Published in final edited form as:

J Exp Child Psychol. 2014 October ; 126: 1–18. doi:10.1016/j.jecp.2014.02.009.

Children's Classification and Lexicalization of Attractiveness, Gender, and Race: Differential Displays of These Concepts and Relatedness to Bias and Flexibility

Jennifer L. Rennels¹ and Judith H. Langlois²

¹University of Nevada, Las Vegas

²The University of Texas at Austin

Abstract

Minimal research has examined children's functional use of attractiveness to classify and label others, an important step in the development of children's biases. This study compared 3- to 11-year-olds' classification, sorting, and labeling of others and themselves based on attractiveness, gender, and race, and investigated whether these abilities and other characteristics predicted children's bias and flexibility. Relative to gender and race, children rarely used attractiveness to spontaneously classify people and were less accurate at sorting and labeling others and themselves by attractiveness, suggesting they have a less explicit concept of attractiveness. Predictors of bias differed depending upon domain and assessment method (forced choice or non-forced choice), showing that children's bias is affected by both individual differences and task characteristics. Predictors of flexibility differed based on whether children were assigning positive or negative traits to target children, demonstrating that the valence of attributes is an important consideration when conceptualizing children's flexibility.

Keywords

attractiveness; gender; race; classification; bias; flexibility

Children's biases based on gender, race, and attractiveness are pervasive and evident as early as 3–4 years of age (e.g., Bigler & Liben, 1993; Dion, 1973; Martin, 1989). Developmental intergroup theory proposes such biases develop because: 1) these cues are perceptually and psychologically salient ways for children to classify others; 2) children functionally use such classifications during social encounters; and 3) children link attributes and affective responses to groups based on these classifications (Bigler & Liben, 2006). In support of the first process, there is evidence that children can categorize others based on all

© 2014 Elsevier Inc. All rights reserved.

Corresponding author: Jennifer L. Rennels, Department of Psychology, 4505 S. Maryland Pkwy, Box 455030, Las Vegas, NV 89154-5030. Jennifer.rennels@unlv.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

three domains as early as infancy (Anzures, Quinn, Pascalis, Slater, & Lee, 2010; Ramsey, Langlois, Hoss, Rubenstein, & Griffin, 2004; Younger & Fearing, 1999). Children's differential assignment of positive and negative attributes based on each of the three domains provides support for the third process (e.g., Bigler & Liben, 1993; Dion, 1973; Powlishta, 1995). For the second process, researchers have examined children's functional classification of others' gender and race via assessments of classification skill and labeling (Bigler & Liben, 1992, 1993; Lam, Guerrero, Damree, & Enesco, 2011; O'Brien et al., 2000; Pauker, Ambady, & Apfelbaum, 2010; Weinraub et al., 1984), but no one has investigated children's spontaneous classification of others' attractiveness and minimal research has examined children's attractiveness labeling skills (Dion, 1973). One goal of this research was to assess and compare children's spontaneous classification and explicit labeling and sorting of individuals based on attractiveness, gender, and race.

Young children often classify others based on gender and race (Bigler & Liben, 1993; Lam et al., 2011; Pauker et al., 2010). Caucasian 4- to 9-year-olds in the United States more commonly used gender, followed by race, to sort faces differing in age, emotional expression, gender, and race (Bigler & Liben, 1993). Caucasian, African, and Asian 3- to 5-year-olds in Britain more commonly used gender than race to spontaneously describe pictures of children varying in clothing color, gender, and race (Lam et al., 2011). Children's use of gender to classify others decreases with age (Serbin & Sprakin, 1986) and their use of race to classify others increases with age (Lam et al., 2011; Pauker et al., 2010).

It is unclear, however, if facial attractiveness is as salient to children as gender and race. The differential treatment attractive and unattractive people receive and the differential behaviors they display can be observed during social interactions (Langlois et al., 2000). Fairy tales and children's media portray glaring distinctions between attractive and unattractive people, which affects children's liking of those characters (Hoffner, 1996). Children may detect these correlations and use attractiveness to classify people in their social world. On the other hand, there are not separate bathrooms for attractive and unattractive people as there are for females and males, and attractive and unattractive people are not segregated into disparate neighborhoods unlike different race individuals. The relatively low functional use of attractiveness in terms of real world institutionalized practices may make it less salient than gender and race and, therefore, the least likely way that children group others during classification tasks.

We were interested not only in children's spontaneous classification of others, but also how they would perform when explicitly requested to sort and label pictures of peers based on gender, race, and attractiveness and to self-identify these attributes. Most children spontaneously produce gender terms by 21 months (Zosuls et al., 2009), correctly produce gender labels for pictures of adults and children by 26–36 months (O'Brien et al., 2000; Weinraub et al., 1984), and accurately point to a pictured girl or boy by 28–36 months (Fagot, Leinbach, & Hagan, 1986; Fagot, Leinbach, & O'Boyle, 1992). Sorting faces based on gender is evident in most children by 36 months (Weinraub et al., 1984). Children's identification of their own gender often precedes the ability to identify others' gender (Katz & Kofkin, 1997).

Compared to gender labeling, children's accuracy in labeling their own and others' race emerges later in development, perhaps because parents mention race substantially less than gender when interacting with their children (Katz & Kofkin, 1997). At 3–5-years-old, 60% of children accurately sorted targets based on racial labels, but accuracy was dependent on race: African children were more proficient than Caucasian children who were more proficient than Asian children (Lam et al., 2011). By 8–11-years-old, children showed high accuracy (91% or greater) of matching pictures to racial labels (Dutton, Singer, & Devlin, 1998).

Although attractiveness labeling may not seem as objective as gender and race labeling, there is high agreement among adults and children about who is and is not attractive (Dion, 1973; Langlois et al., 2000). Using faces adults rated for attractiveness, 3–7-year-olds' choices of which peer within a pair was more attractive based on labels, such as *pretty*, *cute*, or *handsome*, were significantly above chance level (Dion, 1973; Styczynski & Langlois, 1977). Six- and 10-year-olds also showed strong agreement with adults when choosing one of three different sized boxes to signify how pretty or handsome a peer was (Langlois & Stephan, 1977). No one, however, has examined children's proficiency at sorting and labeling peers' attractiveness and minimal research has examined children's identification of their own attractiveness.

Both classification and lexicalization (i.e., having a word for an item or person) are linked to formation of concepts (Gelman, 2009). Attaching a specific label to a person frequently results in children inferring traits about that individual (Gelman & Markman, 1986), and is often indicative of a belief that persons with that label share common, stable characteristics that differ from those of persons with a divergent label (Gelman, 2009; Gelman & Heyman, 1999). Children who accurately classify and label the gender of others are more sex-typed, have greater knowledge of gender stereotypes, and prefer same-sex playmates compared to children who cannot accurately label the gender of others (Fagot et al., 1986; Fagot et al., 1992; Zosuls et al., 2009). Naturalistic observations showed that 3–5-year-olds who used racial labels made stereotypic attributions about the children they were labeling (Van Ausdale & Feagin, 1996). Furthermore, functional use of race to sort faces during a classification task significantly predicted 3–10-years-olds' racial stereotyping (Pauker et al., 2010). Children's consistent use of the same label for all group members may increase their perceptions of intragroup similarity and subsequently escalate their biases (Katz, 1973). Also, children often show in-group biases (e.g., Powlishta, 1995), so self-categorization can influence biases.

Functionally using attractiveness to classify and label others and oneself might also influence children's attractiveness biases, but no one has investigated this possibility. Another goal of this study was to investigate whether these abilities were predictive of children showing positive and negative biases in each domain or showing flexibility. Flexibility refers to children's understanding that members of different social groups can possess similar attributes and can be tested by providing children with the option of stating "both" (e.g., both boys and girls) when asking them to assign attributes to others (Powlishta, Serbin, Doyle, & White, 2004; Serbin & Sprafkin, 1986). Using data with which half our participants were forced to choose a child from one of two social groups (forced choice

assessment) and the other half were also given the options of choosing “both” or “neither” (non-forced choice assessment) when assigning attributes, we explored predictors of bias and flexibility.

Not only are the dimensions with which children classify and labels others related to bias, but so are the number of ways with which they classify others. Children who classify persons on more than one dimension show less bias based on gender and race (Bigler, 1995; Bigler & Liben, 1992, 1993) and may also show less attractiveness bias. If children can group people on multiple dimensions (e.g., gender and hair color), they are less likely to focus solely on a particular characteristic that defines a group to which that person belongs, and thus not rely as heavily on group-related biases. This ability, possessing multiple classification skills, may reflect an environment that encourages flexible thinking about others and tends to increase with age (Bigler & Liben, 1990, 1992, 1993).

The present study compared 3–11-year-olds’ spontaneous categorization, and sorting and labeling of others and their own attractiveness, gender, and race. We examined whether these abilities predicted children’s bias in these domains and their flexibility. Children show evidence of labeling or discriminating between others based on all three of these attributes as early as 3–4 years of age (Dion, 1973; Lam et al., 2011; Weinraub et al., 1984), and their knowledge of social groups and ability to categorize on more than one dimension increases from early to middle childhood (Banse, Gawronski, Rebetez, Gutt, & Morton, 2010; Bigler & Liben, 1990, 1992, 1993), so children within this age range participated.

Consistent with previous research (Pauker et al., 2010; Serbin & Sprafkin, 1986), children should show age-related decreases in classifying by gender and age-related increases in classifying race, but we were unsure whether they would use attractiveness to group others during the classification task. For the sorting and labeling tasks, we expected children to perform best on gender sorting and labeling given the early emergence of this ability in children (Katz & Kofkin, 1997). We were uncertain as to whether children would show similar or divergent levels of sorting and labeling race and attractiveness. People are often uncomfortable discussing race (Katz & Kofkin, 1997) and maxims of beauty (“don’t judge a book by its cover”) may discourage use of explicit labels (Langlois et al., 2000). We did expect there might be individual differences in children’s sorting and labeling abilities. African American children might be better than European American children at sorting and labeling race (Lam et al., 2011) because of its greater relevance to their identity. Girls might be better at sorting and labeling others’ attractiveness than boys because they have a better concept of attractiveness than boys (Dion, 1973), perhaps because appearance is a strong component of children’s gender stereotypes for girls (Miller, Lurye, Zosuls, & Ruble, 2009; Ramsey & Langlois, 2002) and girls have more knowledge of gender stereotypes than boys (O’Brien et al., 2000). Within domains, we expected children’s functional use of an attribute to classify others and their labeling and sorting abilities to positively predict bias, and multiple classification skills to negatively predict bias. In contrast, functional classification, sorting, and labeling abilities might negatively predict flexibility and multiple classification skills might positively predict flexibility.

Method

Participants

Data were from 103 children aged 3 years, 4 months to 11 years, 6 months ($M = 6.27$ years, $SD = 2.01$; 53 females) who also participated in Rennels and Langlois' (2014) study. Research assistants contacted local schools and families via letters and phone calls to recruit their participation. Parents provided written consent for their child to participate and children provided written assent (if 7 years of age or older) or verbal assent to participate. School administrators provided approval for data to be collected at the schools. An additional three children were recruited, but provided no data due to young age or choosing not to participate.

The racial composition of the children was: 51 European American; 27 mixed race; 19 African American; 4 Asian American; and 2 unreported. Eighteen were of Spanish, Hispanic, Mexican, or Latino ethnic background. Most mixed race children were part European American, but varied greatly in terms of their other racial background. Because children of mixed race who are part European American identify most commonly with their minority background when forced to choose a single race or ethnicity (Herman, 2004), we coded mixed race children as belonging to those particular groups (African American, Asian American, or Latino).

Materials

Stimulus faces—Stimuli included digitized, color facial images of 38 children (4–7-year-olds) and eight adults (18–25-year-olds) within the normal range of attractiveness selected from photo databases at two different universities. Racially heterogeneous groups of at least 40 undergraduates, with relatively equal numbers of females and males, rated the images for attractiveness using a 1–5 Likert Scale ($1 = \text{not very attractive}$; $5 = \text{very attractive}$). These ratings were highly reliable (α s $> .90$).

Classification task—This task included 16 3" x 3" laminated facial images. There was one high attractive and one low attractive face for each of the following face types: African American female adults, European American female adults, African American male adults, European American male adults, African American female children, European American female children, African American male children, and European American male children. Adults rated the low attractive faces, $M_{att} = 1.96$, $SD = .38$, as significantly less attractive than the high attractive faces, $M_{att} = 3.56$, $SD = .34$, $t(14) = -8.87$, $p < .0001$. Unsystematic variations among the faces included hair color, eye color, and expression, so there were several dimensions for children to classify the faces.

Sorting and labeling tasks—For the attractiveness, gender, and race sorting and labeling tasks, we used 32 children's faces (two were used in the classification task). For attractiveness sorting and labeling, children saw either eight African American or eight European American faces (face race matched the child's race or was assigned randomly to children who were not African American or European American). For each race, there were four attractive (two female, two male), $M_{att} = 4.04$, $SD = .24$, and four unattractive (two

female, two male) faces, $M_{att}=1.89$, $SD = .26$, that significantly differed in attractiveness, $t(14) = -17.27$, $p < .0001$, but displayed similar emotional expressions. For gender sorting and labeling, there were four female (two African American, two European American) and four male (two African American, two European American) faces similar in attractiveness, $M_{att} = 2.68$, $SD = .17$ for females; $M_{att} = 2.86$, $SD = .25$ for males, $t(6) = -1.20$, $p = .28$. For race sorting and labeling, there were four African American (two female, two male) and four European American (two female, two male) faces similar in attractiveness, $M_{att} = 3.00$, $SD = .43$ for African Americans; $M_{att} = 3.09$, $SD = .46$ for European Americans, $t(6) = -.30$, $p = .77$.

Self-identification tasks—We used children's own photo for the gender self-identification task (Weinraub et al., 1984). For race self-identification, we used a color strip with six different skin tones ranging from lighter to darker (Averhart & Bigler, 1997). For attractiveness self-identification, we used a poster board displaying three different-sized rectangles (one small, one medium, and one large) and a mirror for the child to view him/herself (Downs & Reagan, 1983).

Bias tasks—There were eight picture boards displaying two children's faces shown separately (forced choice assessment) and another eight picture boards displaying the same two faces shown separately, the two faces shown together, and a picture with the two faces cut out and a large X through the picture (non-forced choice assessment). The two faces displayed on these picture boards differed in attractiveness, gender, or race with the other cues and emotional expression held constant. There were two versions each (one displaying African American children and one displaying European American children) of picture boards showing boys differing in attractiveness, girls differing in attractiveness, and children differing in gender. There were two versions (one displaying girls and one displaying boys) of picture boards showing children differing in race. The 16 faces used in this task were also used in the attractiveness, gender, and race sorting and labeling tasks.

Procedure

To begin, each child provided a spontaneous self-description. Next, the experimenter told the child they were going to play several games, she would ask questions, and there were no right or wrong answers. The child then individually completed the classification task and two bias tasks during session 1. The session ended with the experimenter taking the child's picture, if permitted. During session 2 (which occurred within four weeks after session 1), the child completed two more bias tasks, and the sorting, labeling, and self-identification of gender, race, and attractiveness tasks. Each child then provided another spontaneous self-description to determine if completing the tasks caused an increase in the child's use of attractiveness, gender, or race to describe him/herself. Last, because the tasks focused on appearance-based traits, experimenters concluded the study with a follow-up discussion meant to direct the child's attention toward inner traits and behaviors.

Spontaneous self-descriptions—Prior to and at the end of data collection, the child responded to the inquiry, "Tell me about yourself. Tell me what you are like" (Rotenberg, 1982). If the child had any difficulty responding to this inquiry, the experimenter stated,

“Tell me whatever you want about yourself.” Once the child completed the self-description task, the experimenter asked, “Is there anything else you want to tell me about yourself?”

Classification task—First, children practiced sorting with 16 blocks that differed in size, shape, and color. The experimenter asked, “Can you put the blocks that you think go together into one pile, and the other blocks that go together into a different pile?” If the child sorted the blocks, the experimenter asked for a justification. Then, the experimenter mixed up the blocks again, asked the child if there was another way to sort the blocks, and asked for a justification if the child sorted a second time. If the child was unable to sort the blocks either time, the experimenter asked questions related to the blocks’ size, shape, or color to help the child notice similarities among the blocks. This practice ensured children saw that the blocks could be sorted at least two different ways.

For the classification task (Bigler & Liben, 1993), the experimenter randomly placed the 16 faces in front of the child and told the child that this game was just like the one with blocks except there were pictures of people—“Can you put the people that you think go together into one pile, and the other people that go together into a different pile?” If the child sorted, the experimenter asked for a justification, mixed up the photos again, asked the child if there was another way to sort the people, and asked for a justification if the child sorted a second time. If the child was unable to sort at any point, the experimenter moved onto the next task. We assigned scores of 0 if children were unable to sort, 1 if they sorted on one dimension, and 2 if they sorted on two dimensions.

Bias tasks—Children were randomly assigned to either the forced choice or non-forced choice assessment with the constraint that the gender, race, and age distribution of the children was relatively similar across assessment method. Children completed four bias tasks in which they attributed four positive and four negative traits and behaviors to the children displayed on the picture boards. The experimenter asked, “Who do you think is _____ (e.g., nice/nice to other people, naughty/does bad things)?” and requested children to point to one of the two children (forced choice assessment) or to one of the two children, both children, or neither child (non-forced choice assessment). We used different positive and negative attributes for each picture board, randomly assigned and counterbalanced across children. The 32 traits and behaviors were chosen from studies examining attractiveness, gender, or race biases with children ranging in age from 3–12 years (Bigler & Liben, 1993; Dion, 1973; Doyle, Beaudet, & Aboud, 1988; Langlois & Stephan, 1977; Powlishta, 1995; Williams, Bennett, & Best, 1975).

Sorting and labeling others—Children completed the gender, race, and attractiveness sorting and labeling tasks in a randomly assigned order. To determine each child’s sorting ability, the experimenter laid out eight pictures. For gender sorting, the experimenter stated, “Some of these children are boys and some are girls. Can you help me put the girls in one pile and the boys in another pile?” The experimenter took out clip art showing two girls (one African American, one European American), placed it down, and pointed to it saying, “This is the pile for the girls. All the pictures of girls go in this pile.” The experimenter did the same with clip art showing two boys (one African American, one European American). For each picture, the experimenter pointed to it and stated, “Look at this person. Which pile does

this picture go in? Put it in the pile where it goes.” Once the child was done sorting, the experimenter combined the two piles, shuffled the pictures, presented each one individually, and asked, “What kind of person is this?” to assess labeling abilities (Weinraub et al., 1984).

The process was similar for race and attractiveness sorting and labeling, except the experimenter used the terms Black/African American and White/Caucasian or good-looking/look pretty/look handsome and not good-looking/do not look pretty/do not look handsome, respectively, to direct children’s sorting. Clip art was also used for race sorting (an African American boy and a girl paired together and a European American boy and girl paired together), but not for attractiveness sorting because we could not find suitable clip art. Instead the experimenter pointed to a place for children to put the high and low attractive faces and reminded children what the piles were during the task.

Self-identification tasks—For self-sorting of gender, the experimenter showed a picture of the child during the gender sorting task and asked, “Where does this picture go?” For self-labeling of gender, the picture was also included in the pile when the experimenter asked, “What kind of person is this?” (Weinraub et al., 1984).

For self-identification of race (skin tone), the experimenter placed the color strip in front of the child and said, “These colors are some of the very many beautiful skin colors that people have. Can you point to the color that looks most like your skin color?” (Averhart & Bigler, 1997). The experimenter noted on a data sheet, but not to the child, which skin color she thought matched the child and a second coder later used the child’s picture to do the same.

For self-identification of attractiveness, the experimenter showed the child the poster board with three different sized rectangles. The experimenter explained the child could answer questions by pointing to one of the blocks (large block for *very much* [2], medium block for *somewhat* [1], and small block for *not very much* [0]). Then the experimenter pulled out a mirror and ensured the child could see him/herself in the mirror. The experimenter asked, “How strong are you?” so the child could practice using the board. She then asked, “How pretty (for girls)/good-looking (for boys) are you?” reiterating what each of the blocks represented if necessary. Test-retest reliability for this procedure is good, $r = .61$ (Downs & Reagan, 1983). Following data collection, adults rated the children’s pictures for attractiveness on a 5-point Likert scale ($1 = \text{not very attractive}$; $5 = \text{very attractive}$).

Results

The results are presented in two sections. The first section includes information about children’s spontaneous self-descriptions, the number and type of sorts children displayed in the classification task, their accuracy in sorting and labeling peers’ gender, race, and attractiveness, and their self identification of these attributes. The second section examines whether children’s characteristics and performance on the above measures predicted their bias in the various domains and their flexibility across these domains.

Children's Self-Descriptions, Classification, Sorting and Labeling, and Self-Identification

Three children completed only the first spontaneous self-description and classification task and not the sorting and labeling others tasks or the self-identification tasks, so 100 children are included in those analyses. Table 1 displays correlations between children's characteristics and their performance on these tasks.

Spontaneous self-descriptions—At the beginning of the study, only two children referred to appearance attributes when describing themselves: One referred to his gender and another referred to her attractiveness. At the end of the study, seven children referred to appearance attributes when describing themselves. Two were the same ones who referred to appearance attributes at the beginning of the study and they mentioned the same attributes and not any additional ones. Of the other five children, one referred to her gender, three referred to their attractiveness (all female), and one referred to his race (“my skin is a little bit white”). The study tasks' focus on appearance attributes therefore affected 5% of the children's self-descriptions.

Classification task—Only 2 children were unable to sort the faces at all, 40 sorted the faces on only one dimension, and 61 sorted the faces on two dimensions. The number of sorts children made positively correlated with their age, but not with their gender or race (Table 1).

For the first sort, children most frequently grouped faces by race ($n = 52$) followed by gender ($n = 36$), age ($n = 4$), eye color ($n = 3$), facial expression ($n = 3$), and hair color ($n = 3$). There was a negative relationship between age and functional use of gender during the first sort, $\rho = -.20$, $p < .05$; with age, children classified faces less often by gender and more often by race or other cues. For the second sort, children most frequently grouped faces by race ($n = 15$) followed by gender ($n = 14$), hair color ($n = 10$), eye color ($n = 9$), age ($n = 6$), hair length ($n = 3$), facial expression ($n = 1$), shirt color ($n = 1$), attractiveness ($n = 1$), and height ($n = 1$). Altogether, children grouped faces by race 67 times, gender 50 times, and attractiveness 1 time. Children's functional use of race negatively correlated with their functional use of gender (Table 1).

Sorting and labeling others—To determine whether domain affected children's sorting and labeling abilities, we conducted a MANOVA with domain as the within-subjects variable, children's gender and race as between-subjects variables, and their age as a covariate. Multivariate effects were significant for age, $F(2, 271) = 18.02$, $p < .0001$, partial $\eta^2 = .12$, domain, $F(4, 544) = 24.69$, $p < .0001$, partial $\eta^2 = .15$, and race, $F(8, 544) = 2.02$, $p < .05$, partial $\eta^2 = .03$.

Decomposition of the age main effect via Pearson correlations revealed sorting and labeling attractiveness and race, but not gender, improved with age (Table 1), most likely due to the ceiling effects children demonstrated for gender sorting and labeling. We used pairwise comparisons with Sidak adjustments to decompose the main effects of domain and race. Children more accurately sorted and labeled peers' gender compared to race and attractiveness, $ps < .001$. Children also more accurately sorted and labeled peers' race compared to attractiveness, $ps < .001$. Despite the differences in sorting and labeling

abilities across domains, children's ability to sort and label peers in each domain significantly differed from chance responding, a score of 4 (Table 2). Decomposition of the race main effect showed no significant differences in sorting or labeling abilities for children from different racial groups.

Univariate effects replicated the multivariate main effects of age and domain, but not of race, and also showed a main effect of gender on sorting, $F(1, 272) = 4.61, p < .05$, partial $\eta^2 = .02$, and labeling, $F(1, 272) = 4.85, p < .05$, partial $\eta^2 = .02$, that was superseded by a domain $\times$ gender interaction for labeling, $F(2, 272) = 4.01, p < .05$, partial $\eta^2 = .02$.

Analyses for each domain revealed the domain $\times$ gender interaction occurred because girls were better than boys at sorting, $F(1, 98) = 5.40, p < .05$, partial $\eta^2 = .05$, and labeling, $F(1, 98) = 5.47, p < .05$, partial $\eta^2 = .05$, in the domain of attractiveness only (Table 2).

Self-identification tasks—All children accurately sorted and labeled their gender. For skin tone, children and experimenter ratings ranged from 0 to 5 ($M = 1.57, SE = .07$ and $M = 1.59, SE = .08$, respectively). We obtained photos for 77 children. An additional researcher other than the experimenter used the photo to code skin tone. For 92% of these ratings (71 out of 77 children), the experimenter and researcher ratings were the same or within one point on the scale (agreement is similar to Averhart & Bigler, 1997). Agreement between child and experimenter ratings was slightly lower: 85% of the ratings (85 out of 100 children) were the same or within one point on the scale (agreement is 10% higher than Averhart & Bigler, 1997). Eight children (4 African American, 3 mixed race: Hispanic/Filipino, European American/ Mexican, European American/Asian American, and 1 European American) gave themselves a rating that was 2–3 points lower (lighter skin tone) than the experimenter's rating. Seven children (5 European American and 2 mixed race: African American/Vietnamese, European American/ Mexican) gave themselves a rating that was 2–3 points higher (darker skin tone) than the experimenter's rating.

Children used the full scale (0–2) to rate their attractiveness: 74 children gave themselves a rating of 2; 20 gave themselves a rating of 1; and 6 gave themselves a rating of 0 ($M = 1.68, SE = .03$). Undergraduates ($N = 42$, 18 males, 24 females) rated the 77 children's faces for attractiveness. Interrater agreement was high, *Cronbach's alpha* = .96. Adults used the full scale (1–5) and gave children a mean rating of 2.74 ($SE = .08$). There was no significant agreement between children's self-ratings and adult ratings of their attractiveness (Table 1).

Predictors of Children's Bias and Flexibility

Data from the bias tasks were compiled in a manner similar to Powlisha et al. (1994) who also compared children's bias across three domains. They calculated *positive attribution bias*, *negative attribution bias*, and *flexibility*, with each measure reflecting a comparison of traits attributed to in-group relative to out-group members. The discrepancies in children's self-ratings and adult ratings of their attractiveness made it unclear how to define in-group attractiveness status, particularly for children not at the extreme ends of attractiveness. It was also unclear how to define in-group race status for children who were not African American or European American. We therefore compared attributes children assigned to group members they often give higher social status (high attractive, same-sex, or European

American) relative to group members they often give lower social status (low attractive, other-sex, or African American; Dion, 1973; Powlishta, 1995b; Williams, Best, & Boswell, 1975). *Positive attribution bias* refers to the percent of positive attributes assigned to higher social status peers minus the percent of positive attributes assigned to lower social status peers. *Negative attribution bias* refers to the percent of negative attributes assigned to lower social status peers minus the percent of negative attributes assigned to higher social status peers.

Similar to Powlishta et al. (1994), we also calculated flexibility for the non-forced choice assessment. Powlishta et al. (1994) had the option of choosing *both* children in their study, but not the option of choosing *neither* child. Our flexibility measures reflected this difference in method and involved both positive and negative flexibility to parallel the positive and negative attribution bias measures. Flexibility suggests children understand that persons from different social groups can possess similar attributes (Powlishta et al., 1994; Serbin & Sprafkin, 1986), so *positive flexibility* was the percent of positive attributes assigned to both children minus the percent of positive attributes assigned to neither child. *Negative flexibility* was the percent of negative attributes assigned to both children minus the percent of negative attributes assigned to neither child. Table 3 shows means and standard error for the bias and flexibility measures.

We used hierarchical backward regression to assess predictors of bias and flexibility. The hierarchical entry allowed us to proceed theoretically while the backward removal permitted us to reduce noise in the models by removing predictors that did not explain any of the variance. Similar to Pauker et al. (2010), predictor variables included in the models were entered on different blocks: Block 1) child characteristics (gender, race dummy coded-European American/not European American, and adult-rated attractiveness); Block 3) self-perceptions of these characteristics (self-rated skin tone, self-rated attractiveness), if applicable; Block 3 or 5) domain-general developmental factors (age, classification skill); and Block 5 or 7) domain-specific functional use and conceptual knowledge (functional use of race, functional use of gender, race sorting/labeling-an aggregate of race sorting and labeling scores due to their strong correlation, attractiveness sorting/labeling-an aggregate of attractiveness sorting and labeling scores due to their strong correlation). Predictor variables included in the final block depended on the domain being tested for the bias regressions (e.g., attractiveness labeling/sorting to predict attractiveness bias), but all factors were entered for the flexibility regressions. Functional use of race and gender were entered together for the gender and race bias regressions due to their negative correlation. Because children's race and self-rated skin tone showed a strong positive correlation (Table 1), we conducted preliminary analyses to determine which variable explained more of the variance in predicting children's bias and flexibility. Based on these results, we used children's self-rated skin tone to predict bias based on attractiveness and flexibility across domains and children's race to predict bias based on gender and race. See Table 4 for specific information about which factors were entered for each bias analysis. Blocks 2, 4, 6, and 8 removed any factors entered on the previous block that did not contribute to explaining variance in children's bias or flexibility.

We used these characteristics to predict children's bias in the forced choice assessment ($n=50$) and their bias and flexibility in the non-forced choice assessment ($n=52$). One child showed perseverance errors during the bias tasks and was not included in these analyses. Due to significant, positive correlations between children's positive and negative bias within each domain (r s ranged from .58-.71, p s < .001), we created an aggregate bias score by adding these two bias scores together. Similar to Powlishta et al. (1994), bias did not correlate across domains, so we conducted separate regression analyses for each domain in each assessment method. There were no significant correlations between positive and negative flexibility within domains, but there were frequent significant associations between positive flexibility across domains (r s ranged from .32-.55, p s < .05) and between negative flexibility across domains (r s ranged from .29-.57, p s < .05), so we conducted separate regression analyses for positive flexibility and for negative flexibility collapsed across domains. Tests of normality, linearity, and multicollinearity of most variables were satisfactory, but racial sorting/labeling and self-rated attractiveness were negatively skewed. Transformations increased skewness, so we used raw scores.

Predictors of bias (Table 4)

Boys' attractiveness—None of the variables significantly predicted children's bias toward boys based on their attractiveness in the forced choice assessment, but self-rated skin tone was predictive of such bias in the non-forced choice assessment. The darker that children self-rated their skin tone, the more bias they showed based on boys' attractiveness.

Girls' attractiveness—Both self-rated and adult-rated attractiveness were predictive of children's bias based on girls' attractiveness in the forced choice assessment, although the latter finding was only marginally significant. The higher children rated their own attractiveness, the more bias they showed (positive traits attributed to high attractive girls and negative traits attributed to low attractive girls). In contrast, the higher adults rated children's attractiveness, the less bias they showed. In the non-forced choice assessment, more accurate attractiveness sorting and labeling abilities marginally predicted more bias based on girls' attractiveness.

Gender—In the forced choice assessment, functional use of race during the classification task negatively predicted bias based on peers' gender. Functional use of race therefore resulted in children showing less gender bias. In the non-forced choice assessment, classification skill negatively predicted children's bias based on peers' gender. The lower children's classification skill, the more positive traits they attributed to same-sex peers and the more negative traits they attributed to other-sex peers compared to children with higher classification skills.

Race—Children's race and functional use of gender positively predicted children's racial bias in the forced choice assessment. Relative to African American, Asian American, and Latino children, European American children attributed more positive traits to European American peers and more negative traits to African American peers. Moreover, functional use of gender during the classification task resulted in children showing more race bias

relative to those who did not functionally use gender. There were no significant predictors of children's racial bias in the non-forced choice assessment.

Predictors of flexibility (Table 5)

Positive flexibility—Age positively predicted positive flexibility, whereas racial sorting and labeling abilities negatively predicted positive flexibility. Compared to younger children, older children gave more “both” than “neither” responses when assigning positive attributes. Children with lower racial sorting and labeling accuracy also gave more “both” than “neither” responses for positive attributes relative to those with higher racial sorting and labeling accuracy.

Negative flexibility—Age negatively predicted negative flexibility, whereas functional use of gender and race positively predicted negative flexibility. Compared to younger children, older children gave less “both” responses and more “neither” responses when assigning negative attributes. Children who functionally used gender and race to sort faces during the classification task gave more “both” responses relative to “neither” responses when assigning negative attributes.

Discussion

One purpose of this research was to compare children's classification, sorting, labeling, and self-identification of attractiveness, gender and race. Results showed that children rarely used attractiveness to spontaneously classify people compared to gender and race. They were also less accurate at sorting and labeling others by attractiveness than by gender or race and showed no agreement with adults in how they self-rated their attractiveness compared to strong agreement with adults in how they self-rated their gender or skin tone. Children's concept of attractiveness therefore appears to be less explicit and less developed than that of gender and race, even though their perception of attractiveness is at a relatively adult-like level as evidenced by these same children showing strong attractiveness biases (Rennels & Langlois, 2014). These results suggest children may be less aware of their functional use of attractiveness to judge others compared to gender and race, which may impact their ability to control such bias (Devine, 1989).

Another purpose of this research was to explore whether these abilities were predictive of children's domain-specific bias and domain-general flexibility. Child characteristics that predicted bias differed depending upon domain and whether the task was forced choice or non-forced choice, demonstrating that children's bias is affected by both individual differences and task characteristics. Interventions to reduce bias in a particular domain will need to incorporate both situational (available choices) and individual considerations. Domain-specific factors that influenced children's flexibility differed based on whether children were assigning positive or negative attributes to others. Researchers interested in designing developmental interventions should consider how an attribute's valence and individual differences in conceptual knowledge and functional use of categories affect children's ability to engage in flexible thinking.

Children's Classification and Lexicalization of Attractiveness, Gender, and Race

Our impetus for examining children's classification and lexicalization of attractiveness, gender, and race was to compare their functional use of these attributes. Bigler and Liben (2006) proposed that children must perceive these cues as perceptually and psychologically salient ways to classify others and then functionally use such classifications during social encounters prior to linking attributes and affective responses to their groups. Results from the classification task clearly demonstrate gender and race are salient to children and that many children group people based on these attributes. In contrast, attractiveness was not a salient social category for children during the classification task; only one child grouped people based on attractiveness. Children's accuracy in sorting and labeling others' attractiveness was also much lower (69% accuracy) compared to almost perfect sorting and labeling of others' gender and high levels of sorting and labeling of others' race (88% accuracy). Thus, developing explicit classification of and labels for individuals who differ in attractiveness takes more time to develop than gender or race classification and labeling, perhaps due to language input. Individuals may use adjectives more often than nouns to refer to attractiveness ("pretty"), but likely use nouns more often than adjectives to refer to gender ("boy") and race ("African Americans"). Because hearing nouns relative to adjectives facilitates formation of categories and concept acquisition (Graham, Booth, & Waxman, 2012), these differences in language input may cause development of attractiveness concepts to lag behind gender and race concepts. Not explicitly labeling or sorting others by attractiveness may also reflect social desirability. Attractiveness is considered a more evaluative than objective way to classify others relative to gender and race, so children may think it is rude to label a person as "not good-looking." Social desirability, however, increases with age (Clark, Hogevar, & Dembo, 1980) and older children showed higher accuracy in sorting and labeling peers' attractiveness compared to younger children. It is therefore not clear whether social desirability influenced results. Older children's more explicit understanding of attractiveness could increase bias as it does in the gender and race domains (Bigler & Liben, 2006), but it could also provide a means to control stereotype activation and reduce prejudice (Devine, 1989).

We do not think the results related to attractiveness concepts negate process two, functional classification during social encounters, in Bigler and Liben's (2006) developmental intergroup theory. Rather, children's functional use of attractiveness may be displayed in ways other than explicit classification and labeling. For example, by 12 months of age, infants display more positive affect and get more involved in play with an attractive than unattractive stranger (Langlois, Roggman, & Rieser-Danner, 1990). During the preschool years, children play more amicably when they are similar in attractiveness rather than dissimilar in attractiveness (Langlois & Downs, 1979). Moreover, the children in this study showed very strong displays of bias based on attractiveness (Rennels & Langlois, 2014). Adults rated the attractiveness of individuals in these studies, so these differential judgments and behaviors provide strong evidence that very young children perceive attractiveness similarly to adults and make functional use of attractiveness during social interactions; our results show they just do not yet have a well-established concept of this classification.

If children have less developed classification and labeling skills for attractiveness than they do for gender or race, why would they display differential behaviors and biases based on target attractiveness (Langlois et al., 1990; Rennels & Langlois, 2014)? Children and adults may prefer high attractive persons because attractive faces are closer to the population's facial average than less attractive faces, which facilitates processing (Trujillo, Jankowitsch, & Langlois, 2013). This processing fluency should elicit positive affect (Winkielman & Cacioppo, 2001). Dissimilarity to the population's facial average (i.e., low attractiveness) may elicit a disgust reaction and subsequently produce negative affect when viewing low attractive persons (Principe & Langlois, 2011). Spontaneous physiological responses may therefore activate attractiveness biases as opposed to conceptual learning about high and low attractive individuals.

We are not arguing that there is no conceptual learning of attractiveness. Girls showed better attractiveness labeling abilities than boys, supporting Dion's (1973) suggestion that girls have stronger concepts of attractiveness than boys. The notion that attractiveness is more important for girls than boys is part of children's gender stereotypes (Miller et al., 2009; Ramsey & Langlois, 2002). Girls have more knowledge of female stereotypes than boys (O'Brien et al., 2000) and also refer to girls' appearance more so than boys (Miller et al., 2009). These factors may contribute to girls' earlier conceptual learning of attractiveness and better labeling of attractiveness than boys, but even girls' attractiveness labeling is significantly lower than their gender or race labeling, demonstrating developmental differences in acquiring these concepts.

An alternative explanation for children's low accuracy in sorting faces by attractiveness is that the lack of clip art showing examples of high and low attractive individuals hindered their performance. In contrast, the presence of clip art showing examples of boys and girls or African American and European American children may have facilitated children's ability to sort peers based on gender and race. This methodological difference could potentially explain differences in sorting accuracy across domains, but it seems unlikely to explain differences in labeling accuracy across domains because clip art was not used for this task. Moreover, if lack of clip art made the sorting task too difficult, we should not have seen similar age-related increases in both children's attractiveness sorting and labeling abilities.

Children's Self-Identification of Attractiveness, Gender, and Race

Children also showed discrepancies across the three domains in their self-identification. Whereas children were 100% accurate in identifying their gender and 85% accurate in identifying their skin color, children showed minimal accuracy in identifying their attractiveness. Adults are also relatively inaccurate in self-rating their attractiveness, but at least there is some agreement ($r = .24$) between self and other ratings of adults' attractiveness (Feingold, 1992). Similar to Downs and Reagan's (1983) findings, children were more likely to overestimate than underestimate their attractiveness in this study. Such overestimation may reflect children's self-esteem more than attractiveness self-perceptions and overestimation of traits and skills is common among young children (Harter, 1998). Low accuracy in attractiveness self-identification compared to gender and race could also occur because adults occasionally tell children they look pretty or handsome (e.g., when

dressed up for a special occasion) even if the child is not that attractive, whereas children should consistently hear adults accurately label their gender and race. Alternatively, their low accuracy suggests they have a less explicit concept of attractiveness as compared to gender and race. Each explanation proposes attractiveness self-identification entails a qualitatively different form of processing than gender or race self-identification.

Only two children used any of the three domains to spontaneously describe themselves at the beginning of the study, suggesting that these attributes were not highly accessible qualities of the self. This finding contrasts with our result that gender and race were highly salient ways to classify others (although results may have differed if we asked children to describe individual targets rather than classify targets into groups) and may have implications for ways to reduce children's functional use of such attributes. Focusing on internal attributes of others has been successful in reducing children's bias (Bigler & Liben, 1990) and teaching children to transfer how they think about themselves (personal pathway) to how they think about others (attitudinal pathway, Bigler & Liben, 2006) might also reduce bias. The tasks children completed influenced five children's spontaneous self-descriptions at the end of the study. Although it was a small percent of children, adding a final discussion focused on internal attributes, as we did, seems important to include following completion of appearance-based tasks.

Predictors of Bias

A comparison of predictors of bias within each domain between children in the forced choice and non-forced choice assessments showed no overlap in findings. Such results show that both individual differences among children and assessment method affect displays of bias in the attractiveness, gender, and race domains. Others have examined how individual differences, such as classification skill and labeling abilities, are related to children's race and gender biases (e.g., Bigler, 1995; Pauker et al., 2010; Zosuls et al., 2009), but this study was the first to examine predictors of attractiveness biases and to show that children's judgments about others depend on the available choices and differential characteristics weigh into making those judgments.

The predictors also differed depending upon the domain with which children showed bias, which is consistent with the notion that bias reflects domain-specific learning (Banse et al., 2010; Powlishta et al., 1994). For bias based on boys' attractiveness, children's characteristics, at least those we measured, had no impact on the bias displayed in the forced choice assessment. In contrast, darker self-rated skin tone was predictive of children showing more bias based on boys' attractiveness in the non-forced choice assessment. Darker skinned children may have more exposure to males than lighter skinned children due to their greater likelihood of living with or near extended family members or having parents with more egalitarian roles (Kane, 2000). Greater exposure to males may strengthen male facial representations and subsequent display of attractiveness biases even in situations when not forced to show bias. Adults perceive darker-skinned males as more attractive than darker-skinned females (Lewis, 2011). Darker-skinned children may detect this discrepancy from their environment and subsequently apply more attractiveness biases toward males than lighter-skinned children.

The predictors of children's bias toward high and low attractive girls were different from those predicting their bias toward high and low attractive boys, suggesting qualitative differences in children's attractiveness biases based on gender of target (Miller et al., 2009; Ramsey & Langlois, 2002). For the forced choice assessment, the higher children's self-rated attractiveness and the lower their adult-rated attractiveness, the more likely they were to show bias based on girls' attractiveness. Self-perceptions and others' perceptions of attractiveness therefore had opposing effects on children's girl attractiveness biases, but the self-perception effects were stronger and suggest how children view themselves in terms of attractiveness impacts how they view others, a potential in-group bias. For the non-forced choice assessment, children's attractiveness sorting and labeling abilities were marginally predictive of bias based on girls' attractiveness, perhaps because use of such labels is reflective of children's concepts about these attributes (Gelman & Markman, 1986) and results in a stronger focus on people's appearance.

In regard to gender biases, functional use of race negatively predicted bias in the forced choice assessment. Prior to learning about and using race as a way to classify encountered people in their environment, children rely more on gender when assigning attributes to peers. For non-forced choice tasks, children's ability to sort people on more than one dimension during the classification tasks was predictive of less bias. This finding replicates previous work (Bigler, 1995; Bigler & Liben, 1992) and shows that as children learn to focus on more than one attribute of a person, they will show less bias based on peers' gender because it is not their only focus.

There were no predictors of children's race bias in the non-forced choice assessment. In the forced choice assessment, European American children showed more racial bias relative to children from other racial groups, similar to previous research (Williams, Best et al., 1975). Furthermore, functional use of gender marginally predicted children's racial bias in the forced choice assessment. This finding fits with the notion that race encoding is a byproduct of mechanisms designed to encode gender (Kurzban, Tooby, & Cosmides, 2001). Prior to classifying race and displaying bias based on race, children may need to first use gender in a functional manner. Unlike Pauker et al. (2010) who tested a similarly aged sample using a forced choice task, we did not find that functional use of race predicted racial bias. The difference in findings may be because the attributions children made in Pauker et al.'s task were cultural stereotypes, whereas the attributions children made in our task were positive and negative traits and behaviors. Perhaps emergent functional use of race is related to children acquiring the cognitive, but not affective, component of racial stereotypes. Pauker et al.'s sample was also predominantly European American, whereas our sample was more heterogeneous with approximately one-third not belonging to either of the racial groups used in the racial bias task, which could influence results. Last, we might have found different results if we used children's own identification with a racial group as a predictor rather than our assignment of children to particular racial groups.

Predictors of Flexibility

Consistent with previous research (Banse et al., 2010; Doyle et al., 1988), older children showed more positive flexibility (attributed more positive traits to both peers relative to

neither peer) than younger children. A new finding, however, was that older children showed less negative flexibility (attributed less negative traits to both peers relative to neither peer) than younger children. The age differences show that as children get older, they generally increase their positive beliefs about others and decrease their negative beliefs about others, irrespective of social group membership. This latter finding may be important in re-conceptualizing flexibility as both an understanding that members of different social groups can possess similar attributes (Powlishta et al., 2004; Serbin & Sprafkin, 1986) and an understanding that neither member of different social groups has to possess a negative attribute. Alternatively, age differences in allocation of more positive attributes to both children and of less negative attributes to neither child may reflect an increased understanding of socially desirable responding (Clark et al., 1980).

Another new finding was that the less accurately children sorted or labeled a peer's race, the more likely they were to attribute positive traits to both targets across domains during non-forced choice assessments. Sorting and labeling people by race may be negatively correlated to flexible thinking because children who have a label for a person are more likely to think of that person as having stable traits across contexts (Gelman & Heyman, 1999). Belief in trait stability may interfere with processing person information in a cognitively flexible manner (Levy & Dweck, 1999). On the other hand, children who are inflexible in their thinking tend to view group members as relatively homogeneous and as having fixed traits (Levy & Dweck, 1999). Perhaps this viewpoint promotes acquisition of labels and attributions for group members. These explanations may apply only to racial sorting and labeling, as opposed to attractiveness or gender sorting and labeling, because 5–7-year-olds treat gender as more of a natural and accurate way to categorize people than race, which they view as a more subjective and less rigid way to categorize people (Rhodes & Gelman, 2009). Also, attractiveness is relatively stable across development (Pittenger, Mark, & Johnson, 1989), but it can change, which may result in children viewing it as less of a fixed trait relative to gender and race. Perhaps children who use a category (i.e., race) that is relatively subjective to group and label people in their social world do so because of poor cognitive flexibility skills. Although this is a speculative interpretation of our results, this intriguing relation between positive flexibility (a domain-general mechanism) and race sorting and labeling (domain-specific knowledge) merits further attention.

Last, functional use of gender and race during the classification task was predictive of higher levels of negative flexibility (assigning negative traits to both peers relative to neither peer). Such functional use may contribute to rigidity in children's thinking (negative perceptions of other-sex and African American children) and subsequently result in not letting go of such negativity, but rather applying it to both pictured peers. More work, however, is needed to test this interpretation.

Conclusions

Despite children's robust use of attractiveness during interactions and assessments of bias (Dion, 1973; Langlois et al., 1990; Rennels & Langlois, 2014), they less often use attractiveness to spontaneously categorize or overtly sort and label people in their social world compared to gender and race. Within the context of Bigler and Liben's (2006) DIT,

future research should[^] investigate whether advances in conceptual understanding of attractiveness increase bias in this domain. Understanding more about children's acquisition of racial concepts and their ability to think flexibly about positive attributes seems critical to understanding why these two abilities (one of which is domain-specific and other of which is domain-general) are negatively related. Furthermore, work is needed to examine dissociations in the factors that affect children's flexibility when assigning positive and negative attributes to others. Because flexible thinking is necessary to overcome biased responding (Devine, 1989), these are important areas for future research and designing developmental interventions.

References

- Anzures G, Quinn PC, Pascalis O, Slater AM, Lee K. Categorization, categorical perception, and asymmetry in infants representation of face race. *Developmental Science*. 2010; 13(4):553–564. [PubMed: 20590720]
- Averhart CJ, Bigler RS. Shades of meaning: Skin tone, racial attitudes, and constructive memory in African American children. *Journal of Experimental Child Psychology*. 1997; 67:363–388. [PubMed: 9440298]
- Banase R, Gawronski B, Rebetez C, Gutt H, Morton JB. The development of spontaneous gender stereotyping in childhood: Relations to stereotype knowledge and stereotype flexibility. *Developmental Science*. 2010; 13(2):298–306. [PubMed: 20136926]
- Bigler RS. The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development*. 1995; 66(4):1072–1087.
- Bigler RS, Liben LS. The role of attitudes and interventions in gender-schematic processing. *Child Development*. 1990; 61:1440–1452. [PubMed: 2245736]
- Bigler RS, Liben LS. Cognitive mechanisms in children's gender stereotyping: Theoretical and educational implications of a cognitive-based intervention. *Child Development*. 1992; 63:1351–1363. [PubMed: 1446556]
- Bigler RS, Liben LS. A cognitive-developmental approach to racial stereotyping and reconstructive memory in Euro-American children. *Child Development*. 1993; 64:1507–1518.
- Bigler, RS.; Liben, LS. A developmental intergroup theory of social stereotypes and prejudice. In: Kail, RV., editor. *Advances in child development and behavior*. Vol. Vol. 34. Elsevier: San Diego; 2006. p. 39-89.
- Clark A, Hocevar D, Dembo MH. The role of cognitive development in children's explanations of preferences for skin color. *Developmental Psychology*. 1980; 16(4):332–339.
- Devine PG. Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*. 1989; 56(1):5–18.
- Dion KK. Young children's stereotyping of facial attractiveness. *Developmental Psychology*. 1973; 9:183–188.
- Downs, AC.; Reagan, MA. Recognition, development, and correlates of self-defined physical attractiveness among young children. San Francisco, CA: Presented at the Western Psychological Association; 1983.
- Doyle A, Beaudet J, Aboud F. Developmental patterns in the flexibility of children's ethnic attitudes. *Journal of Cross-Cultural Psychology*. 1988; 19(3):3–18.
- Dutton SE, Singer JA, Devlin AS. Racial identity of children in integrated, predominantly Caucasian, and African American schools. *The Journal of Social Psychology*. 1998; 138(1):41–53. [PubMed: 9517311]
- Fagot BI, Leinbach MD, Hagan R. Gender labeling and the adoption of sex-typed behaviors. *Developmental Psychology*. 1986; 22(4):440–443.
- Fagot BI, Leinbach MD, O'Boyle C. Gender labeling, gender stereotyping, and parenting behaviors. *Developmental Psychology*. 1992; 28(2):225–230.

- Feingold A. Good-looking people are not what we think. *Psychological Bulletin*. 1992; 111:304–341.
- Gelman SA. Learning from others: Children's construction of concepts. *Annual Review of Psychology*. 2009; 60:115–140.
- Gelman SA, Heyman GD. Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*. 1999; 10:489–493.
- Gelman SA, Markman EM. Categories and induction in young children. *Cognition*. 1986; 23(3):183–209. [PubMed: 3791915]
- Graham SA, Booth AE, Waxman SR. Words are not merely features: Only consistently applied nouns guide 4-year-olds' inferences about object categories. *Language Learning and Development*. 2012; 8:136–145.
- Harter, S. The development of self-representations. In: Damon, W., Editor-in-chief; Eisenberg, N., editors. *Handbook of child psychology: Vol. 3. Social, emotional, and personality development*. 5th ed.. New York, NY: John Wiley & Sons; 1998. p. 553-617.
- Herman M. Forced to choose: Some determinants of racial identification in multiracial adolescents. *Child Development*. 2004; 75(3):730–748. [PubMed: 15144483]
- Hoffner C. Children's wishful identification and parasocial interaction with favorite television characters. *Journal of Broadcasting & Electronic Media*. 1996; 40(3):38–402.
- Kane CM. African American family dynamics as perceived by family members. *Journal of African American Studies*. 2000; 30(5):691–702.
- Katz PA. Stimulus predifferentiation and modification of children's racial attitudes. *Child Development*. 1973; 44(2):232–237.
- Katz, PA.; Kofkin, JA. Race, gender, and young children. In: Luthar, SS.; Burack, JA.; Cicchetti, D.; Weisz, JR., editors. *Developmental Psychopathology*. Cambridge: Cambridge University Press; 1997. p. 51-74.
- Kurzban R, Tooby J, Cosmides L. Can race be erased? Coalitional computation and social categorization. *Proceedings for the National Academy of Science of the United States of America*. 2001; 98(26):15387–15392.
- Lam V, Guerrero S, Damree N, Enesco I. Young children's racial awareness and affect and their perceptions about mothers' racial affect in a multiracial context. *British Journal of Developmental Psychology*. 2011; 29:842–864. [PubMed: 21199507]
- Langlois JH, Downs AC. Peer relations as a function of physical attractiveness: In the eye of the beholder or behavioral reality? *Child Development*. 1979; 50(2):409–418.
- Langlois JH, Kalakanis L, Rubenstein AJ, Larson A, Hallam M, Smoot M. Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*. 2000; 126(3):390–423. [PubMed: 10825783]
- Langlois JH, Roggman LA, Rieser-Danner LA. Infants' differential social responses to attractive and unattractive faces. *Developmental Psychology*. 1990; 26(1):153–159.
- Langlois JH, Stephan C. The effects of physical attractiveness and ethnicity on children's behavioral attributions and peer preferences. *Child Development*. 1977; 48(4):1694–1698.
- Levy SR, Dweck CS. The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development*. 1999; 70(5):1163–1180.
- Lewis MB. Who is the fairest of them all? Race, attractiveness and skin color sexual dimorphism. *Personality and Individual Differences*. 2011; 50(2):159–162.
- Martin CL. Children's use of gender-related information in making social judgments. *Developmental Psychology*. 1989; 25(1):80–88.
- Miller CF, Lurye LE, Zosuls KM, Ruble DN. Accessibility of gender stereotype domains: Developmental and gender differences in children. *Sex Roles*. 2009; 60:870–881. [PubMed: 19606278]
- O'Brien M, Peyton V, Mistry R, Hruda L, Jacobs A, Caldera Y, Huston A, Roy C. Gender-role cognition in three-year-old boys and girls. *Sex Roles*. 2000; 42(11-12):1007–1025.
- Pauker K, Ambady N, Apfelbaum EP. Race salience and essentialist thinking in racial stereotype development. *Child Development*. 2010; 81(6):1799–1813. [PubMed: 21077865]

- Pittenger JB, Mark LS, Johnson DF. Longitudinal stability of facial attractiveness. *Bulletin of the Psychonomic Society*. 1989; 27(2):171–174.
- Powlisha KK. Gender bias in children's perceptions of personality traits. *Sex Roles*. 1995; 32(1–2): 17–28.
- Powlisha KK, Serbin LA, Doyle A, White DR. Gender, ethnic, and body type biases: The generality of prejudice in childhood. *Developmental Psychology*. 1994; 30(4):526–536.
- Principe CP, Langlois JH. Faces differing in attractiveness elicit corresponding affective responses. *Cognition and Emotion*. 2011; 25(1):140–148. [PubMed: 21432661]
- Ramsey JL, Langlois JH. Effects of the “beauty is good” stereotype on children's information processing. *Journal of Experimental Child Psychology*. 2002; 81:320–340. [PubMed: 11884093]
- Ramsey JL, Langlois JH, Hoss RA, Rubenstein AJ, Griffin AM. Origins of a stereotype: Categorization of facial attractiveness by 6-month-old infants. *Developmental Science*. 2004; 7:201–211. [PubMed: 15320380]
- Rennels JL, Langlois JH. Children's attractiveness, gender, and race biases: A comparison of their strength and generality. *Child Development*. Advance online publication. 2014
- Rhodes M, Gelman SA. A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*. 2009; 59(3):244–274. [PubMed: 19524886]
- Rotenberg KJ. Development of character constancy of self and other. *Child Development*. 1982; 53(2): 505–515.
- Serbin LA, Sprafkin C. The salience of gender and the process of sex typing in three- to seven-year-old children. *Child Development*. 1986; 57:1188–1199.
- Styczynski LE, Langlois JH. The effects of familiarity on behavioral stereotypes associated with physical attractiveness in young children. *Child Development*. 1977; 48:1137–1141.
- Trujillo LT, Jankowitsch JM, Langlois JH. Beauty is in the ease of the beholding: A neurophysiological test of the averageness theory of facial attractiveness. *Cognitive Affective Behavioral Neuroscience*. Advance online publication. 2013
- Van Ausdale D, Feagin JR. Using racial and ethnic concepts: The critical case of very young children. *American Sociological Review*. 1996; 61(5):779–793.
- Weinraub M, Clemens LP, Sockloff A, Ethridge T, Gracely E, Myers B. The development of sex role stereotypes in the third year: Relationships to gender labeling, gender identity, sex-typed toy preference, and family characteristics. *Child Development*. 1984; 55(4):1493–1503. [PubMed: 6488962]
- Williams JE, Bennett SM, Best DL. Awareness and expression of sex stereotypes in young children. *Developmental Psychology*. 1975; 11(5):635–642.
- Williams JE, Best DL, Boswell DA. The measurement of children's racial attitudes in the early school years. *Child Development*. 1975; 46(2):494–500.
- Winkielman PC, Cacioppo JT. Mind at ease puts smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology*. 2001; 81(6):989–1000. [PubMed: 11761320]
- Younger BA, Fearing DD. Parsing items into separate categories: Developmental change in infant categorization. *Child Development*. 1999; 70:291–303.
- Zosuls KM, Ruble DN, Tamis-LeMonda CS, Shrout PE, Bornstein MH, Greulich FK. The acquisition of gender labels in infancy: Implications for gender-typed play. *Developmental Psychology*. 2009; 45(3):688–701. [PubMed: 19413425]

Highlights

- Relative to gender and race, children rarely used attractiveness to classify people
- Children sorted and labeled peers' gender and race more accurately than attractiveness
- Children's self-identification of attractiveness was poor compared to gender and race
- Predictors of bias differed based on domain and task (forced vs. non-forced choice)
- Predictors of flexibility differed based on the valence of the assigned attributes

Table 1
Correlations between Children's Characteristics and Classification, Sorting, Labeling, and Self-Identification Abilities

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Age	-	-.10	.16	.44**	.10	-.01	.35**	.37**	.40**	.40**	.30*	-.14	-.20
2. Gender		-	-.11	-.04	-.08	.01	-.17	-.08	-.23 [†] *	-.22 [†]	-.04	-.08	.00
3. Race			-	.07	-.14	.10	.19	.17	-.07	-.02	.56**	.13	-.07
4. Classification skill				-	.24 [†]	.03	.22 [†]	.13	.28*	.41**	.07	-.15	.07
5. Functional use of race					-	-.47**	.19	.19	.20	.21 [†]	-.15	-.04	.13
6. Functional use of gender						-	-.104	-.18	-.13	-.14	.15	-.03	-.01
7. Race sorting							-	.50**	.27*	.11	.19	.09	.07
8. Race labeling								-	.23 [†]	.19	.16	-.06	-.12
9. Attractiveness sorting									-	.76**	-.03	-.13	-.07
10. Attractiveness labeling										-	.01	-.14	-.05
11. Self-rated skin tone											-	.18	-.11
12. Self-rated attractiveness												-	.03
13. Adult-rated attractiveness													-

Correlations including gender, race, functional use of race, and functional use of gender are *Spearman rho* correlations; all others are *Pearson r* correlations.

[†] p .05

* p .01

** p .001

Table 2

Means and Standard Error of Children's Sorting and Labeling of Others Based on Domain and Children's Gender

<i>Domain</i>	<i>Sorting</i>	<i>Labeling</i>
Attractiveness	5.45 (.15) ^{bd ***}	5.61 (.15) ^{bd ***}
Girls	5.78 (.21) ^e	5.94 (.19) ^e
Boys	5.10 (.21) ^f	5.27 (.22) ^f
Gender	7.99 (.01) ^{a ***}	7.98 (.01) ^{a ***}
Girls	8.00 (.00)	8.00 (.00)
Boys	7.98 (.02)	7.96 (.03)
Race	6.95 (.12) ^{bc ***}	7.15 (.09) ^{bc ***}
Girls	7.14 (.13)	7.22 (.12)
Boys	6.76 (.20)	7.08 (.14)

Scores significantly differed from chance, $p < .001$.

Note: Higher labeling and sorting scores (possible range = 0 – 8) indicate greater accuracy. Within each sorting or labeling column, means with a and b, c and d, or e and f subscripts significantly differ.

Table 3

Means (and Standard Error) of Bias and Flexibility Scores for Each Domain in the Forced and Non-Forced Choice Assessments

Domain and Assessment Method	Positive attribution bias	Negative attribution bias	Positive flexibility	Negative flexibility
Boy attractiveness				
Forced choice	0.22 (0.08)	0.31 (0.08)	-	-
Non-forced choice	0.12 (0.07)	0.09 (0.07)	0.27 (0.06)	-0.22 (0.07)
Girl attractiveness				
Forced choice	0.36 (0.07)	0.45 (0.08)	-	-
Non-forced choice	0.21 (0.05)	0.24 (0.06)	0.29 (0.05)	-0.24 (0.07)
Gender				
Forced choice	0.27 (0.08)	0.14 (0.09)	-	-
Non-forced choice	0.08 (0.06)	-0.01 (0.07)	0.34 (0.06)	-0.18 (0.07)
Race				
Forced choice	0.11 (0.08)	0.20 (0.08)	-	-
Non-forced choice	0.04 (0.06)	0.06 (0.06)	0.26 (0.07)	-0.30 (0.06)

Note: Higher scores (possible range = -1 to 1) indicate more bias or flexibility.

Table 4
Regression Analyses Results for Predictors of Bias in Each Domain for Each Assessment Method

Domain and Variables Entered on Each Block	Final Model Predictors in Each Assessment	Unstandardized Coefficients (Std. Error)	Standardized Coefficients	Significance	Adjusted R ²
<i>Bias based on boys' attractiveness</i>					
<i>Forced choice</i>					
1. Gender & adult-rated attractiveness	Constant	.520 (.169)	-	.004	
3. Self-rated skin tone & self-rated attractiveness	<i>Non-forced choice</i> Constant	-.108 (.206)		.603	.089
5. Classification skill & age	Self-rated skin tone	.204 (.095)	.336	.039	
7. Attractiveness sorting/labeling					
<i>Bias based on girls' attractiveness</i>					
<i>Forced choice</i>					
1. Gender & adult-rated attractiveness	Constant	-.447 (.468)		.345	.218
3. Self-rated skin tone & self-rated attractiveness	Adult-rated attractiveness Self-rated attractiveness	-.359 (.184) .698 (.253)	-.282 .397	.058 .009	
5. Classification skill & age	<i>Non-forced choice</i>				
7. Attractiveness sorting/labeling	Constant	-.732 (.467)		.126	.113
	Age	.049 (.059)	.145	.414	
	Attractiveness sorting/labeling	.078 (.044)	.132	.083	
<i>Bias based on peers' gender</i>					
<i>Forced choice</i>					
1. Gender & race	Constant	.813 (.233)		.001	.078
3. Classification skill & age	Functional use of race	-.672 (.298)	-.312	.029	
5. Functional use of gender & race	<i>Non-forced choice</i>				
	Constant	1.162 (.406)		.006	.121
	Classification skill	-.662 (.236)	-.372	.007	
<i>Bias based on peers' race</i>					
<i>Forced choice</i>					
1. Gender & race	Constant	-.407 (.228)		.080	.224
3. Classification skill & age	Race	.877 (.269)	.420	.002	
5. Functional use of gender & race and race sorting/labeling	Functional use of gender <i>Non-forced choice</i>	.540 (.269)	.259	.051	
	Constant	.106 (.119)	-	.377	

Table 5

Regression Analyses Results for Predictors of Flexibility Across Domains

Type of Flexibility	Final Model Predictors	Unstandardized Coefficients (Std. Error)	Standardized Coefficients	Significance	Adjusted R ²
Positive flexibility	Constant	1.057 (.219)		<.001	.226
	Age	.110 (.018)	.533	<.001	
	Racial sorting/labeling	-.102 (.018)	-.484	<.001	
Negative flexibility	Constant	.230 (.1270)		.072	.230
	Age	-.118 (.018)	-.488	<.001	
	Functional use of race	.210 (.085)	.195	.015	
	Functional use of gender	.239 (.077)	.246	.002	