Developmental Science

Developmental Science 18:3 (2015), pp 469–483 DOI: 10.1111/desc.12228

PAPER

The development of race-based perceptual categorization: skin color dominates early category judgments

Yarrow Dunham,[1] Elena V. Stepanova,[2] Ron Dotsch[3] and Alexander Todorov[3,4]

1. *Department of Psychology, Yale University, USA*
2. *Department of Psychology, University of Southern Mississippi, USA*
3. *Behavioural Science Institute, Radboud University Nijmegen, The Netherlands*
4. *Department of Psychology, Princeton University, USA*

Abstract

Prior research on the development of race-based categorization has concluded that children understand the perceptual basis of race categories from as early as age 4 (e.g. Aboud, 1988). However, such work has rarely separated the influence of skin color from other physiognomic features considered by adults to be diagnostic of race categories. In two studies focusing on Black–White race categorization judgments in children between the ages of 4 and 9, as well as in adults, we find that categorization decisions in early childhood are determined almost entirely by attention to skin color, with attention to other physiognomic features exerting only a small influence on judgments as late as middle childhood. We further find that when skin color cues are largely eliminated from the stimuli, adults readily shift almost entirely to focus on other physiognomic features. However, 6- and 8-year-old children show only a limited ability to shift attention to facial physiognomy and so perform poorly on the task. These results demonstrate that attention to 'race' in younger children is better conceptualized as attention to skin color, inviting a reinterpretation of past work focusing on children's race-related cognition.

Research highlights

- The contribution of skin color and other physiognomic cues to race category discrimination are independently investigated.
- Findings indicate that while adults are highly sensitive to both types of cue, children in early and middle childhood overwhelmingly rely on skin color and have difficulty using other features to make race-based categorizations.
- A central implication is that children are drawing on a category of race that has a different extension from the adult category; past research focusing on children's race-related cognition needs to be reinterpreted in light of this 'category mismatch' between adults and children.

Introduction

For researchers interested in children's understanding of social categories, few topics have attracted as much attention as race. This focus is sensible given the centrality of race in the North American settings in which much such research occurs, and the fact that race-based stereotyping, prejudice, and discrimination remain prominent problems. Thus, many studies have examined children's evaluation of and inferences regarding race (reviewed in Hailey & Olson, 2013; Raabe & Beelmann, 2011). Of course, processes such as race-based stereotyping can only occur if individuals have a prior ability to use perceptual cues to pick out members of the category. That is, the science of race-based stereotyping and prejudice is parasitic on the science of race-based

Address for correspondence: Yarrow Dunham, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205, USA; e-mail: yarrow.dunham@yale.edu

perceptual categorization. Much past research has argued or assumed that even quite young children make adult-like perceptual judgments regarding race. The present study challenges this view by presenting evidence that the process of converging on adult-like perceptual judgments is more protracted than previously reported. More precisely, we focus on the specific case of the Black–White racial distinction, investigating the independent contributions of two primary cues to race category: skin color and facial physiognomy. Our results suggest that attention to physiognomy emerges gradually over the preschool and elementary school years.

It has generally been argued that an understanding of race as a perceptual category emerges quite early in development. For example, 3-month-old infants preferentially attend to faces of the same race as their primary caregivers (Bar-Haim, Ziv, Lamy & Hodes, 2006; Kelly, Quinn, Slater, Lee, Gibson, Smith & Pascalis, 2005), and 9-month-olds habituated to a series of White or Asian faces dishabituate when presented with a face of the other race (Anzures, Quinn, Pascalis, Slater & Lee, 2010), evidencing a basic form of perceptual categorization. When we turn to explicit categorization tasks, data from as far back as the 1940s support a robust ability to categorize by race in children as young as 3 to 4. For example, in the influential Clark Doll Studies, 3–4-year-olds succeeded in over 70% of race categorization trials, while 6–7-year-olds succeeded in over 90%, what we would likely consider ceiling-level performance (Clark & Clark, 1947). Other researchers have reported similar levels of accuracy in 4–5-year-olds (Crooks, 1970; Rice, Ruiz & Padilla, 1974) and 4–8-year-olds (Hraba & Grant, 1970).

Adults are also exquisitely sensitive to racial category variation (reviewed in Maddox, 2004). For example, adults automatically encode race categories during incidental viewing (Fiske, Haslam & Fiske, 1991), and incidental or even subliminal exposure to racial outgroups is sufficient to activate racial prejudice and stereotypes (e.g. Bargh, Chen & Burrows, 1996; Fazio, Jackson, Dunton & Williams, 1995), especially when the faces used as stimuli are racially prototypical in terms of skin color and other facial features (Blair, 2006; Blair, Chapleau & Judd, 2005; Blair, Judd & Chapleau, 2004a; Blair, Judd & Fallman, 2004b; Blair, Judd, Sadler & Jenkins, 2002; Eberhardt, Davies, Purdie-Vaughns & Johnson, 2006; Livingston & Brewer, 2002; see also Maddox, 2004; Maddox & Gray, 2002). In addition, adults rapidly and flexibly categorize faces by race, with categorization decisions influenced by skin color (for review, see Maddox, 2004; Blair & Judd, 2011; Stepanova & Strube, 2009, 2012; Strom, Zebrowitz, Zhang, Bronstad & Lee, 2012) as well as other aspects of facial physiognomy such as nose shape, lip fullness, hair texture, hair quality (Blair & Judd, 2011) and jaw width (Strom *et al.*, 2012). Some research reports interactive effects of skin color and facial physiognomy (Levin & Banaji, 2006; Stepanova & Strube, 2012; Stepanova, Strube & Yablonsky, 2013), while other investigators have suggested that they are relatively independent (Brooks & Gwinn, 2010), but it is increasingly clear that both kinds of features are supported by distinct neural systems (Balas & Nelson, 2010) which potentially emerge quite early in life (Balas, Westerlund, Hung & Nelson, 2011).

Despite these findings, we will argue that children's race categories are not nearly as reliable as this brief review suggests. We begin with the observation that race categories are perceptually complex and do not naturally cluster in any clear way (Cosmides, Tooby & Kurzban, 2003; Farkas, Katic, Forrest, Alt, Bagic, Baltadjiev *et al.*, 2005; Jablonski, 2004), providing reason to think they might represent a relatively difficult domain of perceptual learning. Furthermore, the categorization tasks previously used with children generally employ a two-alternative forced-choice task in which stimuli depicting a Black and a White child are presented in matched pairs, generally holding constant or controlling away other factors such as gender, age, and attractiveness. When the design moves outside that simple context, performance suffers. For example, Hirschfeld (1993) asked children to extend racial labels to classify individuals who could be Black, White, Chinese, or North African; in this case, while still above chance responding, 3–4-year-olds were successful in only about 50% of trials. Similarly, when asked to place an array of eight stimuli that varied by race (Black or White), gender, and shirt color into two bins, one for each race, only 48% of 4-year-olds correctly sorted the stimuli (Guerrero, Enesco, Lago & Rodríguez, 2010). In addition, several investigators have found that children in the preschool and early elementary school years almost never spontaneously mention race in free-description tasks, suggesting that their attention is not habitually drawn to race or that the skills to verbally express race categories are not firmly in place (Semaj, 1981; Lam, Gerrero, Damree & Enesco, 2011; Ramsey, 1991; cf. Ramsey & Myers, 1990, for somewhat higher rates of spontaneous mentioning of race).

Of course, there is nothing particularly surprising in the fact that more complex measures, as well as measures relying heavily on verbal abilities, lead to decreased performance in younger children; this would likely be the case in any research domain. However, two-alternative forced-choice measures are particularly problematic for exploring the Black–White distinction due to polysemy between the category labels and the color terms. That is, when presented with dolls, drawings, or photographs depicting Black and White individuals, we might imagine that an individual fluent in English but entirely ignorant of the *racial* distinction between Black and White labels (ignorant, we might imagine, of even the very idea of kind-like human racial variation) would actually perform quite well, merely by attending to a single dimension of skin color variation. This is certainly true when dolls or drawings that vary solely along a color dimension are employed (as in, e.g. Clark & Clark, 1947), but it is also true when photographic stimuli are used because experimenters have nearly always employed highly prototypical racial exemplars. Such stimuli maximize the difference between categories in terms of both skin color and other aspects of facial phenotype, making the boundary far starker than it frequently is in a multicultural and multiracial world, and presenting the possibility that children could succeed on the task by focusing only on a single dimension of variation (i.e. skin color).

Thus, the pattern of results observed with children cannot necessarily be taken as evidence of an adult-like appreciation of racial phenotypes; it could reflect an understanding of only or primarily skin color, or it could result from generalization from color categories even in the absence of a richer understanding of race. Is there evidence that children are sensitive to the other characteristic patterns of physiognomic variation that differentiate Black from White faces? Two (somewhat dated) studies made efforts to untangle skin color and other physiognomic cues. Gitter and colleagues (Gitter, Mostofsky & Satow, 1972) employed a set of nine dolls crossing three levels of skin color with three levels of facial physiognomy; White and Black 4–6-year-olds were asked, 'Which looks like you?' The researchers reported that children, especially Black children, were somewhat less accurate in matching themselves via physiognomy as compared to skin color, suggesting that they had less understanding of that dimension. However, these results are difficult to interpret given the paucity of detail as to the stimuli, and the fact that their dependent measure involved a comparison between the participants' (experimenter-rated) skin color and physiognomy and the child's choice of matching doll. This opens the possibility of inaccuracies in self-perception, experimenter perception, and/or perception of the dolls.

Sorce (1979) developed a set of eight drawings comprising a parametric 2 (skin color: light or dark) × 2 (hair and eyes: Afrocentric or Eurocentric) × 2 (nose and mouth: Afrocentric or Eurocentric) design; the drawings further shared or differed on one irrelevant dimension, shirt color. Three–five-year-olds were asked whether they could tell apart pairs of stimuli varying on only a single dimension, and made several successive sorts of all eight stimuli into two categories. Children were considerably more likely to notice skin color differences as compared to other differences but were as likely to make sorts based on hair and eye region as skin color (interestingly, shirt color was by far the most common non-idiosyncratic sort). Broadly, these results suggest that skin color might be the more salient dimension for children in this age range, but the results are difficult to interpret given that the task does not involve explicit categorization and that no details of the stimuli are available.

The goal of the present inquiry is to provide a rigorous investigation of children's Black–White race-based categorization when skin color and other aspects of physiognomy independently vary (for ease of exposition, we will use 'physiognomy' to refer to patterns of facial phenotypic variation *other than* skin color). Dozens of prior studies exploring other aspects of children's race-related cognition rest on the tacit assumption that children are making adult-like perceptual judgments regarding race categories. If they are not, such studies will need to be reinterpreted accordingly, for example as showing that children make evaluative judgments based on *skin color* or *verbal color terms*, but not based on an adult-like conception of *race*. To explore this, we borrowed from the adult literature, employing a set of stimuli that independently vary skin color and physiognomy across 10 levels, resulting in 100 items within which the two factors are uncorrelated (adapted from Stepanova & Strube, 2012; see Figure 1). This allows us to independently assess the influence of each factor on children's race-based categorization across the full range of variation. To ensure that we could detect emerging sensitivity to these features, we employed a simple categorization task in which participants placed target faces along a continuum ranging from a prototypically Black face to a prototypically White face, providing us with a continuous dependent measure. We tested adults as well as a relatively wide age range of children, spanning the ages over which previous reports suggest the emergence of ceiling-level performance at race-based categorization, allowing us to ascertain the relative contribution of skin color and other aspects of facial physiognomy across development.

Study 1

Methods

Participants

To capture the range over which race-based categorization appeared to reach ceiling-level performance in past

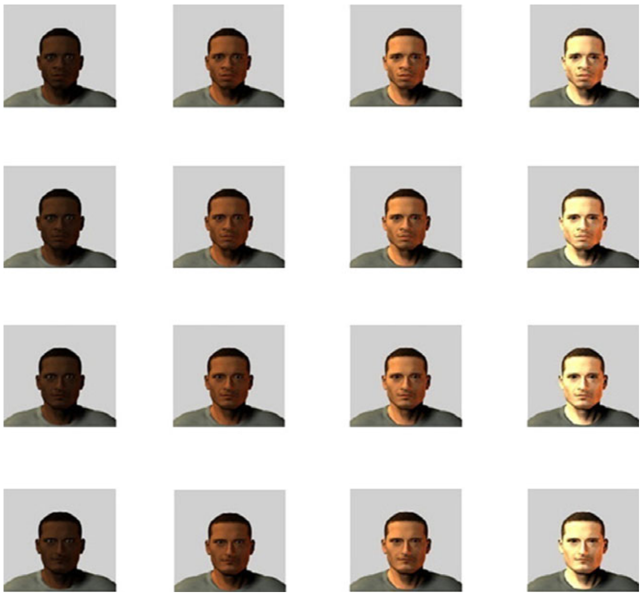

Figure 1 *Example stimuli from one of two sets of faces employed in these studies. The horizontal dimension depicts 4 of 10 levels of skin color in 3-step increments, from darker to lighter when moving left to right (i.e. steps 1, 4, 7, and 10). The vertical dimension depicts 4 of 10 levels of facial physiognomy in 3-step increments from more Afrocentric to more Eurocentric when moving from top to bottom (i.e. steps 1, 4, 7, and 10); the full 10 × 10 array is available in the online supplement. Study 1 employed a random draw of 50 faces from the full set of 100 faces or from a similar second set. In Study 2, participants rated the 20 faces most intermediate in skin color drawn (i.e. from the 5th and 6th levels of skin color) from each face set, for 40 total faces. Adapted from 'The role of skin color and facial physiognomy in racial categorization: Moderation by implicit racial attitudes' by E.V. Stepanova & M.J Strube (2012). Journal Experimental Social Psychology,* ***48****, p. 870. Copyright 2012 by Elsevier. Permission pending.*

studies, we tested 76 children between the ages of 4 and 9 (M = 6.9 years, SD = 1.7 years, *range* = 4.2 to 9.4, male = 49%). Children were recruited from schools in the central New Jersey and central Connecticut areas. Children were primarily White (81%) but also included a small number of Black (11%) and other/multiracial children (8%). Children generally came from middle-class families. Adult participants were recruited via Amazon's Mechanical Turk online marketplace. Consistent with prior work drawing from that population (Buhrmester, Kwang & Gosling, 2011; Rand, 2012), the 54 adult participants were young adult to middle age (M age = 36.4 years, SD = 13.6 years, male = 57%) and were primarily but not exclusively White (67% White, 4% Black, 11% Asian, 13% Hispanic, 5% other/unreported).

Facial stimuli

Facial stimuli were drawn from prior research on the influence of these factors in adult judgments (Stepanova & Strube, 2012). Two unique sets of stimuli created with Poser 6^{TM} software independently varied across 10 levels of skin color and 10 levels of facial physiognomy (see Figure 1 for sample stimuli), for a total of 100 stimuli per set. The 10 levels represented 10 equally spaced intervals (calibrated via the Poser 6^{TM} controls for skin color and physiognomy) along each dimension, with a highly prototypical face anchoring each end of the continuum. Thus, each step represents an independent change of 10% of the total range between the prototypical extremes in each dimension. It is important to note that this does not necessarily mean that step sizes are equal in size *across dimensions*, because if variation in one dimension (e.g. skin color) is more influential overall, then 10% of that dimension will be larger than 10% in the other dimension. Importantly, our design will allow us to empirically assess the total contribution of each dimension, but this is not equivalent to assessing the amount of lower-level perceptual variation associated with each dimension.

Changes in facial physiognomy involved several phenotypic characteristics including width of nose, fullness of lips, and global bone structure. The underlying face model for set one was based on high-fidelity facial imaging of African physiognomy and for set two was based on high-fidelity facial imaging of European physiognomy using *less or more* of the *African* program function for the former and *less or more* of the *European* program function for the latter in Poser 6^{TM}. Pre-testing and detailed process of creation is described in prior work using these stimuli; of importance in the present context, differences in facial physiognomy were clearly apparent to adults at every level of skin color and vice versa (Stepanova & Strube, 2012; Stepanova *et al.*, 2013).

Face rating task

In order to fully sample from this stimulus set while reducing participant burden for younger children, participants rated a random subset of 50 faces, presented one at a time, drawn from one of the two full sets of stimuli. The faces appeared centrally on a computer screen with an approximate angular size ranging from 8° to 10° (depending on the participant's exact seating position), a size sufficient for children to process internal facial features (Lundy, Jackson & Haaf, 2001). Participants placed the face on a linear continuum ranging from a highly prototypical Black target face to a highly prototypical White target face, which were depicted on the bottom left and bottom right of the screen. Prototypical anchor faces were the most extreme White and Black faces (in both skin color and facial physiognomy dimensions) taken from the set of faces from which target images were not drawn, thereby limiting the perceptual similarity between the anchors and the target faces. After training on scale usage, children pointed to a position on the scale, which was converted to a serial score ranging from 1 to 100. Adults selected a position using a slider, which was also converted to a serial 1 to 100 score. Stimulus set and the left/right position of the two anchors were counterbalanced across participants. For adult participants in one of these four side/set conditions, a coding error led to the administration of a single additional trial; thus a subset of adult participants contributed 51 instead of 50 total trials.

Building on our discussion of category-color polysemy, we were concerned that the labels ‘Black’ and ‘White’ might bias participants’ attention towards skin color. However, we reasoned that children were unlikely to be familiar with alternative labels such as ‘African-American’ or ‘European-American’. We therefore employed two instruction conditions as a between-participants factor. In the *label condition* participants were asked to make a judgment as to whether the target faces were ‘White or European-American’ or ‘Black or African-American’. In the *no label condition* the same anchoring images were used, but participants were instead asked to make a judgment as to whether the target faces were ‘this kind of person’ or ‘this kind of person’, accompanied by points to the respective anchor face.

Procedure

Children completed the study via a laptop computer, tested by a White experimenter in a testing room at their school. Children sat in front of the computer and the experimenter explained that they would play a ‘who’s who game’ in which they would see many people and have to decide who was who. The experimenter then advanced the display to a screen showing two anchor faces, i.e. a prototypical Black and White face positioned at opposite ends of a line. Children were told that they would see a number of new faces, which would appear centrally, and that their job was to indicate the position on the line that represented where the new face belonged. They were then verbally instructed in the use of the whole scale as a way of showing variation ranging from a perfect match with one category to a perfect match with the other category (for additional procedural details, see Supplementary Online Materials). At the conclusion of this training phase, the main task began, in which the 50 randomly selected target images drawn from the larger stimulus set were presented one at a time, and children’s points to the line were recorded to indicate their face ratings. Adults completed the entire study on their own computer, and all instructions were written.

Analysis

To respect the possibility of individual variation in attention to skin color and facial features, we analyzed face ratings in a multilevel model with trials nested within participants and treating slopes and intercepts as random effects, fit using Restricted Maximum Likelihood (REML). Following Gelman and Hill (2007), we report estimated regression coefficients and 95% confidence intervals around those estimates. We also report the percentage of modeled variance, which is an analog to explained variance (R^2) for the multi-level context, here focusing on modeled variance at the trial (stimulus) level (Bryk & Raudenbush, 1992; Snijders & Bosker, 1994). Because of the large discontinuity in age between child and adult participants, we explored effects of age within children only by treating age as a continuous variable and compared children to adults by treating age group (child or adult) as a categorical variable; however, the overall pattern of results is very similar if age is treated solely as a continuous variable across our entire sample. Additional analyses, including exploration of potential non-linear relationships between skin color, facial features, and categorization decisions, as well as some additional analyses comparing older children and adults, are described in the Supplementary Online Materials, though these analyses do not qualify the primary results reported here.

Results

Because non-White participants might be thought to exhibit a different pattern of results, for example enhanced sensitivity to race cues, we conducted primary analyses on the full sample as well as independently for White participants; for both adults and children, results were equivalent across these two analysis strategies and so results from the full sample are reported (low power precluded examining non-Whites independently, though descriptively speaking they patterned similarly).

Race categorization in adults

Given the availability of prior data with adults using a very similar procedure (Stepanova & Strube, 2012; Stepanova *et al.*, 2013), we first ensured that we replicated those findings with our modified procedure. The intraclass correlation (*ICC*), calculated from an 'unconstrained' model with random intercepts but no predictors, was 0.08, which in the multilevel modeling context indicates that 8% of the variance was due to between-participant variation, with the remaining variance attributable to trial-level variation. Preliminary analysis revealed no reliable effects of stimulus set or labeling condition, so these factors were not included in reported models. There were some modest effects of trial number, suggesting that attention to different features shifted over the course of the task, but as these findings do not relate to our main questions of interest, they are discussed in the Supplementary Online Materials. As in past work with these stimuli, adults showed clear sensitivity to both skin color, $\beta = 6.1$ [$CI = 5.3$; 6.8] and physiognomy, $\beta = 2.9$ [$CI = 2.3$; 3.5]. Finally, there was a significant interaction, such that the effect of physiognomy was stronger when skin was lighter, $\beta = -0.28$ [$CI = -0.34$; -0.22]. As a whole, this model accounted for 75% of variance in trial-level responding. The main effects are depicted in Figure 2, top panels, and the interaction is displayed in Figure 3. These findings closely parallel past work with adults that employed these same stimuli in a very similar procedure (Stepanova & Strube, 2012; Stepanova *et al.*, 2013).

Race categorization in children

One 4-year-old stopped the task a few trials into the procedure and so was dropped prior to any analysis, leaving 75 usable child participants. The *ICC* was 0.017, indicating that less than 2% of total model variance was attributable to differences between participants and so demonstrating highly consistent responding across participants. Preliminary analysis revealed no reliable effects of stimulus set or labeling condition, so these factors were dropped. There was a modest effect of trial number,

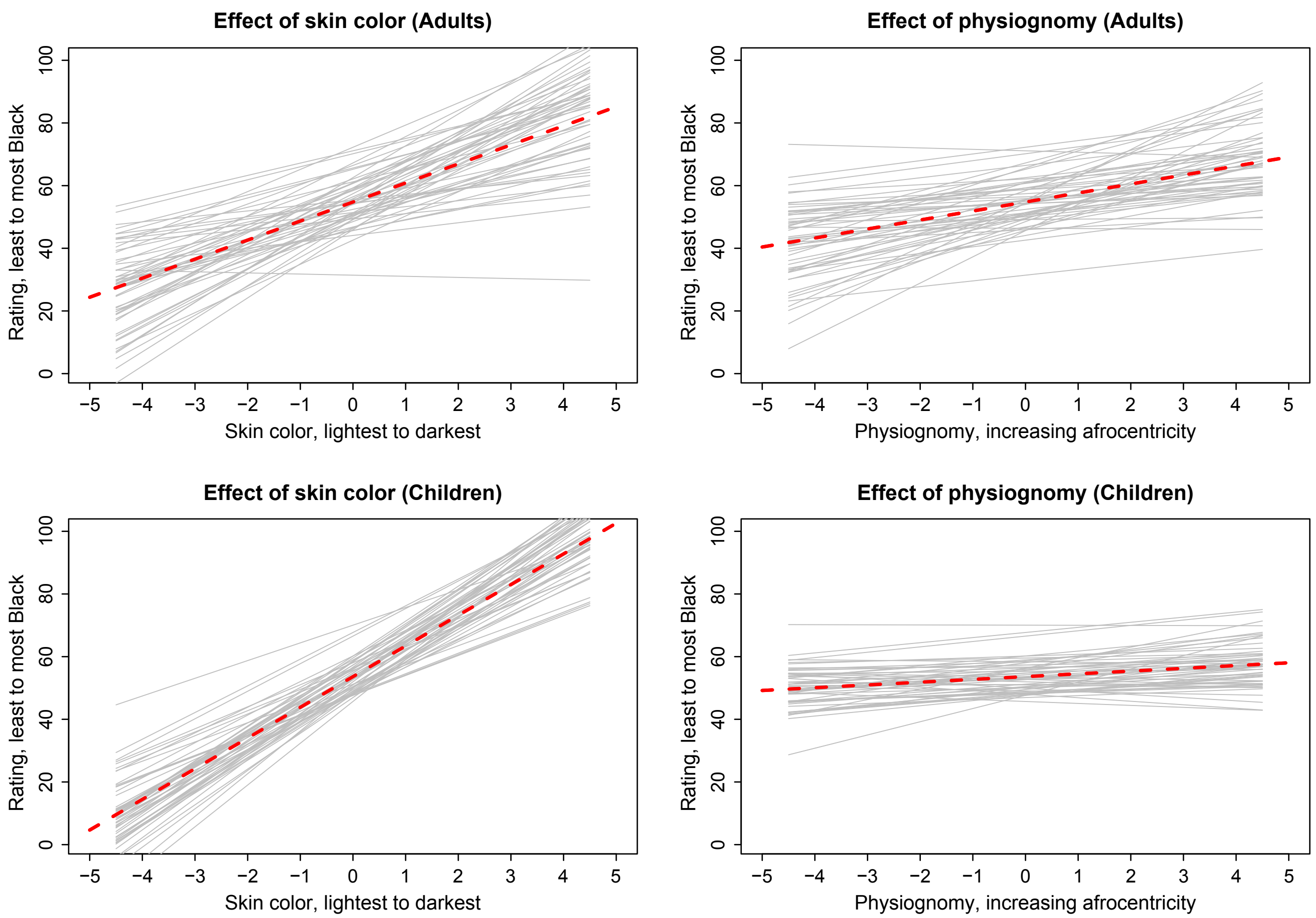

Figure 2 *Main effects of (mean-centered) skin color and facial features as predictors of face ratings for adults (top panels) and children (bottom panels) in Study 1 when the other factor is set to its mean level. Dashed red line represents the overall model and light grey lines represent the predictions for each participant, estimated from the random effects component of the models.*

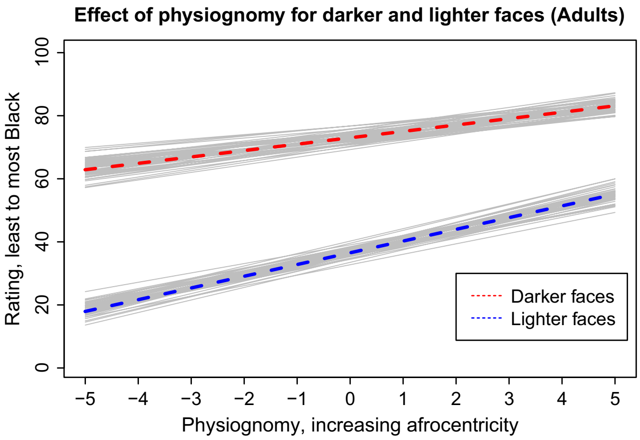

Figure 3 *Interaction between facial features and skin color for adults in Study 1. Lighter and darker faces are depicted at the 3rd and 8th levels of skin color, respectively. Light lines indicate uncertainty in the fitted regression.*

more particularly a decline in the effect of skin color as the task progressed; however, as this finding does not relate to our main questions of interest, it is discussed in the Supplementary Online Materials. Children showed robust attention to skin color, $\beta = 9.8$ [$CI = 9.2; 10.4$] but considerably more modest attention to facial features, $\beta = 0.9$ [$CI = 0.5; 1.3$]. The interaction between these terms, which had been observed in adults, was not present here, $\beta = 0.01$ [$CI = -0.07; 0.09$]. This model accounted for 67% of trial-level variance. The two main effects are depicted in Figure 3, bottom panels. Most notable is the weak effect of physiognomy in children, which is only just visible with the naked eye and is markedly weaker than that in adults.

Age-related change in categorization

The child and adult models were combined by adding a dichotomous variable contrast-coding participants by age (adult or child). There was no main effect of age group, suggesting that average ratings did not differ, $\beta = 1.2$ [$CI = -1.5; 3.8$]. However, higher-order effects were present. First, the main effect of skin color was weaker in adults than in children, $\beta = -3.7$ [$CI = -4.6; -2.8$]. Second, the main effect of physiognomy was stronger in adults than in children, $\beta = 2.0$ [$CI = 1.3; 2.7$]. Finally, the interaction between skin color and physiognomy was stronger in adults than in children, $\beta = -.3$ [$CI = -.4; -.2$]. Given that we were accounting for a similar amount of variance in adults and children considered separately, these results demonstrate that, between childhood and adulthood, participants were reducing their reliance on skin color and increasing their reliance on other facial features.

So far our analyses have collapsed children into a single age group, likely obscuring age-related change within the 4- to 9-year-olds participants in our sample. To explore this possibility, we entered age as an additional (mean-centered) predictor into the child model described above, examining its main effect and interactions with other predictors. The main effect of age was not different from 0, $\beta = -0.2$ [$CI = -1.2; 0.7$], nor was the interaction between skin color and age, $\beta = 0.1$ [$CI = -0.2; 0.4$]. However, the relationship between physiognomy and skin color was positive and was different from 0 at the 90% level, $\beta = 0.2$ [95% $CI = -0.01; 0.42$, 90% $CI = 0.03; 0.38$]. This suggests that children through this age range remain highly reliant on skin color but gradually increase their reliance on facial features. To attempt to identify the age at which attention to facial features becomes reliable, in a follow-up analysis we examined two subsets of younger children. First, for children under the age of 6 ($N = 31$, $M_{age} = 5.2$) the effect of facial features only just exceeded 0, $\beta = 0.6$ [$CI = 0.01; 1.1$], while the effect of skin color remained robust, $\beta = 9.6$ [$CI = 8.7; 10.5$]. Second, for even younger children who had not yet entered elementary school ($N = 18$, $M_{age} = 4.8$), the effect of facial features no longer differed from 0, $\beta = 0.4$ [$CI = -0.3; 1.1$], while the effect of skin color remained robust, $\beta = 9.0$ [$CI = 7.7; 10.2$]. While limited by reduced sample size, these analyses suggest that reliable attention to facial features emerges only around the time that children enter the primary school years and remains surprisingly weak even over the next several years.

Discussion

As demonstrated by past work (Stepanova & Strube, 2012), adults were acutely sensitive to both skin color and facial features, and these two factors and their interaction accounted for 75% of variance in responding. Indeed, the close correspondence between the present results and that prior work is strong evidence that the procedural changes we introduced to make the task more child-friendly did not alter its dynamics as a measure of race categorization, at least for those adult participants. In contrast, children, while quite sensitive to skin color, were much less attentive to facial features and were wholly insensitive to facial features in the preschool years. Thus, for our preschool children, two faces maximally different in facial features were treated identically as long as they were at equivalent levels of skin color. This is a remarkable finding given numerous past reports of respectable categorization accuracy in 4-year-olds (generally in the 70% range) and ceiling-level performance in 6-year-olds. Because past work has not controlled for covariation in skin color and other facial features, such results appear to be driven almost entirely by children's attention to skin color and should not be taken to indicate an adult-like perceptual understanding of race.

Indeed, we can quantify the shift in relative attention to skin color and physiognomy. Our primary models explained 75% of the trial-level variance in adults and 67% in children, but the proportion of this variance that can be attributed to each feature varied dramatically. Of this variance in adults, 59% is attributable to skin color and 15% to other features (with 1% attributable to the interaction of these two predictors). In children, the relevant contribution is 65% and 2%, respectively. Thus, skin color was approximately 4 times as powerful a predictor as other features in adults but was over 30 times as powerful in children. These findings demonstrate the primary role of skin color across development, but also highlight the dramatic age-related increase in attention to other features. A potential consequence of this differential cue weighting is that the actual extension of children's and adults' race categories – that is, the set of exemplars they believe make up each category – may differ, in that some faces that adults placed closer to the 'Black' end of the continuum were placed closer to the 'White' end of the continuum by children, and vice versa (we provide further discussion and some quantitative examinations of these differences in the Supplemental Online Materials). These category mismatches could have serious consequences for how a range of previously reported results are interpreted, a topic to which we return in the general discussion.

An additional developmental difference revealed here is that in adults but not children we observed an interaction between skin color and physiognomy, such that physiognomy exerted a larger influence for lighter faces. In past research (Stepanova & Strube, 2012), this interaction has been interpreted as evidence that very dark skin is independently sufficient to induce categorization as Black, but that lighter skin is not sufficient to induce categorization as White. Thus, in order to disambiguate lighter faces, adults pay relatively more attention to facial physiognomy. One possibility is that this is a socially motivated phenomenon, for example 'ingroup over-exclusion', in which stimuli are increasingly excluded from the ingroup as they grow more ambiguous (Castano, Yzerbyt, Bourguignon & Seron; 2002; Yzerbyt, Leyens & Bellour, 1995). The socially motivated account is bolstered in the present data by its relatively late developmental emergence, though an alternative focusing more on the lower-level dynamics of perceptual categorization for majority and minority categories is also available (Halberstadt, Sherman & Sherman, 2011).

Because we were concerned about the performance impact engendered by polysemy between color terms and racial category labels ('Black' and 'White'), we varied the way in which the task was administered, in one condition using those labels and in the other merely instructing participants to make categorizations based on the 'kind of person'. No effects of this difference in instruction condition were observed, demonstrating that children's attention to skin color in our task is not due to an enhancement of the skin color dimension through its verbal labeling. The lack of an effect of racial labeling is somewhat surprising given the powerful role of category labels in some other developmental intergroup contexts (e.g. Baron, Dunham, Banaji & Carey, 2014; Bigler & Liben, 2007; Waxman, 2010). We suspect that this is because the instructions in the 'no label' condition were sufficient to cue racial categories and thus make that version largely equivalent to the label condition.

We interpret our results as demonstrating that preschool and early elementary school children have not yet acquired a robust perceptual representation of the facial features that underlie Black–White race-based categorization and are primarily relying on the (potentially simpler or more salient) heuristic of skin color. However, an alternative account is that children do possess a richer representation of the perceptual features underlying race but are not drawing on it. This could occur if perceptual features were identifiable but less accessible or salient, either in general or because of specific task features. Of course, given that adults perform as expected, a task-based explanation of our findings with children would need to explain this developmental shift. Nonetheless, in a second experiment we investigated whether we could encourage attention to physiognomic cues by greatly reducing variation in skin color while retaining the full range of variation in facial features.

Study 2

Study 2 was identical to the prior study except for the set from which stimuli were drawn. Rather than sample from the full 10 × 10 space crossing skin color and facial features, we instead selected all stimuli occupying the middle two columns of each of the two stimulus sets, i.e. the most intermediate two levels of skin color (see Figure 1). These 40 faces represented only a single relatively subtle step of variation in skin color while retaining the full 10 steps of variation in facial features.

In adults, we predicted that this would lead to a large-scale shift towards features, and therefore a large increase in the predictive power of the feature dimension accompanied by a large decrease in the predictive power of the skin color dimension. The more crucial question is what children would do; we elected to study 6–8-year-olds, the youngest children from Study 1 who showed at least some reliance on facial features. If the relevant perceptual features are in fact well understood but were not drawn on in our prior procedure, either because of relative accessibility or due to procedural artifacts in the salience of the two dimensions in the stimuli we employed, we would expect children, like adults, to shift their focus dramatically to the facial physiognomy dimension. By contrast, if their different performance did result from a poor command of the (non skin color) perceptual properties underlying this racial distinction, we would expect that their ability to draw on this dimension would remain starkly limited.

Methods

Participants

The study involved 44 6–8-year-old children (*M* age = 7.2 years, *SD* = 1.1 years, male = 48%) and 51 adults (*M* age = 34.2 years, *SD* = 12.0 years, male = 65%). Children were recruited from an elementary school in central Connecticut and were primarily but not exclusively White (75% White) but also included a small number of Asian children (9%), Black children (5%), multiracial children (5%), and children for whom race information was not available (6%). Children were drawn from primarily middle-class families. Adults were recruited from the same online pool as in Study 1 and were primarily but not exclusively White (75% White, 8% Black, 14% Asian, 4% Hispanic).

Face rating task

The task was identical to that described in Study 1 except for a change in stimuli and number of trials. In the prior studies, stimuli varied over 10 levels of skin color; for this study, faces from the most intermediate two levels of color (i.e. the 5th and 6th levels, the middle two columns in Figure 1) were selected. We used all 10 faces from each level and from each stimulus set, resulting in 10 levels of facial features × 2 levels of color × 2 stimulus sets = 40 total stimuli. Participants completed the same task with the same procedure as that described in Study 1 except that we did not vary the labeling condition; instead, the labels 'Black/African-American' and 'White/European-American' were used with all participants, accompanied by the same anchor faces as in Study 1. The anchor faces depicting continuum endpoints were again taken from the most extreme Black and White faces, with the set from which those faces were drawn counterbalanced across participants.

Results and discussion

As in Study 1, we conducted analyses on the full sample as well as independently for White participants. All results were again substantively equivalent across these analyses, hence we report results from the full sample (low power precluded examining non-Whites independently, though descriptively they patterned similarly).

The primary results from Study 2 are depicted in Figure 4. For ease of presentation, we again present results separately for children and adults and then discuss a combined model allowing for direct comparisons across age groups.

Race categorization in adults

In adults, the intraclass correlation was 0.12, again reflecting the need for a multilevel treatment of the data. In adults, facial features were highly predictive of categorization decisions, $\beta = 4.8$ [$CI = 4.0$; 5.5], and skin color was also reliably associated with categorization, $\beta = 3.8$ [$CI = 2.6$; 5.1]. Finally, there was an interaction between these two features, such that the effect of facial features was somewhat stronger in lighter faces, $\beta = -0.4$ [$CI = -0.7$; -0.0], a similar pattern as described in the prior study. As a whole, this model accounted for 65% of trial-level variance, but this was attributable almost entirely to variation in facial features, which accounted for 64% of this modeled variance (in interpreting the parameter estimates reported above, recall that there were 10 levels of variation in features and only one level of variation in skin color). These results demonstrate that heavily constraining the diagnosticity of one of the two available cues in Study 1 did not impair adults' ability to categorize by race, in that they flexibly shifted almost entirely to physiognomy (recall that, in Study 1, we were able to model 75% of trial-level variance, only 10% more than in this study).

Race categorization in children

We observed modest inter-participant variability in children, $ICC = 0.07$. Categorization decisions were influenced by both facial features, $\beta = 2.7$ [$CI = 1.9$; 3.5] and skin color, $\beta = 11.7$ [$CI = 8.0$; 15.4]. The interaction between these two factors did not differ from 0, $\beta = 0.2$ [$CI = -0.6$; 1.0]. Overall, this model

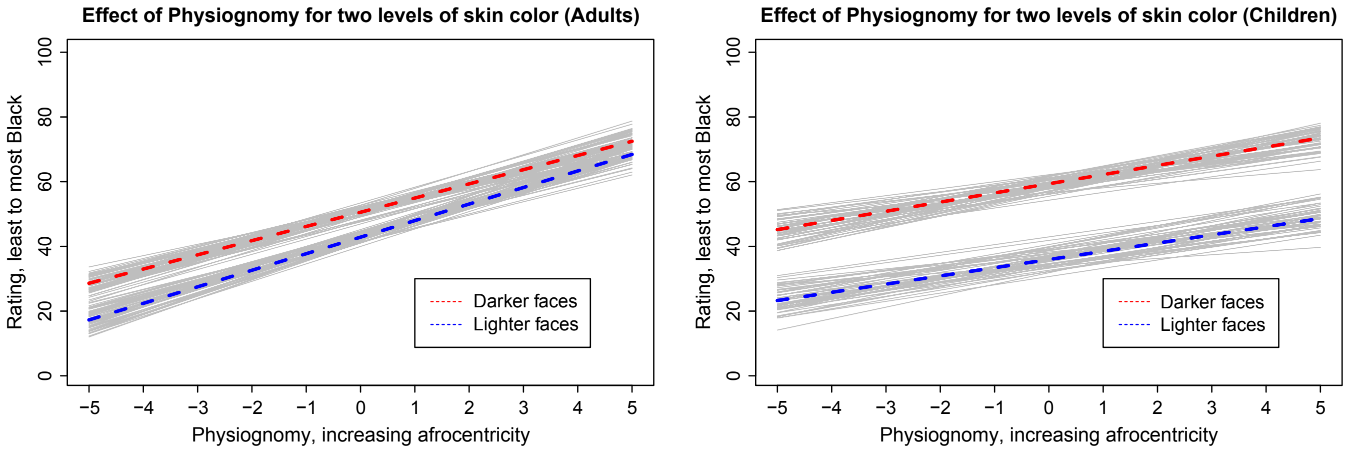

Figure 4 *Predicted effect of (mean-centered) facial features at two levels of skin color for adults (left panel) and children (right panel) in Study 2. Light lines indicate uncertainty in the fitted regression.*

accounted for 24% of trial-level variance, substantially less than we were able to account for in adults, and substantially less than the 67% modeled in children in Study 1. Nonetheless, comparing the estimate of the influence of facial features with that from children in Study 1 ($\beta = 0.9$, [$CI = 0.5$; 1.3]) suggests that the change in procedure did increase attention to facial features. However, children retained a considerable focus on skin color despite the fact that skin color variation was minimal; of the 24% variance modeled, 15% was attributable to variation in facial features and 9% to the (single level) of variation in skin color.

Age-related change in categorization

Turning to the comparison between adults and children, there was no effect of age group itself, suggesting that average ratings did not differ, $\beta = -0.9$ [$CI = -4.3$; 2.4]. However, two interactions with age were revealed, validating the differences suggested by the independent models. Adults were both more sensitive to facial features, $\beta = 2.1$ [$CI = .9$; 3.2] and less influenced by skin color, $\beta = -7.9$ [$CI = -11.5$; -4.2]. The interaction between facial features and skin color was no longer reliable in this model, $\beta = 0.2$ [$CI = -0.4$; 0.7], nor was its interactive effect with age group $\beta = -0.5$ [$CI = -1.3$; 0.3].

In sum, our efforts to channel attention towards facial features by restricting variation in skin color were successful in adults. Features became by far the dominant driver of categorization, accounting for 64% of the variation in trial-by-trial responding, with only a slight (1%) contribution from skin color and the interaction between factors. However, these efforts were considerably less successful in children, within whom only 15% of variation in trial-by-trial responding was predictable from facial features, with an additional 9% attributable to the single level of variability in skin color. Given that only 7% of total variance is attributable to between-child variation (i.e. the *ICC*) and that our stimuli were highly controlled such that other systematic patterns of between-stimulus variation are largely ruled out, it appears that children's responses were essentially becoming more random when skin color cues were reduced. They had only a limited ability to map the more systematically varying dimension of facial features to the Black–White category distinction. This is particularly striking when we consider that physiognomically prototypical Black and White exemplars were present on screen throughout the task. These findings support our interpretation of Study 1, namely that children of this age range primarily conceive of race via skin color and have not yet mastered the other facial features that adults consider category-diagnostic; however, they also add important nuance, namely that children can attend somewhat more systematically to other features when skin color variation is less diagnostic.

General discussion

The overall pattern of results described here is simple. While adults are exquisitely sensitive to both skin color and facial features as determinants of race-based categorization, understanding of facial features emerges slowly over ontogeny. Indeed, preschool-aged children were wholly insensitive to such features, and, even in middle elementary school, physiognomy exerted only a weak influence on categorization. When we adjusted our procedure to greatly increase the category diagnosticity of facial features and greatly decrease that of skin color (Study 2), younger children remained considerably more focused on skin color and less focused on facial features than were adults, who shifted to an almost complete reliance on facial features. These results are consistent with the possibility that children have only a limited understanding of feature-based racial categorization across these age ranges. Why might this be the case?

Scholars from diverse fields have argued that race is a cultural construction with little or no reality as a biological descriptor (for an excellent recent summary, see Maglo, 2011). Others have noted that the perceptual features that mark racial categories do not reliably cluster and indeed vary continuously across world populations (i.e. they reflect a 'clinal' system of human variation; Cosmides *et al.*, 2003; Farkas *et al.*, 2005; Jablonski, 2004). Thus, in increasingly diverse cultural environments, the perceptual learning problem that children face is far from trivial, because the range of variation they are exposed to may not support a 'bottom-up' feature-based emergence of race categories. Instead, the learning process may be 'top-down', requiring children to notice how linguistic categories such as 'Black' and 'White' are used by adults, and in particular how they map on to an ad hoc set of perceptual features. Our results suggest that children begin with a skin color heuristic and only gradually develop an appreciation of other category-diagnostic features.

These findings imply a need for caution in interpreting the large body of research on race-related cognition in children. We noted that studies of the development of racial categorization have tended to use stimuli that confound skin color and other aspects of facial physiognomy. Thus, the extent to which those studies demonstrate perceptual knowledge consistent with adult racial categories is limited. However, stimuli of this same sort are also used in most studies of racial attitudes, stereotypes, and racial essentialism. Thus, our results further imply that across these diverse studies children are not responding to the adult conception of 'race'. Instead, they are likely responding based solely (in the preschool years and earlier) or primarily (in the elementary school years) on skin color. Indeed, this makes it likely that the very extension of race categories in adults and children differs, in that some faces generally categorized as Black by children are generally categorized as White by adults, and vice versa (see Supplementary Online Materials for further investigation of this point). Of course, because skin color is correlated with racial category judgments across the lifespan, the reinterpretation we are urging here is a modest one. Nonetheless, it is important, in part because it may help explain a puzzling disconnect between prejudice and discrimination in early to middle childhood. Why do children so reliably express race bias (Aboud, 1988; Raabe & Beelmann, 2011) many years before clearly discriminating on the basis of race, for example, in playmate preference (Hallinan & Teixera, 1987; Shrum, Cheek & Hunter, 1988)? If studies of discrimination involve adult judgment of racial categories, while children's race-related behavior is driven primarily by attention to skin color, we would be much less likely to observe category-based discrimination simply because we are looking in the wrong place, at least until later childhood when children's categories converge on adult categories. Behavioral discrimination in younger children might be more consistently observed if researchers focused on the dimension of skin color rather than imposing an adult conception of race on children's behavior.

Future work could fruitfully explore the specific features that best predict shifts in categorization. We were unable to do this because the single dimension of physiognomy that varied in our stimuli actually involved a number of facial features changing in tandem. Recently more detailed data-driven methods for exploring mental 'face space' have been developed (e.g. Todorov, Dotsch, Wigboldus & Said, 2011; Todorov, Said, Engell & Oosterhof, 2008), which could allow us to replicate our work with stimuli that independently vary different features, providing a finer-grained picture of the specific content of visual representations of race. Relatedly, it would be helpful to replicate the current results with a broader set of 'base faces' from which stimuli were created, to ensure that results hold across a larger and more realistic range of facial variation; including female faces and other perceptual cues such as hair (e.g. MacLin & Malpass, 2001) would also be useful. In addition, it would be illuminating to explore the perceptual component of other social category distinctions. For example, Asian and White faces are less easily distinguishable by skin color; do children nonetheless attempt to focus on this dimension, or do they make an earlier shift to other facial features? Sex presents another interesting case. Absent other cues such as hair color, sex differences in face structure would seem to be the very sort of features children showed difficulty attending to.

Very broadly, skin color was a much more prominent driver of categorization than other features. Why might this be the case? One possibility is that the range of variation in skin color is simply more salient in terms of low-level perceptual properties, or that it is simply an easier dimension to monitor. Another (non-exclusive) possibility is that perceivers are drawing on a cultural model of race that emphasizes skin color and so have increased the weight assigned to it. While our work does not allow us to settle this issue, we consider the former more plausible because young children are unlikely to have internalized as robust a cultural concept of race, but were even more prone to focus on skin color than were adults.

Our results should also be interpreted in the context of broader developmental shifts in face processing ability. For example, it is generally thought that children gradually increase reliance on configural cues, and that this underlies improvements in face recognition ability (Carey & Diamond, 1977; Mondloch, Le Grand & Maurer, 2002; but cf. Crookes & McKone, 2009). For example, while even preschoolers *can* draw on both featural and configural face cues (Freire & Lee, 2001; Pellicano, Rhodes & Peters, 2006), these abilities are fragile, easily disrupted by competing cues (Freire & Lee, 2001). It is possible that skin color serves as a powerful perceptual lure for young children, pulling them away from greater attention to other features. There is also evidence from face recognition tasks suggesting that other-race faces are processed in a less holistic manner (Michel, Rossion, Han, Chung & Caldara, 2006), plausibly due to less visual experience and/or different visual encoding strategies (Gaither, Pauker & Johnson, 2012; Xiao, Xiao, Quinn, Anzures & Lee, 2013). It is likely that maturation of face processing abilities more broadly is one of the developmental forces driving the changes in categorization that we observe here. We do note, however, that it is important to distinguish results from *face recognition* tasks from *face categorization* tasks. The former will necessarily require strategies focusing at the sub-category level in order to track individual identities, while the latter requires only a mental model of broad category boundaries. Thus, our findings do not necessarily conflict with adult work suggesting that facial morphology is a more powerful force in recognition than is skin color (e.g. Bar-Haim, Saidel & Yovel, 2009; Michel, Rossion, Bülthoff, Hayward & Vuong, 2013). It would be valuable to pursue both categorization and recognition tasks in the same children so that these relationships can be examined empirically.

We interpret our results, increasing age-related integration of featural cues with skin color cues, as an example of perceptual learning in a challenging domain. However, non-perceptual social factors could also be involved. For example, older children become increasingly aware that race can be a charged topic, and are sometimes unwilling to point out skin color differences even when doing so is non-evaluative and highly task-relevant (Apfelbaum, Pauker, Ambady, Sommers & Norton, 2008). Furthermore, general conceptual views of race change over the age ranges covered here, with race becoming both more salient and more essentialized (Pauker, Ambady & Apfelbaum, 2010). More essentialist beliefs about race predict declining memory for racially ambiguous faces (Gaither, Schultz, Pauker, Sommers, Maddox & Ambady, 2014); one intriguing possibility is that increasing racial essentialism motivates children to devote more attention to refining their ability to categorize by race, in this case by broadening their conception of category-diagnostic cues to include physiognomy. At the very least, it would be unwise to consider race as a 'pure' perceptual process; it is surely intimately related to children's cultural views of race, and future work should bring these strands together (for one valuable example in this vein, see Anzures, Quinn, Pascalis, Slater & Lee, 2013).

A final avenue for future research is to broaden the participant base examined. For example, will the same pattern hold for non-White children, and particularly for Black, children? Past work has suggested that Black children show a different pattern with respect to race attitudes (e.g. Davey & Mullin, 1980; Dunham, Chen & Banaji, 2013; Newheiser & Olson, 2012), raising the possibility that they might also differ in terms of sensitivity to race cues. It is worth noting, however, that the broad pattern of results we observed was consistent across White and non-White participants, though small sample sizes of non-White participants preclude strong conclusions here. In addition, a larger and more age-diverse sample would help to clarify the trends reported here, particularly at the younger and older ends of our included ranges.

In short, the present results demonstrate that the perceptual understanding of race is a protracted developmental challenge that is not mastered until well into the elementary school years. Younger children's emerging perceptual category of 'race', as well as the responses they make to race-related stimuli, are highly consistent across children but diverge from adult perceptual representations. Interpretations of prior research on children's understanding of race will need to be reconsidered in this light.

References

Aboud, F.E. (1988). *Children and prejudice*. New York: Blackwell Publishers.

Anzures, G., Quinn, P.C., Pascalis, O., Slater, A.M., & Lee, K. (2010). Categorization, categorical perception, and asymmetry in infants' representation of face race. *Developmental Science*, **13** (4), 553–564. doi:10.1111/j.1467-7687.2009.00900.x

Anzures, G., Quinn, P.C., Pascalis, O., Slater, A.M., & Lee, K. (2013). Development of own-race biases. *Visual Cognition*, **21** (9–10), 1165–1182. doi:10.1080/13506285.2013.821428

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

Balas, B., & Nelson, C.A. (2010). The role of face shape and pigmentation in other-race face perception: an electrophysiological study. *Neuropsychologia*, **48** (2), 498–506. doi:10.1016/j.neuropsychologia.2009.10.007

Balas, B., Westerlund, A., Hung, K., & Nelson, C.A. III (2011). Shape, color and the other-race effect in the infant brain. *Developmental Science*, **14** (4), 892–900. doi:10.1111/j.1467-7687.2011.01039.x

Bargh, J.A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, **71** (2), 230–244. doi: 10.1037/0022-3514.71.2.230

Bar-Haim, Y., Saidel, T., & Yovel, G. (2009). The role of skin colour in face recognition. *Perception*, **38** (1), 145–148. doi:10.1068/p6307

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R.M. (2006). Nature and nurture in own-race face processing. *Psychological Science*, **17** (2), 159–163. doi: 10.1111/j.1467-9280.2006.01679.x

Baron, A.S., Dunham, Y., Banaji, M., & Carey, S. (2014). Constraints on the acquisition of social category concepts. *Journal of Cognition and Development*, **15** (2), 238–268. doi:10.1080/15248372.2012.742902

Bigler, R.S., & Liben, L.S. (2007). Developmental intergroup theory explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, **16** (3), 162–166. doi: 10.1111/j.1467-8721.2007.00496.x

Blair, I.V. (2006). The efficient use of race and Afrocentric features in inverted faces. *Social Cognition*, **24** (5), 563–579. doi: 10.1521/soco.2006.24.5.563

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

Blair, I.V., & Judd, C.M. (2011). Afrocentric facial features and stereotyping. In R.B.J. Adams, N. Ambady, K. Nakayama & S. Shimojo (Eds.), *The science of social vision* (pp. 306–320). New York: Oxford University Press.

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

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

Blair, I.V., Judd, C.M., Sadler, M.S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: judging by features and categories. *Journal of Personality and Social Psychology*, **83** (1), 5–25. doi: 10.1037/0022-3514.83.1.5

Brooks, K.R., & Gwinn, O.S. (2010). No role for lightness in the perception of black and white? Simultaneous contrast affects perceived skin tone, but not perceived race. *Perception*, **39** (8), 1142–1145. doi:10.1068/p6703

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Buhrmester, M., Kwang, T., & Gosling, S.D. (2011). Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, **6** (1), 3–5. doi: 10.1177/1745691610393980

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, **195** (4275), 312–314. doi: 10.1126/science.831281

Castano, E., Yzerbyt, V., Bourguignon, D., & Seron, E. (2002). Who may enter? The impact of in-group identification on in-group/out-group categorization. *Journal of Experimental Social Psychology*, **38** (3), 315–322. doi:10.1006/jesp.2001.1512

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

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, **7** (4), 173–179. doi: 10.1016/S1364-6613(03)00057-3

Crookes, K., & McKone, E. (2009). Early maturity of face recognition: no childhood development of holistic processing, novel face encoding, or face-space. *Cognition*, **111** (2), 219–247. doi: 10.1016/j.cognition.2009.02.004

Crooks, R.C. (1970). The effects of an interracial preschool program upon racial preference, knowledge of racial differences, and racial identification. *Journal of Social Issues*, **26** (4), 137–144. doi:10.1111/j.1540-4560.1970.tb01747.x

Davey, A.G., & Mullin, P.N. (1980). Ethnic identification and preference of British primary school children. *Journal of Child Psychology and Psychiatry*, **21** (3), 241–251.

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

Eberhardt, J.L., Davies, P.G., Purdie-Vaughns, V.J., & Johnson, S.L. (2006). Looking deathworthy: perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, **17** (5), 383–386. doi:10.1111/j.1467-9280.2006.01716

Farkas, L.G., Katic, M.J., Forrest, C.R., Alt, K.W., Bagic, I., Baltadjiev, G., et al. (2005). International anthropometric study of facial morphology in various ethnic groups/races. *Journal of Craniofacial Surgery*, **16** (4), 615–646. doi: 10.1097/01.scs.0000171847.58031.9e

Fazio, R.H., Jackson, J.R., Dunton, B.C., & Williams, C.J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology*, **69** (6), 1013–1027. doi: 10.1037/0022-3514.69.6.1013

Fiske, A.P., Haslam, N., & Fiske, S.T. (1991). Confusing one person with another: what errors reveal about the elementary forms of social relations. *Journal of Personality and Social Psychology*, **60** (5), 656–674. doi: 10.1037/0022-3514.60.5.656

Freire, A., & Lee, K. (2001). Face recognition in 4- to 7-year-olds: processing of configural, featural, and paraphernalia information. *Journal of Experimental Child Psychology*, **80** (4), 347–371. doi: 10.1006/jecp.2001.2639

Gaither, S.E., Pauker, K., & Johnson, S.P. (2012). Biracial and monoracial infant own-race face perception: an eye tracking study. *Developmental Science*, **15** (6), 775–782. doi:10.1111/j.1467-7687.2012.01170.x

Gaither, S.E., Schultz, J.R., Pauker, K., Sommers, S.R., Maddox, K.B., & Ambady, N. (2014). Essentialist thinking predicts decrements in children's memory for racially ambiguous faces. *Developmental Psychology*, **50** (2), 482–488. doi:10.1037/a0033493

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gitter, A.G., Mostofsky, D.I., & Satow, Y. (1972). The effect of skin color and physiognomy on racial misidentification. *Journal of Social Psychology*, **88** (1), 139–143. doi:10.1080/00224545.1972.9922552

Guerrero, S., Enesco, I., Lago, O., & Rodríguez, P. (2010). Preschool children's understanding of racial cues in drawings and photographs. *Cognitive Development*, **25** (1), 79–89. doi:10.1016/j.cogdev.2009.07.001

Hailey, S., & Olson, K.R. (2013). A social psychologist's guide to the development of racial attitudes. *Social and Personality Psychology Compass*, **7** (7), 457–469. doi: 10.1111/spc3.12038

Halberstadt, J., Sherman, S.J., & Sherman, J.W. (2011). Why Barack Obama is black: a cognitive account of hypodescent. *Psychological Science*, **22** (1), 29–33. doi:10.1177/0956797610390383

Hallinan, M.T., & Teixeira, R.A. (1987). Opportunities and constraints: Black-White differences in the formation of interracial friendships. *Child Development*, **58**, 1358–1371. doi: 10.2307/1130627

Hirschfeld, L.A. (1993). Discovering social difference: the role of appearance in the development of racial awareness. *Cognitive Psychology*, **25** (3), 317–350. doi: 10.1006/cogp.1993.1008

Hraba, J., & Grant, G. (1970). Black is beautiful: a reexamination of racial preference and identification. *Journal of Personality and Social Psychology*, **16** (3), 398–402. doi:10.1037/h0030043

Jablonski, N.G. (2004). The evolution of human skin and skin color. *Annual Review of Anthropology*, **33**, 585–623. doi: 10.1146/annurev.anthro.33.070203.143955

Kelly, D.J., Quinn, P.C., Slater, A.M., Lee, K., Gibson, A., Smith, M., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, **8** (6), F31–F36.

Lam, V., Guerrero, S., Damree, N., & Enesco, I. (2011). Young children's racial awareness and affect and their perceptions about mothers' racial affect in a multiracial context. *British Journal of Developmental Psychology*, **29** (4), 842–864. doi:10.1348/2044-835x.002013

Levin, D.T., & Banaji, M.R. (2006). Distortions in the perceived lightness of faces: the role of race categories. *Journal of Experimental Psychology: General*, **135** (4), 501–512. doi:10.1037/0096-3445.135.4.501

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

Lundy, B.L., Jackson, J.W., & Haaf, R.A. (2001). Stimulus properties, attentional limitations, and young children's face recognition. *Perceptual and Motor Skills*, **92** (3), 919–929. doi: 10.2466/pms.2001.92.3.919

MacLin, O.H., & Malpass, R.S. (2001). Racial categorization of faces: the ambiguous race face effect. *Psychology, Public Policy, and Law*, **7** (1), 98–118. doi:10.1037/1076-8971.7.1.98

Maddox, K.B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, **8**, 383–401. doi:10.1207/s15327957pspr0804

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

Maglo, K.N. (2011). The case against biological realism about race: from Darwin to the post-genomic era. *Perspectives on Science*, **19** (4), 361–390. doi:10.1162/POSC_a_00048

Michel, C., Rossion, B., Han, J., Chung, C.-S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, **17** (7), 608–615. doi: 10.1111/j.1467-9280.2006.01752.x

Michel, C., Rossion, B., Bülthoff, I., Hayward, W.G., & Vuong, Q.C. (2013). The contribution of shape and surface information in the other-race face effect. *Visual Cognition*, **21** (9–10), 1202–1223. doi:10.1080/13506285.2013.823141

Mondloch, C.J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, **31** (5), 553–566. doi:10.1068/p3339

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

Pauker, K., Ambady, N., & Apfelbaum, E.P. (2010). Race salience and essentialist thinking in racial stereotype development. *Child Development*, **81** (6), 1799–1813. doi: 10.1111/j.1467-8624.2010.01511.x

Pellicano, E., Rhodes, G., & Peters, M. (2006). Are preschoolers sensitive to configural information in faces? *Developmental Science*, **9** (3), 270–277.

Raabe, T., & Beelmann, A. (2011). Development of ethnic, racial, and national prejudice in childhood and adolescence: a multinational meta-analysis of age differences. *Child Development*, **82** (6), 1715–1737. doi: 10.1111/j.1467-8624.2011.01668.x

Ramsey, P.G. (1991). The salience of race in young children growing up in an all-White community. *Journal of Educational Psychology*, **83** (1), 28–34. doi:10.1037/0022-0663.83.1.28

Ramsey, P.G., & Myers, L.C. (1990). Salience of race in young children's cognitive, affective, and behavioral responses to social environments. *Journal of Applied Developmental Psychology*, **11** (1), 49–67. doi: 10.1016/0193-3973(90)90031-E

Rand, D.G. (2012). The promise of Mechanical Turk: how online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, **299**, 172–179. doi: 10.1016/j.jtbi.2011.03.004

Rice, A.S., Ruiz, R.A., & Padilla, A.M. (1974). Person perception, self-identity, and ethnic group preference in Anglo, black, and Chicano preschool and third-grade children. *Journal of Cross-Cultural Psychology*, **5** (1), 100–108. doi: 10.1177/002202217400500107

Semaj, L.T. (1981). The development of racial-classification abilities. *Journal of Negro Education*, **50** (1), 41–47. doi: 10.2307/2294732

Shrum, W., Cheek, N.H., & Hunter, S.M. (1988). Friendship in school: gender and racial homophily. *Sociology of Education*, **61** (4), 227–239. doi: 10.2307/2112441

Snijders, T.M., & Bosker, R.J. (1994). Modeled variance in two-level models. *Sociological Methods & Research*, **22** (3), 342–363. doi: 10.1177/0049124194022003004

Sorce, J.F. (1979). The role of physiognomy in the development of racial awareness. *Journal of Genetic Psychology: Research and Theory on Human Development*, **134** (1), 33–41. doi:10.1080/00221325.1979.10533396

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

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

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

Strom, M.A., Zebrowitz, L.A., Zhang, S., Bronstad, P.M., & Lee, H.K. (2012). Skin and bones: the contribution of skin tone and facial structure to racial prototypicality ratings. *PLoS ONE*, **7** (7), e41193. doi:10.1371/journal.pone.0041193

Todorov, A., Dotsch, R., Wigboldus, D.H.J., & Said, C.P. (2011). Data-driven methods for modeling social perception. *Social and Personality Psychology Compass*, **5** (10), 775–791. doi: 10.1111/j.1751-9004.2011.00389.x

Todorov, A., Said, C.P., Engell, A.D., & Oosterhof, N.N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, **12** (12), 455–460. doi: 10.1016/j.tics.2008.10.001

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** (4), 593–610. doi:10.1002/ejsp.732

Xiao, W.S., Xiao, N.G., Quinn, P.C., Anzures, G., & Lee, K. (2013). Development of face scanning for own-and other-race faces in infancy. *International Journal of Behavioral Development*, **37** (2), 100–105. doi: 10.1177/0165025412467584

Yzerbyt, V.Y., Leyens, J.P., & Bellour, F. (1995). The ingroup overexclusion effect: identity concerns in decisions about group membership. *European Journal of Social Psychology*, **25** (1), 1–16. doi: 10.1002/ejsp.2420250102

Received: 17 September 2013
Accepted: 23 June 2014