targets ($\beta = .22, p = .013$) and negatively related to children's ratings of honesty in Black targets ($\beta = -.20, p = .02$).

For the Black vs. Asian contrast there was also a significant interaction with age ($p = .004$). Simple slope analyses showed that older children evaluated Black and Asian targets as equally honest ($\beta = -.40, p = .056$), while younger children evaluated Black targets as more honest than Asian targets ($\beta = .45, p = .031$). Moreover, this was driven by a negative association of age with ratings of honesty in Black targets ($\beta = -.15, p = .018$) rather than Asian targets ($\beta = .04, p = .55$).

Overall rating score

Across all the traits, children evaluated White targets more positively than Black targets ($p < .001$) and Asian targets ($p < .001$). Children evaluated Black targets more positively than Asian targets ($p = .015$). None of these effects was moderated by age.

Evaluation of traits

Children rated all traits positively (smart $M = 2.77, SD = 0.46$; sporty $M = 2.47, SD = 0.58$; American $M = 2.49, SD = 0.56$; nice $M = 2.98, SD = 0.20$; rich $M = 2.23, SD = 0.68$; honesty $M = 2.72, SD = 0.54$) and above the midpoint (2) of the scale (all $ps < .001$). Nice was rated more positively compared to all other traits (all $ps < .001$). Followed by honest and smart, that were rated equally positive ($p = .34$) but more positive than being rich ($p = .001$), American, and sporty ($ps < .001$). Finally, being American and sporty were rated equally positive ($p = .70$) and rich was evaluated the least positive of all stereotypes (all $ps < .001$). Last, "mean" was rated negatively and scored below the midpoint of the scale ($M = 1.03, SD = 0.22; p < .001$) and rated more negatively than all traits (all $ps < .001$).

Racial group attitudes

Children liked White targets ($M = 7.77$, $SD = 2.72$) more than Black targets ($M = 4.69$, $SD = 2.42$; $\beta = .92$, $p < .001$) and liked White targets more than Asian targets ($M = 5.56$, $SD = 2.42$; $\beta = .32$, $p = .04$). In addition, children liked Asian targets more than Black targets ($\beta = -.30$, $p = .02$; see supplemental materials). These main effects were not moderated by children's age.

Discussion

The results of Study 2 show that children rated Black and White individuals as equally sporty, nice, rich, and honest (when children were younger), but evaluated Black individuals as less American and less smart than White individuals. These results are in line with adults' stereotypes for three traits: honesty (but only for older children), being American, and intelligence. In addition, children rated Asian individuals as less honest (when children were younger), sporty, rich, American (when children were older) than White individuals, but evaluated Asian and White individuals as equally smart and nice. Children's evaluations of White people as being more sporty, rich, honest and American compared to Asian people is in line with adults' stereotypes. For the comparison of Black and Asian individuals in Study 2, children evaluated Black and Asian targets similarly on 4 out of 6 traits. The two traits for which children's evaluation differed were being rich and honest (i.e., children reported that Black individuals were richer than Asian individuals and younger children indicated Black individuals were more honest than Asian individuals). These results oppose adults' stereotypes for all six traits.

Although how children applied traits to the racial groups in Study 2 diverged from previous studies (e.g., Pauker et al., 2010), the group preference measure showed a clear alignment with previous studies such that children preferred White individuals over both Black and Asian individuals. Such in-group preference resembles previous research on racial attitudes (e.g., Dunham et al., 2006; Raabe & Beelmann, 2011). Finally, children rated all traits we presented to them as highly positive, suggesting the traits were perceived similarly in terms of valence. Taken together, these additional measures in Study 2 show that children in our sample were sensitive to the valence of the traits we asked them about, and held attitudes favoring their in-group consistent with the previous studies.

General discussion

The current research aimed to further illuminate when and how 4- to 8-year-old White American children apply racial stereotypes. Across two studies, White children did not consistently stereotype Asian, Black, and White children in the same way as has been observed in studies of White adults. As such, these results present a more nuanced view on the development of stereotypes than previously assumed (e.g., Bigler & Liben, 2007; Pauker et al., 2010) and give rise to intriguing new questions.

Do young children apply stereotypes like adults do?

First, both studies showed that young White children consistently applied the stereotype that children in the two racial out-groups were less American than White children. Specifically, children of all ages indicated that Black targets were less American than White children and, with age, children also reported that Asian targets were less American than White targets. These results are consistent with adult stereotypes (Devos & Banaji, 2005) as well as responses of somewhat older children in studies of stereotyping (Brown, 2011; Brown et al., 2017; DeJesus et al., 2018). However, note that in our study children thought Asian and Black targets were equally American, whereas previous studies suggests that adults and children see Black people as more American than Asian people (Brown, 2011; Brown et al., 2017; Cheryan & Monin, 2005). Our results thus provide new evidence that even very young children (i.e., from 4 years onward) seem to have intuitions about an American prototype that looks White and does not include Black children, which is somewhat younger than the age reported in previous research on racial stereotyping (i.e., Brown (2011) showed a similar pattern in 5-year-old children). Second, across the two studies, adult-like stereotypes about Asian people consistently emerged for three other traits as well. Namely, children evaluated Asian people as less rich and athletic (only older children in Study 1) than White people. But children indicated Asian and White people were equally honest (although only older children did so in Study 2).

For most traits, however, our results suggest that White American children do not apply cultural stereotypes about racial groups. Specifically, for being sporty, nice, rich, smart, and honest, children's application of the traits to Black people did not reveal a consistent adult-like pattern. Across both studies, children indicated that Black targets were just as sporty, nice, rich, and honest (with the exception of older children in Study 2) as White targets. While children saw Black people as less smart than White people in Study 2, they rated them as equally smart in Study 1. Moreover, the results for how children applied the nice and smart traits to White and Asian targets were mixed across the two studies (i.e., sometimes Asian targets were rated as less nice or smart but also sometimes as equally nice and smart compared with White targets). These results contrast findings from older children and adults that often show, for example, that Black people are seen as more athletic but less wealthy and nice than White people (e.g., Copping et al., 2013; Rowley et al., 2007) or that Asian people are seen as smarter but less athletic than White people (e.g., Zou & Cheryan, 2017). The current findings for children's application of being sporty, nice, wealthy, smart, and honest to Black people are therefore not in line with the cultural stereotypes applied by White adults and neither is children's application of being intelligent and nice to Asian people (hypothesis 1b).

The results from the present studies appear to contradict previous research suggesting the early acquisition of stereotypes (e.g., when using the PRAM, Aboud, 1988; Pauker et al., 2010). That said, when we analyzed our data following the practices of previous researchers – using a forced-choice format and collapsing across all traits (i.e., the overall trait score in Study 2) – we found that children attribute more positive traits to White than to Black and Asian children. Critically, however, when we assessed children's application of traits for each trait separately, our findings contrast with earlier findings, or interpretations thereof, of specific, content-based stereotype application in young children (Pauker et al., 2010).

Do young children apply stereotypes in alignment with their racial attitudes?

Given that children showed little adult-like differentiation in their application of traits to racial groups, did they apply traits based on their racial attitudes as stated in hypothesis 1a? Our data also do not seem to support this notion. The results show a discrepancy between how the White children in our sample rated Black and Asian targets as compared to White targets on the traits and how young White children (in the U.S. and in our sample) feel about Black and Asian targets as compared to White targets (i.e., racial attitudes; Dunham et al., 2006; Raabe & Beelmann, 2011). Moreover, across the two studies, children did not consistently rate White individuals more positively on all traits compared to Asian and Black individuals. In fact, children in our sample rated Black targets similarly to White targets on four traits but, when asked how much they *liked* these groups, they preferred White targets over Black targets (in our sample; see also Raabe & Beelmann, 2011). Children's responses to Asian targets also seem to largely disconfirm hypothesis 1a. Although children in our sample rated Asian targets less positively on most traits compared to White targets and also liked Asian targets slightly less than White targets (a finding that is replicated elsewhere, i.e., Dunham et al., 2006), they never evaluated Asian people more positively on the traits than Black people but they did like Asian people *more* than Black people

Clearly, there is a need for more research on this discrepancy between racial attitudes and the application of traits to racial groups, however these findings suggest that racial attitudes might not drive the application of stereotypes or vice versa (e.g., Stangor, 2016) in young children. The lack of consistent alignment between trait application and racial attitudes also lends support to the notion that there is a strong need for research on the application of stereotypes to better understand how they differentiate from children's general racial attitudes. Note that Study 2 showed that children did perceive all traits similarly in terms of valence (i.e., all were evaluated positively) and therefore discrepancies in perceived trait-valence cannot account for the lack of support for hypothesis 1a.

Why does the stereotype about being American emerge early?

The results from these studies also raise the question of why young White children apply, in particular, a racial stereotype about being American so early in life. One possibility is that children might simply receive more cultural input about this racial stereotype. For example, perhaps stereotypes about being American are communicated more often or more clearly to children than stereotypes about honesty. It is clear that adults consistently and robustly hold stereotypes about being American (Devos & Banaji, 2005); they may communicate this stereotype in many ways to children, for example through their behavior (Brey & Pauker, 2019; Sierksma & Shutts, 2020) and media exposure (see Aboud et al., 2012).

Cultural input alone, however, is likely not the whole story. One possible explanation for the finding that children seem to learn the stereotype about being American earlier than other stereotypes is that American is the only trait that explicitly references group membership. In contrast, the cultural input children receive about traits such as being honest or nice is probably not always tied to racial group membership, but can also indicate meaningful individual differences. Given that young children are able to categorize individuals into racial groups as young as three years (e.g., Nesdale, 2003), it could be that stereotypes

explicitly linked to salient group categories enhance children's learning or application of stereotypes and therefore emerge earlier. Relatedly, the salience of group membership in being American might make this stereotype especially self-relevant for children because they are motivated to view their own group positively (Tajfel & Turner, 1979). This finding raises an interesting avenue for future research that involves testing the role of salient group membership in young children's application of stereotypes (e.g., making the group context more salient might lead to stronger or earlier emerging stereotype application).

How can we explain additional unanticipated findings?

One surprising result is that participants in the present research did not rate White individuals as "richer" than Black individuals. Previous work documents that racial stereotypes about wealth emerge between 3 and 6 years of age (Mandalaywala et al., 2020; Olson et al., 2012; Shutts et al., 2016), but children in the present research did not rate Black individuals as less rich than White individuals. One possible explanation for these divergent findings concerns how we assessed wealth stereotypes in the present research. In many previous studies (Mandalaywala et al., 2020; Olson et al., 2012; Shutts et al., 2016), researchers have assessed children's wealth stereotypes by presenting children with material wealth indicators (e.g., a large, fancy house vs. a smaller, older house) rather than by asking children to apply the trait "rich." Children may believe that different groups tend to possess different levels of material wealth but may fail to fully understand the broader concept or state of being rich. Further, it is worth noting that among children in the present study, "rich" was seen as the least positive trait. Perhaps the trait "rich" has a less positive connotation for young children compared with owning a nice house or toys.

We also observed that children did not clearly and consistently differentiate in their trait application for Asian and Black people. In fact, across the two studies, results for the Asian vs. Black contrast showed these groups were rated similarly for 8 out of the 12 comparisons. This lack of differentiation might stem from the tendency to perceive less variability among out-group members, which is already present in children as young as 5 years (out-group homogeneity effect, e.g., Shilo, Weinsdörfer, Rakoczy, & Diesendruck, 2019). But it could also be that children are simply less familiar with these outgroups and therefore tend to see them as similar. In fact, as estimated by parents, children in our sample predominantly lived in neighborhoods and went to schools with low numbers of Asian people and Black people, and a majority of White people. Lack of contact might influence the extent to which racial groups are highlighted in children's environment or talked about. As a consequence, children might be exposed less to information about the racial groups in their society, including cultural stereotypes. Future research is needed to assess stereotype application in more diverse samples (both minority group participants, and participants who come from environments that are more diverse).

Future directions and conclusions

The investigation of children's application of racial stereotypes can provide a unique window into children's beliefs about the racial groups in their society. The current research suggests that White American children apply stereotypes about being American, but do not yet apply racial stereotypes based on intelligence, niceness, honesty, wealth, and athleticism

in an adult-like manner even though racial attitudes are present at this age. These findings highlight the importance of differentiating children's expression of racial attitudes from their application of specific stereotypes.

The current work also highlights two challenges that future work needs to address. First, it is of critical importance to broaden and deepen developmental work on children's stereotype application: We know very little about how racial stereotypes are represented and understood early in life, and it is therefore important to assess to what extent children's stereotype knowledge might depend on and generalize across specific age samples and across different ethnic and racial groups and cultural context. Eventually, a better understanding of how children's stereotypes emerge over development may provide us with insight into when and how we should intervene to prevent discrimination early in life.

Second, there is a strong need to better understand *how* we should study young children's stereotype application. Here we made the point that historically the field often has not distinguished between racial attitudes and racial stereotypes. However, it is also true that even when one sets out to studying stereotyping per se (as we did here), there are many decisions to be made methodologically (e.g., how to present racial groups or stereotypes). It is not clear what the consequences of those decisions are for how children respond or what we might conclude about the development of stereotype application. In future work, it will be important to carefully examine how stereotype application is influenced by how we present the racial context to children (e.g., groups vs. individual exemplars; absence or presence of group labels; age and gender of the targets; for an in-depth discussion see Pauker, Williams, & Steele, 2016), how we ask our questions (e.g., forced choice, Likert scales) and what information we provide when introducing stereotypic traits (e.g., positively framed traits vs. positive and negative traits; behavioral examples vs. trait concepts; see also Gonzalez et al., 2010). We hope that the present paper serves to encourage more research of both types – i.e., research on different populations, but also studies focused on methods appropriate for studying stereotyping in children.

Acknowledgement

We thank Katharine Scott for providing helpful feedback on an earlier version of this article. We also thank Rachel King and Bailey Immel for their help with recruiting participants, data collection, and coding data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported in part by a core grant to the Waisman Center from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (U54 HD090256) and a Rubicon grant (Nederlandse Organisatie voor Wetenschappelijk Onderzoek 446-16-011) awarded to Jellie Sierksma.

ORCID

Jellie Sierksma  <http://orcid.org/0000-0002-1690-5811>

Elizabeth Brey  <http://orcid.org/0000-0001-7981-4047>

Kristin Shutts  <http://orcid.org/0000-0002-2323-805X>

Data availability statement

The data described in this article are openly available in the Open Science Framework at <https://doi.org/10.1080/15248372.2022.2090945>

Open Scholarship



This article has earned the Center for Open Science badges for Open Data, Open Materials and Preregistered. The data and materials are openly accessible at <https://doi.org/10.1080/15248372.2022.2090945>.

References

- Aboud, F. E. (1988). *Children and prejudice*. Oxford: B. Blackwell.
- 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. doi:10.1037/0012-1649.39.1.48
- Aboud, F. E., Tredoux, C., Tropp, L. R., Brown, C. S., Niens, U., & Noor, N. M. (2012). Interventions to reduce prejudice and enhance inclusion and respect for ethnic differences in early childhood: A systematic review. *Developmental Review*, 32(4), 307–336. doi:10.1016/j.dr.2012.05.001
- Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.
- 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. doi: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. doi:10.1037/a0012835
- Augoustinos, M., & Rosewarne, D. L. (2001). Stereotype knowledge and prejudice in children. *British Journal of Developmental Psychology*, 19(1), 143–156. doi:10.1348/026151001165912
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using eigen and s4. *R Package Version, 1.1-10*. <http://dx.doi.org/10.18637/jss.v067.i01>
- Bigler, R. S., & Liben, L. S. (1993). A Cognitive-developmental approach to racial stereotyping and reconstructive memory in Euro-American children. *Child Development*, 64(5), 1507–1518. doi:10.2307/1131549
- Bigler, R. S., Averbart, C. J., & Liben, L. S. (2003). Race and the workforce: Occupational status, aspirations, and stereotyping among African American children. *Developmental Psychology*, 39(3), 572–580. doi:10.1037/0012-1649.39.3.572

- Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, 16(3), 162–166. doi:10.1111/j.1467-8721.2007.00496.x
- Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology*, 42(3), 500–513. doi:10.1037/0012-1649.42.3.500
- Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, 23(3), 389–411. doi:10.1037/met0000159
- Brey, E., & Pauker, K. (2019). Teachers' nonverbal behaviors influence children's stereotypic beliefs. *Journal of Experimental Child Psychology*, 188, 104671. doi:10.1016/j.jecp.2019.104671
- Brigham, J. C. (1974). Views of black and white children concerning the distribution of personality characteristics. *Journal of Personality*, 42(1), 144–158. doi:10.1111/j.1467-6494.1974.tb00562.x
- Brown, C. S. (2011). American elementary school children's attitudes about immigrants, immigration, and being an American. *Journal of Applied Developmental Psychology*, 32(3), 109–117. doi:10.1016/j.appdev.2011.01.001
- Brown, C. S., Ali, H., Stone, E. A., & Jewell, J. A. (2017). US children's stereotypes and prejudicial attitudes toward Arab Muslims. *Analyses of Social Issues and Public Policy*, 17(1), 60–83. doi:10.1111/asap.12129
- Cheryan, S., & Monin, B. (2005). Where are you really from?: Asian Americans and identity denial. *Journal of Personality and Social Psychology*, 89(5), 717–730. doi:10.1037/0022-3514.89.5.717
- Chiesi, F., & Primi, C. (2006). Italian children's ethnic stereotyping: Age differences among 4–10 year olds. *Review of Psychology*, 13, 3–7. <https://hrcak.srce.hr/9055>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. doi:10.1037/0033-2909.112.1.155
- 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. doi:10.1111/jasp.12061
- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, 82(3), 766–779. doi:10.1111/j.1467-8624.2010.01529.x
- De la Peña, A. G., Ortiz, J. M. C., Berrocal, P. F., & Barrett, M. (2003). Stereotype development in Andalusian children. *The Spanish Journal of Psychology*, 6(1), 28–34. doi:10.1017/S1138741600005187
- DeJesus, J. M., Hwang, H. G., Dautel, J. B., & Kinzler, K. D. (2018). “American= english speaker” before “American= white”: The development of children's reasoning about nationality. *Child Development*, 89(5), 1752–1767. doi:10.1111/cdev.12845
- Devos, T., & Banaji, M. R. (2005). American = White? *American = White?* *Journal of Personality and Social Psychology*, 88(3), 447–466. doi:10.1037/0022-3514.88.3.447
- Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The sage handbook of prejudice, stereotyping and discrimination* (pp. 3–28). London: Sage Publications.
- Dunham, Y., Baron, A. S., & Banaji, M. R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*, 77(5), 1268–1281. doi:10.1111/j.1467-8624.2006.00933.x
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. doi:10.1037/0022-3514.82.6.878
- Gonzalez, C. M., Zosuls, K. M., & Ruble, D. N. (2010). Traits as dimensions or categories? Developmental change in the understanding of trait terms. *Developmental Psychology*, 46(5), 1078–1088. doi:10.1037/a0020207

- Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology*, 77(1), 1–19. doi:10.1006/jecp.1999.2555
- Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47(1), 237–271. <http://dx.doi.org/10.1146/annurev.psych.47.1.237>
- Horwitz, S. R., Shutts, K., & Olson, K. R. (2014). Social class differences produce social group preferences. *Developmental Science*, 17(6), 991–1002. doi:10.1111/desc.12181
- Jaxon, J., Lei, R. F., Shachnai, R., Chestnut, E. K., & Cimpian, A. (2019). The acquisition of gender stereotypes about intellectual ability: Intersections with race. *Journal of Social Issues*, 75(4), 1192–1215. doi:10.1111/josi.12352
- Koenig, M. A., & Jaswal, V. K. (2011). Characterizing children's expectations about expertise and incompetence: Halo or pitchfork effects? *Child Development*, 82, 1634–1647. <https://doi.org/10.1111/j.1467-8624.2011.01618.x>
- Lane, J. D., Wellman, H. M., & Gelman, S. A. (2013). Informants' traits weigh heavily in young children's trust in testimony and in their epistemic inferences. *Child Development*, 84(4), 1253–1268. doi:10.1111/cdev.12029
- LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults. *Frontiers in Psychology*, 5, 1532. doi:10.3389/fpsyg.2014.01532
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40. doi:10.1037/1082-989X.7.1.19
- Mackie, D. M., Hamilton, D. L., Susskind, J., & Rosselli, F. (1996). Social psychological foundations of stereotype formation. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 41–78). New York: The Guilford Press.
- Mandalaywala, T. M., Tai, C., Tai, C., & Butler, L. P. (2020). Children's use of race and gender as cues to social status. *PLoS One*, 15(6), e0234398. doi:10.1371/journal.pone.0234398
- McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development*, 74(2), 498–515. doi: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. doi:10.1111/j.1467-8624.2009.01359.x
- Murayama, K., Usami, S., & Sakaki, M. (2022). Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. *Psychological Methods*. <https://doi.org/10.1037/met0000330>
- Nasir, N. S., McKinney de Royston, M., O'Connor, K., & Wischnia, S. (2017). Knowing about racial stereotypes versus believing them. *Urban Education*, 52(4), 491–524. doi:10.1177/0042085916672290
- Olson, K. R., Shutts, K., Kinzler, K. D., & Weisman, K. G. (2012). Children associate racial groups with wealth: Evidence from South Africa. *Child Development*, 83(6), 1884–1899. doi:10.1111/j.1467-8624.2012.01819.x
- Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). Race salience and essentialist thinking in racial stereotype development. *Child Development*, 81(6), 1799–1813. doi:10.1111/j.1467-8624.2010.01511.x
- Pauker, K., Williams, A., & Steele, J. R. (2016). Children's racial categorization in context. *Child Development Perspectives*, 10(1), 33–38. doi:10.1111/cdep.12155
- Persson, A., & Musher-Eizenman, D. R. (2003). The impact of a prejudice-prevention television program on young children's ideas about race. *Early Childhood Research Quarterly*, 18(4), 530–546. doi:10.1016/j.ecresq.2003.09.010
- Raabe, T., & Beelmann, A. (2011). Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta-analysis of age differences. *Child Development*, 82(6), 1715–1737. doi:10.1111/j.1467-8624.2011.01668.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. doi:10.1111/j.1467-9507.2007.00376.x

- Shilo, R., Weinsdörfer, A., Rakoczy, H., & Diesendruck, G. (2019). The out-group homogeneity effect across development: A cross-cultural investigation. *Child Development, 90*(6), 2104–2117. doi:[10.1111/cdev.13082](https://doi.org/10.1111/cdev.13082)
- Shutts, K., Brey, E. L., Dornbusch, L. A., Slywotzky, N., & Olson, K. R. (2016). Children use wealth cues to evaluate others. *PLoS One, 11*(3), 3. <https://dx.doi.org/10.1371/journal.pone.0149360>
- Sierksma, J., & Shutts, K. (2020). When helping hurts: Children think groups that receive help are less smart. *Child Development, 91*(3), 715–723. doi:[10.1111/cdev.13351](https://doi.org/10.1111/cdev.13351)
- Sierksma, J., & Shutts, K. (2021). Competence-based helping: Children's consideration of need when providing others with help. *Journal of Experimental Child Psychology, 210*, 105206. doi:[10.1016/j.jecp.2021.105206](https://doi.org/10.1016/j.jecp.2021.105206)
- Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review, 13*(2), 147–183. doi:[10.1006/drev.1993.1007](https://doi.org/10.1006/drev.1993.1007)
- Stangor, C. (2016). The study of stereotyping, prejudice, and discrimination within social psychology: A quick history of theory and research. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 3–27). Psychology Press.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Utah, United States: Brooks-Cole.
- Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of Personality and Social Psychology, 112*(5), 696–717. doi:[10.1037/pspa0000080](https://doi.org/10.1037/pspa0000080)