Now you see race, now you don't: Verbal cues influence children's racial stability judgments

ABSTRACT

Research suggests that young children do not consistently believe that race is stable (e.g., that a Black child will grow up to be a Black adult). Here, we tested the strength of White and minority children's beliefs by testing whether verbal cues influenced the extent to which they believed in the relative stability of race versus emotional expression. We presented participants (5-6 years, 9-10 years, adults) with images of children who were Black or White, and happy or angry, and asked them to indicate which of two adults each child would grow up to be: one matching on emotion but not race, or one matching on race but not emotion. Verbal cuing had strong effects on both younger and older children's choices (e.g., cuing emotional expressions strengthened emotion-matches; cuing skin color strengthened race-matches), although racial minorities were less susceptible to cuing. Verbal cues had weaker effects on children's judgments about the stability of gender. These results show that race concepts vary across age and racial groups, that verbal framing influences younger and older children's beliefs about racial stability, and that belief in racial stability is relatively weak between the ages of 5 and 10.

Keywords: cognitive development, race concepts, social cognition, language, social group differences

Now you see race, now you don't: Verbal cues influence children's racial stability judgments

Lauer: Let me just ask you the question in simple terms again because you've sent mixed signals over the years, are you an African American woman?

Dolezal: I identify as Black.

Lauer: You identify as Black. Let me put a picture up of you in your early twenties...when you see this picture...is she an African-American woman or a Caucasian woman?

Dolezal: I would say that visibly she would be identified as White, by people who see her.

Lauer: But at the time were you identifying yourself as African-American?

Dolezal: In that picture, during that time, no.

-Interview of Rachel Dolezal by Matt Lauer on the *Today* show (June 16, 2015)

1. Introduction

Rachel Dolezal had two self-identified White parents, identified as White as a child, and identified as Black as an adult. Because of her self-professed racial identity, Rachel became a headline in 2015, with many reacting with outrage that she would choose to adopt a different racial identity than she was assigned at birth. Embedded in this complex case is a core issue that is the heart of the current studies: Is it possible for an individual to change their race (see Hobbs, 2014)? We examine the question from a developmental perspective, to ask when, and with what degree of certainty, children treat a person's race as stable over time.

Understanding the development of racial stability concepts is important for understanding the development of essentialist beliefs more broadly (i.e., the belief that category membership is objective, absolute, inductively potent, immutable, and stable; Gelman, 2003; Prentice & Miller, 2006; Rothbart & Taylor, 1992). Adults often essentialize race: they believe that race reflects real

and objective “kinds” of people, is important for making inferences about individuals, has always existed throughout human history, and is an unchangeable aspect of an individual's identity (Haslam, Rothschild, & Ernst, 2000; Haslam & Whelan, 2008).

In contrast to studies with adults, research with children suggests that racial essentialism may develop with age and as a function of social experiences. For example, during kindergarten, children in both rural and urban communities treat racial categories as determined subjectively and by convention, whereas by age 10, children in rural communities (but not in urban communities) treat racial categories as determined objectively and by nature (Rhodes & Gelman, 2009). U.S. children are more likely than Israeli children to develop the concept that race is objective, thus indicating cross-cultural variation in racial essentialism (Diesendruck, Goldfein-Elbaz, Rhodes, Gelman, & Neumark, 2013). Moreover, the extent to which children categorize Multiracial children into different “kinds” of people varies as a function of their age, race, and experiences with inter-group contact (Roberts & Gelman, 2015; Roberts & Gelman, 2016a). Regarding the belief that race is inductively potent, although by age 4, children use race to make inferences about a person's wealth (Olson, Shutts, Kinzler, & Weisman, 2012; Shutts, 2015), social relationships (Shutts, Pemberton, Roben, & Spelke, 2013), and language (Hirschfeld & Gelman, 1997), the extent of these inferences varies as a function of children's awareness of racial hierarchies (e.g., that White individuals have higher socioeconomic status than Black individuals; Olson et al., 2012) and their own racial group membership (e.g., Multiracial children make different inferences than White children; Roberts, Williams, & Gelman, in press). We turn next to studies examining the belief that race is stable across an individual's lifespan, which is the central focus of the present research.

1. 1. Children's Concept of Racial Stability

In several independent investigations, researchers examined children's concepts of racial stability, though because of differences in methodologies, populations, and results, the extent to which this concept varies across age and racial clear remained unclear. Hirschfeld (1995) presented children with a triad task that assessed the extent to which they believed that race was more stable across the lifespan than occupation or body build (e.g., on occupation trials, children were shown a Black adult in a police uniform, paired with a Black child in plain clothing [race match] and a White child in a police uniform [occupation match], and were then asked to indicate which of the two children was a picture of the adult when as a child). At age 4, although children judged race to be more stable than occupation, they did not judge race to be more stable than body build. At age 7, children judged race to be more stable than both occupation and body build.

Similarly, Kinzler and Dautel (2012) showed White 5- to 6-year-olds, White 9- to 10-year-olds, and Black 5- to 6-year-olds triads consisting of one target child and two adults. One adult matched the target child in language but not race (language-match), and the other adult matched the target child in race but not language (race-match). On each trial, children were first introduced to a target child and told, "Here is a child [pointing], he/she sounds like this." The target child was then shown, a voice clip was played, and the target child was then concealed again. This was then repeated for two adult response options, and children were then asked which adult each child would grow up to be. White 9- to 10-year-olds and Black 5- to 6-year-olds chose the race-match, thus judging race to be stable. In contrast, White 5- to 6-year-olds chose the language-match, thereby reasoning that a person could change race across their lifespan. This was found for children living in urban and racially heterogeneous contexts as well as children living in rural and racially homogeneous contexts. These data show the role of

development and social group membership in racial stability judgments, such that older children and minority children were more likely to think of race as stable (likely as a result of increased race-based experiences).

Pauker and colleagues (2010, 2015) presented children with a three-item test that assessed children's reasoning about racial stability: what a child would look like as an adult (e.g., White child matched with White adult and Black adult choices), what an adult looked like as a child (e.g., Black adult matched with White child and Black child choices), and whether someone could change their skin color. Overall, older children (ages 7 to 11) conceptualized race as more stable than younger children (ages 3 to 6), and among older children, the belief that race was stable was greater among children from a homogenous context (i.e., Massachusetts) compared to children from a more racially diverse context (i.e., Hawai'i).

Most recently, Roberts and Gelman (2016b) argued that the research reviewed above, although important, did not provide clear insight into when children, particularly those of different racial backgrounds, reason that race is stable. For example, Roberts and Gelman suggested that children who judged that race was more stable than occupation (see Hirschfeld, 1995) could have done so because occupation was not sufficiently salient, and that children who judged that race was less stable than language (Kinzler & Dautel, 2012) may have reflected their beliefs about language more than their beliefs about race (for a fuller critique, see Roberts & Gelman, 2016b, p. 888). Therefore, to provide a further test of children's racial stability concepts, Roberts and Gelman (2016b) examined children's stability judgments of race relative to emotional expressions, as children understand emotions to be salient, but also temporary and non-stable (Flavell, Flavell, & Green, 2001; Lagattuta, Elrod, & Kramer, 2016; Lagattuta & Wellman, 2001). White participants (5-6 years, 9-10 years, and adults) and minority children (5-6

years) viewed triads consisting of one target child and two adults: one who matched the target child in race but not emotion (race match) and one who matched the target child in emotion but not race (emotion match). White 9- to 10-year-olds and adults, and minority 5- to 6-year-olds reasoned that race was more stable than emotion, whereas White 5- to 6-year-olds selected race- and emotion-matches equally, suggesting that young White children did not conceptualize race as more stable than emotion.

1. 1. The Present Studies

Prior work suggests that the extent to which children conceptualize race as stable develops between the ages of 5 and 10, and that White children conceptualize race as stable at a later age than do racial minority children. In the present studies, we assessed the strength of these beliefs. To do this, we deliberately varied the language in which the questions were presented and tested whether or how emphasizing one dimension or another would influence children's beliefs about racial stability. Following the method used by Roberts and Gelman (2016b), we presented participants with a series of target children who were happy or angry, and Black or White. In Studies 1A (White participants) and 1B (minority participants), we emphasized the temporariness of emotional expressions--that they reflected *feelings* and that they were *momentary* (e.g., "he is feeling angry right now"). We predicted that this emphasis would encourage children, even White 5- to -6-year-olds, to think of emotional expressions as temporary and therefore to conceptualize race as more stable. In Studies 2A (White participants) and 2B (minority participants), we emphasized skin color (e.g., "he has dark skin"). We predicted that this emphasis would encourage children to attend to race and treat it as more stable. In Study 3 (White 5- to 6-year-olds) we tested the influence of simultaneous race-based and emotion-based verbal cues. In each of these studies, following the introduction of the target, participants

were introduced to two images of adults: one who matched the target in race but not emotion (race match) and one who matched the target in emotion but not race (emotion match).

Participants were then asked to indicate which adult the target child would grow up to be. In Study 4, to ensure that children's responses did not reflect a simple struggle with relative tasks, we tested children's beliefs about the stability of emotion relative to gender. In Study 5, we again tested children's beliefs about the stability of emotion relative to gender, but here, we verbally cued emotion. Overall, as mentioned previously, although these studies were designed to examine children's beliefs about racial stability, they have important broader implications for understanding the origins of racial essentialism.

2. Study 1A

2. 1. Method

2. 1. 1. Participants

This study included three groups of English-speaking White U.S. participants: 5- to 6-year-olds ($N = 24$, 38% female; M age 5.58, range 5.01–6.96), 9- to 10-year-olds ($N = 26$, 62% female, M age 9.58, range 9.04–10.82), and adults ($N = 26$, 58% female; M age 20, range 18–26). Children were recruited in the Midwestern U.S. at museums affiliated with a university lab. Adults were recruited through Amazon's Mechanical Turk. Children's race was reported by their parents and adults reported their own race.

2. 1. 2. Materials

All materials were validated for use in this paradigm in Roberts and Gelman (2016b). Stimuli consisted of eight child faces and 16 adult faces (50% female/male; 50% angry/happy; 50% Black/White). Child images were drawn from the Child Affective Facial Expression Set (Lobue & Thrasher, 2014) and adult images were drawn from the NimStim Set (Tottenham et al.,

2009). On each of eight trials, one child and two adult faces of the same gender were displayed on a computer using PowerPoint: one adult matched the child in emotion but not race, and the other matched the child in race but not emotion (e.g., an angry White child paired with a happy White adult and an angry Black adult). On four of the trials, the target was a girl; on the other half, the target was a boy. The lateral position of the adult faces was counterbalanced within and across participants. Trials were presented in random order. All possible pairings of emotion and race were presented and gender was always kept constant within a set. Each target was randomly assigned one of two pairs of same-gender adult faces (i.e., one angry White adult and one happy Black adult; one angry Black adult and one happy White adult), counterbalanced across participants.

2. 1. 3. Procedure

All participants first saw two practice trials depicting cartoon characters with differently shaped schematic bodies (i.e., triangle person, square person). For each practice trial, participants were shown a cartoon child and two cartoon adults (e.g., circle child, square adult, circle adult). All characters in the practice trials had no facial expressions and were uniformly red in color. Participants were then asked to identify which adult the child would grow up to be. Feedback was provided only if the response was incorrect. Across all studies, feedback was provided for only 3.2% of participants, and the experimenter only proceeded to the test trials after successful completion of the practice trials. On the eight test trials, the experimenter revealed a target child, pointed to the screen, and emphasized the emotional expression of the target by saying, "Here is a child [pointing]. He/she is feeling happy/angry right now. When this child grows up, which grown-up will he/she be?" The experimenter then revealed two images of adults and said (while

pointing), "Will this child grow up to be this grown-up or this grown-up?" No emotion-based verbal cues were provided for the images of adults.

2. 2. Results

All analyses were Bonferroni-corrected and results are depicted in Figure 1. There were no effects of target gender, emotion, or race, or for participant gender, so the data were collapsed over these variables. A univariate ANOVA with age group (5-6, 9-10, adult) as the independent variable and the number of same-race matches as the dependent variable yielded a significant effect, $F(2, 73) = 34.88, p < .001, \eta_p^2 = .49$. Pairwise comparisons indicated that White 5- to 6-year-olds ($M = 1.75, SE = .54$) and White 9- to 10-year-olds ($M = 1.96, SE = .57$) were significantly less likely than adults ($M = 6.96, SE = .37$) to make same-race matches ($ps < .001$). Responses for White 5- to 6-year-olds and White 9- to 10-year-olds were not significantly different ($p = 1$). One-sample t -tests indicated further that White 5- to 6-year-olds made same-race matches at below chance levels (i.e., 4), $t(23) = -4.17, p < .001, d = .85$, as did White 9- to 10-year-olds, $t(25) = -3.60, p = .001, d = .70$. Adults made same-race matches at above chance levels, $t(25) = 7.67, p < .001, d = 1.50$. Non-parametric Wilcoxon signed-ranks tests confirmed these tests further. In the group of White 5- to 6-year-olds, 19 children chose more same-emotion matches, 4 children chose more same-race matches, and 1 child chose the same-emotion and same-race matches at equal rates ($Z = 2.99, p = .003, r = .61$). For White 9- to 10-year-olds, 19 children chose more same-emotion matches, 5 children chose more same-race matches, and 2 children chose the same-emotion and same-race matches at equal rates ($Z = 2.85, p = .004, r = .57$). For adults, 24 chose more same-race matches, 1 adult chose more same-emotion matches, and 1 adult chose the same-emotion and same-race matches at equal rates ($Z = 4.16, p < .001, r = .81$).

Figure 1. Mean frequencies (out of 8) for participants' choice of emotion- and race-matches in Study 1. Bars depict standard error.

2. 3. Discussion

Previous research suggests that White 5- to 6-year-old children do not conceptualize race as more stable than emotion, whereas White 9- to 10-year-old children do (Roberts & Gelman, 2016b). We hypothesized that emphasizing the temporary nature of emotions may scaffold children's understanding and provide a context in which they might conceptualize race as stable at a younger age. To do so, we presented children with a task that used verbal cues to emphasize the temporary nature of emotional expressions (e.g. "Here is a child. She is *feeling* angry *right now*"). We expected that this wording would encourage White 5- to 6-year-old children to think of emotional expressions as non-stable, and therefore race as more stable. Contrary to this expectation, however, White 5- to 6-year-olds made more emotion-based matches, thereby indicating a persistent view of race as not stable. Even more surprisingly, White 9- to 10-year-olds made more emotion- than race-based matches, suggesting that they, too, treated race as non-

stable. These results could not be explained by a simple auditory matching strategy, because the images of the adult response options were never labeled.

An alternative explanation for these surprising results could be that verbally labeling emotions (and not race) focused children's attention toward emotion and away from race (see Diesendruck & Weiss, 2015; Gelman & Davidson, 2013; Waxman & Markow, 1995; Waxman, 2010, for evidence of children's sensitivity to labels). That is, these results may indicate that children's attention is drawn toward emotions when they are verbally highlighted, rather than that they necessarily believe race to be non-stable. We return to this issue in Studies 3, 4, and 5. In any case, these results are striking for demonstrating that even at 9-10 years of age, White children do not seem firmly committed to the idea that race is stable over development.

3. Study 1B

Because prior work showed an earlier conceptualization of race as stable among minority children than among White children (Kinzler & Dautel, 2012; Roberts & Gelman, 2016b), it is critical to examine the influence of verbal cues with non-White samples. This was the purpose of Study 1B. We predicted that minority children, compared to same-aged White children in Study 1A, would be less influenced by emotion-based verbal cues, thereby reflecting an earlier belief that race is stable. We focused only on racial minority children (ages 5-6 and 9-10), reasoning that if they were less susceptible to emotion-based cues than their White peers, adults would be as well. We included racial minority participants from a variety of racial minority backgrounds (i.e., both Black and non-Black), as previous research indicates that by virtue of their minority status, racial minority children think about racial categories sooner than White children (e.g., Bigler & Liben, 2006; Kinzler & Dautel, 2012; Roberts & Gelman, 2016b; Quintana 1998).

3. 1. Method

3. 1. 1. Participants

Study 1B included two groups of English-speaking racial/ethnic minority participants: 5- to 6-year-olds ($N = 24$, 42% female; M age 6.17, range 5.01–6.99; 7 African American, 5 Asian, 5 Multiracial, 4 Latino, 2 Arabic, 1 Native American), and 9- to 10-year-olds ($N = 26$, 50% female, M age 9.42, range 9.01–10.99; 13 African American, 7 Asian, 1 Arabic, 1 Latino, 1 Multiracial, 1 Other). Children were recruited from the same sources as those in Study 1A. Children's race/ethnicity was reported by their parents. Notably, White 5- to 6-year-olds in Study 1A were on average slightly older than racial minority 5- to 6-year-olds in Study 1B, though a univariate ANOVA with age as the dependent variable and race (White, racial minority) as the independent variable showed that this was not a significant age difference ($p = .72$).

3. 1. 2. Materials, Design, and Procedure

The materials, design, and procedure were identical to those in Study 1A.

3. 2. Results

Results are presented in Figure 1. There were no effects of target gender, emotion, or race, or for participant gender, so the data were collapsed over these variables. An independent samples t -test with age group (5-6, 9-10) as the independent variable and the number of same-race matches as the dependent variable yielded no significant effect, $t(48) = .91$, $p = .37$, $d = .25$. One-sample t -tests indicated that both groups made same-race matches at chance levels (i.e., 4; 5-to 6-year-olds: $M = 3.71$, $SE = .62$, $t(23) = .47$, $p = .64$, $d = .10$; 9-to 10-year-olds: $M = 2.96$, $SE = .76$, $t(23) = -1.38$, $p = .18$, $d = .29$). According to non-parametric Wilcoxon signed ranks tests, in the group of minority 5- to 6-year-olds, 13 children chose more same-emotion matches, 8 children chose more same-race matches, and 3 children chose the same-emotion and same-race

matches at equal rates ($Z = .30, p = .76, r = .06$). For minority 9- to 10-year-olds, 16 children chose more same-emotion matches and 8 children chose more same-race matches ($Z = 1.45, p = .15, r = .30$). Notably, all effects held when looking at only Black children.

3. 2. 1. Study Comparison (1A vs. 1B)

To compare responses of White children with those of minority children, we conducted a two-way ANOVA with race (White, minority) and age (5-6, 9-10) as the independent variables and the number of race matches as the dependent variable. This analysis yielded a significant main effect of race, $F(1, 96) = 5.22, p = .025, \eta_p^2 = .05$, indicating that minority participants ($M = 3.27, SE = .44$) made more same-race matches than White participants ($M = 1.87, SE = .43$). There was no main effect of age group or an interaction between age group and race ($ps < .40$). Thus, in the presence of emotion-based verbal cues, minority children were overall more likely than same-aged White children to conceptualize race as stable.

3. 3. Discussion

In Study 1B, both minority 5- to 6-year-olds and minority 9- to 10-year-olds made same-race and same-emotion matches at comparable rates, thus not predicting race to be stable, in contrast to prior research. These results are similar to those of Study 1A with White children. Importantly, however, a cross-study comparison showed that minority children made more same-race matches than same-aged White children, thus showing further the role of social group membership in race-based concepts (Dunham, Chen, & Banaji, 2013; Quintana, 1998; Roberts & Gelman, 2015, 2016a, 2016b). Critically, these data also suggest that although minority children and older White children are more likely than White 5- to 6-year-olds to treat race as stable (Kinzler & Dautel, 2012; Roberts & Gelman, 2016b), they may not be fully committed to this concept because even they, in the present study, were swayed by verbal cues (unlike White

adults). That is, a large body of work shows that verbal cues tell children that categories are real, salient, informative, and useful for inferential judgments (e.g., Bigler & Liben, 2006; Diesendruck & HaLevi, 2006; Diesendruck & Weiss, 2015; Gelman & Davidson, 2013; Gelman & Heyman, 1999; Gelman & Markman, 1986; Heyman & Gelman, 2000a; Waxman, 2010). Thus, it appears that because we provided participants with emotion-based verbal cues but not race-based verbal cues, we oriented children away from race and toward emotion. Importantly, this effect was not found across the board, but rather selectively used more by White than minority children. We hypothesized that such task features were most relevant for participants who were uncertain or noncommittal (e.g., White 5- to 6-year-olds).

4. Study 2A

Although an intuition that race is stable is present by age 10, the results of the prior studies suggest it may not be firmly established until middle childhood (i.e., their judgments are easily swayed by verbal cues). If children are indeed uncertain regarding the stability of race, they may also be susceptible to the influence of verbal cues in the opposite direction, that is, to treat race as more stable than emotion when race-based cues are provided. Studies 2A (White participants) and 2B (minority participants) provided a test of this hypothesis.

4. 1. Method

4. 1. 1. Participants

Study 2A included three groups of English-speaking White U.S. participants: 5- to 6-year-olds ($N = 34$, 48% female; M age 5.50, range 5.05–6.83), 9- to 10-year-olds ($N = 24$, 67% female, M age 9.33, range 9.03–10.75), and White adults ($N = 24$, 54% female; M age 23, range 18–32). Children and adults were recruited from the same sources as those in Study 1A.

Children's race was reported by their parents and adults provided their own race. In each study,

our a priori goal was to recruit at least 24 participants per each age group (with the exception of Study 3, in which we tested for order effects). However, due to experimenter error, we over-recruited 4- to 6-year-olds in Study 2A.

4. 1. 2. Materials, Design, and Procedure

The materials, design, and procedure paralleled those in Study 1A. However, on each test trial, the experimenter revealed a target child, pointed to the screen, and emphasized the race of the target by saying, "Here is a child [pointing]. He/she has dark skin/light skin. When this child grows up, which grown-up will he/she be?" As in Study 1A, the experimenter then revealed two images of adults and said (while pointing), "Will this child grow up to be this grown-up or this grown-up?" No race-based cues were provided for the images of the adult response options.

4. 2. Results

Results are depicted in Figure 2. There were no effects of target gender, emotion, or race, or for participant gender, so the data were collapsed over these variables. A univariate ANOVA with age group (5-6, 9-10, and adult) as the independent variables and the number of same-race matches as the dependent variable yielded a non-significant trend, $F(2, 79) = 3.04, p = .054, \eta_p^2 = .07$, with no significant pairwise comparisons ($ps \geq .064$). One-sample t -tests revealed that all age groups made same-race matches at above chance rates (i.e., 4; 5- to 6-year-olds: $M = 6.94, SE = .34, t(33) = 8.27, p < .001, d = 1.42$; 9- to 10-year-olds: $M = 7.92, SE = .06, t(23) = 67.96, p < .001, d = 13.99$; Adults: $M = 7.63, SE = .29, t(23) = 12.34, p < .001, d = 2.52$). Non-parametric Wilcoxon signed ranks tests indicated that in the group of White 5- to 6-year-olds, 31 children chose more same-race matches and 3 children chose more same-emotion matches ($Z = 4.83, p < .001, r = .83$). In the group of White 9- to 10-year-olds, all 24 children made more same-race

matches ($Z = 4.74, p < .001, r = .97$). In the group of White adults, 23 adults made more same-race matches, and 1 adult made more same-emotion matches ($Z = 4.64, p < .001, r = .95$).

Figure 2. Mean frequencies (out of 8) for participants' choice of emotion- and race-matches in Study 2. Bars depict standard error.

4. 3. Discussion

Unlike White children in Study 1A, and consistent with our hypothesis, White participants in Study 2A (5-6 years, 9-10 years, and adults) made same-race matches when race-based cues were provided, thereby showing further children's sensitivity to verbal cues in their racial stability judgments.

5. Study 2B

We next turned to minority 5- to 6-year-olds, whom we expected to show the same pattern as same-aged White children in Study 2A. We focused only on racial minority 5- to 6-year-olds, reasoning that older racial minority children and adults would select on the basis of race, as did the same-aged White participants in Study 2A.

5. 1. Method

5. 1. 1. *Participants*

Study 2B included English-speaking racial/ethnic minority 5- to 6-year-olds ($N = 25$, 40% female; M age 5.52, range 5.02–6.95; 12 Asian, 5 Multiracial, 4 Latino, 3 African American, 1 Other). Children were recruited from the same sources as those in Study 1A and their race/ethnicity was reported by their parents.

5. 1. 2. *Materials, Design, and Procedure*

The materials, design, and procedure were identical to those in Study 2A.

5. 2. *Results*

Results are depicted in Figure 2. There were no effects of target gender, emotion, or race, or for participant gender, so the data were collapsed over these variables. A one-sample t -test revealed that minority 5- to 6-year-olds made same-race matches at above chance levels, $M = 7.04$, $SE = .46$, $t(24) = 6.66$, $p < .001$, $d = 1.33$. A non-parametric Wilcoxon signed ranks tests indicated that 22 children made more same-race matches, 2 children made more same-emotion matches, and 1 child made same-race and same-emotion matches at equal rates ($Z = 3.72$, $p < .001$, $r = .74$).

5. 2. 1. *Study Comparison (White 5- to 6-year-olds vs. minority 5- to 6-year-olds)*

To compare the extent to which the White and minority 5- to 6-year-olds used race-based cues, we conducted an independent samples t -test comparing the responses of the White 5- to 6-year-olds in Study 2A to the minority 5- to 6-year-olds in Study 2B. This analysis revealed no significant differences across the two groups, suggesting that in the presence of race-based verbal cues, minority children and White children were equally likely to conceptualize race as stable.

5. 2. 2. *Study Comparison (Emotion Cues vs. RaceCues)*

Although children displayed sensitivity to verbal cues of both emotion and race, we compared the relative strength of each cue. We therefore conducted a series of independent samples *t*-tests, comparing the number of cue-consistent matches across studies (emotion-based matches in Studies 1A and 1B, race-based matches in Studies 2A and 2B). White 5- to 6-year-olds made equivalent emotion cue-consistent matches and race cue-consistent matches, $t(56) = 2.15$, $p = .27$, $d = .29$. In contrast, both White 9- to 10-year-olds, $t(48) = 3.17$, $p = .003$, $d = .45$, and minority 5- to 6-year-olds, $t(47) = 3.59$, $p = .001$, $d = .52$, made more race cue-consistent matches than emotion-cue consistent matches.

5. 3. Discussion

As expected, when race-based cues were provided, minority 5- to 6-year-olds made same-race matches at higher rates than same-emotion matches. However, the cross-study comparisons show both developmental and race-of-child differences in children's responses. Whereas the younger White children (5-6 years) treated verbal cues regarding race and emotion as equivalent, both younger minority children (5-6 years) and older White children (9-10 years) treated verbal cues to race as more informative than verbal cues to emotion. These data thus provide converging evidence that older White children and younger minority children are more sensitive to the belief that race is stable than younger White children.

6. Study 3

An important question is why White 5- to 6-year-olds performed differently from all the other groups. Recall that these children were the only participants to reason that race was *less* stable than emotion (Study 1A), and the only group that did not favor race cues versus emotion cues on a cross-study comparison (Studies 1A vs. 2A). One possibility is that they needed more direct verbal support and scaffolding for recognizing the relative stability of race. If provided

with a task in which they hear both emotion-based and race-based verbal cues, such that both emotion and race are made salient, they may selectively treat race as more stable. However, an alternative possibility is that even when the two cues are provided side-by-side, this group would still not conceptualize race as stable. We tested these alternative predictions in Study 3.

6. 1. Method

6. 1. 1. Participants

Participants were English-speaking White 5- to 6-year-olds ($N = 43$, 42% female, M age 6.03, range 5.02-6.99). We increased the sample size in order to be able to test for any potential order effects (i.e., emotion first, race first). Children were recruited from the same sources as the previous studies and their race was reported by their parents.

6. 1. 2. Materials, Design, and Procedure

The materials, design, and procedure paralleled those in Studies 1A and 2A. However, both emotion and skin color were labeled. Thus, on each test trial, the experimenter revealed a target child, pointed to the screen, and emphasized both the emotional expression and the skin color of the target child. The order in which each feature was emphasized was counterbalanced across participants. Specifically, children were told, "Here is a child [pointing]. He/she has dark skin/light skin and he/she is feeling happy/angry right now", or "Here is a child [pointing], He/she is feeling happy/angry right now and he/she has dark skin/light skin." Following the labeling, all children heard, "When this child grows up, which grown-up will he/she be?" The experimenter then revealed two images of adult response options and said (while pointing), "Will this child grow up to be this grown-up or this grown-up?" No emotion-based or race-based cues were provided for the images of the adult response options.

6. 2. Results

There were no effects of target gender, emotion, or race, or for the order in which the verbal cues were provided (i.e., emotion first; race first), so the data were collapsed over these variables. A one-sample t -test revealed that overall, White 5- to 6-year-olds made same-race matches at chance levels (i.e., 4), $M = 3.44$, $SE = .52$, $t(42) = .29$, $p = .29$, $d = .16$, and a non-parametric Wilcoxon signed ranks test indicated that 24 children made more same-race matches, 18 children made more same-emotion matches, and 1 child made same-race and same-emotion matches at equal rates ($Z = 1.08$, $p < .28$, $r = .16$). Additionally, an independent sample t -test with participant gender as the independent variable and the number of same-race matches as the dependent variable yielded a significant effect, $t(41) = 2.61$, $p = .013$, $d = .40$, indicating that boys ($M = 4.52$, $SE = .71$) made more same-race matches than girls ($M = 1.94$, $SE = .63$). Indeed, one sample t -tests indicated that boys made same-race matches at chance levels, $t(24) = .74$, $p = .47$, $d = .15$, whereas girls were significantly below chance in selecting same-race matches, $t(17) = -3.27$, $p = .005$, $d = .73$. That is, girls selected same-emotion matches above chance. Non-parametric Wilcoxon signed ranked tests indicated that 11 boys made more same-emotion matches, and 14 boys made more same-race matches ($Z = .70$, $p = .48$, $r = .14$), whereas 13 girls made more emotion-based matches, 4 girls made more race-based matches, and 1 girl made same-emotion and same-race matches at equal rates ($Z = 2.61$, $p = .009$, $r = .62$).

6. 3. Discussion

Overall, when both emotion cues and race verbal cues were provided, White 5- to 6-year-olds made same-emotion and same-race matches equally. These data parallel those reported by Roberts and Gelman (2016b), who provided verbal cues to neither emotion nor race and also found no preference for race versus emotion stability among White 5- to 6-year-olds. Importantly, the present study verbally cued both race and emotion and contrasted them directly

(side-by-side), thus potentially scaffolding a direct comparison and making both emotions and race maximally salient. Even with this optimal context, White children in this age group did not consistently reason that race was stable, thereby further supporting the conclusion that they do not yet have robust beliefs concerning the stability of race. Additionally, we obtained a gender difference (which was not the case in any of the previous studies), such that girls were more likely to match based on emotion than were boys. Although unexpected, these data may have reflected girls' greater sensitivity toward emotional expressions, as previous research suggests that during this age period, girls are better than boys at labeling and understanding the causes of emotions (Bosacki & Moore, 2004; Zajdel, Bloom, Fireman, & Larsen, 2012). Additional research is needed to test this possibility. Nonetheless, the important point for the present context is that neither boys nor girls judged race to be stable (i.e., boys were at chance, and girls were below chance in their selection of race matches).

One important question is what the current data, which assess concepts of race using a relative measure (race *versus* emotion), can reveal about concepts of race in an absolute sense. Diesendruck, Birnbaum, Deeb, and Segall (2013) proposed that early on, children think of a broad range of social categories, including ethnicity, as stable (in an absolute sense), but that they are simply unclear which social categories are more stable than others (a relative judgment). Extending this framework to the current studies, perhaps young White children conceptualized both race and emotion as stable, and they were simply unsure which was more stable. Although this is possible, we think that it is unlikely, for two reasons. First, when the two cues were put side-by-side (Study 3), White 5- to 6-year-old girls actually selected emotion as *more* stable than race. Second, as noted earlier, children understand emotion to be malleable and temporary (e.g., Flavell et al., 2001; Lagattuta & Wellman, 2001). Thus, given that children do not view race as

more stable than emotion, this indicates that they likewise consider race to be malleable and temporary. Put somewhat differently, even if it were the case that children were unsure whether race is more stable than emotion, this would indicate that the absolute sense in which race is judged to be stable is exceedingly weak.

7. Study 4

As a more concrete test of whether children struggled with relative stability judgments, rather than racial stability judgments specifically, we tested White children's reasoning about the stability of emotion relative to gender, which is a category that they conceptualize as informative, salient, inductively potent, and stable (e.g., Blakemore, 2003; Conry-Murray & Turiel, 2012; Diesendruck, Goldfein-Elbaz, Rhodes, Gelman, & Neumark, 2013; Liben, Bigler, & Krogh, 2001; Rhodes & Gelman, 2009; Taylor, Rhodes, & Gelman, 2009). Additionally, recent work suggests that young White children may even conceptualize gender as more informative than race (Shutts, 2015; Shutts, Pemberton Roben, & Spelke, 2013). For these reasons, we predicted that White 5- to 6-year-olds, who in the absence of any verbal cues reasoned that emotion was as stable as race (Roberts & Gelman, 2016b), would reason that emotion was *less* stable than gender. Because children often perceive racial out-group faces as perceptually homogenous (e.g., Anzures et al., 2013; Bar-haim, Ziv, Lamy, & Hodes, 2006) we included only faces of White adults and children, which would therefore ensure that the stimuli were maximally distinctive with regard to gender. Critically, though, we used the same faces as those used in the previous studies.

7. 1. Method

7. 1. 1. Participants

Participants were English-speaking White 5- to 6-year-olds ($N = 25$, 48% female, M age 5.96, range 5.01-6.96). Children were recruited from the same sources as the previous studies and their race was reported by their parents.

7. 1. 2. Materials, Design, and Procedure

The materials, design, and procedure paralleled those in Studies 1A and 2A. However, instead of contrasting emotion with race, we contrasted emotion with gender. Also, because we used only the faces of White adults and children (the exact same faces used in the previous studies), there were a total of four trials (presented in random order). On each trial, one Whitechild and two Whiteadult faces were displayed: one adult matched the child in emotion (but not gender) and the other adult matched the child in gender (but not emotion). Specifically, on each trial, the experimenter first introduced the target child, "Here is a child [pointing]" followed by two adults, "Here are two grown-ups [pointing]." Next, the experimenter said, "When this child [pointing] grows up, which grown-up will it be. Will it grow up to be this grown-up [pointing] or this grown-up [pointing]"? We intentionally did not use gender pronouns (e.g., he/she) as they could have oriented children toward gender.

7. 2. Results

Because there were no effects for target gender, target emotion, or participant gender, data were collapsed over these variables. As predicted, a one-sample t -test comparing children's gender-based matches to chance (i.e., .5) revealed that overall, White 5- to 6-year-olds made same-gender matches at above chance levels, $M = 2.72$, $SE = .32$, $t(24) = 2.26$, $p = .033$, $d = .46$, and a non-parametric Wilcoxon signed ranks test indicated that 16 children made more same-gender matches, 7 children made more same-emotion matches, and 2 children made same-gender and same-emotion matches at equal rates ($Z = 2.16$, $p = .03$, $r = .44$).

7. 3. Discussion

Thus, as predicted, White 5- to 6-year-olds, who had been previously shown to conceptualize race as not more or less stable than emotion, did indeed conceptualize gender as more stable than emotion, demonstrating that these children do not simply struggle with *relative* stability judgments per se, but more specifically, that they struggle with relative *racial* stability judgments.

8. Study 5

We next tested the strength of children's belief that gender was more stable than emotion, by providing children emotion-based cues, which would provide further insight into children's beliefs. That is, if children, treat gender as more stable than emotion even in the presence of emotion-based verbal cues, this would provide additional evidence that the previous results did not stem from a task demand or a broader reasoning difficulty, but rather, from a race-based difficulty more specifically. We again included a group of White 9- to 10-year-olds, because they, too, were influenced by emotion-based verbal cues in Study 1A.

8. 1. Method

8. 1. 1. Participants

Study 5 included two groups of English-speaking White U.S. participants: 5- to 6-year-olds ($N = 24$, 46% female; M age 6.06, range 5.22–6.92) and 9- to 10-year-olds ($N = 24$, 58% female, M age 10.22, range 9.03–10.99) recruited from the same sources as those in Study 1. Children's race was reported by their parents.

8. 1. 2. Materials, Design, and Procedure

The materials, design, and procedure paralleled those in Study 4. However, on each test trial, the experimenter revealed a target child, pointed to the screen, and emphasized the emotion

of the target by saying, "Here is a child [pointing]. This child is feeling angry/happy right now. When this child [pointing] grows up, which grown-up will it be? Will this child grow up to be this grown-up [pointing] or this grown-up [pointing]?" No emotion-based cues were provided for the images of the adults, and gender pronouns were never used.

8. 2. Results

There no effects for target gender or target emotion, so the data were collapsed over these variables. Atwo-way ANOVA with age group (5-6, 9-10) and participant gender (male, female) as the independent variables and the number of same-gender matches as the dependent variable yielded a significant effect of participant gender, $F(1, 44) = 3.58, p = .024, \eta_p^2 = .11$, but no significant effect of age group, $F(1, 44) = .05, p = .82, \eta_p^2 = .001$.¹ One-sample t -tests showed further that both 5- to 6-year-olds, $t(23) = .54, p = .54, d = .13$, and 9- to 10-year-olds, $t(23) = .10, p = .92, d = .02$, made same-gender matches at chance levels (i.e., 2). Non-parametric Wilcoxon signed ranks tests indicated that in the group of White 5- to 6-year-olds, 14 children chose more same-emotion matches and 10 children chose more same-gender matches ($Z = .47, p = .64, r = .10$). For White 9- to 10-year-olds, 12 children chose more same-gender matches, 11 children chose more same-raceemotion matches, and 1 child chose same-gender and same-emotion matches at equal rates ($Z = .02, p = .99, r < .01$).

8. 2. 1. Study Comparison (Strength of Emotion Cues)

¹ Girls ($M = 2.50, SE = .38$) made more same-gender matches than boys ($M = 1.22, SE = .39$), though one-sample sample t -tests comparing gender matches to chance (i.e., 2) showed that both girls, $t(24) = 1.42, p = .17, d = .28$, and boys, $t(22) = -2.20, p = .059, d = .42$, made same-gender matches at chance levels. Non-parametric tests showed that in the group of boys, 16 made chose more same-emotion matches, and 7 chose more same-gender matches ($Z = 1.88, p = .06, r = .39$). In the group of girls, 15 chose more same-gender matches, 9 chose more same-emotion matches, and 1 chose both same-gender and same-emotion matches equally ($Z = 1.49, p = .14, r = .30$). These patterns held across both age groups, though note that the cell sizes were underpowered for this analysis. Because these effects were not expected, central to the current research, or below the $p < .05$ level, they are not discussed further.

We next tested whether children made more emotion cue-consistent matches when the competing feature was race (Study 2A) compared to when the competing feature was gender (Study 5). Because the number of trials varied across studies (8 trials in Study 2A, 4 trials in Study 5), we calculated the proportion of trials on which children made emotion-based matches (potential range = 0 to 1). A two-way ANOVA with age group (5-6, 9-10) and Study (Study 2A – emotion vs. race, Study 5 emotion vs. gender) yielded a significant main effect of Study, $F(1, 94) = 7.99, p = .006, \eta_p^2 = .08$, showing that children made more emotion cue-consistent matches in Study 2A (race vs. emotion; $M = .77, SE = .06$) than in Study 5 (gender vs. emotion; $M = .53, SE = .06$).

8. 3. Discussion

When emotion was pitted against gender, emotion cues influenced White 5- to 6-year-olds and White 9- to 10-year-olds to reason that gender was *equally* as stable as emotion. This finding was somewhat surprising, especially given the salience of gender in children's social cognition (e.g., Blakemore, 2003; Conry-Murray & Turiel, 2012; Diesendruck et al., 2013; Liben et al., 2001; Rhodes & Gelman, 2009; Taylor et al., 2009), and does suggest that children's responses stemmed, at least in part, from a task demand. Critically, though, the effects detected here were not completely explained by a task demand. Recall that when emotion was pitted against race (Study 2A), emotion cues influenced White 5- to 6-year-olds and White 9- to 10-years old to reason that race was *less* stable than emotion. Indeed, when we directly compared how often children in Study 2A made emotion-cue consistent matches to those in Study 5, we found that those in Study 2A were swayed more by emotion cues. Taken together, these data suggest further that children's reasoning about race is different from their reasoning about

gender, such that verbal cues can more powerfully override racial stability judgments than gender stability judgments.

9. General Discussion

In five experiments, we examined children's judgments about racial stability, which adults--particularly those who hold an essentialist notion of race--often hold as intuitive (Haslam et al., 2000). Specifically, we tested the strength of children's racial stability judgments by examining how such judgments were influenced by verbal cues that emphasized emotions and/or race. Overall, we found that throughout elementary school, children, unlike adults, did not consistently conceptualize race as more stable than emotion, children's judgments were influenced by verbal cues, and minority children's judgments were more adult-like than those of same-aged White children. In Study 1A, focusing on White participants, we found that when emotion cues were highlighted verbally, White children (ages 5-6 and 9-10), but not White adults, judged that race was less stable than emotion. In Study 1B we used the same task and found that minority children (ages 5-6 and 9-10) judged that race was stable to a greater extent than same-aged White children, although they did not match by race above chance (i.e., they made same-emotion and same-race-matches at equal rates). These findings were surprising given that the temporary nature of emotion was highlighted. We speculate that a more powerful and extended reminder about the changing nature of emotions may be more effective in reducing children's selections based on emotional expression (e.g., presenting participants with a vignette in which a child's race remains stable, but their emotional expressions change throughout a day).

Following Study 1A and 1B, we hypothesized that the use of emotion words per se may have encouraged children to attend to emotion and disregard race. Indeed, when we verbally highlighted race, both White (Study 2A) and minority children (Study 2B) judged that race was

more stable than emotion. When examining the strength of the verbal cues, we found that minority children 5-6 years of age and White children 9-10 years of age did indicate some sensitivity to the relative importance of race (they made more use of race-based verbal cues), compared to White children 5-6 years of age, who again showed complete lack of differentiation, even when both cues were provided side-by-side to facilitate their comparison (Study 3).

In Study 4, children judged gender as more stable than emotion, which demonstrated that children's lack of racial stability judgments stemmed from a difficulty with reasoning about race, rather than from a difficulty with reasoning on relative tasks. That is, when the task involved emotion relative to race, White 5-to -6-year-olds reasoned that emotion was not more (or less) stable than race (Roberts & Gelman, 2016b), yet when the task involved emotion relative to gender, White 5-to 6-year-olds reasoned that emotion was indeed less stable than gender. In Study 5, when emotions were highlighted verbally, White children (ages 5-6 and 9-10) made same-emotion and same-gender matches at equal rates. Importantly, this finding demonstrated that children were not simply matching on the basis of whichever feature the experimenter cued. Had this been the case, we would have found emotion cues to have the same effect regardless of whether the competing feature was race or gender. Rather, emotion cues had significantly stronger effects when the competing feature was race (Study 1A), and significantly weaker effects when the competing feature was gender (Study 5). In other words, it was easier to sway children away from race-based judgments than away from gender-based judgments. Taken together, Studies 4 and 5 suggest that children's judgments about racial stability were not artifacts created by relative task demands, nor were they only the product of verbal cuing.

The present data have important implications for understanding the development of racial stability judgments, and the development of racial essentialism more broadly. First, as noted

above, White 5- to 6-year-olds do not judge that race is stable over development, and this judgment differs from that of their same-aged minority peers. Second, verbal cues shift children's racial stability judgments, which highlights the powerful role that language plays in children's judgments about social categories (see Diesendruck & HaLevi, 2006; Heyman & Gelman, 2000a; 2000b; Waxman & Markow, 1995). Third, even White 9- to 10-year-olds and minority children (at ages 5-6 and 9-10) indicated some uncertainty about the stability of race over time, as demonstrated by their shifts in responses as a function of the verbal cues (though racial minorities were less susceptible to verbal highlighting than their White peers), suggesting that their conceptualization of race as stable is less firmly established than detected in previous research. Racial stability judgments may indeed develop well into adolescence, which is when children begin to associate race with various social consequences (e.g., identity, social status, relationship patterns, experiences with inequality and discrimination; Bigler, Averhart, & Liben, 2003; Quintana, 1998, 2008; Shutts, 2015). If so, this would suggest important developmental changes are taking place in race concepts during this period. Research with older children and adolescents could detect when children's racial stability judgments become resistant to verbal cues. Such research could also address the important question of causal direction: does learning about the stability of race cause children to subsequently attend to the social consequences of race, or does attending to the social consequences of race cause children to understand that race is stable? Until then, these data document further the development of race-based concepts across age groups and social groups, and show how verbal cues can yield surprising judgments.

Future work is needed to determine the mechanism by which verbal cues influence children's judgments. One possibility is that children are sensitive to pragmatic implications, such that verbal cues point to the relevant task features. Even if children are aware of both race

and emotional expression in the experimental stimuli, those who are uncertain about the stability of race or are not committed may be more likely to use verbal cues to decide which feature to select. A second possibility could involve category salience, such that verbal cues orient children toward one feature (e.g., emotion) and away from another (e.g., race). That is, the verbal cues may influence children to consider only one feature to the exclusion of another. The present studies do not completely rule out this latter possibility. For instance, in Study 4, children reliably indicated that gender was more stable than emotion (see also Taylor et al., 2009), but in Study 5, when emotion-based cues were provided, both younger and older children indicated that gender and emotion were equally stable, suggesting that the emotion-based cues made emotions more (and gender less) salient.

Critically, though, regardless of the mechanism (i.e., racial stability vs. racial salience), the present data provide several important conclusions about the development of race concepts. First, White 5- to 6-year-olds were swayed more by verbal cues than were minority 5- to 6-year-olds and White 9- to 10-year-olds, suggesting that minority children have stronger race concepts than White children, and that older children have stronger race concepts than younger children (see also Kinzler & Dautel, 2012; Roberts & Gelman, 2016b). Second, children in Study 3 were provided with a task in which emotion and race were both maximally salient; they received verbal cues to both features, yet roughly only half of the sample made more race-based matches. Again, this suggests that White 5- to 6-year-olds do not have a firm understanding of race concepts as stable. Third, as mentioned above, emotion-based cues held more sway when the competing feature was race (Study 1A) than when it was gender (Study 5), suggesting that children's understanding of race may emerge later than their understanding of gender. Thus, under either interpretation of the present tasks (i.e., as assessing race salience vs. race stability),

the present studies document that race concepts vary with age and racial group membership, are influenced by verbal input, and are relatively weak compared to gender concepts (see also Shutts, 2015).

Undoubtedly, future research will be needed to systematically disentangle children's beliefs about racial stability from the salience of these social dimensions. Until then, the present research suggests that understanding race as stable varies across development and racial group membership, such that even by age 10, children do not have robust beliefs concerning the stability of race (but minority children do to a greater extent than White children). Interestingly, these data suggest that although adults may question a person who adopts a different racial identity than they were assigned at birth (e.g., Rachel Dolezal), young children may not give it a second thought.

Acknowledgments

References

- Bigler, R. S., Averhart, C. J., Liben, L. S. (2003). Race and the workforce: Occupational status, aspirations, and stereotyping among African American children. *Developmental Psychology*, 39, 572-580. doi: 10.1037/0012-1649.39.3.572
- Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.). *Advances in Child Development and Behavior* (pp. 39–89). San Diego, CA: Elsevier Academic Press.
- Bosacki, S. L., & Moore, C. (2004). Preschoolers' understanding of simple and complex emotions: Links with gender and language. *Sex Roles*, 50, 659–675. doi: 10.1023/B:SERS.0000027568.26966.27
- Diesendruck, G., Birnbaum, D., Deeb, I., & Segall, G. (2013). Learning what is essential: Relative and absolute changes in children's beliefs about the heritability of ethnicity. *Journal of Cognition and Development*, 14, 546–560. doi:10.1080/15248372.2012.691142
- Diesendruck, G., Goldfein-Elbaz, R., Rhodes, M., Gelman, S. A., & Neumark, N. (2013). Cross-cultural differences in children's beliefs about the objectivity of social categories. *Child Development*, 84, 1906-1917. doi: 10.1111/cdev.12108
- Diesendruck, G., & HaLevi, H. (2006). The role of language, appearance, and culture in children's social induction. *Child Development*, 77, 539–553. doi:10.1111/j.1467-8624.2003.00889.x
- Dolezal, R (2016, June 16th). Rachel Dolezal breaks her silence on TODAY: 'I identify as black' (M. Lauer, Interviewer) [video recording]. Retrieved from:

<http://www.today.com/news/rachel-dolezal-speaks-today-show-matt-lauer-after-naacp-resignation-t26371>

- Dunham, Y., Chen, E. E., & Banaji, M. R. (2013). Two signatures of implicit intergroup attitudes: Developmental invariance and early enculturation. *Psychological Science*, 24, 860–8. doi: 10.1177/0956797612463081
- Flavell, J. H., Flavell, E. R., & Green, F. L. (2001). Development of children's understanding of connections between thinking and feeling. *Psychological Science*, 12, 430–432. doi:10.1111/1467-9280.00379
- Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, 66, 327–353. doi:10.1016/j.cogpsych.2013.02.001
- Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*, 10, 489–493. doi: 10.1111/1467-9280.00194
- Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *The British Journal of Social Psychology*, 39, 113–27. doi: 10.1348/014466600164363
- Haslam, N., & Whelan, J. (2008). Human natures: Psychological essentialism in thinking about differences between people. *Social and Personality Compass*, 2/3, 1297-1312. doi: 10.1111/j.1751-9004.2008.00112.x
- Heyman, G. D., & Gelman, S. A. (2000a). Preschool Children ' s Use of Trait Labels to Make Inductive Inferences. *Journal of Experimental Child Psychology*, 19, 1–19. doi:10.1006/jecp.1999.2555

Heyman, G. D., & Gelman, S. A. (2000b). Preschool children's use of novel predicates to make inductive inferences about people. *Cognitive Development*, 15, 263–280.

doi:10.1016/S0885-2014(00)00028-9

Hirschfeld, L. A. (1995). Do children have a theory of race? *Cognition*, 54, 209–52.

doi:10.1016/0010-0277(95)91425-R

Hirschfeld, L. A., & Gelman, S. A. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development*, 12, 213–238.

doi: 10.1016/S0885-2014(97)90014-9

Hobbs, A. (2014). *A chosen exile: A history of racial passing in the American Life*. Cambridge, MA: Harvard University Press.

Kinzler, K. D., & Dautel, J. B. (2012). Children's essentialist reasoning about language and race.

Developmental Science, 15, 131–138. doi:10.1111/j.1467-7687.2011.01101.x

Lagattuta, K. H., & Wellman, H. M. (2001). Thinking about the past: early knowledge about links between prior experience, thinking, and emotion. *Child Development*, 72, 82–102.

doi:10.1111/1467-8624.00267

Lagattuta, K. H., Elrod, N. M., & Kramer, H. J. (2016). How do thoughts, emotions, and decisions align? A new way to examine theory of mind during middle childhood and beyond. *Journal of Experimental Child Psychology*. Advance online publications. doi:

10.1016/j.jecp.2016.01.013

Lobue, V., & Thrasher, C. (2014). The Child Affective Facial Expression (CAFÉ) set.

Databrary. Retrieved from <http://databrary.org/volume/30>

- Olson, K. R., Shutts, K., Kinzler, K. D., & Weisman, K. G. (2012). Children associate racial groups with wealth: Evidence from South Africa. *Child Development*, 83, 1884-1899. doi: 10.1111/j.1467-8624.2012.01819.x
- Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). Race salience and essentialist thinking in racial stereotype development. *Child Development*, 81, 1799–1813. doi: 10.1111/j.1467-8624.2010.01511.x
- Pauker, K., Xu, Y., Williams, A., & Morris-Biddle, A. (2015). Racial essentialism and social contextual differences in children's racial stereotyping. *Child Development*, 1, 1–35. doi: 10.1017/CBO9781107415324.004
- Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science*, 16, 202-206. doi: 10.1111/j.1467-8721.2007.00504.x
- Quintana, S. M. (1998). Children's developmental understanding of ethnicity and race. *Applied and Preventive Psychology*, 7, 27–45. doi:10.1016/S0962-1849(98)80020-6
- Quintana, S. M. (2008). Racial perspective taking ability: Developmental, theoretical, and empirical trends. In S. M. Quintana & C. McKown (Eds.). *Handbook of race, racism, and the developing child* (pp. 16-36). Hoboken, New Jersey: John Wiley and Sons.
- Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, 59, 244-274. doi: 10.1016/j.cogpsych.2009.05.001
- Roberts, S. O., & Gelman, S. A. (2015). Do children see in black and white? Children's categorizations of multiracial individuals. *Child Development*, 86, 1830-1847. doi: 10.1111/cdev.12410

- Roberts, S. O., & Gelman, S. A. (2016a). Multiracial children's and adults' categorizations of multiracial individuals. *Journal of Cognition and Development*. Advance online publication.doi: 10.1080/15248372.2015.1086772
- Roberts, S. O., & Gelman, S. A. (2016b). Children's reasoning about the stability of emotion and race. *Developmental Psychology*, 52, 887-893. doi: 10.1037/dev000132
- Roberts, S. O., Williams, A. D., & Gelman, S. A. (in press). Children's and adults' predictions of Black, White, and Multiracial friendship patterns. *Journal of Cognition and Development*.
- Ruble, D. N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., & Shrout, P. E. (2007). The role of gender constancy in early gender development. *Child Development*, 78, 1121-1136. doi: 10.1111/j.1467-8624.2007.01056.x
- Shutts, K. (2015). Young children's preferences: Gender, race, and social status. *Child Development Perspectives*, 9, 262-266. doi: 10.1111/cdep.12154
- Shutts, K., Pemberton Roben, C. K., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships, 14, 35-62, doi: 10.1080/15248372.2011.638686
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Marcus, D. J.,...Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research assistants. *Psychiatry Research*, 168, 242-249. doi: 10.1016/j.psychres.2008.05.006
- Waxman, S. R. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, 40, 593–610. doi:10.1002/ejsp
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence

from 12- to 13-month-old infants. *Cognitive Psychology*, 29, 257-302. doi:

10.1006/cogp.1995.1016

Zajdel, R. T., Bloom, J. M., Fireman, G., & Larsen, J. T. (2012). Children's understanding and experience of mixed emotions: the roles of age, gender, and empathy. *The Journal of Genetic Psychology*, 174, 582–603. doi: 10.1080/00221325.2012.732125