Psychological Reports
2021, Vol. 124(2) 809–839

DOI: 10.1177/0033294120916867
journals.sagepub.com/home/prx
SAGE

What Drives Racial Attitudes in Elementary School Children: Skin Tone, Facial Physiognomy, or Both?

Elena V. Stepanova
The University of Southern Mississippi, Hattiesburg, MS, USA

Yarrow Dunham
Yale University, New Haven, CT, USA

Maya Rex
The University of Southern Mississippi, Hattiesburg, MS, USA

Nao Hagiwara
Virginia Commonwealth University, Richmond, VA, USA

Abstract

This work examines whether racial attitudes—when measured by both explicit and implicit measures— are driven primarily by skin tone, facial physiognomy, or both in 5 to 12-year-old children. Participants evaluated faces varying in skin tone (from dark to light) and facial physiognomy (from Afrocentric to Eurocentric). In an explicit task, children rated how much they liked each face. In an implicit task, participants completed a child-friendly version of the Affect Misattribution Procedure, where they rated a Chinese character as "good" or "bad" following a racial prime. Results suggest that pro-White attitudes (especially those measured by the explicit task)

Corresponding Author:
Elena Stepanova, School of Psychology, The University of Southern Mississippi, Owings-McQuagge Hall, 118 College Drive, #5025, Hattiesburg, MS 39406, USA.
Email: Elena.Stepanova@USM.edu

are driven by both factors, vary by perceivers' race, and are present in both White and non-White children, though skin tone exerts a larger influence than other features, at least in explicit evaluations. Our results also raise the possibility that pro-White biases might be more internalized by non-White children in the American South.

Keywords
Racial attitudes, implicit and explicit measures, skin color, facial physiognomy, social cognitive development

Introduction

A long history of developmental psychological research has focused on the emergence of racial attitudes in children, both that are assessed with self-reported, "explicit" measures (Aboud, 1988; Bigler & Liben, 2006), and more recently, with automatic or "implicit" measures (reviewed in Dunham et al., 2008; Olson & Dunham, 2010). Generally speaking, in White children, anti-Black attitudes assessed with explicit measures show a gradual age-related decline in strength after peaking in early to middle childhood, around age 7 (for review, see a meta-analysis by Raabe & Beelmann, 2011). By contrast, pro-White attitudes assessed with implicit measures appear to emerge equally early but then remain largely stable throughout an individual's lifetime (e.g., Dunham et al., 2008). However, to date, results have focused on children's attitudes toward broad racial categories (e.g., "Black") and have not yet investigated how specific physical features that signal racial group membership (i.e., that signal a degree of racial phenotypicality), such as skin tone and other facial features (i.e., facial physiognomy), contribute to these forms of attitudes. This is important because advancements in theory (e.g., racial phenotypicality model, see Maddox, 2004) and empirical work with adults (e.g., Stepanova & Strube, 2012b) suggest that beyond simple category membership the degree of racial phenotypicality can also influence racial attitudes. Therefore, the main goal of this work was to investigate how the degree of racial phenotipicality achieved by manipulation of certain racial cues (i.e., skin tone and facial physiognomy) affects racial biases assessed with implicit and explicit measures in children.

Phenotypicality can only affect racial attitudes if observers are sensitive to phenotypic cues. It is well documented in the literature that skin tone and facial physiognomy drive racial categorization (Stepanova & Strube, 2012a) and racial attitudes (e.g., Hagiwara et al., 2012; Stepanova & Strube, 2012b) in adults. Whether these findings can be generalized to children is an empirical question. In the recent research, several researchers have examined the role of skin tone and facial physiognomy in racial categorization in children. For example, Dunham et al. (2015) investigated whether children's racial category judgments (between the ages of 4 and 9) are sensitive to varied phenotypic cues including skin tone and other aspects of facial physiognomy that distinguish Black from White individuals. They employed a categorization task in which children made continuous racial category judgments (on a scale from *White or European American* to *Black or African American*), as well as nonlabeled judgments of similarity (on a scale from *this kind of person* to *this kind of person*), when scales were anchored by an extremely prototypical Black or White face on each end. Participants rated faces that independently varied in skin color and facial physiognomy. Dunham et al. found that categorization decisions in early childhood are driven almost exclusively by attention to skin tone, with attention to other physiognomic features emerging only in middle childhood. Interestingly, when categorizing faces along the White-Asian continuum (Dunham et al., 2016), younger children were sensitive to both skin tone and facial physiognomy. These findings indicate that understanding of racial category boundaries in younger children is dependent upon the specific comparison in question (e.g., White-Black or White-Asian), suggesting that at least under some conditions children can extract both skin tone and other physiognomic cues when making judgments of racial categories. How these cues are used by children when affectively evaluating individuals from different racial groups is yet to be examined.

The present study extends this developmental work on weighting of racially phenotypic cues in racial *categorization* to the domain of *racial attitudes*: how do skin tone and other aspects of facial physiognomy affect racial attitudes assessed with explicit and implicit measures in both White and non-White elementary school children? We focus on the Black–White racial distinction and test whether explicit and implicit racial attitudes are driven primarily by skin tone, facial physiognomy, or both. We additionally test for possible developmental change by comparing children of two age groups: 5 to 7-year-olds and 8 to 12-year-olds.

Racial attitudes in children

Attitudes are measured by both explicit and implicit measures. Explicit measures are instruments based on self-reports such as questionnaires (Bohner & Dickel, 2011). Implicit measures are not based on self-reports and include, among others, priming tasks where participants' responses to certain targets (pictures or words) are influenced by primes such as pictures or words representing various social groups (for review, see De Houwer, 2006; Fazio & Olson, 2003).

Racial attitudes assessed with explicit measures

While negative racial attitudes toward Black Americans in adults assessed with explicit measures have declined over the years in general (for review, see Dovidio et al., 2000), many White American adults still continue to hold negative views of Blacks (e.g., Schmidt & Nosek, 2010). When do these attitudes emerge? Children are aware of racial categories from the age of 3 to 4 (Aboud, 1988; Clark & Clark, 1947; Hirschfeld, 1995) and begin to form positive ingroup racial attitudes soon thereafter (for reviews, see Aboud, 1988; Bigler & Liben, 2006; Cameron et al., 2001). Negative outgroup attitudes toward Blacks and other racial/ethnic minority outgroups appear to peak a few years later (5–7 years) before undergoing a gradual decline into adolescence (Raabe & Beelmann, 2011), perhaps due to children's desire to conform to egalitarian social norms and their emerging awareness of the possible consequences of nonadherence to those norms (Rutland et al., 2005).

Racial attitudes assessed with implicit measures

Past work with White adults generally suggests that they hold negative attitudes toward Black Americans when assessed with implicit measures (e.g., Dovidio et al., 1997; Payne, 2001; Wittenbrink et al., 2001), with most recent work even suggesting that such negative attitudes are on the rise in the United States (e.g., Lane & Jost, 2011; A. L. Skinner & Cheadle, 2016). While originally conceptualized as "slow learned" internalizations of statistical regularities (e.g., Greenwald & Banaji, 1995), early developmental work suggested that even White children as young as 3 to 4 years old hold pro-White/anti-Black racial attitudes that are comparative in strength with those held by adults (Baron & Banaji, 2006; Dunham et al., 2006, 2013). The broad conclusion is that racial attitudes assessed with implicit measures are acquired early in life and remain stable throughout development (though cf. Degner & Wentura, 2010; Williams & Steele, 2019, for alternative perspectives).

The general pattern of early emerging negative attitudes toward racial/ethnic outgroups assessed by implicit measures has now been documented in a wide range of contexts including Singaporean Chinese (bias toward Indians; Setoh et al., 2017), mainland Han-Chinese (bias toward Blacks, Whites, and Africans; Qian et al., 2017; Xiao et al., 2015), White Americans (bias toward Blacks) and Taiwanese (bias toward Whites; Dunham et al., 2013), Canadian Whites (bias toward Blacks; Gonzalez et al., 2017), Indians (bias toward low castes and religious outgroups; Dunham et al., 2015), and Japanese (bias toward Blacks and Whites; Dunham et al., 2006). Despite this extensive work in a variety of populations, racial attitudes assessed with implicit measures in non-White children, especially Black/African American children have been understudied (although see Newheiser & Olson, 2012 for an important exception). In addition, prior research on racial attitudes (both explicit and implicit) in children has focused almost exclusively on attitudes toward racial categories such as "Black" and "White." However, evidence suggests that there is relatively large variation in attitudes toward individuals within any given racial category, which we now review.

Skin tone and facial physiognomy

Variations in racially phenotypical physical features

Accounting for the complexity of contemporary racial categories requires acknowledging that in addition to the between-category variation, there is also substantial within-category variation in racially prototypical physical features. This variation can be simplified to two dimensions: skin tone and facial physiognomy. For example, a face may have dark skin with Eurocentric facial physiognomy, or light skin with Afrocentric physiognomy. In addition to evaluating broad racial categories, in some cases observers also evaluate *features* that are associated with racial categories. In an early demonstration of this phenomenon, Livingston and Brewer (2002) showed, using an affective priming task, that Black faces with strong Afrocentric features (i.e., darker skin tone and eye color, wider nose, thicker lips, and coarse hair) were evaluated more negatively than Black faces with weaker Afrocentric features, even when the latter faces were readily identified as Black. When participants performed a racial categorization task that facilitated racial category activation *prior* to the affective priming task, both types of faces received negative implicit evaluations, regardless of the varying level of Afrocentric features. These findings provided initial evidence that racial attitudes can be driven by both racial categories and specific phenotypic properties, with each potentially contributing independently. This work has since inspired a new model of racial phenotypicality (see Maddox, 2004) that argues that racially phenotypical physical cues, even in the absence of explicit category activation, contribute to racial attitudes. Variations in racially phenotypical physical features have been recently found to be systematically associated with the effects of such feature-based bias (e.g., Blair et al., 2002, 2005; Blair, Judd, & Chapleau, 2004; Blair, Judd, & Fallman, 2004; Eberhardt et al., 2006; Kahn & Davies, 2011). However, such studies usually treat feature-based typicality as one predictor where skin tone and facial features occur or are manipulated simultaneously (e.g., darker faces also have more Afrocentric facial features) or examine the role of skin tone alone in adults (e.g., Dixon & Maddox, 2005; Klonoff & Landrine, 2000; Maddox & Gray, 2002) or children (Averhart & Bigler, 1997).

Racial categorization and attitudes in adults

Several recent studies attempted to disentangle the specific effects of skin tone and facial physiognomy on both racial categorization and affective reactions toward individuals. For example, Stepanova and Strube (2009) conducted a study with adults while manipulating skin tone (three levels: dark, medium, and light) and facial physiognomy (three levels: Afrocentric, mixed, and Eurocentric). Their results showed that skin tone and facial physiognomy independently affected the ratings of racial prototypicality and racial categorizations. They followed these findings up with a more sensitive manipulation of skin tone (10 levels, from dark to light) and facial physiognomy (10 levels, from very Afrocentric to very Eurocentric), revealing a more complex pattern of contribution of these two factors to racial categorization (Stepanova & Strube, 2012a). Specifically, when skin tone was dark, participants largely ignored facial physiognomy and categorizing a person as an African-American. As skin tone got lighter, participants relied on both skin tone and facial physiognomy and their categorization decisions became more variable. Adapting this approach to the study of racial prejudice, Stepanova and Strube (2012b) also examined the relative contributions of skin tone and facial physiognomy to affective reactions toward individuals' faces that systematically varied on skin tone and facial across the same three by two facial grid used in prior work (Stepanova & Strube, 2009). In an affective priming task, participants evaluated faces with lighter skin tone and stronger Eurocentric facial physiognomy more positively than faces with darker skin tone and stronger Afrocentric facial physiognomy. These findings were replicated with photographs of actual faces (Hagiwara et al., 2012). Taken together, these results indicated that skin tone and facial physiognomy can differentially contribute to both racial categorization and racial attitudes at the explicit as well as the implicit levels. But to date, no research has investigated the development of this critical phenomenon.

Racial categorization in children

A reasonable number of studies have also examined the racial categorization process in children (for review, see Pauker et al., 2017). However, the majority of the existing studies have employed highly racially prototypical and unambiguous facial stimuli that grossly simplify the process of racial categorization. One early exception was Gitter and Satow (1969) who manipulated skin tone and facial physiognomy independently to investigate racial self-identification in children. They found that skin tone and facial physiognomy independently affected children's judgment of their own racial group membership. However, more recently, several studies have started to investigate how skin tone and facial physiognomy contribute to the process of racial categorization in children. For example, Dunham et al. (2015) investigated both children's and adults' Black–White racial categorizations while varying skin tone and other aspects of physiognomy independently (see also Dunham et al., 2016 for Asian-White categorization) across the same 10×10 stimulus space used in prior work (Stepanova & Strube, 2012a). This study replicated the results from Stepanova and Strube (2012a), showing that adults used both skin tone and facial physiognomy to make judgments of racial group membership. In contrast, children were more sensitive to skin tone than to facial physiognomy. Specifically, pre-school children had primarily relied on skin tone when making racial categorization judgments. That is, if two faces had the same skin tone, they were categorized into the same racial group even if they varied dramatically on the facial physiognomy (e.g., a dark face with Eurocentric features and a dark face with Afrocentric features were perceived as equally Black). This suggests that younger children's understanding of the Black–White racial boundary is grounded primarily in skin tone, raising the additional possibility that variation in skin tone will exert a larger influence on racial bias than will variation in other facial physiognomy, at least in younger children. The next step is to determine how skin tone and facial physiognomy affect *racial attitudes* in children of various age groups, which is investigated in this study.

The current study

We tested whether racial attitudes are driven primarily by skin tone, facial physiognomy, or both in younger (5–7-year-old) and older (8–12-year-old) elementary school children of various racial backgrounds by using both explicit and implicit measures. The present study extends the current literature by including a racially diverse sample of children. We investigated the developmental trajectory of racial attitudes by including children from two age groups: younger (5–7 year olds) versus older (8 and older) children. To examine whether findings can be generalized to various racial populations, participants in the present study included both White and non-White children.

Children were recruited from the American South with its distinct history of slavery, segregation, and continuing racial tensions. In this region, current exposure to racial outgroups is high, and both Black and White adult residents show the strongest ingroup bias on implicit and explicit measures among all the U.S. states (Rae et al., 2015). While participants from previous studies on racial and ethnic prejudice in children were recruited from many different parts of the world, as we previously reviewed, most studies utilizing implicit measures in Northern America recruit participants from Canada (e.g., Gonzalez et al., 2017), the American Midwest (e.g., Perszyk et al., 2019), or the Northeast (e.g., Dunham et al., 2013). This study allowed us to investigate racial biases in a region—the American South— that has been understudied but might provide some insights into differences in racial attitudes of White and Black children.

Given the limited number of empirical studies addressing independent and interactive effects of skin tone and facial physiognomy, the present study used the same stimuli from previous studies (Dunham et al., 2015; Stepanova & Strube, 2012a). This approach enables fair comparison between the previous findings and findings from the present study.

To examine how skin tone and facial physiognomy contribute to children's racial attitudes using an explicit measure, we employed a modified version of continuous racial categorization task used in Dunham et al. (2015) study, converting it to a preference task by eliciting a continuous judgment of liking on a 100-point slider; we reasoned that this would facilitate comparisons with that past work and, with the larger rating space, provide more opportunities to pick up subtle differences driven by individual features.

To examine how skin tone and facial physiognomy contribute to children's racial attitudes at the implicit level, we adapted the Affect Misattribution Procedure (AMP; Payne et al., 2005). In the AMP, participants were presented on each trial with a facial prime varying on skin tone and facial physiognomy, followed by a target (a Chinese character) about which they made a "good" or "bad" judgment. The version we used is specifically tailored for and well validated in children with insect–flower primes (Dunham & Emory, 2014; Williams et al., 2016, Study 1) and cute–aggressive animals and happy–sad faces as primes (Williams et al., 2016, Studies 2–3). A recent meta-analysis indicated that the AMP, as other sequential priming tasks, predicts behavior well (Cameron et al., 2012).

One of most well-known implicit measures is the Implicit Association Task (IAT), developed by Greenwald et al. (1998). Importantly for the present purposes, the IAT is a *category-level* measure in that it seeks to measure attitudes toward a general category such as *Black* or *White*. This makes it less suitable for understanding how attitudes vary *within a category,* for example, as stimulus features vary. Other measures, in particular a variety of *sequential priming* tasks, are better able to capture attitudinal variability as a function of the specific individual targets (e.g., visual stimuli such as faces), and so can be used to explore how features of those targets influence evaluation (De Houwer, 2009).

Thus, while several recent studies have adapted implicit tasks to measure attitudes in children (e.g., IAT, Baron & Banaji, 2006; Dunham et al., 2006; the "angry = outgroup" paradigm, Dunham, 2011; Dunham et al., 2013; Xiao et al., 2015; the AMP, Dunham & Emory, 2014; A. C. Skinner et al., 2017; Williams et al., 2016; Williams & Steele, 2019; and the Implicit Racial Bias Test, Setoh et al., 2017; Qian et al., 2016, 2017), to our knowledge no research has yet explored how individual features of target stimuli affect the implicit responses that result (but see Perszyk et al., 2019, for an application of the AMP to investigate bias based on race and gender). Thus, this is the first investigation of a developmental trajectory of attitudes in diverse children in the American South using facial stimuli varying on skin tone and facial physiognomy and implicit and explicit measures.

Method

Participants

This study recruited elementary school children of both genders and various racial backgrounds (recruited $N = 120$, age range of 5–12 years old) from several elementary school and the Young Men's Christian Association (YMCA) locations in a town in the Southern United States. The racial demographics of the schools is the following: one school reports 64.9% African American students; 28.0% White students, and 4.7% multiracial; the other school reports 65.0% White students, 30.4% African American students, and 3.4% multiracial students. While we were not able to obtain exact statistics from the YMCA locations, one branch of the YMCA serves primarily Black population and one branch of the YMCA serves primarily White population (personal communication, YMCA official, December 8th, 2015).

Eight children withdrew at various stages of the experiment. Since one participant did not provide information on age, their data were excluded from the analyses. In addition, two participants only used one response option ("good") on all trials in the AMP and one used all extreme options on the explicit rating task (ERT).

As a result, we excluded data from four additional participants, leaving us with an analyzable sample of 108 participants ($M_{\text{age}} = 7.73$, $SD_{\text{age}} = 2.13$, 52 girls and 55 boys, 1 participant of unknown gender). The ethnic/racial breakdown of the final sample was the following: 33 Black (30.6%), 62 White (57.4%), 10 Mixed race (9.3%), 2 Hispanic (1.9%), and 1 Other (.9%). All but 18 participants provided racial self-identification (i.e., they chose "Not Sure"). For them, racial identification was done by experimenters. There were no children who self-identified as Asian and no participants, including mixed race children, were visually identified by experimenters as of Asian heritage. None of the children in the study spoke Chinese to ensure the validity of the AMP that uses Chinese characters as targets of judgment.

In addition, many mixed race children were identified by experimenters with a high degree of certainty as biracial Black–White (5 of the 10). The rest of the mixed race children ($n = 5$) were classified in that category on the basis of non-Eurocentric (but not Asian) facial features and darker skin tone. Historically, mixed Black–White individuals in the United States have been considered Black (Hollinger, 2003). Thus, we grouped children into two race groups: White ($n = 62$) and non-White children ($n = 46$). While the majority of our non-White children (79%) were Black, given that most of the mixed race children are likely to be considered Black, which is especially true in the region of the country the study was conducted, our non-White sample would be likely classified as Black.

Given that previous work has suggested that age 8 is an upper limit for the development of relatively consistent racial category judgments in Black and White children (Aboud, 1988), participants were divided into two age groups: a younger children group (5–7 year olds, $M_{\text{age}} = 5.91$, $SD_{\text{age}} = .78$, $n = 55$) and an older children group (8 year olds and older, $M_{\text{age}} = 9.62$, $SD_{\text{age}} = 1.23$, $n = 53$).

Materials

Stimuli were 100 computer-generated male faces varying in skin tone (gradually changing from dark to light, 10 levels) and facial physiognomy (gradually changing from Afrocentric to Eurocentric, 10 levels; see Figure 1), which were created with Poser 6^{TM} software and employed in the previous studies examining various face-based categorization judgments (Chen et al., 2018; Dunham et al., 2015; Stepanova & Strube, 2012a, 2018; Stepanova et al., 2013). The facial physiognomy manipulation comprised of multiple phenotypic characteristics (e.g., width of the nose, fullness of the lips, bone structure, etc.). These features were manipulated simultaneously in the Poser using one built-in control that globally modifies the faces to make them "less/more African" (for European faces) or "less/more European" (for African faces; see Stepanova & Strube, 2012a for a more detailed description explaining creation and pre-testing of the facial stimuli).

Each child was exposed to 50 faces that were randomly selected from the larger stimulus set. To increase generalizability, we used 2 sets of 100 faces from

Figure 1. Facial stimuli used in both explicit and implicit tasks (a sample set). Skin tone varies from 1 (dark) to 10 (light), left to right, and physiognomy varies from 1 (Afrocentric) to 10 (Eurocentric), top to bottom. Adapted from Stepanova & Strube, 2012a. Copyright 2012 by Elsevier. Reprinted with permission.

Stepanova and Strube (2012a), with face set counterbalanced across participants (i.e., each participant saw faces drawn from only one of two sets).

Procedure

Data were collected in 2015 to 2016. Following the institutional review board approval, the investigators received approval from a local public school district. Researchers approached school officials from two local schools and two local branches of the YMCA and received permission to recruit participants there. Recruitment letters and consent forms were sent to parents. Children whose parents returned signed consent forms (thus, indicating their consent to participate) were enrolled in the study.

In general, children were tested in groups of two to four at a time in a space provided by the school/program. Each child was assigned to an individual experimenter. Nine individuals acted as experimenters after undergoing an extensive training. While experimenters varied in gender and race (three White males, four White females, and two Black females), we attempted to avoid systematic biases in assigning children to experimenters. Experimenters arrived to testing sites in groups of two to six. In most cases, experimenters worked in pairs, with one reading the script (see Supplemental Material online) and another one providing assistance with charging and starting the laptop, escorting children from the waiting area to testing stations and back, ensuring that the appropriate parental consent forms were signed, and filling out demographic information (i.e., race of participants) and keeping record of those who completed participation. It was randomly determined which child would be assigned to which pair of experimenters among children who were available for testing on that date. We attempted to keep the experimenter pairs heterogeneous on gender and race as much as possible (e.g., a Black female with a White female or a Black female with a White male) so that participants will be exposed to diverse individuals, however, it was not always possible to do so due to logistical reasons (e.g., several participants were attended by a Black experimenter alone and several participants were run by two White experimenters of different genders).

First, children were told that they were going to play a computer game to help us understand how children view and categorize different objects, including faces, and were asked to provide verbal assent. During the experiment, children completed two different computer tasks: an ERT and an AMP. The order of these two tasks was counterbalanced. In each task, participants were presented with 50 faces randomly drawn from a larger set of 100 faces. The same 50 faces were used in both tasks. Participants were trained on the task and completed several practice trials before starting each task.

In the ERT, participants indicated their liking toward each face presented on the computer screen in a random order by sliding an anchor on a scale that had a frowny face ("do not like") on one end and a smiley face ("like") on the other end. We counterbalanced across participants the placement of anchors (i.e., some participants were presented with "do not like" anchor on the left, and some on the right). Participants were instructed to use their finger to show where they wanted to move the anchor on the screen. The experimenter used the mouse and moved the anchor where the participant pointed, consulting the participant on the placement; the position on the scale was automatically converted to a 1 to 100 rating. Most of the children complied. However, several older children requested a permission to use the mouse and place the anchor themselves, which was granted. Clicking on "finish" progressed the task to the next trial.

We used a modified version of AMP (Payne et al., 2005), which was previously and successfully adapted for children (Dunham & Emory, 2014). Participants were told that they will see Chinese characters that have either bad or good meaning: "For each Chinese letter, you will have to guess if it means something good or something bad by choosing either a smiley face (good) or a frowny face (bad)." On each trial, participants briefly saw a face (1 of the 50 faces described above) for 250 ms, which was followed by an intertrial interval (a blank screen) for 125 ms and then by a Chinese character, which stayed on the screen until they made a judgment. The AMP is designed to assess spontaneous evaluative responses to facial pictures (primes). The premise is that if participants feel positive toward a prime, they will rate neutral, unrelated stimuli, such as Chinese letters, more positively, whereas if they feel negatively toward a prime, participants will rate the neutral stimuli more negatively. To indicate their feelings toward Chinese letters, participants were asked to press a smiley face button if they thought the letter was a "good letter" and a frowny face button if they thought it was a "bad letter" on the keyboard. To remind them of the response options, each side of the screen had a depiction of the smiley and frowny faces (see Figure 2 for schematic depictions of the tasks). We counterbalanced across participants the placement of the smiley and frowny face keys (i.e., some participants had the frowny face on the right and some on the left). Participants completed 75 randomized trials: 50 experimental trials and 25 neutral trials, where primes were neutral grey squares.

Upon completion of both tasks, children were asked to indicate their gender, age, and race/ethnicity (Black, White, Asian, Latino/Hispanic, Mixed, Other, and Not Sure), if they were capable of doing so. If children could not provide information about their race/ethnicity, the experimenters attempted to code for participants' racial/ethnic background to the best of their ability. At the end of the experiment, participants were thanked, debriefed, and dismissed. All institutions received a small donation ($1 for each participating child). The entire study took about 15 minutes to complete and was administered on laptop computers using Inquisit 4 (2015).

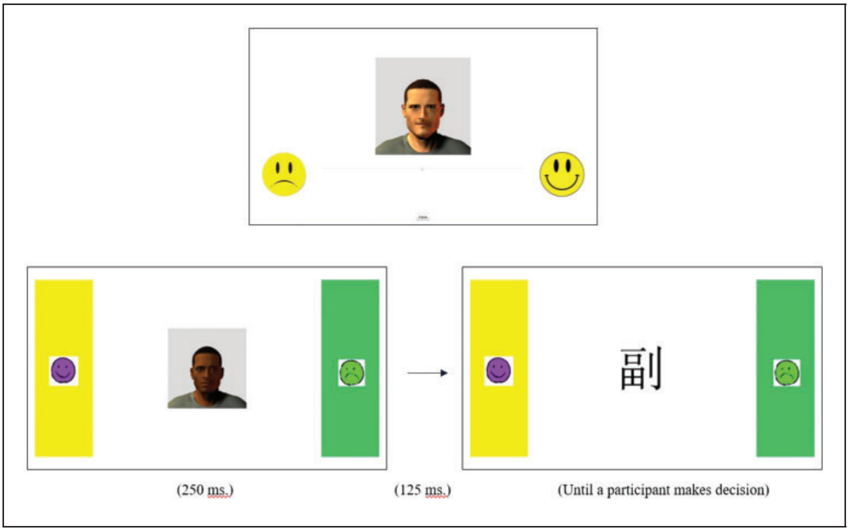

Figure 2. Explicit Rating Task (above) and Affect Misattribution Procedure Task (below).

Results

The analyses reported below are based on the final sample of 108 participants. Three participants experienced a computer malfunction in the beginning of the experiment. The study was restarted for them, and data from the complete session were used in the subsequent analyses.

In analyzing our data, we divided the 10 levels of skin tone into three groups: dark (three darkest skin tones), medium (middle four skin tones), and light (three lightest skin ones). Next, we divided the 10 levels of facial physiognomy into 3 groups: Afrocentric (3 faces with strongest Afrocentric features), mixed (middle 4 faces with mixture of Afrocentric and Eurocentric features), and Eurocentric (3 faces with strongest Eurocentric features). Crossing three levels of skin tone with three levels of facial physiognomy resulted in the nine unique combinations of skin tone and facial physiognomy (see Figure 1). Our decision to group levels of skin tone and facial physiognomy in this way followed from the fact that children saw a random draw of 50 faces from a larger set, such that at the stimulus level, the design was unbalanced (i.e., individual children only saw a random subset of half the faces). By collapsing to three levels of each dimension, we ensured an adequate number of responses from each child in each of the nine cells.

For each participant, we computed the average liking score (for the ERT) and the proportion of "good" responses (i.e., number of "good" responses ÷ number of total responses; for the AMP) for each of the nine stimuli groups (i.e., 3 Skin Tone × 3 Facial Physiognomy). To examine the relative contributions of skin tone and facial physiognomy to racial attitudes, a 3 (Skin Tone: dark, medium, light) $\times$ 3 (Facial Physiognomy: Afrocentric, mixed, or Eurocentric) $\times$ 2 (Age: younger children vs. older children) $\times$ 2 (Race: White vs. non-White) mixed analysis of variance, with the first two factors manipulated within participants, was conducted for each outcome.[1] We conducted all follow-up comparisons with Bonferroni corrections to control for inflation of the Type I error rate.

ERT

The main effects of Skin Tone and Facial Physiognomy were both significant, $F(2, 208) = 14.96$, $p < .001$, $\eta_p^2 = .12$ and $F(2, 208) = 4.96$, $p = .008$, $\eta_p^2 = .05$ correspondingly, with the effect size of Skin Tone considerably larger than that of Facial Physiognomy. Participants rated faces of light ($M = 52.84$, $SE = 1.75$) and medium skin tone ($M = 53.84$, $SE = 1.59$) more favorably than faces of dark skin tone ($M = 45.10$, $SE = 2.10$, all $ps \leq .001$). Participants rated faces with mixed physiognomy ($M = 52.35$, $SE = 1.56$) more favorably than faces with Afrocentric physiognomy ($M = 48.86$, $SE = 1.66$, $p = .006$); ratings of faces with Eurocentric physiognomy ($M = 50.57$, $SE = 1.68$) were not significantly different from either mixed or Afrocentric faces. There was a marginally significant main effect of participant Race: White participants ($M = 53.11$, $SE = 1.96$) rated faces somewhat more positively than non-White participants ($M = 48.08$, $SE = 2.28$), $F(1, 104) = 2.81$, $p = .097$, $\eta_p^2 = .03$.

However, the Skin Tone and Facial Physiognomy main effects were further qualified by several interactions. A significant Skin Tone $\times$ Facial Physiognomy interaction emerged, $F(4, 416) = 4.37$, $p = .002$, $\eta_p^2 = .04$. Analysis of a simple main effect of skin tone at three different levels of facial physiognomy indicated the following pattern. When judging faces with Eurocentric and mixed physiognomy, participants rated medium and light skin tone faces more favorably than dark skin tone faces; when judging faces of Afrocentric physiognomy, participants rated faces of medium skin tone more favorably than faces of dark skin tone (see Table 1 and Figure 3).

The main effect of Race was further qualified by an interaction between Skin Tone and Race, $F(2, 208) = 3.44$, $p = .03$, $\eta_p^2 = .03$. The significant interaction was further probed by analyzing a simple main effect of skin tone for White participants and non-White participants separately. The analysis revealed that there were differences in ratings of three levels of skin tone in non-White children, $F(2, 103) = 8.30$, $p < .001$, $\eta_p^2 = .13$, such that non-White children rated faces of medium skin tone most favorably, followed by light and then dark skin tone faces. There were also differences in ratings of skin tone in White children, $F(2, 103) = 8.35$, $p < .001$, $\eta_p^2 = .14$. White children rated light skin tone faces most favorably, followed by medium and dark skin tone faces (see Figure 4). No main effect or interactions involving age emerged.

Table 1. Favorability ratings of faces as a function of skin tone and facial physiognomy (*M*, *SE*, and *F* tests for simple effects of skin within each facial physiognomy level).

	Dark skin *M (SE)*	Medium skin *M (SE)*	Light skin *M (SE)*	*F*
Afrocentric physiognomy	45.06[a]	52.22[a]	49.31	$F(2, 103) = 4.95$,
	(2.37)	(1.94)	(2.06)	$p = .009$, $\eta_p^2 = .09$
Mixed physiognomy	48.21[a]	55.02[a]	53.82	$F(2, 103) = 4.94$,
	(2.27)	(1.81)	(1.85)	$p = .009$, $\eta_p^2 = .09$
Eurocentric physiognomy	42.04[a,b]	54.28[b]	55.39[a]	$F(2, 103) = 15.02$,
	(2.30)	(1.81)	(2.30)	$p < .001$, $\eta_p^2 = .23$

[a]These comparisons within each Facial Physiognomy are significant at the .05 level.
[b]These comparisons within each Facial Physiognomy are significant at the .05 level.

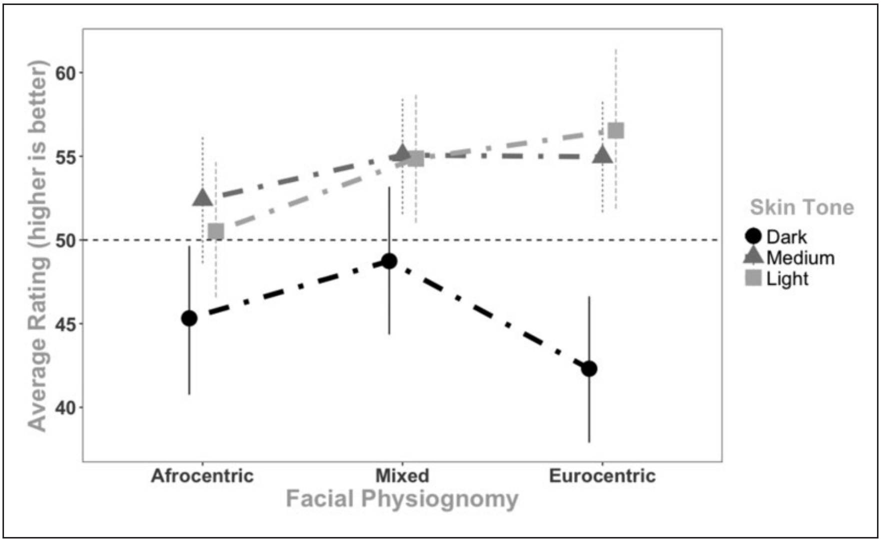

Figure 3. Average self-reported favorability ratings of faces as a function of facial physiognomy and skin tone. Error bars are standard error of the mean; dashed line reflects chance responding.

AMP

No significant main effects for Skin Tone and Facial Physiognomy emerged, $F(2, 208) = .07$, $p = .93$, $\eta_p^2 = .001$ and $F(2, 208) = 1.08$, $p = .34$, $\eta_p^2 = .01$, correspondingly. A main effect for Race emerged, $F(1, 104) = 5.78$, $p = .02$, $\eta_p^2 = .053$, with White children having a higher proportion of "good" responses ($M = .53$, $SE = .02$) than non-White children ($M = .48$, $SE = .02$).

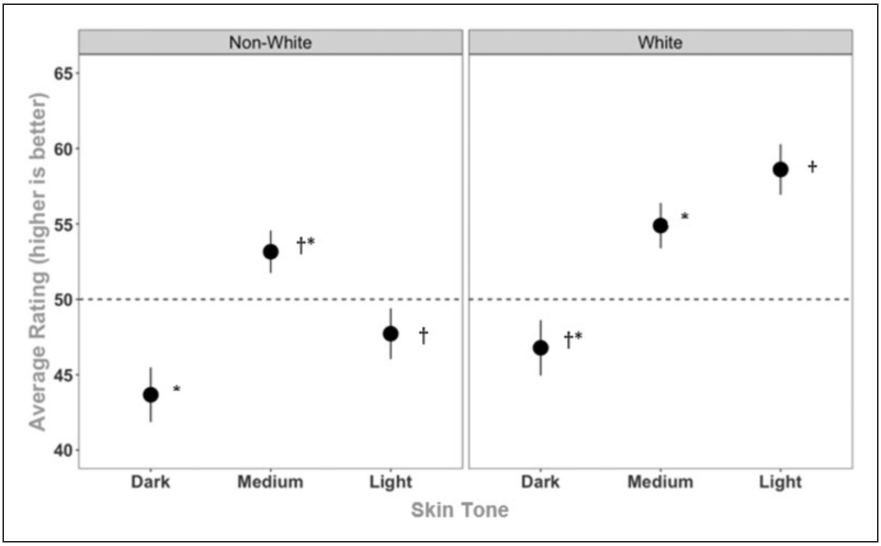

Figure 4. Average self-reported favorability ratings of faces as a function of skin tone, faceted by participant race. Error bars are standard error of the mean; dashed line reflects chance responding. *These comparisons within each Race are significant at the.05 level. †These comparisons within each Race are significant at the .05 level.

There was also a marginally significant Skin Tone × Facial Physiognomy interaction, $F(4,\ 416) = 2.23$, $p = .066$, $\eta_p^2 = .02$, further qualified by a Facial Physiognomy × Skin Tone × Race interaction, $F(4,\ 416) = 2.53$, $p = .04$, $\eta_p^2 = .02$. To understand the nature of this three-way interaction, we examined an effect of Race within each combination of Skin Tone and Facial Physiognomy (see Table 2). Specifically, for Afrocentric facial physiognomy primes with medium skin tone, a proportion of "good" responses was lower in non-White than White children. For mixed facial physiognomy primes with dark and medium skin tone, a proportion of "good" responses was also lower in non-White than White participants (see Figure 5 and Table 2). Age did not figure in any significant effects, with the strongest (though still nonsignificant) trend being the Skin × Age Group interaction, $F(2,\ 208) = 2.16$, $p = .12$, $\eta_p^2 = .02$, suggesting that younger children might have been somewhat less biased against dark skin tone faces (see Table 3).

To further examine a relative importance of both factors (Skin Tone and Facial Physiognomy) in the ERT and AMP, we examined the proportions of variance (η_p^2) that each factor accounts for at levels of the other factor (see Figure 6). The proportions of the variance explained for simple effects of skin tone within each level of physiognomy, and, the proportions of the variance

Table 2. Proportion of "Good" responses in the affective priming task as a function of skin tone and facial physiognomy of primes and race of participants (*M*, *SE*, and associated *F* tests for the effects of race within each level combination of facial physiognomy and skin tone).

	Race				
	White		Non-White		
	M	*SE*	*M*	*SE*	*F*
Facial physiognomy					
Skin tone					
Afrocentric					
Dark	.52	.03	.50	.04	$F(1, 104) = .16, p = .69, \eta_p^2 = .002$
Medium	.51[a]	.03	.41[a]	.03	$F(1, 104) = 4.96, p = .03, \eta_p^2 = .05$
Light	.55	.03	.45	.04	$F(1, 104) = 3.66, p = .06, \eta_p^2 = .03$
Mixed					
Dark	.58[a]	.03	.45[a]	.03	$F(1, 104) = 8.20, p = .005, \eta_p^2 = .07$
Medium	.56[a]	.03	.47[a]	.03	$F(1, 104) = 4.34, p = .04, \eta_p^2 = .04$
Light	.50	.03	.55	.04	$F(1, 104) = 1.10, p = .30, \eta_p^2 = .01$
Eurocentric					
Dark	.50	.03	.46	.04	$F(1, 104) = .65, p = .42, \eta_p^2 = .006$
Medium	.56	.03	.55	.03	$F(1, 104) = .15, p = .70, \eta_p^2 = .001$
Light	.53	.04	.47	.04	$F(1, 104) = .98, p = .32, \eta_p^2 = .009$

[a]White to non-White comparisons significant at the .05 level.

explained for simple effects of physiognomy within each level of skin tone were plotted for performance in each of the tasks. While for the ERT, the effects of skin tone at the levels of facial physiognomy were generally larger than (or similar to in one instance) the effects of physiognomy at the levels of skin tone, no clear pattern emerged for the AMP.[2]

Relation between ERT and AMP measures

We also explored whether participants' responses in the ERT and the AMP were related within each skin tone and facial physiognomy combinations (i.e., the liking score was correlated with the proportion of "good" responses for each of the nine unique combinations of skin tone and facial physiognomy) based on the final sample of 108 participants; see Table 4. In general, ratings on one measure were positively related to ratings on the other measure, but this trend was significant for only three of nine skin tone/facial physiognomy combinations (Afrocentric features with dark and medium skin tone and mixed features with dark skin tone). The average correlation for nine skin tone/facial physiognomy combinations was .13, suggesting that implicit and explicit evaluations are only weakly related.

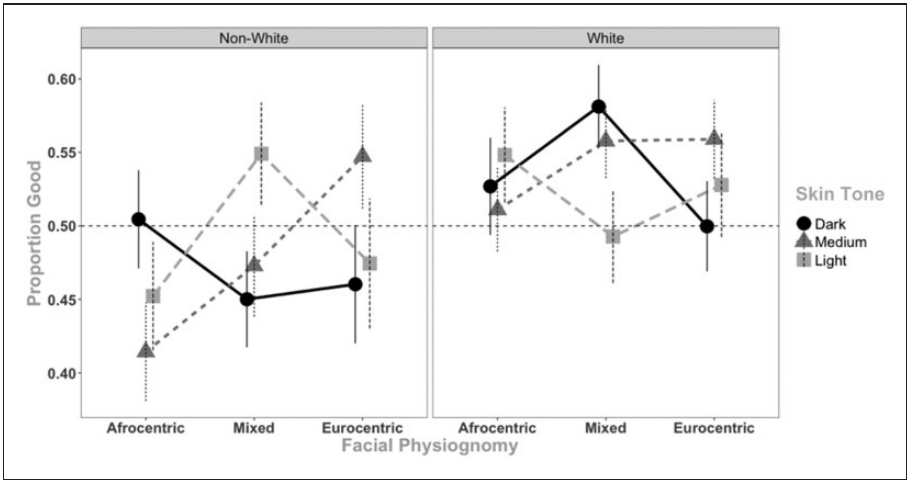

Figure 5. Proportion of positive interpretations of targets in the AMP as a function of facial physiognomy and skin tone, faceted by participant race. Error bars are standard error of the mean; dashed line reflects chance responding.

Table 3. Proportion of "Good" responses in the Affective priming task as a function of skin tone and age of participants (*M*, *SE*, and *F* tests for simple effects of age within each skin tone level).

	Younger children (5–7)		Older children (8–12)		
Skin tone	*M*	*SE*	*M*	*SE*	*F*
Dark	.52	.02	.49	.02	$F(1,\ 104) = 1.32$, $p = .25$, $\eta_p^2 = .01$
Medium	.49	.02	.53	.02	$F(1,\ 104) = 1.53$, $p = .22$, $\eta_p^2 = .01$
Light	.49	.03	.52	.03	$F(1,\ 104) = .77$, $p = .38$, $\eta_p^2 = .01$

Discussion

Racial attitudes assessed with an explicit measure

The results of this study support earlier findings that children develop negative racial attitudes assessed by explicit measures from early ages, at least by the time they attend kindergarten. Overall, children's racial evaluations were heavily influenced by skin tone, with darker skin tone faces evaluated most negatively by both White and non-White participants. Furthermore, racial differences emerged for the ERT, with non-White children rating medium skin tone faces as the most favorable, and White children rating light skin tone faces as the most favorable.

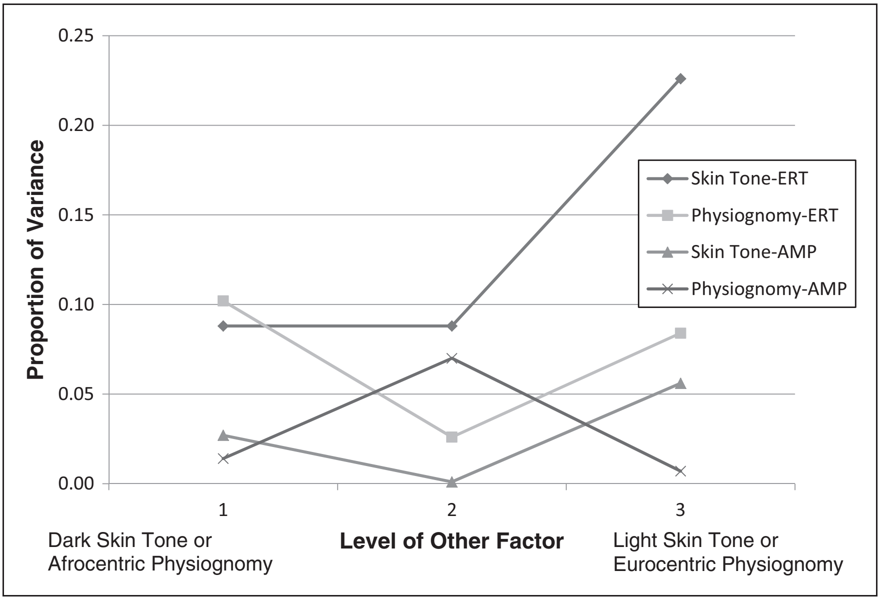

Figure 6. Proportion of variance explained (η_p^2) for simple effects of skin tone within levels of facial physiognomy and simple effects of facial physiognomy within skin tone, for the ERT and AMP. Skin tone levels varied from dark (1) to light (3). Facial physiognomy levels varied from Afrocentric (1) to Eurocentric (3).

Table 4. Correlations between proportion of good responses in the AMP and ERT ratings.

	Dark skin	Medium skin	Light skin
Afrocentric physiognomy	.28[a]	.21[b]	.13
Mixed physiognomy	.27[a]	.02	-.04
Eurocentric physiognomy	.09	.12	.05

[a]Correlation is significant at the .01 level.
[b]Correlation is significant at the .05 level.

Preference for medium skin tone in Black children has been previously observed (Porter, 1991). When asked what skin tone would make a hypothetical child happy with his/her skin tone and presented with six swatches of skin tone ranging from very dark brown to very light yellow, most Black American children between the ages of 6 and 13 picked "honey brown," which was in the middle range of skin tones, thus Porter's findings are consistent with results of our study. Based on extensive work on colorism, an internalized bias for lighter skin tone, preference for lighter skin tone is perpetuated by media and reinforced in schools (for review, see Ward et al., 2017). Thus, our findings might reflect White and non-White children's preference for certain social status and positive qualities associated with lighter skin tone. Yet, while dark skin tone was never preferred, children's own race/ethnicity was associated with their preferred skin tone in a meaningful way. Specifically, Black children preferred skin tone that, although lighter than dark tone, was darker than White children's preferences.

Importantly, this finding would not have been captured by measures that focus only on highly prototypical Black and White faces; on such a measure non-White children would probably have preferred White faces, missing the critical nuance that, when given the option, they show a stronger preference for medium skin tone faces.

What are some potential mechanisms of such findings? White children might have *a higher degree of exposure* to light skin tone than non-White children. Analogously, non-White children might have a higher degree of exposure to medium skin tone than their White counterparts. Thus, preference for light skin tone in White children and for medium skin tone in non-White children might indicate a familiarity effect.

While children tended to prefer lighter skin tone faces and faces with Eurocentric features overall, and skin tone and facial physiognomy also interacted to affect children's evaluations of faces. Specifically, Eurocentric faces and mixed physiognomy faces were perceived most favorably with lighter or medium skin tones. In contrast, Afrocentric faces were perceived most favorably with medium skin tone. Again, no faces were perceived favorably with dark skin tone. We cautiously interpret these findings as expectancy or matching effects based on participants' knowledge that Afrocentric features and darker skin tone *or* Eurocentric features and lighter skin tone usually cooccur. Yet, the absence of preference for dark skin tone when matched with Afrocentric physiognomy is certainly telling and an exception to this proposed account: dark skin tone is not explicitly favored by children even when it might be perceived as appropriately matching Afrocentric features.

Overall, while both skin tone and facial physiognomy contributed to explicit liking, skin tone played a more important role, as indicated by (a) more interactive effects involving skin tone, (b) a larger effect size for the skin tone main effect, and (c) a larger proportion of variance explained by skin tone within each level of physiognomy than by physiognomy within each level of skin tone (with only one instance when they were similar). These results dovetail with findings from prior explicit racial *categorization* tasks, in which skin tone also exerted a substantially larger effect on categorization decisions than did other facial physiognomy in both children and adults (Dunham et al., 2015; Stepanova & Strube, 2012a). Perhaps, the explicit nature of these judgments (liking and racial categorization) allowed us to replicate previous findings, thereby revealing close links between explicit categorization and explicit evaluative judgments.

Racial attitudes assessed with an implicit measure

Turning to the AMP, the current study indicated no main effects for either skin tone or facial physiognomy. Previous research employing implicit tasks assessing racial attitudes in adults (though, admittedly, not the AMP) found robust main effects for both skin tone and facial physiognomy (Hagiwara et al., 2012; Stepanova & Strube, 2012b). Perhaps, a unique task (the AMP) and stimuli differences from those previously employed in other studies precluded us from detecting such effects. Our examination of effect sizes associated with the skin tone and facial physiognomy effects at the levels of the other factor suggests, however, that at the implicit level, both factors did contribute somewhat to evaluative responses.

Somewhat surprisingly, non-White (majority of self-identified or identifiable Black heritage, 79%) children in this study showed more negative implicit biases toward faces with Afrocentric physiognomy and dark skin tones than did White children. We acknowledge a greater variability in the AMP data, as indicated by larger standard errors in relation to the corresponding means in the AMP than ERT, and therefore interpret these results cautiously. With that in mind, these results suggest that non-White participants have internalized a subtle form of negativity toward faces such as their own. Previous attitudinal research using adult samples has shown that pro-White biases at the implicit level are also documented in Blacks (Correll et al., 2002; Nosek et al., 2002, 2007). In addition, past developmental data do suggest that attitudes (assessed with implicit measures) favoring a high-status majority can emerge early, for example, in non-White South African children (Dunham et al., 2014; Newheiser et al., 2013). While past work with Black American children has generally suggested no relative preference in either direction for White versus Black comparisons (Dunham et al., 2013; Newheiser & Olson, 2012), the fact that the present research was conducted in the American South rather than the American Northeast could underlie this difference given the possibility that prevailing patterns of bias are stronger in the South, where protesters regularly waive Confederates flags in front of the main campus entrance at a local university. It is possible that our Black participants might have been overexposed to racial negativity, including in their own communities, as many Black families reinforce colorism (e.g., Wilder & Cain, 2011) leading to greater internalization of negative in-group attitudes than would not be observed elsewhere. It would therefore be valuable to conduct similar research in other regions of the United States and beyond.

Interestingly, results somewhat similar to ours were obtained in Black adults with the First Person Shooter Task (Study 2, Kahn & Davies, 2011) where more phenotypically Black targets (darker skin tone and more Afrocentric facial physiognomy) received a pattern of responses indicative of stronger implicit racial bias. Thus, our results might signify a faster developmental internalization of pro-White/anti-Black bias in Black and mixed children, especially for faces that more strongly portray Afrocentric features. However, further studies will need to confirm this effect.

Results of the AMP with White children were somewhat more difficult to interpret. One reason is perhaps because White children showed an overall positivity bias, judging considerably more characters as good than bad overall, regardless of skin tone and facial physiognomy. While our results were somewhat unexpected, recent work (Williams & Steele, 2019) has shown that 5 to 12-year-old White children do not show negativity when primed with outgroup Black primes in the AMP (but do show racial bias in the IAT). This, perhaps, underscores the fact that the AMP taps into more-specific feature-based rather than category biases.

In addition, while it did not achieve significance and so should be interpreted cautiously, younger children had more positive implicit evaluations of darker skin tones, while older children had more positive evaluations of lighter skin tones. This is suggestive of the emergence of increasingly negative attitudes over the age range tested here.

Racial attitudes assessed with the explicit versus implicit measures

The differences observed in children's responses to the ERT and the AMP, as well as the relatively low within-child correlation between them, may reflect the dichotomy between explicit and implicit processes widely discussed in the attitudinal literature (for review, see Bohner & Dickel, 2011; De Houwer, 2006; Fazio & Olson, 2003). On this model, attitudes assessed with implicit measures are based on repeated associations between pairing stimuli in the environment (e.g., as children are exposed to more "light = good" and "Eurocentric = good" portrayals). Our results suggest that internalizing these associations at the fine-grained level of individual facial physiognomy that was investigated here might take a longer time than internalizing the more general associations between valence and broad categories that have been explored in past work (reviewed in Dunham et al., 2008; Olson & Dunham, 2010). Supporting this contention, one study (Degner & Wentura, 2010; but see Williams & Steele, 2019, for a different pattern of results in the Canadian context) has found that implicit bias toward ethic outgroups in Europe emerged earlier when those attitudes were assessed at the level of ethnic categories with the IAT than at the level of individual faces with an evaluative priming task. Our task could be conceptualized as even subtler than a variant of task used by Degner and Wentura, given that we manipulated different aspects of facial physiognomy across multiple levels ranging across the entire Afrocentric to Eurocentric space. The somewhat unclear pattern with children, especially White children, could be due to this complexity and does suggest that adult-like sensitivity to features takes a long time to develop (cf. Dunham et al., 2015, 2016).

Limitations and future directions

Several limitations of the present work should be mentioned. First, there are several stimuli-specific limitations. We used male computer-generated face images rather than real-life photographs. We opted to use them in light of the previous work on the development of social categorization that could serve as a reference point (Dunham et al., 2015), and because it is more straightforward to vary features continuously using computer generated faces, but future researchers should consider employing photographs of both female and male faces, and finally, photographs of children. That said, it is unclear whether these changes would yield different results. For example, using photographs of young Black boys as primes in the Weapons Identification Task (Todd et al., 2016) produces similar racial bias effects in adult subjects when adult faces are used. Two major features that were examined in our work are skin tone and facial features; yet, it is certainly not the only way in which racial cues can be pulled apart, for example, physiognomy can be theoretically broken down further into an array of features (e.g., lips, nose, face and head shape, etc.). Such manipulations are worthy of further exploration.

Second, there are limitations related to our populations in question. We relied on children's racial self-identification and researcher's identification, and not on parental race identification. Encouragingly, researchers' identification matched children's identification when racial identification was available from children, suggesting that our proxy identification worked. It is also worth noting that while there was variability in skin tone and facial physiognomy in faces presented to children, we have not coded for such features in children themselves, instead relying on somewhat coarser reported and/or observed racial categorization. Examining how these finer-grained aspects of the participants' own appearances affect results could be an interesting avenue for future work. Furthermore, attitudes toward and among other racial groups should be investigated as well, especially given past work suggesting that children are somewhat more sensitive to facial physiognomy with some other category contrasts such as "White-Asian" (Dunham et al., 2016).

Third, we must consider potential experimenter effects. While we do not have data on which experimenter was assigned to which child, concerns could be raised about the experimenters' race or gender effects; it could be argued that White children might feel uncomfortable showing Black experimenters that they do not like faces with dark skin and/or Afrocentric features or Black children experiencing similar effects with light faces/faces with Eurocentric features in the presence of White experimenters. We attempted to ameliorate this problem by having experimenter pairs that were heterogonous on gender and race to maintain exposure to diversity in participants; yet, it was not always possible to do so due to logistical reasons. Yet, the presence of attitudinal biases in our data (especially ones obtained with the explicit measure) suggests that the experimenter effect is not strong enough to mask them. Experimenter effects in measures of implicit attitudes have not been previously reported in work on racial/ethnic biases in children, perhaps due to the fact that such measures might be less susceptible to conscious control and the effects of social desirability (though it is not clear to what extent and requires further investigations, Olson & Dunham, 2010). In addition, most recently, Sacco et al. (2019) reported no experimenter effects when the race of experimenter (Black or White) was manipulated when measuring racial biases using both explicit and implicit measures in children in Brazil, ages 6 to 14 years old.

Finally, our results could have been influenced by the unique cultural context of the American South, though this limitation could be also conceptualized as a strength of this study in that most prior research has focused on children from other regions of the United States, primarily the Northeast. While this work was not a true regional comparative study, it was conducted in an area where such studies have not been previously done. Since research in adults has found that racial biases are highest in more racially diverse areas, where exposure to racial outgroups is high (Rae et al., 2015), *replicating this work in other parts of the country might shed light onto unique developmental differences across regions*.

Conclusion

Explicit and implicit evaluations of faces are influenced by a complex interactive pattern of both skin tone and facial physiognomy as well as participant race. Skin tone exerts larger influences (at least in explicit evaluations), and darker skin tones are judged more negatively. From at least the elementary school years, children are building a map of the social world that penalizes physical features that are common in many non-White individuals, providing evidence of a disturbing aspect of early socialization in both White and non-White children.

Authors' Note

The authors would like to thank Principal Quan O'Neal at North Forrest Elementary, Principal Karen Vines at Dixie Attendance Center, and Lisa Blackwell at the Family YMCA of Southeast Mississippi for allowing them to collect data at their sites. The authors also acknowledge Cristobal Cruz, James Gaddy, Katherine Jackson, Kathleen Ramsey, Christine Miller, Joshua Polk, and Ja'Nisha Robinson for their assistance with data collection.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Support for this research was provided by an award from the Office of the Vice Provost for Research to the first author and by a grant from the Eagle Scholars Program for Undergraduate Research by the Drapeau Center for Undergraduate Research to the third author at The University of Southern Mississippi.

ORCID iD

Elena V. Stepanova https://orcid.org/0000-0003-2447-9717

Notes

1. We have considered employing a multilevel modeling approach with 10 levels of each of the 2 main predictors, age and race of participants plus potentially additional terms to account for the nonlinearity and found that such analyses are not easily interpretable. For the sake of simplicity, we decided to present results as is.
2. Please note that effect sizes for skin tone effects within all the Race × Facial Physiognomy and the physiognomy effects within all the Race × Skin Tone combinations were highly variable with no factor (i.e., Skin Tone and Physiognomy) consistently accounting for a higher proportion of the variance.

Supplemental Material

Supplemental material for this article is available online.

References

Aboud, F. E. (1988). *Children and prejudice*. Basil Blackwell.

Averhart, C. J., & Bigler, R. S. (1997). Shades of meaning: Skin tone, racial attitudes, and constructive memory in African American children. *Journal of Experimental Child Psychology*, *67*(3), 363–388. https://doi.org/10.1006/jecp.1997.2413

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*, 53–58. https://doi.org/10.1111/j.1467-9280.2005.01664.x

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. *Advances in Child Development and Behavior*, *34*, 39–89. https://doi.org/10.1016/S0065-2407(06)80004-2

Blair, I. V., Chapleau, K. M., & Judd, C. M. (2005). The use of Afrocentric features as cues for judgment in the presence of diagnostic information. *European Journal of Social Psychology*, *35*(1), 59–68. https://doi.org/ 10.1002/ejsp.232

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

Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, *87*, 763–778. https://doi.org/10.1037/0022-3514.87.6.763

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, *83*, 5–25. https://doi.org/10.1037/0022-3514.83.1.5

Bohner, G., & Dickel, N. (2011). Attitudes and attitude change. *Annual Review of Psychology*, 62391–62417. https://doi.org/10.1146/annurev.psych.121208.131609

Chen, J. M., de Paula Couto, M. C. P., Sacci, A., & Dunham, Y. (2018). To be or not to be (Black or Multiracial or White): Cultural variation in racial boundaries. *Social Psychological and Personality Science*, *9*(7), 763–772. https://doi.org/10.1177%2F1948550617725149

Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2001). Children's lay theories about ingroups and outgroups: Reconceptualizing re-search on prejudice. *Personality and Social Psychology Review*, *5*, 118–128.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A Meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, *16*(4), 330–350. https://doi.org/10.1177/1088868312440047

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

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

De Houwer, J. (2006). What are implicit measures and why are we using them? In R. W. Wiers, A. W. Stacy, R. W. Wiers, & A. W. Stacy (Eds.), *Handbook of implicit cognition and addiction* (pp. 11–28). SAGE. https://doi.org/10.4135/9781412976237.n2

De Houwer, J. (2009). Comparing measures of attitudes at the functional and procedural level: Analysis and implications. In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 361–390). Psychology Press.

Degner, J., & Wentura, D. (2010). Automatic prejudice in childhood and early adolescence. *Journal of Personality and Social Psychology*, *98*(3), 356–374. https://doi-org.lynx.lib.usm.edu/10.1037/a0017993

Dixon, T. L., & Maddox, K. B. (2005). Skin tone, crime news, and social reality judgments: Priming the stereotype of the dark and dangerous black criminal. *Journal of Applied Social Psychology*, *35*(8), 1555–1570. https://doi.org/10.1111/j.1559-1816.2005.tb02184.x

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2000). Reducing contemporary prejudice: Combating explicit and implicit bias at the individual and intergroup level. In S. Oskamp (Ed.), *"The Claremont Symposium on Applied Social Psychology" Reducing prejudice and discrimination* (pp. 137–163). Lawrence Erlbaum Associates.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540. https://doi.org/10.1006/jesp.1997.1331

Dunham, Y. (2011). An angry = Outgroup effect. *Journal of Experimental Social Psychology*, *47*(3), 668–671. https://doi.org/10.1016/j.jesp.2011.01.003

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. https://doi.org/10.1111/j.1467-8624.2006.00933.x

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*(7), 248–253. https://doi.org/10.1016/j.tics.2008.04.006

Dunham, Y., Chen, E. E., & Banaji, M. R. (2013). Two signatures of implicit intergroup attitudes: Developmental invariance and early enculturation. *Psychological Science*, *24*(6), 860–868. https://doi.org/10.1177/0956797612463081

Dunham, Y., & Emory, J. (2014). Of affect and ambiguity: The emergence of preference for arbitrary ingroups. *Journal of Social Issues*, *70*, 81–98. https://doi.org/10.1111/josi.12048

Dunham, Y., Dotsch, R., Clark, A. R., & Stepanova, E. V. (2016). The development of White-Asian categorization: Contributions from skin color and other physiognomic cues. *PLoS One*, *11*(6), e0158211. https://doi.org/10.1371/journal.pone.0158211

Dunham, Y., Newheiser, A. K., Hoosain, L., & Merrill, A. (2014). From a different vantage: Intergroup attitudes among children from low-and intermediate-status racial groups. *Social*, *32*(1), 1–21. https://doi.org/10.1521/soco.2014.32.1.1

Dunham, Y., Stepanova, E. V., Dotsch, R., & Todorov, A. (2015). The development of race-based perceptual categorization: Skin color dominates early category judgments. *Developmental Science*, *18*(3), 469–483. https://doi.org/10.1111/desc.12228

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthly: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*, 383–386. https://doi.org/10.1111/j.1467-9280.2006.01716.x

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, 54297–54327. https://doi.org/10.1146/annurev.psych.54.101601.145225

Gitter, A. G., & Satow, Y. (1969). Color and physiognomy as variables in racial misidentification among children. *Proceedings of the Annual Convention of the American Psychological Association*, *4*(Pt. 2), 677–678.

Gonzalez, A. M., Steele, J. R., & Baron, A. S. (2017). Reducing children's implicit racial bias through exposure to positive out-group exemplars. *Child Development*, *88*(1), 123–130. https://doi.org/10.1111/cdev.12582

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Hagiwara, N., Kashy, D. A., & Cesario, J. (2012). Independent effects of skin tone and facial features on Whites' affective reactions to Blacks. *Journal of Experimental Social Psychology*, *48*, 892–898. https://doi.org/10.1016/j.jesp.2012.02.001

Hirschfeld, L. A. (1995). Do children have a theory of race? *Cognition*, *54*(2), 209–252. https://doi.org/10.1016/0010-0277(95)91425-R

Hollinger, D. A. (2003). Amalgamation and hypodescent: The question of ethnoracial mixture in the history of the United States. *American Historical Review*, *108*, 1363–1390.

Inquisit 4 [Computer software]. (2015). *Precision software for cognitive, social, neurophysiological, and online psychological experiments*. http://www.millisecond.com

Kahn, K. B., & Davies, P. G. (2011). Differentially dangerous? Phenotypic racial stereotypicality increases implicit bias among ingroup and outgroup members. *Group Processes & Intergroup Relations*, *14*(4), 569–580. https://doi.org/10.1177/1368430210374609

Klonoff, E. A., & Landrine, H. (2000). Is skin color a marker for racial discrimination? Explaining the skin color–hypertension relationship. *Journal of Behavioral Medicine*, *23*(4), 329–338.

Lane, K. A., & Jost, J. T. (2011). Black man in the White House: Ideology and implicit racial bias in the age of Obama. In. G. S. Parks, M. W. Hughey, G. S. Parks, & M. W. Hughey (Eds.), *The Obamas and a (post) racial America*? (pp. 48–69). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199735204.003.0003

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, *82*(1), 5–18. https://doi.org/10.1037/0022-3514.82.1.5

Maddox, K. B. (2004). Perspectives on racial phenotypicality Bias. *Personality and Social Psychology Review*, *8*(4), 383–401. https://doi.org/10.1207/s15327957pspr0804_4

Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of Black Americans: Reexploring the role of skin tone. *Personality and Social Psychology Bulletin*, *28*, 250–259. https://doi.org/10.1177/0146167202282010

Newheiser, A.-K., & Olson, K. R. (2012). White and Black American children's implicit intergroup bias. *Journal of Experimental Social Psychology*, *48*(1), 264–270. http://doi.org/10.1016/j.jesp.2011.08.011

Newheiser, A.-K., Dunham, Y., Merrill, A., Hoosain, L., & Olson, K. R. (2013). Preference for high status predicts implicit outgroup bias among children from low-status groups. *Developmental Psychology*, http://doi.org/10.1037/a0035054

Nosek, B. A., Banaji, M., & Greenwald, A. (2002). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics*, *6*(1), 101–115. https://doi.org/10.1037/1089-2699.6.1.101

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 1836–1888. https://doi.org/10.1080/10463280701489053

Olson, K. R., & Dunham, Y. (2010). The development of implicit social cognition. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. Guilford Press.

Pauker, K., Williams, A., & Steele, J. R. (2017). The development of racial categorization in childhood. In A. Rutland, D. Nesdale, & C. S. Brown (Eds.), *The Wiley-Black handbook of group processes in children and adolescents*. John Wiley & Sons Ltd.

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., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277. https://doi.org/10.1037/0022-3514.89.3.277

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*, e12788. https://doi.org/10.1111/desc.12788

Qian, M. K., Heyman, G. D., Quinn, P. C., Messi, F. A., Fu, G., & Lee, K. (2016). Implicit racial biases in preschool children and adults from Asia and Africa. *Child Development*, *87*, 285–296. https://doi.org/10.1111/cdev.12442

Qian, M. K., Quinn, P. C., Heyman, G. D., Pascalis, O., Fu, G., & Lee, K. (2017). Perceptual individuation training (but not mere exposure) reduces implicit racial bias in preschool children. *Developmental Psychology*, *53*(5), 845–859. https://doi.org/10.1037/dev0000290

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*, 1715–1737. https://doi.org/10.1111/j.1467-8624.2011.01668.x

Rae, J. R., Newheiser, A.-K., & Olson, K. R. (2015). Exposure to racial out-groups and implicit race bias in the United States. *Social Psychological and Personality Science*, *6*(5), 535–543. https://doi.org/10.1177/1948550614567357

Porter, C. P. (1991). Social reasons for skin tone preferences of Black school-age children. *American Journal of Orthopsychiatry*, *61*(1), 149–154. https://doi.org/10.1037/h0079219

Sacco, A. M., de Paula Couto, M. C. P., Dunham, Y., Santana, J. P., Nunes, L. N., & Koller, S. H. (2019). Race attitudes in cultural context: The view from two Brazilian states. *Developmental Psychology*, *55*(6), 1299–1312. https://doi.org/10.1037/dev0000713

Schmidt, K., & Nosek, B. A. (2010). Implicit (and explicit) racial attitudes barely changed during Barack Obama's presidential campaign and early presidency. *Journal of Experimental Social Psychology*, *46*(2), 308–314. https://doi.org/10.1016/j.jesp.2009.12.003

Setoh, P., Lee, K. J., Zhang, L., Qian, M. K., Quinn, P. C., Heyman, G. D., & Lee, K. (2017). Racial categorization predicts implicit racial bias in preschool children. *Child Development*, https://doi.org/10.1111/cdev.12851

Skinner, A. L., & Cheadle, J. E. (2016). The 'Obama effect'? Priming contemporary racial milestones increases implicit racial bias among whites. *Social Cognition*, *34*(6), 544–558. https://doi.org/10.1521/soco.2016.34.6.544

Skinner, A. C., Payne, K., Perrin, A. J., Panter, A. T., Howard, J. B., Bardone-Cone, A., Bulik, C. M., Steiner, M. J., & Perrin, E. M. (2017). Implicit weight bias in children age 9 to 11 years. *Pediatrics*, *140*(1), e20163936. https://doi.org/10.1542/peds.2016-3936

Stepanova, E. V., & Strube, M. J. (2009). Making of a face: Role of facial physiognomy, skin tone, and color presentation mode in evaluations of racial typicality. *Journal of Social Psychology*, *149* (1), 66–81. https://doi.org/10.3200/SOCP.149.1.66-81

Stepanova, E. V., & Strube, M. J. (2012a). The role of skin color and facial physiognomy in racial categorization: Moderation by implicit racial attitudes. *Journal of Experimental Social Psychology*, *48* (4), 867–878. https://doi.org/10.1016/j.jesp.2012.02.019

Stepanova, E. V., & Strube, M. J. (2012b). What's in a face? The role of skin tone, facial physiognomy, and color presentation mode of facial primes in affective priming effects. *The Journal of Social Psychology*, *152*(2), 212–227. https://doi.org/10.1080/00224545.2011.597797

Stepanova, E. V., & Strube, M. J. (2018). Attractiveness as a function of skin tone and facial features: Evidence from categorization studies. *The Journal of General Psychology*, *145*, 1–20. https://doi.org/10.1080/00221309.2017.1394811

Stepanova, E. V., Strube, M. J., & Yablonsky, G. S. (2013). Now one is Russian, now one is not: Ethnic attitudes and categorization effects in the Russian Federation. *International Perspectives in Psychology: Research, Practice, Consultation*, *2*(2), 85–99. https://doi.org/10.1037/a0032339

Todd, A. R., Thiem, K. C., & Neel, R. (2016). Does seeing faces of young Black boys facilitate the identification of threatening stimuli? *Psychological Science*, *27*(3), 384–393. https://doi.org/10.1177/0956797615624492

Ward, J. V., Robinson-Wood, T. L., & Boadi, N. (2017). Resisting everyday colorism in schools: Strategies for identifying and interrupting the problem that won't go away. In C. Monroe (Ed.), *Race and colorism in education* (pp. 5–23). Routledge.

Wilder, J., & Cain, C. (2011). Teaching and learning color consciousness in Black families: Exploring family processes and women's experiences with colorism. *Journal of Family Issues*, *32*(5), 577–604. https://doi.org/10.1177/0192513X10390858

Williams, A., & Steele, J. R. (2019). Examining children's implicit racial attitudes using exemplar and category-based measures. *Child Development*, *90*, e322–e338. https://doi.org/10.1111/cdev.12991

Williams, A., Steele, J. R., & Lipman, C. (2016). Assessing children's implicit attitudes using the Affect Misattribution Procedure. *Journal of Cognition and Development*, *17*(3), 505–525. https://doi.org/10.1080/15248372.2015.1061527

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*(5), 815–827. https://doi.org/10.1037/0022-3514.81.5.815

Xiao, W. S., Fu, G., Quinn, P. C., Qin, J., Tanaka, J. W., Pascalis, O., & Lee, K. (2015). Individuation training with other-race faces reduces preschoolers' implicit racial bias: A link between perceptual and social representation of faces in children. *Developmental Science*, *18*(4), 655–663. https://doi.org/10.1111/desc.12241

Author Biographies

Elena V. Stepanova is an Associate Professor of Psychology and the director of Social Cognition and Behavior Lab in the School of Psychology at the University of Southern Mississippi. Her broad areas of interest are social cognition, specifically conceptualization of social categories, and cross-cultural research, particularly explicit and implicit ethnic and racial attitudes.

Yarrow Dunham is an Associate Professor of Psychology and Cognitive Science and the director of the Social Cognitive Development Lab at Yale University. His research focuses on studying how knowledge of social groups is acquired, both in cognitively mature adults and in developing children. His research combines a range of experimental and cross-cultural methodologies.

Maya Rex is currently in her third year at Tulane Law focusing on criminal indigent defense. Her research interests include social psychology, immigration, race psychology, identity, and abolition.

Nao Hagiwara is an Associate Professor of Health Psychology and Social Psychology and the director of Discrimination and Health Research Lab at Virginia Commonwealth University. Her overall research interest is to better understand the intrapersonal and interpersonal processes involved in intergroup bias from the perspectives of both the targets and the instigators of social inequality.