Diverse Face Images (DFI): Validated for racial representation and eye gaze

Charisse B. Pickron[1] · Alexia J. Brown[2] · Caitlin M. Hudac[3] · Lisa S. Scott[2]

Abstract

Face processing is a central component of human communication and social engagement. The present investigation introduces a set of racially and ethnically inclusive faces created for researchers interested in perceptual and socio-cognitive processes linked to human faces. The Diverse Face Images (DFI) stimulus set includes high-quality still images of female faces that are racially and ethnically representative, include multiple images of direct and indirect gaze for each model and control for low-level perceptual variance between images. The DFI stimuli will support researchers interested in studying face processing throughout the lifespan as well as other questions that require a diversity of faces or gazes. This report includes a detailed description of stimuli development and norming data for each model. Adults completed a questionnaire rating each image in the DFI stimuli set on three major qualities relevant to face processing: (1) strength of race/ethnicity group associations, (2) strength of eye gaze orientation, and (3) strength of emotion expression. These validation data highlight the presence of rater variability within and between individual model images as well as within and between race and ethnicity groups.

Keywords Face processing · Racial diversity · Naturalistic stimuli set

Introduction

How do you know where someone is looking? What they are feeling? Or whether you recognize them? Faces are a central source of input for quick social evaluation of other people. Eye gaze captures elements of action, prediction, goal-direction, and facial expressions portraying underlying mental states (e.g., emotions, recognition). Within milliseconds of

✉ Charisse B. Pickron
cpickron@umn.edu

✉ Lisa S. Scott
lscott@ufl.edu

Alexia J. Brown
axb1596@med.miami.edu

Caitlin M. Hudac
chudac@mailbox.sc.edu

1 Institute of Child Development, University of Minnesota, Minneapolis, MN, USA

2 Department of Psychology, University of Florida, Gainesville, FL, USA

3 Department of Psychology and Center for Autism and Neurodevelopment Research Center, University of South Carolina, Columbia, SC, USA

seeing a face, infants to adults can make socially motivated and informed decisions about a person (Jessen & Grossmann, 2016, 2019; Qian et al., 2016; Todorov & Oh, 2021; Willis & Todorov, 2006; Yovel & Belin, 2013). The salience and centrality of social information associated with and learned from faces drive researchers to study face processing across the lifespan.

The significance of understanding the way faces are perceived, processed, and responded to has far-reaching implications found in almost all aspects of our social lives. Researchers interested in public policy to neurodivergent development consider individual-level and/or system-level phenomena relevant to processing faces. From the justice system (Chen et al., 2021; Eberhardt et al., 2004, 2006; Golby et al., 2001), to health outcomes (Greenwood et al., 2020; Hardeman et al., 2016), education (Kumar et al., 2015; Williams et al., 2019), and socialization (Halberstadt, 2020; McKone et al., 2021), researchers demonstrate the way faces are quickly categorized and interpreted, inform our explicit and implicit judgments.

Disturbances in face processing capabilities across the lifespan are believed to be foundational indicators of a variety of psychiatric and neurodevelopmental disorders (Costa et al., 2021; Griffin et al., 2021; Killgore et al., 2014; Monk

et al., 2006; Saarinen et al., 2021). From a developmental perspective, the first years of life include dramatic shifts in face processing capacities (Markant & Scott, 2018; Scherf & Scott, 2012; Scott & Arcaro, 2023; Scott & Fava, 2013; Scott et al., 2007). These and other shifts in face processing throughout the lifespan are believed to be driven by both top-down and bottom-up processes (Hadley et al., 2014) and shaped by one's environment or culture (Liu et al., 2015; Rennels & Davis, 2008; Sugden et al., 2014). It is even hypothesized that face processing biases are a cornerstone of implicit racial biases (Lee et al. 2017a, 2017b), and improving face recognition for unfamiliar race faces can reduce implicit associations in both children and adults (Lebrecht et al., 2009; Qian et al., 2019; Xiao et al., 2015).

Due to ongoing interest in the early-emerging and robust face processing system, there are numerous face stimuli databases using computer-generated faces (Matheson & McMullen, 2011; Roesch et al., 2011) and dynamic recordings (Krumhuber et al., 2017, 2021) as well as photographs of adults (Gross, 2005; Workman & Chatterjee, 2021) and children (Bijsterbosch et al., 2021; Dalrymple et al., 2013; Egger et al., 2011; LoBue & Thrasher, 2015; Prada et al., 2018) in either simple controlled or more natural environments. However, a much smaller subset of these validated and published databases includes Models of Color as well as White models (Chen et al., 2021; Conley et al., 2018; LoBue & Thrasher, 2015; Ma et al., 2015, 2021; Sacco et al., 2016; Strohminger et al., 2016; Tottenham et al., 2009; Ueda et al., 2019). These validated stimuli sets more accurately reflect racial diversity and globally growing multiracial populations (Bureau, 2021; Chen et al., 2021; Ma et al., 2021). Additionally, these racially diverse stimuli sets contribute to the needed systemic shift away from implicitly equating a White racial category and White faces as the norm and/or baseline for other comparisons. Having racially diverse validated and publicly available face databases enables a richer characterization of face perception and the influences of experience. Many of these data sets provide images of direct and profile views along with a range of positive and negative facial expressions to provide researchers with different ways to examine face-related perceptual expertise. However, most are missing variation in a key social communication attribute, eye gaze orientation. The aim of the Diverse Face Images (DFI) is focused not only on including racial and ethnic representation, but also on providing images of direct and averted eye gaze.

Gaze orientation is reliant on high-quality face images. Eye gaze and gaze following are critical communication cues that infants learn about in their first year (Akhtar & Gernsbacher, 2008; Itier & Batty, 2009; Renfrew et al., 2008). Infants learn that gaze orientation is not merely meaningless movement, but instead a purposeful cue indicative of shifting one's attention (Frischen et al., 2007;

Okumura et al., 2013a, 2013b; Senju et al., 2008; Striano & Reid, 2006). Gaze following is related to socio-cognitive skills such as joint attention, language, and even theory of mind, supporting our ability to learn about people, objects, and events in our visual world (Cleveland et al., 2007; Flom et al., 2017; Johnson et al., 2007; Reid & Striano, 2005; Reid et al., 2004; Striano et al., 2006). Disruption or atypical response to facial communication cues are linked to many psychiatric and neurodevelopmental disorders, such as major depressive disorder (Gaffrey et al., 2013) and anxiety disorders (Monk et al., 2006). In fact, interventions involving gaze detection and training gaze shifts are proposed as a critical step for autism spectrum disorder (Krstovska-Guerrero & Jones, 2016; Webb et al., 2014), a disorder for which varied or atypical perception and response to gaze is observed early in development (Bedford et al., 2012; Leekam et al., 1998; Stallworthy et al., 2022). Even within neurotypical populations, successful gaze following and processing of gaze-cued objects is influenced by several social factors (Dalmaso et al., 2020; Hadders-Algra, 2022). For example, more efficient gaze following and cued object processing has been found when viewing faces of both individual familiarity (Barry-Anwar et al., 2017; Hoehl et al., 2012) and group-level familiarity, such as familiar-race faces (Pavan et al., 2011; Pickron et al., 2017; Xiao et al., 2018). Other cues such as affective valence (Hoehl & Striano, 2010; Hoehl et al., 2008) and social status (Ciardo et al., 2013) are found to also influence gaze processing. The intersection of group membership, social evaluation, and gaze following further indicates the importance of having publicly available diverse stimuli sets of faces with both direct and averted eye gaze.

There are some published face databases available which include models with direct and averted gaze (Courset et al., 2018; Langner et al., 2010) as well as eye gaze toward peripheral objects in complex scenes (Bill et al., 2020; Recasens et al., 2015). The currently available large stimuli sets that highlight eye gaze orientation offer many strengths that fulfill the respective authors' aims, yet there are a few missed opportunities for generalized applications or uses of these stimuli sets that DFI aims to address. For example, some of the current stimuli sets have only one (all White; Langner et al., 2010) or two ethnicities included or no validation data evaluating quality of eye gaze orientation (Courset et al., 2018). The current stimuli set supplements these other sets by including women, multiple races and ethnicities, and data on perceived eye gaze orientation quality. Additionally, current stimuli sets that offer eye gaze following within naturalistic scenes include distant, indirect, partially obstructed, or even backward-facing images of human faces (Bill et al., 2020; Recasens et al., 2015). These qualities are strengths for studying perception of scenes but may have limitations for specifically investigating face

processing. The current stimuli set includes close-crop views of faces that can be used for eye gaze following as well as for many other research aims.

Many studies related to gaze following and object processing report manually manipulating photographs of models with direct eye gaze into averted gaze by moving the appearance of the iris using image editing software (Ciardo et al., 2013, 2021; Hoehl & Striano, 2010; Hoehl et al., 2008; Richeson et al., 2008; Weisbuch et al., 2017). This type of manipulation results in no longer using the exact faces that were validated in the published stimuli dataset. Despite the effectiveness researchers can achieve by manipulating a photographed face's direct gaze to averted gaze, utilizing existing databases of stimuli with already averted gaze will likely cut down on stimuli development time and reduce the risk of between-stimuli editing variation. Alternatively, some researchers utilize computer-generated faces (Pavan et al., 2011), which provides flexibility in face race and eye gaze orientation but may be reduced in ecological validity. It may be that some of the existing databases need to be more widely publicized or that researchers have requirements (e.g., racial or gender diversity) not completely fulfilled by what is presently available.

Specifically, the present paper contributes norming data of female models who self-identified as one of five different racial identities. This database of faces expands the existing databases through the actual faces that are included and the type of norming data collected. The faces include multiple exemplars of female models and images of direct and averted eye gaze orientation that have been equated for low-level visual differences. The rating data for the stimuli aim to go beyond accuracy of categorization by including in-depth ratings of each model across three major themes: (1) racial and ethnic group associations, (2) eye gaze orientation, and (3) emotion expression. Researchers will have access not only to averaged data across all models, but critically, to summary data about each individual model included in our database. By including individual model data, we offer opportunities for researchers to consider variability in evaluations within and between the different groups of faces.

Methods

Procedures for recording the face models were approved by the University of Massachusetts Amherst's Institutional Review Board in 2015. All face models provided written informed consent for their images to be used and distributed for study participation, future research, and publication purposes. Validation data collection methods were approved by the Institutional Review Board of the University of Florida (2022). Participants recruited through the University of Florida provided written informed consent to participate in data collection procedures and have de-identified data published. All research conducted through each institution was performed in accordance with the Declaration of Helsinki.

Stimulus development

Model recruitment and demographics

Individuals were recruited from the University of Massachusetts Amherst and surrounding communities to participate in our stimuli dataset. Information on age, gender, and racial and ethnic identity was included in our recruitment materials. Specifically, we advertised for adults between 18 and 34 years of age who self-identified as male or female and as African-American/Black, White, Chinese/Chinese-American/Taiwanese American, Chinese/Vietnamese, South Asian/Indian/Indian-American, or Hispanic/Latiné. Participants were paid $5.00 in cash for coming in for a single 30-min recording session. A total of 41 female models were video-recorded (7 East Asian, 7 Southeast Asian, 10 Black/African-American, 10 White, and 7 Hispanic/Latiné). Six of the seven women who ethnically identified as Hispanic/Latiné also self-identified with multiple racial groups, which is commonly reported among this population (Araujo Dawson & Quiros, 2014; Cruz-Janzen, 2002; Umaña-Taylor et al., 2014). We included these models in the Hispanic/Latiné group as their primary ethnic identity. The goal of the project is to provide both the model's self-identity data and raters' perceptions of racial and ethnic identity to give researchers who are using the dataset as much information as possible and to leave it up to the individual researcher to choose which faces to use and how to incorporate race/ethnicity demographic diversity. Only eight men volunteered to participate in the stimuli recording. This sample size did not provide enough representation across race and ethnicity groups, and therefore we decided not to include male faces for this iteration of our face stimuli set.

After recording all the models, our research group created still frame images of each female model expressing direct eye gaze and averted eye gaze to the left and the right. Our team then completed a preliminary visual inspection of the quality of each model's still frame image. We looked for clear facial expressions, direction of eye gaze, and centrality of head position. The final sample of models included 41 self-identified women ($M_{age} = 25.2$, $SD_{age} = 3.43$). Additional demographic metadata (e.g., further racial and/or ethnic identity details if provided by the individual) are included in Fig. 1 (see column headings "Self-identified ethnicity group" and "Self-identified ethnicity subgroup").

Fig. 1 DFI face stimuli. Note. * In the questionnaire, South Asian and Southeast Asian ethnicities were collapsed. Figure includes every model, their self-identified race and ethnicity, and the multiple examples of each model based on facial and eye gaze expression

Model recording procedure

Each model wore the same black T-shirt and little to no makeup or jewelry, and removed eyeglasses, hats, and any other items that may have obscured the view of their faces. A Canon XF100 camera was used and positioned directly in front of the models. Fluorescent box lights and softboxes were used for the lighting setup. Light fixtures were positioned to the left and right of the model. Each model was recorded from their shoulders up, and all sat in the same chair during recording sessions. The height of the camera was adjusted based on the height of each participant.

Models were instructed to maintain a pleasant, relaxed, but neutral expression, keeping their mouth closed during recording. For some models (see Fig. 1), representative images of the category were not available from the video. For example, several models did not have a closed-mouth happy image that could be extracted. To control between-model variation in eye movement, we had models follow a PowerPoint presentation of a ball projected on a wall across

from them. The ball moved from the center of the wall to either the left or right. Models were instructed to track the ball with their eyes while maintaining a still head, neck, and torso (i.e., keeping the rest of their body facing forward toward the camera). This ensured that eye movement timing and duration were consistent.

Post-recording image processing

Within Photoshop, faces were centered within an oval shape measuring 875×1387 pixels, with a white background masking the other parts of the image. All faces were centralized such that eyes were level within a boundary (875×246 pixels) that was 123 pixels above the center. A vertical boundary (246×1387 pixels) served as a guide to ensure that eyes were equally spaced horizontally. Faces were resized to ensure that both the chin and part of the hairline were visible. Any additional distinguishing features (e.g., nose ring) were blurred or removed to reduce implicit attention. Faces were converted to grayscale, because performance on

Self-identified ethnicity group	Model	Self-identified ethnicity subgroup	Gaze Direction			
			Looking right	Looking left	Happy	Calm
Black	BF_2	Black; Nigerian				
	BF_3	Black; Nigerian				
	BF_4	Black; American, Barbadian				
	BF_5	Black; Nigerian			-	
	BF_6	Black, Hispanic; Puerto Rican/Honduran				-
	BF_7	Black; Cameroonian			-	
	BF_8	Black				
	BF_9	Black				
	BF_10	Black				

Fig. 1 (continued)

face recognition tasks is stronger for grayscale than color (R. Russell et al., 2007), and low-level perceptual differences (e.g., luminance and contrast) were reduced by averaging dark and light pixel contrasts and equating the standard deviations of the luminance distribution across all faces using SHINE (Spectrum, Histogram, and Intensity Normalization and Equalization; Willenbockel et al., 2010) in MATLAB. Within the SHINE toolbox source folder, the faces were placed in the SHINE Input folder, and no template images were used in the SHINE Template folder. In MATLAB, the path was set for the SHINE toolbox folder, and lines 69–72 of the main m-file SHINE were edited to reflect the correct image file type (e.g., TIFF), and the locations of the input, output, and template folders. The main m-file, SHINE, was run to obtain mean luminance matching on the whole images, using custom options. The matching mode used was luminance, the luminance option used was lumMatch, and the matching region used was the whole image. The final

Self-identified ethnicity group	Model	Self-identified ethnicity subgroup	Gaze Direction			
			Looking right	Looking left	Happy	Calm
Hispanic	HF_1	Hispanic				
	HF_2	Hispanic; Puerto Rican				
	HF_3	Hispanic; Puerto Rican				
	HF_4	Hispanic; Puerto Rican			-	
	HF_5	Hispanic; White				
	HF_6	Hispanic; White				
	HF_7	Hispanic, Black, White; 1/2 Puerto Rican, 1/2 Colombian				
	HF_8	Hispanic				

Fig. 1 (continued)

shined faces were automatically generated into the SHINE output folder. Individual faces were then edited for unnatural patches or distortions using Adobe Photoshop. See Fig. 1 for examples of finalized model stimuli.

Rater data collection

Participants

A total of 327 adults consented to participate in the face model rating study; 38 exited the survey after consent or after completing demographic information and were subsequently excluded. Of the final 288 participants, the majority completed the entire survey ($n = 241$, 83.7%); the remaining participants completed a range of 10–55%, but their data were still included in analyses given a planned analytic solution to account for missing data. Most participants were aged 18–25 years (279, 96.9%) and an additional six participants indicated being 25–45 years of age. Participants self-reported their race and ethnicity identity (see Table 1). The majority of participants identified as women ($n = 219$, 76%), with an additional 63 participants identifying as men (21.9%), transgender ($n = 1$), nonbinary ($n = 2$), or preferring not to answer ($n = 3$). See Supplemental Table S1 for additional demographic information related to highest education level, annual household

Self-identified ethnicity group	Model	Self-identified ethnicity subgroup	Gaze Direction			
			Looking right	Looking left	Happy	Calm
South Asian*	SAF_1	South Asian				
	SAF_2	South Asian				
	SAF_3	South Asian				
	SAF_4	South Asian				
	SAF_5	South Asian; Indian				
	SAF_6	South Asian				
	SAF_7	South Asian				

Fig. 1 (continued)

income, and additional self-reported details regarding gender identity.

Data collection procedure

Sona Systems, the participant pool management systems for universities, was used to recruit adults taking undergraduate psychology courses at the University of Florida (Copyright 1997–2023, Sona Systems Ltd.). An online study page was created in the University of Florida's Department of Psychology Research Participation Credit Manager, Sona, as a platform for adults to access our stimulus set. Adults opted to participate by registering for one time slot per person, after which they received the link to our survey. We tested as many students as possible willing to sign up for the online study during a time window of a single academic semester. Face model ratings were collected using an online survey through Qualtrics software (release February 2023; Copyright 2023, Provo, UT, USA). Each participant rated all the available images for a total of 149 trials and images. A randomized list of the stimuli was generated in Excel and then used for the presentation order in Qualtrics. Each face image was presented by itself in the center of the screen. Below the image, raters were given instructions about the type of rating they were being asked to complete.

The primary goal of our study is to offer validation data evaluating three major categories of our included models. Categories included (1) racial and ethnic group associations, (2) eye gaze orientation, and (3) emotional facial expressions. Participants were asked multiple sub-questions within each of these three major categories, and the model's image was visible to participants for all of the questions. These three categories were selected for evaluation as there is

Self-identified ethnicity group	Model	Self-identified ethnicity subgroup	Gaze Direction			
			Looking right	Looking left	Happy	Calm
White	WF_1	White				
	WF_2	White				
	WF_3	White				
	WF_4	White				
	WF_5	White				
	WF_7	White; Polish				
	WF_8	White				
	WF_9	White				
	W_10	White				

Fig. 1 (continued)

extensive evidence that each quality both uniquely and interconnectedly influences face processing across the lifespan (Adams et al., 2010; Farroni et al., 2004; Gregory et al., 2020; Hoehl & Striano, 2008; Quinn et al., 2018; Richeson et al., 2008; Trawalter et al., 2008).

Race and ethnicity categories For the race and ethnicity group association, participants responded to the prompt: "How strong of an association does this face have with the following five racial or ethnic groups?" The five racial and ethnic groups included Asian (Chinese, Japanese, Korean, etc.), Black (African-American, Ethiopian, Haitian, Nigerian, etc.), Hispanic and/or Latin American (Mexican, Puerto Rican, Cuban, Dominican, etc.), South and Southeast Asian (Indian, Indonesian, Thai, Malaysian, etc.), and White (Dutch, English, Irish, European, Norwegian, etc.). Participants used a five-point Likert scale, coded as follows: 0 = no association, 1 = very little association, 2 = not sure,

Table 1 Race and ethnicity self-identity from 288 participants who rated the stimuli

Participant race and ethnicity self-identity	No	Percent	Further identifications
American Indian or Alaska Native	8	2.8%	All indicated American Indian
Asian	57	19.8%	Asian Indian ($n=17$), Cambodian ($n=2$), Chinese ($n=18$), Filipino ($n=8$), Japanese ($n=3$), Korean ($n=2$), Pakistani ($n=6$), Vietnamese ($n=7$), None of these fully describe me ($n=5$): Indonesian and Chinese, Kazakh, Mongolian, Pakistani-Indo-Guyanese, Thai
Black, African American, or African	27	9.4%	African American ($n=16$), Caribbean ($n=7$), Haitian ($n=0$), Jamaican ($n=3$), Nigerian ($n=1$)
Hispanic, Latino, or Spanish	72	25.0%	Colombian ($n=6$), Cuban ($n=18$), Dominican ($n=4$), Ecuadorian ($n=3$), Honduran ($n=2$), Mexican or Mexican American ($n=11$), Puerto Rican ($n=14$), Salvadoran ($n=1$), Spanish ($n=7$), None of these fully describe me ($n=26$): Bolivian ($n=2$), Brazilian ($n=3$), Chilean, Ecuadoran, Indigenous mixed with Spanish bloodlines and Chinese ancestry, Nicaraguan ($n=6$), Guatemalan ($n=3$), Peru ($n=4$), Venezuelan ($n=8$)
Middle Eastern or North African	5	1.7%	Egyptian ($n=3$), Iraqi ($n=1$), Lebanese ($n=1$), Moroccan ($n=2$)
White ⌘	176	61.1%	English ($n=53$), European ($n=61$), French ($n=19$), German ($n=51$), Irish ($n=61$), Italian ($n=41$), Norwegian ($n=0$), Polish ($n=14$), Scottish ($n=13$), Spanish ($n=17$), None of these fully describe me ($n=17$): American ($n=3$), Israeli ($n=2$), Eastern Europe, Portuguese, Russia ($n=2$), Swedish, Mostly Swiss, South African
None of these categories describe me	2	0.7%	

⌘ Additional comments included: *White is just the color of my skin, i.e. race ($n=1$), I do not know the heritage of my white side. As my dad is white but is from America ($n=1$), I've always reported my race to be white, but I am ethnically Hispanic ($n=1$), American? My grandparents are Italian, but I don't feel like that describes me ($n=1$), I don't know my family history/specific country of origin very well. Just your standard white American-born Floridian ($n=1$)*

3 = strong association, and 4 = extremely strong association. For every face, participants used the same Likert scale to answer the association prompt for each of the five listed racial and ethnic groups. There were no explicit instructions about how many races a model could be associated with. By asking for an association rating for each race/ethnicity category for every face, we allowed raters to give a range of association values. For example, a model could be rated as having an extremely strong association for the racial category of Black, and no association for racial categories of White, Asian, or South or Southeast Asian, but could also be rated as having a strong association with the ethnicity category of Hispanic/ Latiné.

Using this rating technique, we obtained more nuanced information about the way each model's face was perceived racially and ethnically, instead of a simple racial categorization response. We acknowledge that race and corresponding racialized categories are social constructs that can be used to perpetuate false associations of biological underpinnings and meaningful distinctions between people (Salter et al., 2018; Smedley & Smedley, 2005). Despite racial categories being products of racial ideation, particularly in the United States, people have been historically socialized to use racialized categories as marking real differences (Hochman, 2021; Roberts & Rizzo, 2021). Perceptual categorization is driven by many factors including experience, socio-cognitive bias, cultural and individual definitions of race and corresponding

racial groups, emotional expression, and eye gaze orientation. In our study, we are not focused on the "accuracy" of categorizing faces. That is, we aim to provide information about individual variability in perceiving racial category associations, not whether any given person's categorization matches the stated racial identity of the model. An individual's racial identity can be fluid, is constructed from experience, and often does not match with how the rest of the human population perceives and categorizes them (Albuja et al., 2018; Davenport, 2020; Douglass et al., 2016). We focus on perception instead of racial identity accuracy because perceptual processes, especially experience-driven biased perceptions, underlie fast-acting neural and behavioral responses following the presentation of a face. These responses are important and are often what is being measured by researchers using human faces in their studies. Thus, collected data offer a descriptive range of the way the faces in this set are evaluated and the strength of associations individuals have between different socially constructed racial categories and the presented faces. Association data is used as a way to offer a description for how people (U.S.-based population) view the faces and what this may mean for the types of decisions researchers need to make when choosing which face to include in their own research.

Eye gaze orientation Participants responded to the prompt: "Where do you think this person is looking?" Responses were as follows: 0 = difficult to determine,

1 = directly at the participant, 2 = away from the participant. Models were prompted to display direct and averted eye gaze (to the left and to the right). These naturalistic eye movements come with individual variability in how obvious it is where someone is looking.

Emotional facial expressions Including ratings regarding facial expressions is critical when creating a database of face stimuli. We included this evaluation for three reasons. First, the perception of a facial expression is contextually influenced; that is, there could be spillover effects in rating one face to another (Albohn & Adams, 2021; Russell & Fehr, 1987). Second, despite the instructions to express a calm or neutral expression, there is individual variability in the way this type of expression is executed. Third, calm or neutral expressions can be perceived as having emotional messages, particularly a more negative valence (Albohn et al., 2019; Lee et al., 2008). The interpretation of a neutral facial expression can be racialized based on the race of the viewer and the presented face (Hu et al., 2017). Participant raters responded to the prompt: "Based on the image above: How ___ does this person look in this image?" The emotional expressions included happy, calm, angry, and neutral (no expression). Participants used a five-point Likert scale for each of the five emotional expressions listed. The four emotion category questions were listed in table format, such that there were four rows with one row per emotion category and five columns, one for each point on the Likert scale. Ratings ranged as follows: 0 = not at all, 1 = somewhat, 2 = average, 3 = very, and 4 = extremely. For example, a model could be rated as "very" calm, "very" neutral, and "not at all" for happy and angry. Similar to the racial and ethnic group associations, we aimed to collect data to highlight individual variability instead of simple yes/no or categorization data.

Data analysis

Descriptive statistics are included for future users of the stimuli set to understand the means and standard deviations of ratings. In addition, analyses were conducted to better understand whether these ratings were influenced by the stimuli (i.e., differences due to stimuli race, orientation, or emotional expression) or rater race. These analyses were conducted in R (version 4.3.1) using the linear mixed-effects package lme4 (Bates et al., 2015) to show the association strength for each judgment (e.g., race/ethnicity, eye gaze orientation, and emotion). All analytic models include a random intercept for each photo model to account for the multiple ratings across different images. This approach also helps account for missing item-level data (e.g., empty survey answers) by utilizing available item-level data without reducing deductive power (Mazza et al., 2015). Estimated

marginal means (EMM) were used to describe fixed effects (specified for each model in each respective results section), and Bonferroni correction was applied to pairwise comparisons.

Results

Analysis of the presented data is focused on descriptive statistics for three areas of evaluation: (1) race and ethnicity groups association, (2) eye gaze orientation, and (3) emotional facial expressions.

Validity of race groups

Validity ratings for each of the five race and ethnicity groups are presented in Fig. 2 (see Supplemental Tables S2 and Supplemental Figs. S2–S5 for mean association ratings of the race and ethnicity group association for each of the model's multiple face exemplars). Linear mixed-effects models predicting the mean rating of association strength included a fixed effect of what race was selected by the raters, a fixed effect of rater's race (same as or different from actor's race), and the interaction between the two fixed effects.

Full statistical model details are available in the Supplemental Material for Model 1. Considering target responses (i.e., response matching photo model's self-identified race; e.g., association strength of Asian faces for self-identified Asian photo model), raters' association strength did vary based upon the photo model's self-identified race, $F(4, 34) = 23.51$, $p < 0.0001$, such that, overall, estimated marginal mean (EMM) ratings were lower for Hispanic faces ($EMM = 2.16$, $SE = 0.12$) than all other faces, $p < 0.001$. Ratings for target responses to Southeast Asian faces ($EMM = 2.78$, $SE = 0.13$) were also lower than Black faces ($EMM = 3.63$, $SE = 0.11$), $p = 0.002$. As indicated in Fig. 4, the variability in ratings was higher for Hispanic faces ($EMM = 0.39$, $SE = 0.069$) than for Asian faces ($EMM = 0.068$, $SE = 0.079$; $p = 0.0004$), Black faces ($EMM = 0.11$, $SE = 0.066$; $p = 0.0005$), and Southeast Asian faces ($EMM = 0.31$, $SE = 0.073$, $p = 0.042$), as confirmed using post hoc Levene's test based upon absolute deviation from the mean, statistic = 7.23, $p < 0.001$. See Fig. 3 for mean ratings of associations for each face race and ethnicity group by rater's self-identified race.

Overall, association strength was weaker when the photo model's race was different from the rater's self-identified race, $F(4,1, 34) = 10.69$, $p = 0.003$; however, a significant interaction, $F(4, 34) = 14.34$, $p < 0.0001$, indicated that this was only true for Black faces, $p < 0.0001$, and not for the other face races, $p > 0.34$. In other words, non-Black raters had weaker association strength to Black faces.
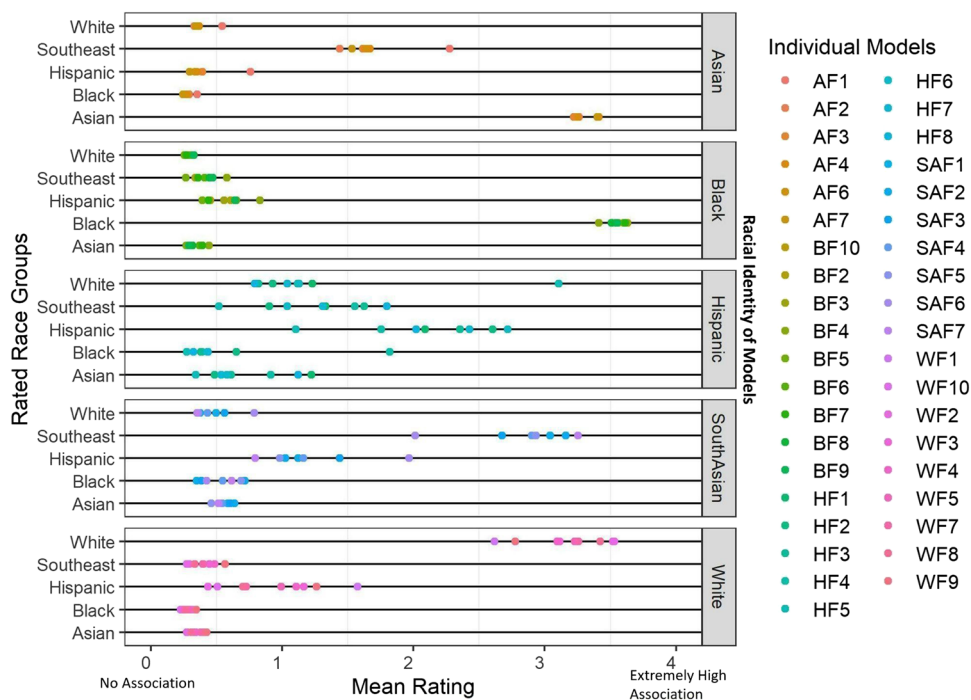
Fig. 2 Averaged race and ethnicity group association by model number. Note. Figure depicts the mean rating of race and ethnicity group associations for each model. These data are collapsed across the specific model examples (e.g., averted left eye gaze, averted right eye gaze, calm, and happy) and grouped by the self-identified racial group of the models

Validity of eye gaze orientation

Generally, participants were accurate in reporting gaze orientation (range = 74.4–99.5%). Full statistical model details are available in the Supplemental Material for Model 2. Linear mixed-effects models predicting percent gaze orientation included a fixed effect of gaze condition [averted (collapsed across left and right) and direct (collapsed across happy and calm)], a fixed effect of photo model race, a fixed effect of rater's race (same as or different from actor's race), and full-factorial interactions between all fixed effects. Raters were more accurate for direct gaze ($EMM = 94.9\%$, $SE = 0.4\%$) than averted gaze ($EMM = 90.2\%$, $SE = 0.4\%$), $F(1, 102) = 83.84$, $p < 0.0001$. An interaction between photo model race, rater race, and gaze direction, $F(4, 102) = 4.64$, p 0.0018, indicated that Black raters were less accurate than non-Black raters for gaze orientation of Black faces, but only for direct-facing stimuli, $p = 0.045$. No other pairwise comparisons were significant, $p > 0.064$. See Fig. 4 for the accuracy ratings identifying direct and averted eye gaze orientation for each actor race.

Validity of facial expressions

Full statistical model details are available in the Supplemental Material for Model 3. Linear mixed-effects models predicting the mean rating included a fixed effect of emotion condition for each stimulus, a fixed effect of photo model race, a fixed effect of rater's race (same as or different from photo model's race), and full-factorial interactions between all fixed effects (see Supplemental Table S3 for the average rating of emotional facial expression for all face models and each iteration of their image).

Overall, ratings were lowest for angry ($EMM = 0.46$, $SE = 0.027$) relative to other emotions, $p < 0.001$, including calm ($EMM = 1.77$, $SE = 0.027$), happy ($EMM = 1.48$, $SE = 0.027$), and neutral ($EMM = 1.70$, $SE = 0.027$). Happy ratings were also lower relative to neutral, $p < 0.0001$. Rating of facial expressions was not related to rater race, $p > 0.17$. An interaction between photo model race and emotional expression, $F(12, 238) = 4.88$, $p < 0.0001$, indicated that, first, Asian and Hispanic faces had stronger happy associations than Black faces, $p < 0.022$, and second, Black faces had stronger neutral associations than White faces, $p = 0.036$. See Fig. 5 for depiction of facial expression findings.

Discussion

The present project introduces a new face stimuli database along with validation data rating three key themes of face processing: racial and ethnicity group association, eye gaze orientation, and emotional facial expressions. The DFI

How strong of an association does this face have with __?
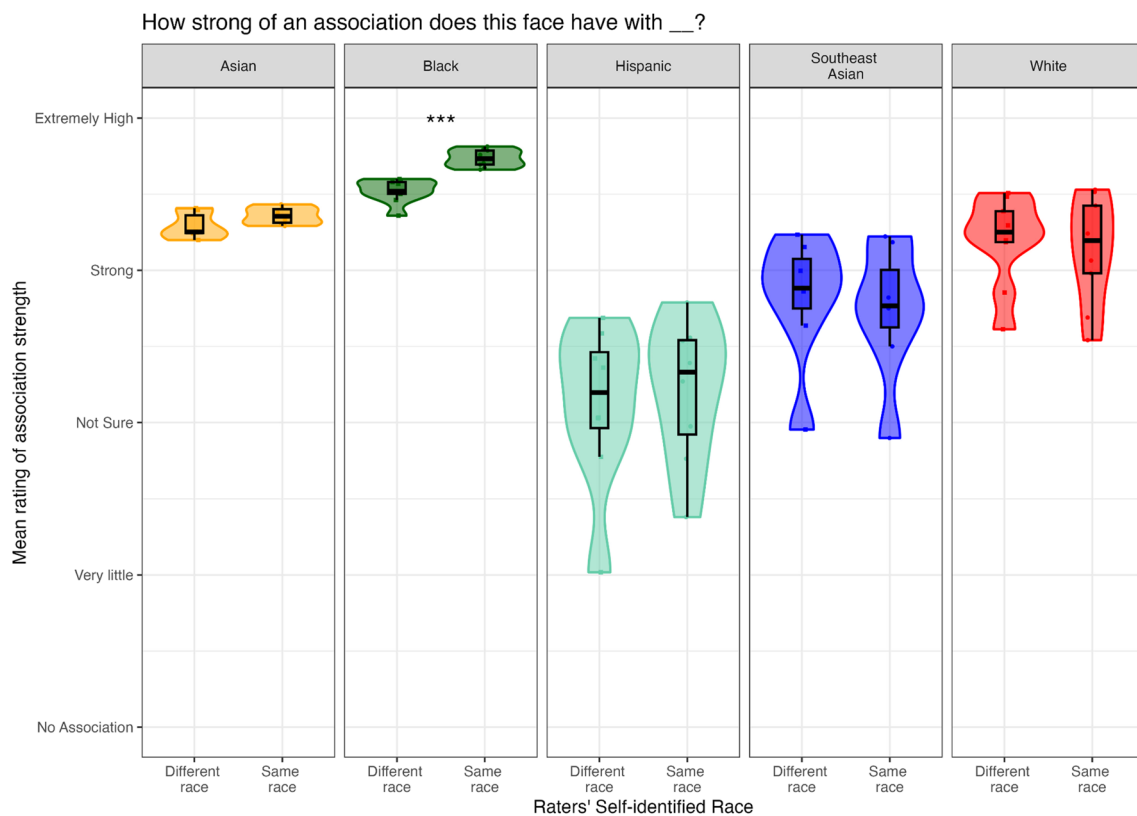
Fig. 3 Mean rating of the strength of association (i.e., strength of stimulus association to each of the five race and ethnicity groups) was scored on a Likert scale from 1 (no association) to 5 (extremely high).

Analysis was conducted on numeric means but is represented by category here for illustrative purposes. *** $p < 0.0001$. Note. Violin plot highlights relative density, and boxplots illustrate quartile values

stimuli set will be an open-access tool that includes images of racially diverse female-identifying adults with direct and averted eye gaze orientation, as captured from dynamic video recordings to ensure ecological validity.

The current validation rating data extend beyond the question of accurately categorizing faces on a single criterion. A unique feature of our stimuli rating procedure was that it allowed participants to individually rate faces on each element of interest. Our aim is to provide researchers with data that give a richer characterization of how the faces in our database are perceived. From the reported descriptive statistics, researchers will have access to generate additional analyses to fit their specific needs when making stimuli selection choices. We were particularly interested in highlighting the variability in adults' perceptions of racial and ethnic group associations. Specifically, we provide descriptive data that highlight the variability in model rating within and between racial and ethnic groups. We offer a general conclusion from each of the three face category evaluations of interests. First, adult participants had more variability in race and ethnicity group associations for models who self-identified as Hispanic/Latiné, relative to Asian and Black models. Images with averted

eye gaze were clearly perceived as such by our participant raters. Overall, faces were rated as relatively calm or neutral. This rating of the model's emotional facial expression is consistent with the original instructions to the models during the video recording session, which was to express a calm, but pleasant expression.

Broadly, our results suggest consistent findings across the rater's self-identified race, despite the fact that the ratings of these models come from a sample of adults from a majority White sample (60.5%). There were two unexpected rater race effects. First, non-Black raters had slightly weaker race-rating association strength to Black faces (3.52) than Black raters (3.74). This weaker association made by non-Black raters was somewhat unexpected given the robust "other-race categorization advantage" that is reflected by a stronger or faster race categorization for unfamiliar-race faces (Caharel et al., 2011; de Lissa et al., 2021; Feng et al., 2011; Sekimoto, 2018; Zhao & Bentin, 2008). It is possible that non-Black raters' reduced categorization association was a result of seeing both direct and averted eye gaze of Black models, as eye gaze orientation is reported to impact different elements of face processing (Adams et al., 2010; Sessa & Dalmaso, 2016). However, it is unclear why this
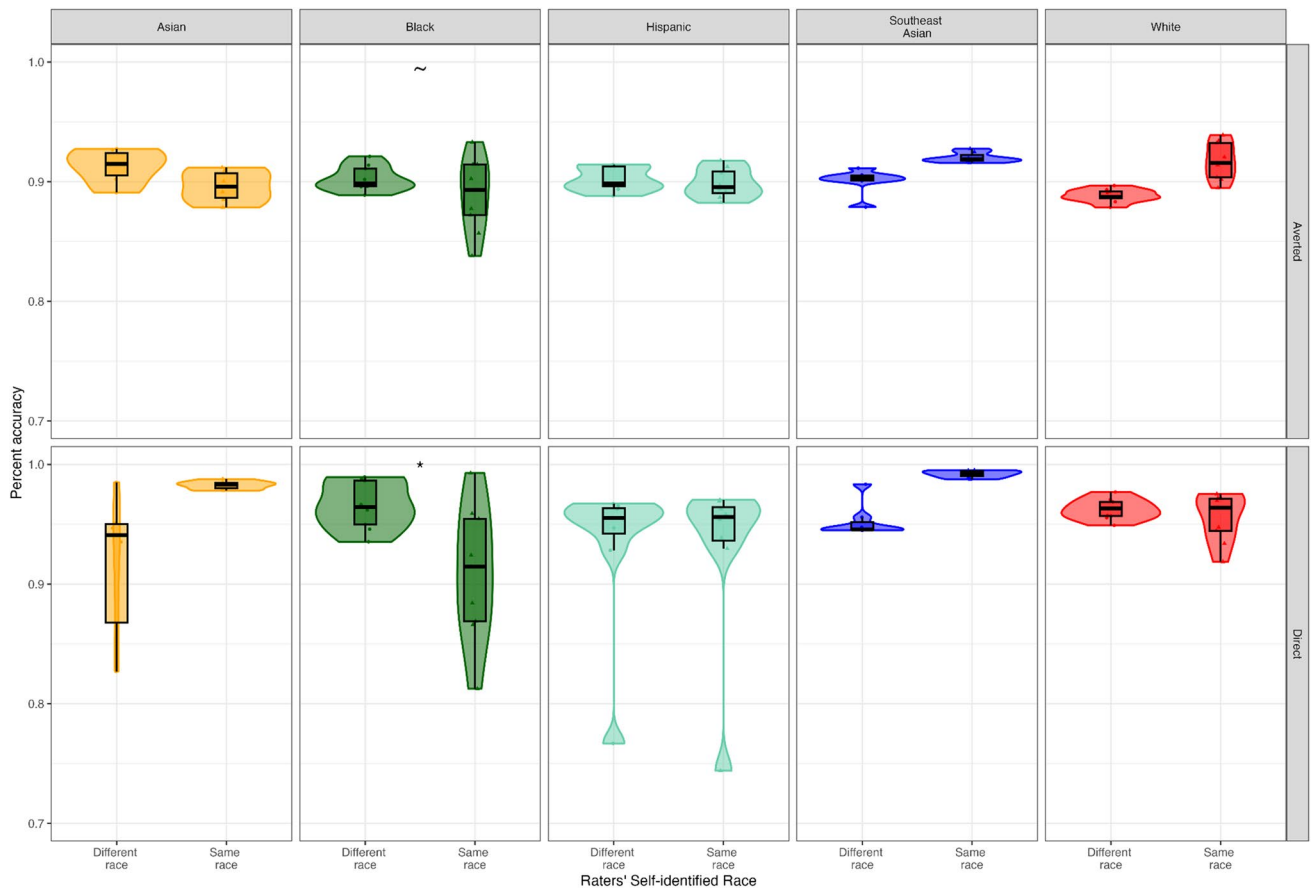
Fig. 4 Accuracy of eye gaze orientation. Note. Percent accuracy of eye gaze orientation categorization (e.g., accurately indicated that averted stimulus is looking away). Violin plot highlights relative density, and boxplots illustrate quartile values. * p < 0.05

finding was only specific for Black models and non-Black raters and not any model that was of a different race than the rater. Second, Black raters were slightly less accurate in rating gaze orientation (89.9%) relative to non-Black raters (93.4%). This finding may be related to prior work suggesting differences in visual scanpath strategies for Black observers (i.e., lower half of the face) compared to White observers (upper half of the face) (Hills & Pake, 2013). However, broadly, all raters were accurate (> 70%), and this subtle difference requires more work before considering or concluding race group differences in categorizing direct and averted eye gaze.

The faces included in DFI will be particularly useful for researchers and practitioners examining processes related to areas such as intergroup bias, face perception, attention orientation, and communication cues. With increasing racial and ethnic participant diversity represented in developmental, socio-cognitive, and neuroscience-based studies, this database will support researchers' efforts in maximizing inclusivity. One example of using this new database of faces is to support testing of the Interactive

Model of Attention and Perception (I-MAP). I-MAP predicts that with development comes increasing control over attention which directs perceptual learning and supports top-down selective attention biases for familiar faces (Markant & Scott, 2018). The I-MAP model hypothesizes that the interaction between perceptual learning and attention results in increased anterior-to-posterior neural processing and increasingly right-lateralized occipitotemporal face processing during the first year of life. To this end, the DFI face database will support such research with its racial and ethnic representation, multiple images of each model, and controls for low-level visual cues.

Limitations and future directions

Despite the strengths of the DFI stimuli set, there are two key limitations to acknowledge. The first is the demographics of the participants who rated the models. The participants who completed the stimuli rating questionnaire were primarily White and female. However, the participant sample is representative of the population where
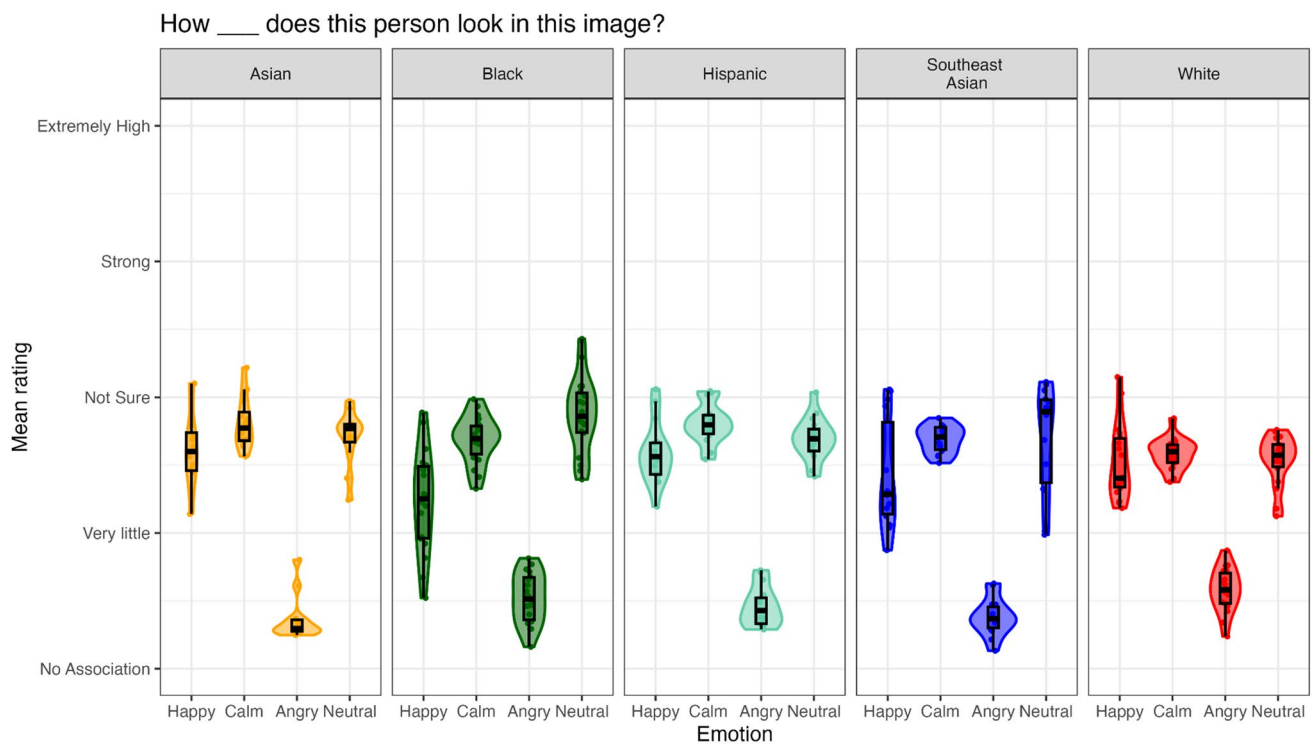
Fig. 5 Evaluation of facial expression. Note. Violin plot highlights relative density and boxplots illustrate quartile values. Mean rating (i.e., strength of stimulus association to each of the four emotions) was scored on a Likert scale from 1 (not at all) to 5 (extremely). Analysis was conducted on numeric means but is represented by category here for illustrative purposes

the data were collected in north-central Florida. Despite obtaining detailed race and ethnicity identity from participating raters, we failed to obtain details about the racial diversity of raters' daily lives. This type of information will be beneficial for future stimuli development projects to gain a richer characterization of the experiential context that may be influencing raters' evaluations. In the future we would also like to increase the racial and ethnic diversity within our participant sample as well as cross-cultural evaluations. Of particular interest is increasing the sample size of raters whose racial and ethnic identity match those of the models included in the DFI stimuli set. This may be especially important for Black-identified raters to obtain a fuller representation of individual- and group-level differences for reviewing eye gaze orientation. Researchers conducting cross-cultural work may find it especially useful to complete additional validation checks with intended participant populations (i.e., those outside of northern Florida, USA). The second limitation is that we were unable to recruit enough male models, and not all racial and ethnic groups are represented within our stimuli set, including multiracial-identifying models. In the future, the authors plan to make the original videos of each model publicly available as well as adding male-identifying models to the DFI database.

Conclusion

Development of the DFI was motivated by the need to have high-quality, racially representative images of faces with averted eye gaze to investigate the way early experiences shape face perception. Responses to eye gaze are a key developmental skill that may have transdiagnostic implications for neurodiverse development and social communication capabilities. This publicly available stimuli database and rater validation descriptive statistics compliment and extend face stimuli resources that researchers and practitioners have access to. The included images give researchers a path forward in efforts toward decentering Whiteness as a standard in studying processes related to human faces.

visualization, and drafting the manuscript. A.B. participated in stimuli image editing, questionnaire development, data collection, data visualization, and manuscript editing. C.H. participated in questionnaire development, stimuli image editing, and data visualization, and took the lead on all analyses and drafting the manuscript. L.S.S. participated in the development of stimuli video recording, questionnaire development, data visualization feedback, and drafting the manuscript. All authors critically revised and approved the final manuscript.

Data availability Data, analysis syntax, database images, and supplemental materials will be available via OSF (https://osf.io/7u2vc/?view_only=90ade4e855b144e2849c5e7299a99208). On the OSF page there are three subfolders: (1) "Data_Codebook_Syntax" contains model demographics data, rater data, analysis syntax, and a codebook, (2) "Face_Images" contains all of the images of the models included in the database, and (3) "Publication" contains the supplemental materials associated with this manuscript.

Code availability Code used in software R will be made available upon publication of this manuscript.

Declarations

Ethics approval Procedures for recording the face models were approved by the University of Massachusetts Amherst's Institutional Review Board in 2015. Validation data collection methods were approved by the Institutional Review Board of the University of Florida (2022). All research conducted through each institution was performed in accordance with the Declaration of Helsinki.

Consent to participate/Consent for publication All face models provided written informed consent for their images to be used and distributed for study participation, future research, and publication purposes. Participants recruited through the University of Florida provided written informed consent to participate in data collection procedures and have de-identified data published.

Conflicts of interests All authors declare no competing interests for this manuscript.

Open Practices Statement Data, code, rater data, and stimuli images will become available upon publication of this manuscript. This study was not preregistered.

References

Adams, R. B., Pauker, K., & Weisbuch, M. (2010). Looking the other way: The role of gaze direction in the cross-race memory effect. *Journal of Experimental Social Psychology, 46*(2), 478–481. https://doi.org/10.1016/j.jesp.2009.12.016

Akhtar, N., & Gernsbacher, M. A. (2008). On Privileging the Role of Gaze in Infant Social Cognition. *Child Development Perspectives, 2*(2), 59–65. https://doi.org/10.1111/j.1750-8606.2008.00044.x

Albohn, D. N., & Adams, R. B. (2021). Emotion Residue in Neutral Faces: Implications for Impression Formation. *Social Psychological and Personality Science, 12*(4), 479–486. https://doi.org/10.1177/1948550620923229

Albohn, D. N., Brandenburg, J. C., & Adams, R. B. (2019). Perceiving Emotion in the "Neutral" Face: A Powerful Mechanism of Person Perception. In U. Hess & S. Hareli (Eds.), *The Social Nature of Emotion Expression: What Emotions Can Tell Us About the World* (pp. 25–47). Springer International Publishing. https://doi.org/10.1007/978-3-030-32968-6_3

Albuja, A. F., Sanchez, D. T., & Gaither, S. E. (2018). Fluid racial presentation: Perceptions of contextual "passing" among biracial people. *Journal of Experimental Social Psychology, 77*, 132–142. https://doi.org/10.1016/j.jesp.2018.04.010

Araujo Dawson, B., & Quiros, L. (2014). The effects of racial socialization on the racial and ethnic identity development of Latinas. *Journal of Latina/o Psychology, 2*(4), 200–213. https://doi.org/10.1037/lat0000024

Barry-Anwar, R. A., Burris, J. L., Estes, K. G., & Rivera, S. M. (2017). Caregivers and strangers: The influence of familiarity on gaze following and learning. *Infant Behavior and Development, 46*, 46–58.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). *Fitting Linear Mixed-Effects Models using lme4* (arXiv:1406.5823). arXiv. https://doi.org/10.48550/arXiv.1406.5823

Bedford, R., Elsabbagh, M., Gliga, T., Pickles, A., Senju, A., Charman, T., Johnson, M. H., & the BASIS team. (2012). Precursors to Social and Communication Difficulties in Infants At-Risk for Autism: Gaze Following and Attentional Engagement. *Journal of Autism and Developmental Disorders, 42*(10), 2208–2218. https://doi.org/10.1007/s10803-012-1450-y

Bijsterbosch, G., Mobach, L., Verpaalen, I. A. M., Bijlstra, G., Hudson, J. L., Rinck, M., & Klein, A. M. (2021). Validation of the child models of the Radboud Faces Database by children. *International Journal of Behavioral Development, 45*(2), 146–152. https://doi.org/10.1177/0165025420935631

Bill, G., Whyte, E., Griffin, J. W., & Scherf, K. S. (2020). Measuring sensitivity to eye gaze cues in naturalistic scenes: Presenting the eye gaze FoCuS database. *International Journal of Methods in Psychiatric Research, 29*(4), e1833. https://doi.org/10.1002/mpr.1833

Bureau, U. C. (2021). *2020 Census Illuminates Racial and Ethnic Composition of the Country*. Census.Gov. https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html. Accessed 2/21/2022.

Caharel, S., Montalan, B., Fromager, E., Bernard, C., Lalonde, R., & Mohamed, R. (2011). Other-race and inversion effects during the structural encoding stage of face processing in a race categorization task: An event-related brain potential study. *International Journal of Psychophysiology, 79*(2), 266–271. https://doi.org/10.1016/j.ijpsycho.2010.10.018

Chen, J. M., Fine, A. D., Norman, J. B., Frick, P. J., & Cauffman, E. (2021). Out of the Picture: Latinx and White Male Youths' Facial Features Predict Their Juvenile Justice System Processing Outcomes. *Crime & Delinquency, 67*(6–7), 787–807. https://doi.org/10.1177/0011128721999347

Ciardo, F., De Angelis, J., Marino, B. F. M., Actis-Grosso, R., & Ricciardelli, P. (2021). Social categorization and joint attention: Interacting effects of age, sex, and social status. *Acta Psychologica, 212*, 103223. https://doi.org/10.1016/j.actpsy.2020.103223

Ciardo, F., Marino, B. F., Rossetti, A., Actis-Grosso, R., & Ricciardelli, P. (2013). Face age and social status exert different modulatory effects on gaze following behaviour. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(35).

Cleveland, A., Schug, M., & Striano, T. (2007). Joint attention and object learning in 5- and 7-month-old infants. *Infant and Child Development, 16*(3), 295–306. https://doi.org/10.1002/icd.508

Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, N., & Casey, B. (2018). The racially diverse affective expression (RADIATE) face stimulus set. *Psychiatry*

Research, 270, 1059–1067. https://doi.org/10.1016/j.psychres.2018.04.066

Costa, C., Cristea, I. A., Dal Bò, E., Melloni, C., & Gentili, C. (2021). Brain activity during facial processing in autism spectrum disorder: An activation likelihood estimation (ALE) meta-analysis of neuroimaging studies. *Journal of Child Psychology and Psychiatry, 62*(12), 1412–1424. https://doi.org/10.1111/jcpp.13412

Courset, R., Rougier, M., Palluel-Germain, R., Smeding, A., Jonte, J. M., Chauvin, A., & Muller, D. (2018). The Caucasian and North African French Faces (CaNAFF): A Face Database. *International Review of Social Psychology, 31*(1), Article 1. https://doi.org/10.5334/irsp.179

Cruz-Janzen, M. I. (2002). Lives on the Crossfire: The Struggle of Multiethnic and Multiracial Latinos for Identity in a Dichotomous and Racialized World. *Race, Gender & Class, 9*(2), 47–62.

Dalmaso, M., Castelli, L., & Galfano, G. (2020). Social modulators of gaze-mediated orienting of attention: A review. *Psychonomic Bulletin & Review, 27*(5), 833–855. https://doi.org/10.3758/s13423-020-01730-x

Dalrymple, K. A., Gomez, J., & Duchaine, B. (2013). The Dartmouth Database of Children's Faces: Acquisition and Validation of a New Face Stimulus Set. *PLoS ONE, 8*(11), e79131. https://doi.org/10.1371/journal.pone.0079131

Davenport, L. (2020). The Fluidity of Racial Classifications. *Annual Review of Political Science, 23*(Volume 23, 2020), 221–240. https://doi.org/10.1146/annurev-polisci-060418-042801

de Lissa, P., Sokhn, N., Lasrado, S., Tanaka, K., Watanabe, K., & Caldara, R. (2021). Rapid saccadic categorization of other-race faces. *Journal of Vision, 21*(12), 1. https://doi.org/10.1167/jov.21.12.1

Douglass, S., Wang, Y., & Yip, T. (2016). The Everyday Implications of Ethnic-Racial Identity Processes: Exploring Variability in Ethnic-Racial Identity Salience Across Situations. *Journal of Youth and Adolescence, 45*(7), 1396–1411. https://doi.org/10.1007/s10964-015-0390-1

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. https://doi.org/10.1111/j.1467-9280.2006.01716.x

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing Black: Race, Crime, and Visual Processing. *Journal of Personality and Social Psychology, 87*(6), 876. https://doi.org/10.1037/0022-3514.87.6.876

Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., & Angold, A. (2011). The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): A new set of children's facial emotion stimuli. *International Journal of Methods in Psychiatric Research, 20*(3), 145–156. https://doi.org/10.1002/mpr.343

Farroni, T., Johnson, M. H., & Csibra, G. (2004). Mechanisms of Eye Gaze Perception during Infancy. *Journal of Cognitive Neuroscience, 16*(8), 1320–1326. https://doi.org/10.1162/0898929042304787

Feng, L., Liu, J., Wang, Z., Li, J., Li, L., Ge, L., Tian, J., & Lee, K. (2011). The other face of the other-race effect: An fMRI investigation of the other-race face categorization advantage. *Neuropsychologia, 49*(13), 3739–3749. https://doi.org/10.1016/j.neuropsychologia.2011.09.031

Flom, R., Lee, K., & Muir, D. (2017). *Gaze-Following: Its Development and Significance*. Psychology Press.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694.

Gaffrey, M. S., Barch, D. M., Singer, J., Shenoy, R., & Luby, J. L. (2013). Disrupted amygdala reactivity in depressed 4-to 6-year-old children. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*(7), 737–746.

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience, 4*(8), Article 8. https://doi.org/10.1038/90565

Greenwood, B. N., Hardeman, R. R., Huang, L., & Sojourner, A. (2020). Physician–patient racial concordance and disparities in birthing mortality for newborns. *Proceedings of the National Academy of Sciences, 117*(35), 21194–21200.

Gregory, S. E. A., Langton, S. R. H., Yoshikawa, S., & Jackson, M. C. (2020). A cross-cultural investigation into the influence of eye gaze on working memory for happy and angry faces. *Cognition and Emotion, 34*(8), 1561–1572. https://doi.org/10.1080/02699931.2020.1782353

Griffin, J. W., Bauer, R., & Scherf, K. S. (2021). A quantitative meta-analysis of face recognition deficits in autism: 40 years of research. *Psychological Bulletin, 147*(3), 268–292. https://doi.org/10.1037/bul0000310

Gross, R. (2005). Face databases. In *Handbook of face recognition* (pp. 301–327). Springer.

Hadders-Algra, M. (2022). Human face and gaze perception is highly context specific and involves bottom-up and top-down neural processing. *Neuroscience & Biobehavioral Reviews, 132*, 304–323. https://doi.org/10.1016/j.neubiorev.2021.11.042

Hadley, H., Rost, G. C., Fava, E., & Scott, L. S. (2014). A mechanistic approach to cross-domain perceptual narrowing in first year of life. *Brain Sciences, 4*, 613–634. https://doi.org/10.3390/brainsci4040613

Halberstadt, A. G. (2020). Racialized emotion recognition accuracy and anger bias of children's faces. *Emotion, undefined*(undefined), undefined. https://doi.org/10.1037/emo0000756

Hardeman, R. R., Medina, E. M., & Kozhimannil, K. B. (2016). Structural Racism and Supporting Black Lives—The Role of Health Professionals. *New England Journal of Medicine, 375*(22), 2113–2115. https://doi.org/10.1056/NEJMp1609535

Hills, P. J., & Pake, J. M. (2013). Eye-tracking the own-race bias in face recognition: Revealing the perceptual and socio-cognitive mechanisms. *Cognition, 129*(3), 586–597. https://doi.org/10.1016/j.cognition.2013.08.012

Hochman, A. (2021). Janus-faced race: Is race biological, social, or mythical? *American Journal of Physical Anthropology, 175*(2), 453–464. https://doi.org/10.1002/ajpa.24169

Hoehl, S., & Striano, T. (2008). Neural Processing of Eye Gaze and Threat-Related Emotional Facial Expressions in Infancy. *Child Development, 79*(6), 1752–1760. https://doi.org/10.1111/j.1467-8624.2008.01223.x

Hoehl, S., & Striano, T. (2010). The development of emotional face and eye gaze processing. *Developmental Science, 13*(6), 813–825.

Hoehl, S., Wahl, S., Michel, C., & Striano, T. (2012). Effects of eye gaze cues provided by the caregiver compared to a stranger on infants' object processing. *Developmental Cognitive Neuroscience, 2*(1), 81–89.

Hoehl, S., Wiese, L., & Striano, T. (2008). Young infants' neural processing of objects is affected by eye gaze direction and emotional expression. *PLoS ONE, 3*(6), e2389.

Hu, C. S., Wang, Q., Han, T., Weare, E., & Fu, G. (2017). Differential emotion attribution to neutral faces of own and other races. *Cognition and Emotion, 31*(2), 360–368. https://doi.org/10.1080/02699931.2015.1092419

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews, 33*(6), 843–863. https://doi.org/10.1016/j.neubiorev.2009.02.004

Jessen, S., & Grossmann, T. (2016). Neural and Behavioral Evidence for Infants' Sensitivity to the Trustworthiness of Faces. *Journal of Cognitive Neuroscience, 28*(11), 1728–1736. https://doi.org/10.1162/jocn_a_00999

Jessen, S., & Grossmann, T. (2019). Neural evidence for the subliminal processing of facial trustworthiness in infancy. *Neuropsychologia, 126*, 46–53. https://doi.org/10.1016/j.neuropsychologia.2017.04.025

Johnson, S. C., Ok, S.-J., & Luo, Y. (2007). The attribution of attention: 9-month-olds' interpretation of gaze as goal-directed action. *Developmental Science, 10*(5), 530–537.

Killgore, W. D. S., Britton, J. C., Schwab, Z. J., Price, L. M., Weiner, M. R., Gold, A. L., Rosso, I. M., Simon, N. M., Pollack, M. H., & Rauch, S. L. (2014). Cortico-Limbic Responses to Masked Affective Faces Across Ptsd, Panic Disorder, and Specific Phobia. *Depression and Anxiety, 31*(2), 150–159. https://doi.org/10.1002/da.22156

Krstovska-Guerrero, I., & Jones, E. A. (2016). Social-Communication Intervention for Toddlers with Autism Spectrum Disorder: Eye Gaze in the Context of Requesting and Joint Attention. *Journal of Developmental and Physical Disabilities, 28*(2), 289–316. https://doi.org/10.1007/s10882-015-9466-9

Krumhuber, E. G., Küster, D., Namba, S., & Skora, L. (2021). Human and machine validation of 14 databases of dynamic facial expressions. *Behavior Research Methods, 53*(2), 686–701. https://doi.org/10.3758/s13428-020-01443-y

Krumhuber, E. G., Skora, L., Küster, D., & Fou, L. (2017). A Review of Dynamic Datasets for Facial Expression Research. *Emotion Review, 9*(3), 280–292. https://doi.org/10.1177/1754073916670022

Kumar, R., Karabenick, S. A., & Burgoon, J. N. (2015). Teachers' implicit attitudes, explicit beliefs, and the mediating role of respect and cultural responsibility on mastery and performance-focused instructional practices. *Journal of Educational Psychology, 107*(2), 533.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual Other-Race Training Reduces Implicit Racial Bias. *PLoS ONE, 4*(1), e4215. https://doi.org/10.1371/journal.pone.0004215

Lee, E., Kang, J. I., Park, I. H., Kim, J.-J., & An, S. K. (2008). Is a neutral face really evaluated as being emotionally neutral? *Psychiatry Research, 157*(1), 77–85. https://doi.org/10.1016/j.psychres.2007.02.005

Lee, K., Quinn, P. C. ., & Heyman, G. D. (2017). Rethinking the Emergence and Development of Implicit Racial Bias: A Perceptual-Social Linkage Hypothesis. In N. Budwig, E. Turiel, & P. D. Zelazo (Eds.), *New Perspectives on Human Development* (pp. 27–46). Cambridge University Press. https://doi.org/10.1017/CBO9781316282755.004

Lee, K., Quinn, P. C., & Pascalis, O. (2017b). Face Race Processing and Racial Bias in Early Development: A Perceptual-Social Linkage. *Current Directions in Psychological Science, 26*(3), 256–262. https://doi.org/10.1177/0963721417690276

Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and Cues: Gaze-following in Children with Autism. *Journal of Child Psychology and Psychiatry, 39*(7), 951–962. https://doi.org/10.1111/1469-7610.00398

Liu, S., Xiao, N. G., Quinn, P. C., Zhu, D., Ge, L., Pascalis, O., & Lee, K. (2015). Asian infants show preference for own-race but not other-race female faces: The role of infant caregiving arrangements. *Frontiers in Psychology, 6*. https://www.frontiersin.org/articles/https://doi.org/10.3389/fpsyg.2015.00593

LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults. *Frontiers in Psychology, 5*. https://www.frontiersin.org/article/https://doi.org/10.3389/fpsyg.2014.01532

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Ma, D. S., Kantner, J., & Wittenbrink, B. (2021). Chicago Face Database: Multiracial expansion. *Behavior Research Methods, 53*(3), 1289–1300. https://doi.org/10.3758/s13428-020-01482-5

Markant, J., & Scott, L. S. (2018). Attention and Perceptual Learning Interact in the Development of the Other-Race Effect. *Current Directions in Psychological Science, 27*(3), 163–169. https://doi.org/10.1177/0963721418769884

Matheson, H. E., & McMullen, P. A. (2011). A computer-generated face database with ratings on realism, masculinity, race, and stereotypy. *Behavior Research Methods, 43*(1), 224–228. https://doi.org/10.3758/s13428-010-0029-9

Mazza, G. L., Enders, C. K., & Ruehlman, L. S. (2015). Addressing Item-Level Missing Data: A Comparison of Proration and Full Information Maximum Likelihood Estimation. *Multivariate Behavioral Research, 50*(5), 504–519. https://doi.org/10.1080/00273171.2015.1068157

McKone, E., Dawel, A., Robbins, R. A., Shou, Y., Chen, N., & Crookes, K. (2021). Why the other-race effect matters: Poor recognition of other-race faces impacts everyday social interactions. *British Journal of Psychology*, *n/a*(n/a). https://doi.org/10.1111/bjop.12508

Monk, C. S., Nelson, E. E., McClure, E. B., Mogg, K., Bradley, B. P., Leibenluft, E., Blair, R. J. R., Chen, G., Charney, D. S., & Ernst, M. (2006). Ventrolateral prefrontal cortex activation and attentional bias in response to angry faces in adolescents with generalized anxiety disorder. *American Journal of Psychiatry, 163*(6), 1091–1097.

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013a). Infants understand the referential nature of human gaze but not robot gaze. *Journal of Experimental Child Psychology, 116*(1), 86–95.

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013b). The power of human gaze on infant learning. *Cognition, 128*(2), 127–133.

Pavan, G., Dalmaso, M., Galfano, G., & Castelli, L. (2011). Racial Group Membership Is Associated to Gaze-Mediated Orienting in Italy. *PLoS ONE, 6*(10), e25608. https://doi.org/10.1371/journal.pone.0025608

Pickron, C. B., Fava, E., & Scott, L. S. (2017). Follow My Gaze: Face Race and Sex Influence Gaze-Cued Attention in Infancy. *Infancy, 22*(5), 626–644. https://doi.org/10.1111/infa.12180

Prada, M., Garrido, M. V., Camilo, C., & Rodrigues, D. L. (2018). Subjective ratings and emotional recognition of children's facial expressions from the CAFE set. *PLoS ONE, 13*(12), e0209644. https://doi.org/10.1371/journal.pone.0209644

Qian, M. K., Heyman, G. D., Quinn, P. C., Messi, F. A., Fu, G., & Lee, K. (2016). Implicit Racial Biases in Preschool Children and Adults From Asia and Africa. *Child Development, 87*(1), 285–296. https://doi.org/10.1111/cdev.12442

Qian, M. K., Quinn, P. C., Heyman, G. D., Pascalis, O., Fu, G., & Lee, K. (2019). A Long-Term Effect of Perceptual Individuation Training on Reducing Implicit Racial Bias in Preschool Children. *Child Development*, *90*(3). https://doi.org/10.1111/cdev.12971

Quinn, P. C., Lee, K., & Pascalis, O. (2018). Perception of Face Race by Infants: Five Developmental Changes. *Child Development Perspectives, 12*(3), 204–209. https://doi.org/10.1111/cdep.12286

Recasens, A., Khosla, A., Vondrick, C., & Torralba, A. (2015). *Where are they looking?* 10.

Reid, V. M., & Striano, T. (2005). Adult gaze influences infant attention and object processing: Implications for cognitive neuroscience.

European Journal of Neuroscience, 21(6), 1763–1766. https://doi.org/10.1111/j.1460-9568.2005.03986.x

Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *NeuroReport, 15*(16), 2553–2555.

Renfrew, C., Frith, C., Malafouris, L., & Frith, C. D. (2008). Social cognition. *Philosophical Transactions of the Royal Society b: Biological Sciences, 363*(1499), 2033–2039. https://doi.org/10.1098/rstb.2008.0005

Rennels, J. L., & Davis, R. E. (2008). Facial experience during the first year. *Infant Behavior and Development, 31*(4), 665–678. https://doi.org/10.1016/j.infbeh.2008.04.009

Richeson, J. A., Todd, A. R., Trawalter, S., & Baird, A. A. (2008). Eye-Gaze Direction Modulates Race-Related Amygdala Activity. *Group Processes & Intergroup Relations, 11*(2), 233–246. https://doi.org/10.1177/1368430207088040

Roberts, S. O., & Rizzo, M. T. (2021). The psychology of American racism. *American Psychologist, 76*(3), 475–487. https://doi.org/10.1037/amp0000642

Roesch, E. B., Tamarit, L., Reveret, L., Grandjean, D., Sander, D., & Scherer, K. R. (2011). FACSGen: A Tool to Synthesize Emotional Facial Expressions Through Systematic Manipulation of Facial Action Units. *Journal of Nonverbal Behavior, 35*(1), 1–16. https://doi.org/10.1007/s10919-010-0095-9

Russell, J. A., & Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *Journal of Experimental Psychology: General, 116*(3), 223–237. https://doi.org/10.1037/0096-3445.116.3.223

Russell, R., Biederman, I., Nederhouser, M., & Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Research, 47*(2), 157–165. https://doi.org/10.1016/j.visres.2006.11.002

Saarinen, A., Keltikangas-Järvinen, L., Jääskeläinen, E., Huhtaniska, S., Pudas, J., Tovar-Perdomo, S., Penttilä, M., Miettunen, J., & Lieslehto, J. (2021). Early Adversity and Emotion Processing From Faces: A Meta-analysis on Behavioral and Neurophysiological Responses. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 6*(7), 692–705. https://doi.org/10.1016/j.bpsc.2021.01.002

Sacco, A. M., de Paula Couto, M. C. P., & Koller, S. H. (2016). Construction and validation of the White, Pardo, and Black Children Picture Set (BIC-Multicolor). *Psychology & Neuroscience, 9*(1), 68–78. https://doi.org/10.1037/pne0000040

Salter, P. S., Adams, G., & Perez, M. J. (2018). Racism in the Structure of Everyday Worlds: A Cultural-Psychological Perspective. *Current Directions in Psychological Science, 27*(3), 150–155. https://doi.org/10.1177/0963721417724239

Scherf, K. S., & Scott, L. S. (2012). Connecting developmental trajectories: Biases in face processing from infancy to adulthood. *Developmental Psychobiology, 54*(6), 643–663. https://doi.org/10.1002/dev.21013

Scott, L. S., & Arcaro, M. J. (2023). A domain-relevant framework for the development of face processing. *Nature Reviews Psychology, 2*(3), Article 3. https://doi.org/10.1038/s44159-023-00152-5

Scott, L. S., & Fava, E. (2013). The own-species face bias: A review of developmental and comparative data. *Visual Cognition, 21*(9–10), 1364–1391. https://doi.org/10.1080/13506285.2013.821431

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A Domain-General Theory of the Development of Perceptual Discrimination. *Current Directions in Psychological Science, 16*(4), 197–201. https://doi.org/10.1111/j.1467-8721.2007.00503.x

Sekimoto, S. (2018). Race and the Senses: Toward Articulating the Sensory Apparatus of Race. *Critical Philosophy of Race, 6*(1), 82–100. https://doi.org/10.5325/critphilrace.6.1.0082

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition, 108*(2), 303–319.

Sessa, P., & Dalmaso, M. (2016). Race perception and gaze direction differently impair visual working memory for faces: An event-related potential study. *Social Neuroscience, 11*(1), 97–107. https://doi.org/10.1080/17470919.2015.1040556

Smedley, A., & Smedley, B. D. (2005). Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. *American Psychologist, 60*(1), 16–26. https://doi.org/10.1037/0003-066X.60.1.16

Stallworthy, I. C., Lasch, C., Berry, D., Wolff, J. J., Pruett, J. R., Marrus, N., Swanson, M. R., Botteron, K. N., Dager, S. R., Estes, A. M., Hazlett, H. C., Schultz, R. T., Zwaigenbaum, L., Piven, J., & Elison, J. T. (2022). Variability in Responding to Joint Attention Cues in the First Year is Associated With Autism Outcome. *Journal of the American Academy of Child & Adolescent Psychiatry, 61*(3), 413–422. https://doi.org/10.1016/j.jaac.2021.03.023

Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology, 3*(3), 289–299. https://doi.org/10.1080/17405620600879779

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in Cognitive Sciences, 10*(10), 471–476.

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods, 48*(3), 1197–1204. https://doi.org/10.3758/s13428-015-0641-9

Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective: Infants' Daily Exposure to Faces. *Developmental Psychobiology, 56*(2), 249–261. https://doi.org/10.1002/dev.21183

Todorov, A., & Oh, D. (2021). Chapter Four—The structure and perceptual basis of social judgments from faces. In B. Gawronski (Ed.), *Advances in Experimental Social Psychology* (Vol. 63, pp. 189–245). Academic Press. https://doi.org/10.1016/bs.aesp.2020.11.004

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Trawalter, S., Todd, A. R., Baird, A. A., & Richeson, J. A. (2008). Attending to threat: Race-based patterns of selective attention. *Journal of Experimental Social Psychology, 44*(5), 1322–1327. https://doi.org/10.1016/j.jesp.2008.03.006

Ueda, Y., Nunoi, M., & Yoshikawa, S. (2019). Development and Validation of the Kokoro Research Center (krc) Facial Expression Database. *Psychologia, 61*(4), 221–240. https://doi.org/10.2117/psysoc.2019-A009

Umaña-Taylor, A. J., Quintana, S. M., Lee, R. M., Cross Jr., W. E., Rivas-Drake, D., Schwartz, S. J., Syed, M., Yip, T., Seaton, E., & Ethnic and Racial Identity in the 21st Century Study Group. (2014). Ethnic and Racial Identity During Adolescence and Into Young Adulthood: An Integrated Conceptualization. *Child Development, 85*(1), 21–39. https://doi.org/10.1111/cdev.12196

Webb, S. J., Jones, E. J. H., Kelly, J., & Dawson, G. (2014). The motivation for very early intervention for infants at high risk for autism spectrum disorders. *International Journal of Speech-Language Pathology, 16*(1), 36–42. https://doi.org/10.3109/17549507.2013.861018

Weisbuch, M., Pauker, K., Adams, R. B., Jr., Lamer, S. A., & Ambady, N. (2017). Race, power, and reflexive gaze following. *Social Cognition, 35*(6), 619–638.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The

SHINE toolbox. *Behavior Research Methods, 42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Williams, M. J., George-Jones, J., & Hebl, M. (2019). The face of STEM: Racial phenotypic stereotypicality predicts STEM persistence by—and ability attributions about—students of color. *Journal of Personality and Social Psychology, 116*(3), 416.

Willis, J., & Todorov, A. (2006). First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science, 17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Workman, C. I., & Chatterjee, A. (2021). The Face Image Meta-Database (fIMDb) & ChatLab Facial Anomaly Database (CFAD): Tools for research on face perception and social stigma. *Methods in Psychology, 5*, 100063. https://doi.org/10.1016/j.metip.2021.100063

Xiao, N. G., Wu, R., Quinn, P. C., Liu, S., Tummeltshammer, K. S., Kirkham, N. Z., Ge, L., Pascalis, O., & Lee, K. (2018). Infants Rely More on Gaze Cues From Own-Race Than Other-Race Adults for Learning Under Uncertainty. *Child Development, 89*(3), e229–e244. https://doi.org/10.1111/cdev.12798

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

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences, 17*(6), 263–271. https://doi.org/10.1016/j.tics.2013.04.004

Zhao, L., & Bentin, S. (2008). Own- and other-race categorization of faces by race, gender, and age. *Psychonomic Bulletin & Review, 15*(6), 1093–1099. https://doi.org/10.3758/PBR.15.6.1093