Testing the functional basis of first impressions: Dimensions for children's faces are not the same as for adults' faces

Jemma R. Collova*, Clare A.M. Sutherland and Gillian Rhodes

*Corresponding Author: Jemma Collova

Email: jemma.collova@research.uwa.edu.au

ARC Centre of Excellence in Cognition and its Disorders,

School of Psychological Science,

University of Western Australia, Crawley, Australia, WA 6009,

To be cited as: Collova, J. R., Sutherland, C. M. & Rhodes, G. (in press). Testing the functional basis of first impressions: Dimensions for children's faces are not the same as for adults' faces. *Journal of Personality and Social Psychology*

Acknowledgements

This research was supported by PhD funding from the University of Western Australia, School of Psychological Science, funding from the ARC Centre of Excellence in Cognition and Its Disorders (CE110001021) an ARC Discovery project (DP170104602) and an ARC Discovery Early Career Researcher Award to CS (DE190101043). We are grateful to Nichola Burton for her help coding the Study 1 descriptions, the PSYC3310 class for their help recruiting participants, and Gilles Gignac for his helpful statistical insight.

Abstract

Despite warnings not to "judge a book by its cover", people rapidly form facial impressions. In Oosterhof and Todorov's (2008) two-dimensional model of facial impressions, trustworthiness and dominance underlie impressions and primarily function to signal the potential threat of others. Here, we test a key assumption of these models, namely that these dimensions are functional, by evaluating whether the adult-face dimensions apply to young children's faces. Although it may be functional for adults to judge adult faces on dimensions that signal threat, adults associate different social goals with children, and these goals are likely to impact the impressions adults make of such faces. Thus, a functional approach would predict that the dimensions for children's faces are not threat focused. In Studies 1-2, we build a data-driven model of Caucasian adults' impressions of Caucasian children's faces, finding evidence for two dimensions. The first dimension, *niceness*, is similar (although not identical) to the adult dimension of trustworthiness. However, we find a second dimension, *shyness*, that is clearly dissociable from dominance (Study 3), and critically, is not focused on threat. We demonstrate that adults are sensitive to subtle facial manipulations of these dimensions (Studies 4-5) and that these impressions impact adults' behavioral expectations of children (Study 6). Finally, we show that niceness and shyness dimensions generalize to an independent sample of ambient images, demonstrating their robustness (Study 7). Our results suggest that social goals have the power to drive functional impressions and highlight the flexibility of our visual system when forming such inferences.

Keywords: "impression formation" "person perception" "face perception" "social cognition" "first impressions"

Testing the functional basis of first impressions: Dimensions for children's faces are not the same as for adults' faces

People rapidly form first impressions from a glimpse of a face. Based on facial information, adults form consistent judgments of traits such as trustworthiness, competence, dominance and intelligence (for a review see Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). Whether accurate or not, these impressions influence real-life behavior towards adults (Ert, Fleischer, & Magen, 2015; Olivola, Sussman, Tsetsos, Kang, & Todorov, 2012; Porter, ten Brinke, & Gustaw, 2010) and children (Barocas & Black, 1974; Berkowitz & Frodi, 1979; Salvia, Algozzine, & Sheare, 1977).

Recently, Oosterhof and Todorov (2008) have proposed an influential two-dimensional model of facial impressions, whereby trustworthiness and dominance are argued to underlie trait impressions of adult faces. Crucially, Oosterhof and Todorov (2008) employed a data-driven approach whereby they used adults' unconstrained descriptions of adult faces to model the dimensions underlying impressions. This data-driven approach is theoretically powerful because it allows patterns to be drawn from otherwise highly-dimensional and complex judgments and stimuli, without imposing *a priori* assumptions.

To date, theory has conceptualized trustworthiness and dominance as universal evaluations, which are functional because together they signal the potential threat of others (Oosterhof & Todorov, 2008). Here, we test the key underlying assumption that the dimensions are functional. Currently, research has found that similar dimensions emerge across culture (Sutherland, Liu, et al., 2018; Wang, et al., 2019), sex (South Palomares & Young, 2017), and different social stimuli (Abele & Wojciszke, 2007; Fiske, Cuddy, & Glick, 2007; Rosenberg, Nelson, & Vivekananthan, 1968; Sutherland et al., 2013). These lines of evidence suggest that these dimensions may indeed be universal and importantly, have a functional basis. However, a more powerful test of this functional assumption would

be to find an *exception* to these dimensions, by testing these dimensions on a population of faces for whom it would not be adaptive to assess threat.

Young children comprise just such a population. In general, adults associate different social tasks and goals with adults and children. Adult-adult relationships are associated with social goals such as competing for mates and resources, selecting a mate, and establishing friendships and alliances (Geary & Bjorklund, 2000; Havighurst, 1948). Detecting threat from adult faces appears functional in the context of these social goals. For example, when discriminating friend from foe it might be important to gauge the potential threat of other adults. Likewise, it might be important to determine the potential threat of one's opponents before deciding whether or not to compete with them for a mate. In contrast, adult-child interactions are primarily driven by different social goals, such as establishing a caregiving and nurturing attachment (Bowlby, 1969). Detecting threat from children's faces does not appear directly adaptive. If the functional perspective is correct, then crucially, we would not expect the same threat-based dimensions to appear for children's faces. Importantly, this result would not undermine previous models of impressions (Oosterhof & Todorov, 2008; Sutherland, et al., 2013), but rather support the functional assumption underlying them. In support of this idea, different social goals have previously been shown to shape another aspect of face processing, namely facial recognition, in distinctive ways for child and adult faces (Picci & Scherf, 2016; Scherf & Scott, 2012).

Here, we provide an important theoretical test of the functionality of impressions by examining whether different dimensions emerge for faces that adults associate with different social goals, i.e. young children's faces. In Study 1 and 2, we take a data-driven approach to build a dimensional model for adult's impressions of young children's faces. In Study 3, we quantify how these dimensions relate to the adult face dimensions of trustworthiness and dominance (Oosterhof & Todorov, 2008) and the potential cues driving such impressions

(e.g. emotion overgeneralization; Zebrowitz, 2017; Zebrowitz & Montepare, 2008). In Study 4, we validate our dimensional model by investigating whether we can visually manipulate the child face dimensions. Additionally, we create a new set of validated children's face images that have been digitally manipulated to vary along these psychologically relevant dimensions, for use in future research. In Study 5, we apply a strict test of the independence of the child face dimensions by testing whether adults are able to discriminate subtle differences between the dimensions. In light of our new theoretical understanding, we consider the practical implications of these results by examining potential behavioral consequences of first impressions of children's faces (Study 6). Finally, we provide a stringent test of the robustness of the child face dimensions by examining whether they generalize to an independent sample of naturalistic children's faces (Study 7).

Facial impressions are socially important

Regardless of whether first impressions are accurate (Carré & McCormick, 2008; Penton-Voak, Pound, Little, & Perrett, 2006; Stirrat & Perrett, 2010), or not (Olivola & Todorov, 2010b; Rule, Krendl, Ivcevic, & Ambady, 2013), they can be powerful determinants of real-world behavior for both adults and children. For adults, facial appearance influences people's likelihood of being elected as political leaders (Olivola et al., 2012; Olivola & Todorov, 2010a), likelihood of having their accommodation booked on Airbnb (Ert et al., 2015), sentences received within the criminal justice system (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Porter et al., 2010), and predicts organizational performance (Rule & Ambady, 2008). For children, facial appearance influences how positively they are treated (Langlois et al., 2000; Zebrowitz & Lee, 1999), the positivity of their school reports (Salvia et al., 1977), severity of discipline received from adults (Berkowitz & Frodi, 1979; Zebrowitz, Kendall-Tackett, & Fafel, 1991), likelihood of receiving "helpful" referrals from

teachers (Barocas & Black, 1974), and how popular, motivated and bright teachers assume them to be (Clifford & Walster, 1973; Kenealy, Frude, & Shaw, 1988). Given that first impressions have significant social consequences throughout our lifetime, it is important that we understand how and why we form these impressions.

A model of first impressions of adult faces

With almost 18,000 English words to describe character judgments (Allport & Odbert, 1936), how do we know which judgments are most important in driving our impressions? Recently, a new theoretical model of facial impressions has shown that numerous trait impressions from faces can be explained by trustworthiness and dominance evaluations, comprising two dimensions of fundamental, adaptive importance. To achieve this understanding, Oosterhof and Todorov (2008) asked participants to rate unfamiliar adult faces on a range of traits that were produced when forming unconstrained, spontaneous impressions. They found that these facial trait inferences could be captured through two dimensions, trustworthiness and dominance, which together explained over 80% of the variance in the original judgments.

In many ways, this dimensional model has theoretically structured the field of social face perception. It has inspired studies investigating the facial cues (e.g. Said, Sebe, & Todorov, 2009), neural correlates (Mende-Siedlecki, Said, & Todorov, 2012), conscious processing (Abir, Sklar, Dotsch, Todorov, & Hassin, 2017), and developmental trajectory (Cogsdill, Todorov, Spelke, & Banaji, 2014) of these dimension judgments, as well work investigating how such judgments might vary across individuals (Hehman, Sutherland, Flake, & Slepian, 2017) and cultures (Sutherland, Liu, et al., 2018; Wang, et al., 2019).

Empirically, Oosterhof and Todorov's (2008) discovery pioneered the use of a data-driven approach in *face* perception. This work builds on the use of data-driven approaches in

person perception, which can be traced back to over 50 years of research (e.g. Asch, 1946; Rosenberg et al., 1968; Secord, 1958). Unlike hypothesis-driven approaches, data-driven approaches do not rely on variables selected according to researchers' predictions. Instead, the focus is on the discovery of new relationships and models as well as understanding complex, non-linear patterns from high-dimensional stimuli, such as faces. Data-driven approaches, such as the factor reduction analysis employed by Oosterhof and Todorov (2008), have increasingly gained traction in empirical papers (e.g. Abir et al., 2017; Sutherland, Liu, et al., 2018; Sutherland, Rhodes, & Young, 2017) and in psychology more widely (for reviews see Adolphs, Nummenmaa, Todorov, & Haxby, 2016; Jack & Schyns, 2017).

Since the initial development of the facial first impressions model, there has been robust evidence for the dimensions of trustworthiness and dominance. These two dimensions have been replicated using a variety of face images (Sutherland, Liu, et al., 2018; Walker & Vetter, 2009; Wolffhechel et al., 2014), including naturalistic photographs (Sutherland et al., 2013; note that in some of these face samples, an attractiveness-age dimension emerges). Interestingly, these dimensions also mirror well-established dimensions found across other fields of social psychology and other stimuli, beyond face images (Abele & Wojciszke, 2007; Fiske et al., 2007; Oosterhof & Todorov, 2008; Rohner & Rohner, 1981; Rosenberg et al., 1968). For example, trustworthiness and dominance show distinct similarities to the dimensions of warmth and competence (Fiske et al., 2007), respectively (although there is less similarity between dominance and competence, than between trustworthiness and warmth; Sutherland, Oldmeadow, & Young, 2016). Warmth and competence underlie impressions of group stereotypes and are argued to be universal dimensions of social cognition (Fiske et al., 2007). Trustworthiness and dominance also broadly resemble dimensions from early work in person perception (social-goodness and intellectual-goodness,

respectively; Rosenberg et al., 1968) and dimensions of semantic judgments (evaluation and potency, respectively; Osgood, 1952). It is interesting that these similar dimensions consistently emerge within social perception, as it suggests that these dimensions are likely to have an important functional role.

In line with an ecological perspective (Zebrowitz & Collins, 1997), impressions based on trustworthiness and dominance are thought to be particularly important due to selection pressures which advantage the detection of these traits (Oosterhof & Todorov, 2008). Trustworthiness is based on subtle resemblances to emotional expressions and signals the appropriateness of avoiding or approaching someone, so that a happy expression looks trustworthy, whereas anger looks untrustworthy. Dominance is sensitive to facial cues of physical strength, and signals that person's ability to cause harm. Critically, Oosterhof and Todorov (2008) argue that people overgeneralize these judgments to make inferences about the potential threat of others, giving these dimensions a functional basis. The idea that impressions may be overgeneralizations of responses to adaptively significant facial cues can be traced back to Secord (1958) and has strongly influenced our understanding of facial impressions (in particular, driven by Zebrowitz and her colleagues: e.g. Zebrowitz and Collins, 1997). From this perspective, trait judgments do not necessarily have to be accurate in order to have a functional basis.

Do social goals drive functional facial impressions?

Although research highlights the functionality of trustworthiness and dominance judgments for adult evaluations of adult faces, no study has yet considered the possibility that different social goals might drive different dimensions for certain faces. If the dimensions are indeed functional, then they should be flexible depending on the social goals associated with populations of faces (as also suggested by Sutherland et al., 2017). This perspective raises an

interesting question. Are there populations of faces for which detecting potential threat is not appropriate, and does this context influence the sorts of impressions made for such faces? One obvious example would be adults' impressions made to children's faces. In particular, it seems unlikely that children's faces should be evaluated on potential threat given that in most contexts, threat is not a primary concern in the social developmental tasks and goals associated with adult-child interactions.

Social developmental tasks associated with children (Havighurst, 1948, 1972) are centered on establishing a caregiving attachment with parent figures (Bowlby, 1969). In adulthood, developmental goals become focused on peer relationships (e.g. mate selection, establishing friendships and mate competition: Havighurst, 1972). Interestingly these different social goals influence face processing abilities for adult and child faces (Picci & Scherf, 2016; Scherf & Scott, 2012). Specifically, face processing skills favor the recognition of faces that align with specific social goals. For adults, identification of other adults is most important with respect to the social goals associated with adulthood. Consequentially, adults' face processing skills favor the recognition of adult (i.e. peer faces) over child faces (Picci & Scherf, 2016; Scherf & Scott, 2012). Adolescents' face processing skills also favor the recognition of peer (i.e. adolescent) faces over both child and adult faces. Importantly, this peer bias emerges during adolescence (specifically, puberty; Picci & Scherf, 2016), when developmental goals become centered on establishing new friendships with peers (Havinghurst, 1972). Together, this evidence highlights the ability of specific social tasks and goals to shape the "goals" of the perceptual system (Picci & Scherf, 2016).

To date, research on first impressions has conceptualised trustworthiness and dominance as important dimensions for all faces (e.g. Oosterhof & Todorov, 2008; Sutherland et al., 2013; Walker & Vetter, 2016) and has focused on testing the universality of such dimensions (Sutherland, Liu, et al., 2018; Wang, et al., 2019). Here, we propose that

children's faces are likely to be evaluated on qualitatively different dimensions to adults' faces, driven by a corresponding difference in social goals.

A functional approach would predict that the dimensions for children's faces would not be focused on detecting threat. For example, adults might be sensitive to signals of cooperativeness and vulnerability in children's faces, rather than potential capability to harm. However, a data-driven approach allows us to remain open to any potential alternative dimensional structure. For example, we may find the same dimensions emerge for children's faces as have been found previously for adults' faces. This result would support the suggestion that trustworthiness and dominance are universal evaluations, but would question whether their primary function is to signal threat. Alternatively, we may find a simpler dimensional structure emerges (e.g. just one valence dimension), possibly because children's faces are physically less developed than adults' (Enlow & Moyers, 1982). Nevertheless, the data-driven approach we employ will allow us to consider all of these possibilities.

Investigating adults' impressions of children's faces is particularly important for both theoretical and applied reasons. As mentioned above, it provides a test of a key theoretical claim of the functional model. If the dimensions are indeed functional (Oosterhof & Todorov, 2008), then they should be different across faces associated with different social goals. It is also important to understand how impressions of children's faces are structured because researchers are starting to examine impressions of children's faces with the assumption that the adult dimensions of trustworthiness and dominance apply to these faces (Cogsdill & Banaji, 2015; Q. Li, Heyman, Mei, & Lee, 2017). Of course, when explicitly requested, adults are able to make consistent judgments about trustworthiness in children's faces (e.g. Cogsdill & Banaji, 2015; Q. Li et al., 2017). However, it is not clear that impressions of trustworthiness and dominance represent the most important judgments inferred from children's faces, or that these judgments are spontaneously inferred from children's faces.

Through employing a data-driven approach, we can determine the important dimensions for impressions of children's faces without imposing *a priori* assumptions. Past research has considered how facial impressions may change across the lifespan, focusing on specific, hypothesized variables. For example, Zebrowitz and Montepare (1992) investigated how variations in babyfaceness across the lifespan (i.e. from infancy to adulthood) influence judgments of five broad dimensions (maturefaced, social autonomy, naiveté, physical weakness and warmth). These five impressions were selected based on their *a priori* hypothesized importance in social perception (Berry & McArthur, 1986). Indeed, babyfaced individuals were perceived to have childlike tendencies as measured by these five trait dimensions and this relationship was true for faces across the lifespan (Zebrowitz & Montepare, 1992). Although this research provided important insight into the stability of impressions over time, the method employed was not sensitive to detecting whether new or different dimensions exist for faces across the lifespan. In contrast, our data-driven approach can extend past research by examining whether entirely new trait dimensions emerge for children's faces.

Understanding adults' impressions of children's faces is also important from an applied perspective, because children's facial appearance does influence adults' expectations and behaviors towards them (Barocas & Black, 1974; Clifford & Walster, 1973; Kenealy et al., 1988; Ritts, Patterson, & Tubbs, 1992; Salvia et al., 1977; Zebrowitz et al., 1991; Zebrowitz & Montepare, 1992; Zebrowitz, Voinescu, & Collins, 1996). In previous research investigating the behavioral consequences of children's facial appearance, stimuli have been selected and manipulated according to researchers' hypotheses, which might underestimate the effects of these consequences. In particular, research has focused on cues that are useful in guiding adaptive behavior, namely facial babyfaceness and attractiveness. Facial characteristics that signal infantile qualities and low fitness are so important that they are

overgeneralized to individuals whose facial appearance merely resembles such qualities (i.e. overgeneralization effects; see Zebrowitz & Montepare, 2008 for a review). Indeed, children who look babyfaced are assumed to have child-like qualities (e.g. Zebrowitz, Brownlow, & Olson, 1992; Zebrowitz et al., 1991; Zebrowitz, Olson, & Hoffman, 1993), while children who are more attractive are perceived more positively than their peers on a variety of traits (e.g. Clifford & Walster, 1973; Langlois et al., 2000) and are treated more favorably by adults (Barocas & Black, 1974; Berkowitz & Frodi, 1979). Although it is clear that these cues have important social consequences for children, there are likely to be other qualities that also drive adult behavior towards children. Here, we use a data-driven approach to help identify any other theoretically important variables.

Finally, it is important to note that the social consequences of facial impressions might be particularly influential during childhood, a crucial time of development (see Phillips & Shonkoff, 2000). For example, adults might behave in ways that reinforce the behavioral expectations associated with children's appearance (e.g. self-fulfilling prophecies; Zebrowitz, Collins, & Dutta, 1998) and this behavior is likely to have long lasting consequences on children's development. Considering these lines of evidence, it is clear that understanding impressions of children's faces is important from both a theoretical and an applied perspective.

Present Research

We apply the data-driven methods used to derive the dimensions underlying adult impressions of adult faces (Oosterhof & Todorov, 2008; Sutherland, Liu, et al., 2018; Sutherland et al., 2013; Walker & Vetter, 2016) to determine the dimensions underlying adult impressions of young children's faces for the first time. In Study 1, we collect unconstrained descriptions of children's faces to capture the traits spontaneously used to describe children.

In Study 2, we ask new participants to rate the same children's faces on the most commonly mentioned traits taken from Study 1. We then run a factor analysis on these trait ratings to determine the dimensions underlying spontaneous first impressions of young children's faces. This approach is important to testing the original model, because it allows us to model adults' impressions of children's faces without imposing researcher-driven assumptions about which traits are most important.

One potential concern is that adults may use distinct words to describe child as compared to adult faces, but these descriptions may not actually tap conceptually distinct traits. By examining the overlap of the child dimensions with other important social dimensions, we can rule out this less interesting possibility. Therefore, in Study 3, we examine the nature of the child face dimensions by directly comparing them to the two dimensions underlying impressions of adult faces, trustworthiness and dominance (Oosterhof & Todorov, 2008). We also compare them to other theoretically important dimensions emerging from a broader social cognition literature (e.g. warmth and competence from the Stereotype Content Model: Fiske, et al., 2017). In Study 3, we also consider potential physical correlates of the child face dimensions, guided by Zebrowitz' ecological approach (e.g. emotion overgeneralization, babyfaceness; see Zebrowitz, 2017; Zebrowitz & Montepare, 2008).

In Studies 4 and 5, we borrow the logic of previous facial impression studies with adult faces (Sutherland, Liu, et al., 2018; Sutherland et al., 2013; Walker & Vetter, 2009, 2016) and use image manipulation techniques to validate the dimensions underlying children's faces. In Study 4, we investigate whether adults are sensitive to subtle facial manipulations that enhance/reduce the saliency of the dimensions. In Study 5, we check that the dimensions are dissociable (following Walker & Vetter, 2016). Specifically, we check that adults can distinguish between the dimensions even though they may be correlated. We

also build a validated set of stimuli that systematically vary along the dimensions found important for children's faces, which we intend to make available for use by other researchers.

In Study 6, we ask whether physically varying children's faces along the dimensions influences adults' hypothetical behavior towards and expectations of children. These facial impressions are likely to have real world consequences for children's lives, just as has been found for other traits such as attractiveness and babyfaceness (e.g. Berkowitz & Frodi, 1979; Langlois et al., 2000; Zebrowitz et al., 1991; Zebrowitz & Lee, 1999). Here, we investigate the potential social consequences of evaluating children on the dimensions we find to be theoretically important and that are responsible for most of the variance in spontaneous impressions. Although we examine behavioral expectations, we see our study as providing a foundation that will structure future endeavors into the behavioral consequences of first impressions.

Finally, in Study 7 we provide a strict test of the robustness of the dimensions by examining whether they generalize to a larger and more heterogeneous, independent sample of children's faces. Importantly, these images are "ambient" images, and vary highly in image properties, pose and expression, reflecting the variable conditions under which we see faces in everyday life. Ambient images are now widely used in face perception research, complementing the traditional focus on highly controlled and homogeneous images (e.g., Jenkins, White, Van Montfort, & Burton, 2011; Sutherland, Young & Rhodes, 2017).

It is important to note that here, we test Caucasian adults' impressions of Caucasian children's faces, to avoid any potential *other-race effects*. The classic other-race effect refers to the well-established phenomenon that individuals are poorer at recognizing other-race than own-race faces (for reviews; Anzures, Quinn, Pascalis, Slater, & Lee, 2013; Meissner & Brigham, 2001). Critically, other-race effects extend beyond recognition to a range of social

information extracted from a face, including judgments of sex (O'Toole, Peterson, & Deffenbacher, 1996), emotional expression (Elfenbein & Ambady, 2002) and gaze discrimination (Collova et al., 2017). We were concerned that these other-race effects might also impact the quality of impressions adults made of own- vs. other-race faces. To avoid this possibility, we limit our study to Caucasian participants and faces (here and throughout). We return to this point in the General Discussion.

We also note, that here, we specifically look at adults' impressions of *young* children's faces (i.e. between the ages of around 4 – 6 years old). We chose a younger age range as it is possible that older children, who might be approaching puberty, are associated with different social goals. In a functional theoretical account, these differences could drive different dimensions for such faces. We also return to this point in the General Discussion.

STUDIES 1 AND 2: Developing a Dimensional Model for Children's Faces

Study 1: Free descriptions

The purpose of Study 1 was to determine which trait descriptors best capture adults' impressions of young children's faces for the first time, with the intention to use these trait judgments in Study 2 to form a new model of first impressions for children's faces. We therefore collected unconstrained descriptions of children's faces from adult participants. This step was important in our data-driven approach because we wanted to make sure our model was based on trait words that adults spontaneously use to describe young children's faces.

Method

Participants

We recruited 84 Caucasian adult participants (42 female, $M = 36.7$, $SD = 11.7$, range = 19 - 65 years) from the online Amazon Mechanical Turk (M-Turk) American population. An additional 61 participants took part, but were excluded as they did not self-identify as Caucasian ($N = 52$), were not in America during the time of testing ($N = 2$), did not complete the entire task ($N = 5$), or did not understand the task instructions (e.g. entered the same response for every face: $N = 2$).

Stimuli and Materials

Photographs of young, Caucasian children's faces were obtained from The Child Affective Facial Expression Set (CAFE; LoBue & Thrasher, 2014; LoBue & Thrasher, 2015). We included faces of all children aged between 4-6 years (inclusive) who were Caucasian, had direct gaze and a neutral expression (to ensure equal numbers of male and female faces, we excluded one additional female face from the final set). This method left us with a total of 56 children's faces (28 female, $M = 5$ years 4 months, $SD = 8$ months)[1]. There was no significant difference in age between the female ($M = 5$ years 5 months, $SD = 8$ months) and male ($M = 5$ years, 4 months, $SD = 7$ months) faces in the final set, $t(54) = 0.67$, $p = .506$, Cohen's $d = .178$. (See Appendix for a list of facial IDs included).

The children photographed appear in frontal view, against an off-white background and with off-white sheets covering their clothes. Very few (five) of the faces had adornments (earrings, headband or hairclips). Comparable to the adult stimuli used originally by Oosterhof and Todorov (2008; cf. Lundqvist, Flykt, & Öhman, 1998), images were colored,

[1] Note that the CAFE database does contain additional children's faces outside of this age range. We chose 4-6 year old faces as we wanted to investigate adults' impressions of *young* children's faces, not children near puberty (as different social goals are associated with pubescent individuals and this difference could drive different dimensions for such faces) and we wanted to minimize the age range of faces.

taken under controlled conditions (i.e. lighting, camera used) and did not have their hair masked. Photographs were cropped to a standard size (419 x 419 pixels, dpi = 240, approximately 4.5 x 4.5cm in size on screen) but were otherwise unmodified.

Procedure

The 56 children's faces were pseudo-randomly split into 7 groups of 8 faces, balanced for sex. Each group contained at least one face of each age (4, 5 and 6 years). Participants were randomly assigned to each face group, with 10-15 participants in each group. Participants were prompted to, "Write down your first impressions. Write whatever comes to mind" for each of the child face images (following Sutherland, Liu, et al., 2018). Participants typed their responses into a blank text box presented under each photograph. Participants were informed the task was about first impressions, and that they should move on to the next face when they felt they were no longer being spontaneous, so that responses reflected genuine first impressions (following Oosterhof & Todorov, 2008; Sutherland, Liu, et al., 2018). The anonymous nature of the experiment was emphasized, and honesty was encouraged.

The experiment started with two practice trials (one 7-year old female and male face, from the CAFE database) to familiarize participants with the procedure. Responses to the practice faces were not analyzed. Participants then responded to the eight experimental faces presented in a random order, different for each participant.

We wanted to make sure our sample was representative of both parents and non-parents, as we wanted to check that parental status did not affect our results. Therefore, after the face description task, participants answered a series of additional demographic questions to assess their contact with and attitudes towards children (See Supplementary Materials page 1), and questions regarding age, sex and ethnicity.

The experiment was administered online via Qualtrics (2009), and M-Turk participants completed the tasks on a computer at a time and location convenient to them. The experiment took ~10min to complete, for which participants received a small honorarium (US 40c).

Results and Discussion

Our data consisted of participants' descriptions of the children's faces, which we split into words or short phrases. In total, participants provided 2,075 descriptions, mostly falling into the categories of trait (e.g. 'bratty'), age (e.g. 'kid'), sex (e.g. 'girl'), emotional state (e.g. 'sad'), and appearance-based statements (e.g. 'chubby cheeks'). Two researchers independently classified the descriptions into each of these categories, with a third rater resolving any disagreement. Cohen's κ revealed strong agreement between the two researchers within all categories (trait; $\kappa = .792$, $p < .001$, age; $\kappa = .949$, $p < .001$, sex; $\kappa = .950$, $p < .001$, emotion; $\kappa = .880$, $p < .001$, appearance; $\kappa = .891$, $p < .001$). Based on this classification, descriptions most commonly consisted of statements regarding traits (28%), confirming that trait impressions are spontaneously inferred. Descriptions were also classified by appearance (20%), emotions (15%), age (7%), and sex (5%).

Trait descriptors

We did not predict any participant sex differences given previous results for dimensions of facial impressions for adult faces (South Palomares, Sutherland, & Young, 2017; Sutherland, Young, Mootz, & Oldmeadow, 2015). However, it is possible that males and females have different caring stances towards children (Paquette, 2004), so to be conservative we performed initial analyses separate for each sex of participant. Overall, female and male participants mentioned 260 and 227 traits, respectively (excluding repetition of the same trait by the same participant). To ensure that the frequencies of important concepts were not underestimated, traits with the same root word were combined (e.g.

"intelligence" and "intelligent") and two raters independently sorted these traits into categories of similar concepts (e.g. "smart" "intelligent" and "slow" into one category, defined by the most frequently mentioned trait word). A third rater resolved disagreements between trait classifications (following Oosterhof and Todorov 2008's approach).

Twenty trait categories accounted for 78% and 93% of the total traits mentioned by female and males, respectively (Table 1). We calculated the frequency with which unique female and male participants mentioned traits within each of the categories (Table 1). Two trait categories (imaginative and selfish) contained traits that were mentioned by less than 10% of either male or female participants, and therefore were excluded from further analyses. This procedure left us with 18 trait categories (trouble-maker, nice, intelligent, shy, sweet, energetic, funny, curious, tough, serious, innocent, quiet, kind, friendly, timid, cheerful, confident, and lonely) to submit to the factor analysis in Study 2.

Table 1.

Frequency with which unique participants mentioned traits captured by each trait category.

Trait Category	Female Frequency	Male Frequency	Total Frequency
Trouble-maker	12	21	33
Nice	12	13	25
Intelligent	12	9	21
Shy	11	9	20
Sweet	11	8	19
Energetic	7	11	18
Funny	8	10	18
Curious	7	10	17
Tough	10	7	17
Serious	11	5	16
Innocent	5	10	15
Quiet	8	6	14
Kind	10	4	14
Friendly	7	6	13
Timid	6	7	13
Cheerful	7	4	11
Confident	8	3	11
Lonely	7	3	10
Imaginative*	3	2	5
Selfish*	0	1	1

* Less than 10% of either female or male participants mentioned traits within this category. We omitted these categories from further studies.

To check that these 18 categories accurately captured our data and to be as objective as possible, we also took a frequency-based approach by calculating the frequency with which different participants mentioned the same root *trait word* (following the approach taken by Sutherland, Liu, et al., 2018). Reassuringly, this process confirmed that our initial

process had captured all of the variance in impressions already (See Table S1 and page 1 in the Supplementary Materials).

To be conservative, we also looked for any participant sex differences in the traits used to describe the children, but found no differences (See Supplementary Materials, page 26). We also did not find any differences based on the amount of contact participants had with children (See Supplementary Materials page 26-27). We therefore created a word cloud (wordle.net) to visualize the spontaneous trait descriptions based on data from all participants. This word cloud depicts the frequency of trait attributions mentioned by unique participants, with larger fonts representing more frequent descriptors (Figure 1). Note that only words or short phrases (< 5 words) were included, so that the word cloud captured the main concepts of our dataset, rather than unrepresentative long phrases (following Sutherland, Liu, et al., 2018; Sutherland et al., 2015).

Figure 1. Word clouds depicting adults' spontaneous trait impressions made to young children's faces. Larger fonts represent the traits more frequently mentioned by adults.

STUDY 2: Factor Analysis of Face Trait Ratings

In Study 1, we found that 18 trait categories (trouble-maker, nice, intelligent, shy, sweet, energetic, funny, curious, tough, serious, innocent, quiet, kind, friendly, timid, cheerful, confident and lonely) best captured adults' unconstrained impressions of young, children's faces. In Study 2, we submitted ratings of these 18 traits to a factor analysis to build a model of the dimensions underlying impressions of children's faces. We included "attractive" in this model as a 19th category, because the most frequently mentioned appearance descriptors were indeed attractiveness judgments (e.g. 'cute', 'beautiful', 'pretty eyes'; See Supplementary Materials Table S2: note that a factor analysis excluding attractiveness judgments yielded almost identical results; Supplementary Materials Table S12). Moreover, including attractiveness as a trait paralleled the method of leading adult face models (Oosterhof & Todorov, 2008; Sutherland et al., 2013). Including 19 attributes should allow enough variables to test models up to four dimensions (Fabrigar, Wegener, MacCallum, & Strahan, 1999), and is comparable in scope to Oosterhof and Todorov's (2008) original study.

Method

Participants

We recruited a new group of 570 adult Caucasian participants from M-Turk (285 females, M = 38.4, SD =13.02, range = 18 – 74 years). Thirty adults (15 female) rated each of the 19 attributes (total N = 570). As Study 1, we excluded an additional 311 participants (182 non-Caucasian / not born in a majority Caucasian country, 17 not in America during the time of testing, 10 who reported we should not use their data, 33 who gave the same trait rating for at least an entire block, and 69 who did not pass the attention check). The decision to exclude these participants was made *a priori*, and therefore their data were never analyzed.

Stimuli and Procedure

The same images of young Caucasian children's faces were used as in Study 1 (images were also the same size as in Study 1). Each participant rated all 56 children's faces on one of the 19 attributes derived from Study 1, randomly assigned. For each face, participants were asked to rate, "How [trait] is this child?" on a scale of 1 (Not at all [trait]) to 9 (Extremely [trait]). Participants were encouraged to respond honestly and spontaneously, and were reminded that there were no right or wrong answers. Each image was presented in the center of the screen, below the question and above the scale. Stimuli remained on screen until participants responded. Each face was presented three times, with each repetition in a separate block in which face order was randomized. The task started with the same two practice trials as those used in Study 1.

Participants also provided demographic information and answered questions assessing their contact with and attitudes towards children, as in Study 1. As a final step, participants completed an attention check, where they were asked to report the trait they had rated. To pass, typed responses needed to match the trait word exactly. The experiment took participants ~15min, for which participants were reimbursed (US 60c).

Data Analysis

To determine the dimensional structure for adults' impressions of children's faces, we performed a factor analysis (following Sutherland, Liu, et al., 2018; Sutherland et al., 2013). As an initial step, we checked intra-rater and inter-rater reliability separately for female and male participants. We examined intra-rater reliability by correlating each participant's trait ratings across the three blocks, for each trait. On average, there was a large correlation between all block combinations for both participant sexes (all rs > .49, ps < .001). As intra-

rater reliability was good, we averaged trait ratings across the three blocks for each participant. We then used Cronbach's alpha to examine the reliability of the overall ratings, i.e. averaged across the group. This method is commonly used to capture reliability in the person perception literature when group-based judgments are used (e.g. Oosterhof & Todorov, 2013 Sutherland et al., 2013). Inter-rater reliability was good (>.7 Nunnally, 1978) for all trait judgments made by both the female and male participants (except one trait rated by male participants; see Supplementary Materials Table S3; this trait was reliable when data were combined across sex). We note that we intentionally left male and female faces in the same model as this allowed us to understand whether sex was a cue to any dimension that emerged (following Oosterhof & Todorov, 2008; Sutherland et al., 2013; Walker & Vetter, 2009). Importantly, this method also mirrored that of Oosterhof and Todorov (2008), (as well as adult face literature, more broadly; e.g. Sutherland et al., 2013; Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013; Walker & Vetter, 2009, 2016), allowing us to more directly compare our child dimensions with this previous literature.

To determine the dimensional structure, we performed a factor analysis on the average ratings of the 19 traits, with oblique rotation. There were no obvious participant sex differences, so we ran an initial factor analysis on the data as a whole (see Supplementary Materials Table S6 for separate analyses). The data were suitable for factor analysis (almost all correlation coefficients > .3: see Table 2; Bartlett's test of sphericity < .001; Kaiser-Meyer-Olkin > .6: Kaiser, 1970; Kaiser & Rice, 1974) and our sample size of faces was found to be adequate for factor analysis (see Mundfrom, Shaw, & Ke, 2005; See Supplementary Materials Table S4 for communalities).

To determine the number of factors to extract, we examined Kaiser criterion (eigenvalue < 1; Kaiser, 1960), the scree plot, and parallel analysis (Horn, 1965). Kaiser's criterion is commonly used in face perception studies (e.g. Oosterhof & Todorov, 2008; Walker &

Vetter, 2009), the scree test gives information about additional factors, and parallel analysis gives a stricter test of factor genuineness as it formally tests the likelihood that the factors extracted are due to chance (Fabrigar & Wegener, 2011). All three sources supported the extraction of two factors (see Supplementary Materials Table S5). Therefore, we extracted a two-dimensional model. Note, data for this study and all following studies is available as supplementary material.

Table 2.

Correlational matrix: inter correlations for the trait variables (and estimated 95% CI's based on standardized beta weights) from Study 2.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
1. Attractive	-																		
2. Cheerful	.34* (.09, .60)	-																	
3. Confident	.24** (-.02, .51)	.76** (.58, .94)	-																
4. Curious	.49** (.25, .73)	.81* (.66, .97)*	.44** (.20, .69)	-															
5. Energetic	.33** (.08, .59)	.89** (.76, 1.0)	.77** (.59, .94)	.78** (.61, .95)	-														
6. Friendly	.46** (.22, .70)	.93** (.83, 1.0)	.54** (.31, .77)	.87** (.74, 1.0)	.76** (.58, .93)	-													
7. Funny	.25* (-.15, .51)	.82** (.66, .98)	.52** (.29, .75)	.82** (.66, .97)	.84** (.69, .99)	.76** (.59, .94)	-												
8. Innocent	.54** (.30, .77)	.71** (.51, .90)	.21 (-.06, .47)	.81** (.65, .97)	.53** (.30, .76)	.87** (.74, 1.0)	.64** (.43, .85)	-											
9. Intelligent	.74** (.55, .92)	.64** (.43, .85)	.35* (.10, .61)	.73** (.54, .92)	.50** (.26, .73)	.78** (.61, .95)	.48** (.24, .72)	.79** (.62, .95)	-										
10. Kind	.53** (.29, .76)	.84** (.70, .99)	.39* (.14, .64)	.89** (.76, 1.0)	.70** (.51, .90)	.95** (.87, 1.0)	.75** (.57, .93)	.93** (.83, 1.0)	.81** (.64, .97)	-									
11. Lonely	-.41** (-.66, -.16)	-.90** (-1.0, -.78)	-.85** (-.99, -.70)	-.74** (-.92, -.56)	-.91** (-1.0, -.79)	-.79** (-.96, -.62)	-.77** (-.95, -.60)	-.57** (-.79, -.34)	-.61** (-.83, -.40)	-.70** (-.90, -.51)	-								
12. Nice	.56** (.34, .79)	.80** (.64, .97)	.33* (.07, .59)	.85** (.71, 1.0)	.63** (.42, .85)	.94** (.85, 1.0)	.69** (.50, .89)	.95** (.86, 1.0)	.81** (.65, .97)	.96** (.88, 1.0)	-.64** (-.85, -.43)	-							
13. Quiet	.25* (-.02, .51)	-.35* (-.60, -.09)	-.67** (-.87, -.47)	-.03 (-.30, .24)	-.51** (-.74, -.27)	-.05 (-.40, .23)	-.30* (-.56, -.04)	.27* (.01, .54)	.22 (-.05, .49)	.08 (-.20, .25)	.50** (.26, .74)	.20 (-.07, .47)	-						
14. Serious	-.30* (-.56, -.03)	-.98** (-1.0, -.92)	-.71** (-.90, -.51)	-.82** (-.98, -.67)	-.86** (-1.0, -.76)	-.92** (-1.0, -.81)	-.82** (-.98, -.67)	-.72** (-.91, -.53)	-.61** (-.83, -.39)	-.84** (-.99, -.70)	.88** (.75, 1.0)	-.80** (-.96, -.64)	.34* (.09, .60)	-					
15. Shy	-.09 (-.36, .18)	-.69** (-.89, -.49)	-.94** (-1.0, -.84)	-.36* (-.61, -.10)	-.76** (-.93, -.58)	-.44** (-.69, -.20)	-.52** (-.75, -.28)	-.07 (-.35, .20)	-.20 (-.47, .07)	-.28* (-.55, -.02)	.80** (.64, .97)	-.21 (-.47, .06)	.78** (.60, .95)	.65** (.44, .86)	-				
16. Sweet	.71** (.52, .90)	.58** (.36, .80)	.16 (-.11, .43)	.75** (.57, .93)	.46** (.21, .70)	.76** (.59, .94)	.54** (.31, .77)	.88** (.74, 1.0)	.77** (.59, .94)	.85** (.70, .99)	-.46** (-.70, -.22)	.88** (.75, 1.0)	.36* (.10, .61)	-.57** (-.79, .35)	-.01 (-.28, .27)	-			
17. Timid	.12 (-.16, .39)	-.40** (-.65, -.15)	-.81** (-.97, -.65)	-.03 (-.30, .24)	-.48** (-.72, -.25)	-.11 (-.38, .16)	-.21 (-.47, .06)	.28* (.02, .54)	.09 (-.18, .36)	.05 (-.22, .33)	.53** (.30, .76)	.14 (-.13, .41)	.82** (.66, .98)	.36* (.10, .61)	.90** (.78, 1.0)	.31* (.05, .57)	-		
18. Tough	-.51** (-.74, -.27)	-.62** (-.83, -.41)	-.08 (-.35, .20)	-.81** (-.97, -.64)	-.47** (-.71, -.23)	-.82** (-.98, -.66)	-.63** (-.84, -.41)	-.96** (-1.0, -.87)	-.76** (-.94, -.59)	-.90** (-1.0, -.78)	.47** (.23, .71)	-.92** (-1.0, -.82)	-.36* (-.61, -.11)	.64** (.43, .85)	-.07 (-.34, .21)	-.88** (-1.0, -.75)	-.42** (-.67, -.17)	-	
19. Trouble	-.54** (-.77, -.33)	-.47** (-.71, -.23)	.04 (-23, .31)	.67** (-.87, -.47)	-.27* (-.53, -.01)	-.73** (-.91, -.54)	-.41** (-.66, -.16)	-.92** (-1.0, -.81)	-.78** (-.95, -.61)	-.81** (-.97, -.65)	.33* (.07, .59)	-.87** (-1.0, -.73)	.53** (-.76, -.30)	.49** (.25, .73)	-.19 (-.46, .08)	-.86** (-1.0, -.73)	-.49** (-.73, -.26)	.94** (.84, 1.0)	-

* $p < .05$, ** $p < .001$

Results and Discussion

To evaluate our two-dimensional model, we interpreted the structure matrix to examine trait loadings above .4 (as suggested by Field, 2009; see Table 3). Positive judgments loaded positively (e.g. nice, kind, innocent, intelligent), and negative judgments loaded negatively (e.g. trouble-maker, tough) onto the first dimension. Hence, the first child face dimension can be characterized as *niceness*, because judgments of this trait loaded most strongly onto this dimension.

The second dimension had high loadings from shy and confident judgments. The second dimension for children's faces can be characterized as *shyness*, because judgments of this trait loaded most strongly on this dimension. The two dimensions were not correlated highly (r = -.17).

Table 3.

Principal axis factor analysis on the 19 traits for children's faces with oblique rotation, as derived from the structure matrix. Numbers are factor loadings (and can be interpreted as correlations between the dimensions and original trait variables). The dotted line divides the traits which load most onto each dimension. Factor loadings > .4 are in bold.

Trait	Dimension 1: Niceness	Dimension 2: Shyness
Nice	**.98**	-.18
Kind	**.97**	-.28
Innocent	**.96**	-.05
Tough	**-.95**	-.07
Friendly	**.93**	**-.43**
Sweet	**.90**	.04
Curious	**.89**	-.37
Trouble-maker	**-.88**	-.24
Intelligent	**.84**	-.14
Cheerful	**.79**	**-.70**
Serious	**-.79**	**.67**
Funny	**.70**	**-.55**
Attractive	**.58**	-.02
Shy	-.19	**.98**
Confident	.33	**-.91**
Timid	.17	**.88**
Quiet	.20	**.83**
Lonely	**-.67**	**.80**
Energetic	**.65**	**-.78**
Variance explained: equivalent varimax PCA	**53.5%**	**33.5%**

Model robustness

We performed several additional analyses to determine the robustness of the original two-dimensional model. To be conservative, we performed separate factor analyses for female and male faces, and female and male participants (to check for potential sex differences) as well as on parent and non-parent participants (in case parents made more complex impressions requiring additional dimensions). The same two dimensions (niceness and shyness) emerged in all of these samples (See Supplementary Materials Table S6, S7, S8, S9, S11 and page 28 for additional information). In some of these samples, a third dimension (attractiveness-intelligence) emerged (see Supplementary Materials Table S7). For these samples, we ran parallel analyses to test whether the third dimension existed beyond chance level. Parallel analyses supported the two-dimensional model over a three-dimensional model, in all samples (see Supplementary Materials Table S5, S10). The same two dimensions also emerged given a principal components analysis (PCA) with varimax rotation (as used by Oosterhof and Todorov, 2008; see Supplementary Materials Table S9).

In summary, we built the first dimensional model of facial impressions of children. We found strong evidence that two dimensions, niceness and shyness, underlie adults' impressions of young children's faces. Similar dimensions consistently emerged irrespective of face sex, participant sex, whether or not participants were parents, and regardless of the statistical techniques used.

STUDY 3: Comparing the Child Face Dimensions to Models in Social Psychology and Examining Potential Facial Cues

The purpose of Study 3 was two-fold. First, we wanted to examine how the niceness and shyness dimensions found here, relate to two other important models in the literature. By directly relating our new model to past theories, we contribute to theoretical knowledge about

these social dimensions, more broadly. Foremost, we compared the dimensions we found for young children's faces to the dimensions for adult faces, trustworthiness and dominance (Oosterhof & Todorov, 2008), and to the traits most closely associated with the adult face dimensions (trustworthiness: trustworthy, responsible, emotionally stable, likeable; dominance: dominant, aggressive, competent, and facial maturity/babyfaceness; Oosterhof & Todorov, 2008; Rule & Ambady, 2008). This comparison is important because we wanted to rule out the possibility that adults simply use distinct language to describe child and adult faces, which nevertheless indexes the same conceptual dimensions. We also compared the child-face dimensions to dimensions found more broadly in social cognition, i.e. dimensions important in the social stereotype literature (e.g. warmth, competence and sociability; Fiske et al., 2007; Landy, Piazza, & Goodwin, 2016; Leach, Ellemers, & Barreto, 2007).

Second, we examined the physical facial cues underlying these dimensions. Past research has focused on emotional expression (Montepare & Dobish, 2003; Oosterhof & Todorov, 2008; Said, Haxby, & Todorov, 2011), health (Boothroyd, Jones, Burt, & Perrett, 2007), masculinity (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Walker & Wänke, 2017), babyfaceness (Montepare & Zebrowitz, 1998; Zebrowitz & Montepare, 2008) and attractiveness (Zebrowitz, et al., 2002; Zebrowitz, et al., 1996) as facial cues for trait impressions in adult faces (see Zebrowitz, 2017). Here, we assessed how these facial cues are related to the dimensions found for children's faces. To this end, we collected judgments of emotion (happy, sad, angry, fearful, and disgust), perceived health (health and weight), masculinity, babyfaceness and attractiveness. We also measured facial width to height ratio (fWHR), which has been negatively linked to shy behavior in children (Arcus & Kagan, 1995; Zebrowitz, Franklin, & Boshyan, 2015) and to impressions of dominance in adult faces (see Geniole, Denson, Dixson, Carré, & McCormick, 2015).

Method

Participants

A total of 324 Caucasian adult participants were recruited from M-Turk (181 females, 171 parents, $M = 37.1$, $SD = 12.0$, range = 19 – 86 years). An additional 212 participants took part in the task but were excluded as before (18 non-Caucasian, 122 did not pass the attention check, 12 not in America during time of testing, 13 repeated same rating for at least one whole block, 1 did not understand the task, 29 already completed one of our previous studies, and 17 who had technical issues / said not use their data). Participants were recruited until a minimum of 10 participants (i.e. five of each sex) had rated each trait or facial cue (see Supplementary Materials Table S13). This sample size has been found sufficient for good reliability at the group level, under similar experimental conditions (e.g. Sutherland et al., 2013; Sutherland et al., 2016).

Stimuli and Procedure

Participants were randomly assigned to rate the 56 Caucasian children's faces (same as in Study 2) on one of 10 trait attributes, important in social psychology literature (i.e. trustworthy, responsible, emotionally stable, likeable, dominant, aggressive, valence, warm, competent, or sociable) or one of 10 facial cues (i.e. happy, sad, angry, fearful, disgust, health, weight, masculinity, babyfaceness, or attractiveness). The same procedure was followed as in Study 2. Participants rated how [attribute] each child looked, on a 9-point scale. To evaluate valence, participants were asked how positive their impression of each child was and responded on a scale ranging from very negative to very positive (Sutherland et al., 2015). Judgments of babyfaceness were made on a scale of mature-faced to baby-faced (Zebrowitz et al., 1993; Zebrowitz et al., 1996), judgments of weight were made on a scale of underweight to overweight (Coetzee, Perrett, & Stephen, 2009), and judgments of femininity-

masculinity were made on a scale of feminine to masculine (Oosterhof & Todorov, 2008; Sutherland et al., 2013). All other attribute judgments were made on a scale ranging from not at all [attribute] to extremely [attribute]. For emotion ratings, participants were informed that all faces were emotionally neutral but could show subtle variations in emotional information (Caulfield, Ewing, Burton, Avard, & Rhodes, 2014) and responded on a scale of not at all [emotion] to moderately [emotion].

Following the ratings, participants answered demographic questions and questions regarding their contact with and attitudes towards children, as in Study 1 and 2. Participants were reimbursed for their time (US 60c).

In addition to these trait ratings, we measured fWHR, calculated as facial width (i.e. the maximum horizontal distance between the left and right face border) divided by height (i.e. vertical distance between the top of the upper eyelid and the top of the centre upper lip: (Carré & McCormick, 2008; Lewis, Lefevre, & Bates, 2012).

Data-Analysis

To investigate how the two dimensions underlying impressions of children's faces compare to other important models and facial cues, we calculated factor scores on the niceness and shyness dimension for each face (from Study 2) using the regression method (factor scores were weighted by trait loadings). We then correlated these scores with the ratings collected/measured here.

Given that there were no apparent sex differences in impression formation of children's faces (Study 1, 2), we analyzed data collapsed across participant sex (here and in all following studies). Intra-rater reliability was calculated and found adequate (see Supplementary Materials Table S13 for correlations), as in Study 2. Therefore, we averaged

across face ratings and calculated inter-rater reliability (which was also adequate; See Supplementary Materials Table S13 for Cronbach's alpha).

Results and Discussion

First, we correlated the factor scores for the child face dimensions to 1) important models in literature and 2) potential facial cues. Correlations are shown in Table 4 and visually represented in the heat maps shown in Figure 2 and 3 (see Supplementary Materials Table S14 for inter-correlations between all variables collected here).

1) Important models in literature:

Facial impressions models (trustworthiness and dominance)

There was a large, positive correlation between judgments of trustworthiness and the niceness dimension for children's faces (r = .947, p <.001). Correlations were also large between trustworthiness and the individual traits that loaded most onto the niceness dimension (See Supplementary Materials Table S15). There was also a large, positive correlation between valence (positivity) ratings and niceness for children's faces (r = .884, p < .001). Therefore, the first dimension for children's faces is highly similar to that reported previously for adult faces (Oosterhof & Todorov, 2008; Sutherland et al., 2015). However, the niceness dimension for children's faces also correlated highly with judgments of dominance (r = -.868, p < .001). This correlation highlights a noteworthy difference between the niceness dimension for children's faces (which has a strong negative correlation with dominance), and the trustworthiness dimension for adults' faces (which does not correlate highly with dominance; Oosterhof & Todorov, 2008).

If the second dimension for the adult (dominance) and child (shyness) models are similar, we would expect a large, negative correlation between these variables. In contrast,

the correlation between dominance judgments and the shyness dimension was not significant (r = -.239, p = .076; also see correlations between dominance and individual traits that loaded most onto the shyness dimension, Supplementary Materials Table S15). This result does not provide clear support for an association between dominance and shyness. However, it remains possible that dominance and shyness may both be tapping into the same general dimension.

Considering the above, we therefore wanted to more robustly test the difference between the child and adult face dimensions. First, we ruled out the possibility that the dominance dimension did not emerge because important judgments comprising this dimension were not measured. To do so, we ran a factor analysis including ratings of the traits spontaneously used to describe children faces (from Study 2: N = 19) and traits most relevant to the dominant and trustworthiness adult face dimensions (i.e. trustworthiness: trustworthy, responsible, emotionally stable and likeable; dominant: dominant, aggressive, competent and facial maturity/babyfaceness; Oosterhof & Todorov, 2008; Rule & Ambady, 2008). Two dimensions emerged that were almost identical to those from the original factor analysis from Study 2 (niceness and shyness; for analysis and additional information, see Supplementary Materials page 16-17). Critically, all of the variables associated with the dominance dimension for adult faces (dominant, aggressive, competent, and low babyfaceness) loaded substantially onto the niceness, rather than the shyness, dimension (see Supplementary Materials, Table S17). Thus, shyness appears to be an independent dimension for children's faces, that is dissociable from the dominance dimension found for adult faces. We also ran a PCA (with varimax rotation, as Oosterhof & Todorov, 2008) including all 19 trait ratings from Study 2 and the 10 trait ratings from Study 3, and again found the same two dimensions (see Supplementary Materials Table S18).

In summary, although adults showed consensus in dominance judgments of children's faces (i.e. they agreed with each other regarding which children looked dominant; Cronbach's alpha = .88), these judgments did not best capture the second dimension for children's faces. This result is striking, because for the first time it indicates that dominance does not represent a substantive dimension for children's faces.

Social stereotyping models (warmth, competence, and sociability)

We found a large, positive correlation between warmth judgments and the niceness dimension for children's faces (r = .946, p < .001). This correlation mirrors the relationship between warmth and the adult face dimension of trustworthiness (Sutherland et al., 2016; also see Supplementary Materials Table S16 for correlations of the individual traits which loaded most onto the niceness dimension). There was a medium, negative correlation between competence and the shyness dimension (r = -.371, p = .005), but this correlation was significantly smaller than the correlation between competence and the niceness dimension for children's faces (r = .712, p < .001). Together, these results suggest that the niceness dimension for children's faces taps into a concept similar to the warmth dimension in social stereotypes (Fiske et al., 2007), and to a lesser degree, the shyness dimension relates to the competence dimension (Fiske et al., 2007).

Recently, research has argued that the warmth dimension actually conflates two distinct variables: morality (i.e. trustworthiness) and sociability (Bauman & Skitka, 2012; Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Goodwin, Piazza, & Rozin, 2014; Landy et al., 2016; Leach et al., 2007). Here, we found that sociability correlated with both the niceness (r = .763, p <.001) and shyness (r = -.696, p <.001) dimensions, whereas trustworthiness (cf. morality) only correlated with niceness (r = .947, p <.001). This difference between sociability and trustworthiness suggests they may indeed be independent

concepts (cf. Bauman & Skitka, 2012; Brambilla et al., 2011; Goodwin et al., 2014; Landy et al., 2016; Leach et al., 2007).

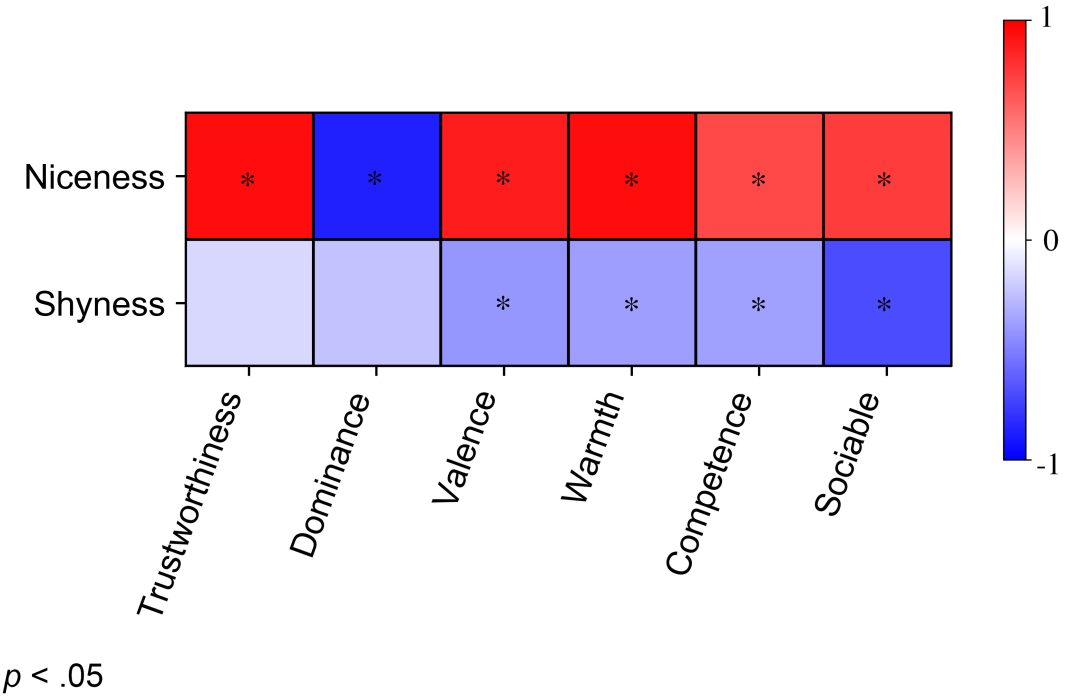

Figure 2. Heat maps representing the correlations between the factor scores for each dimension (niceness and shyness) and variables important in theoretical models describing perceptions of adults, from across social psychology. Pearson's $r = 0$ (white), 1 (red), or -1 (blue). *$p < .05$ (**To be printed in color**).

2) Potential Facial Cues

Emotion judgments were an important cue for both the niceness and shyness dimensions in opposite directions, but contributed absolutely more to the niceness than the shyness dimension (except for judgments of fear which correlated more with shyness: See Table 4 for correlations and Figure 3 for heat map). Babyfaceness was a cue for niceness, but did not correlate with the shyness dimension. Masculinity correlated negatively with both the

niceness and shyness dimensions. Cues of perceived health correlated positively with niceness and negatively with shyness. Over-weight judgments correlated negatively with both niceness and shyness. FWHR correlated negatively with shyness, but did not correlate with niceness. Finally, attractiveness correlated positively with niceness but did not correlate with shyness (See Table 4 and Figure 3).

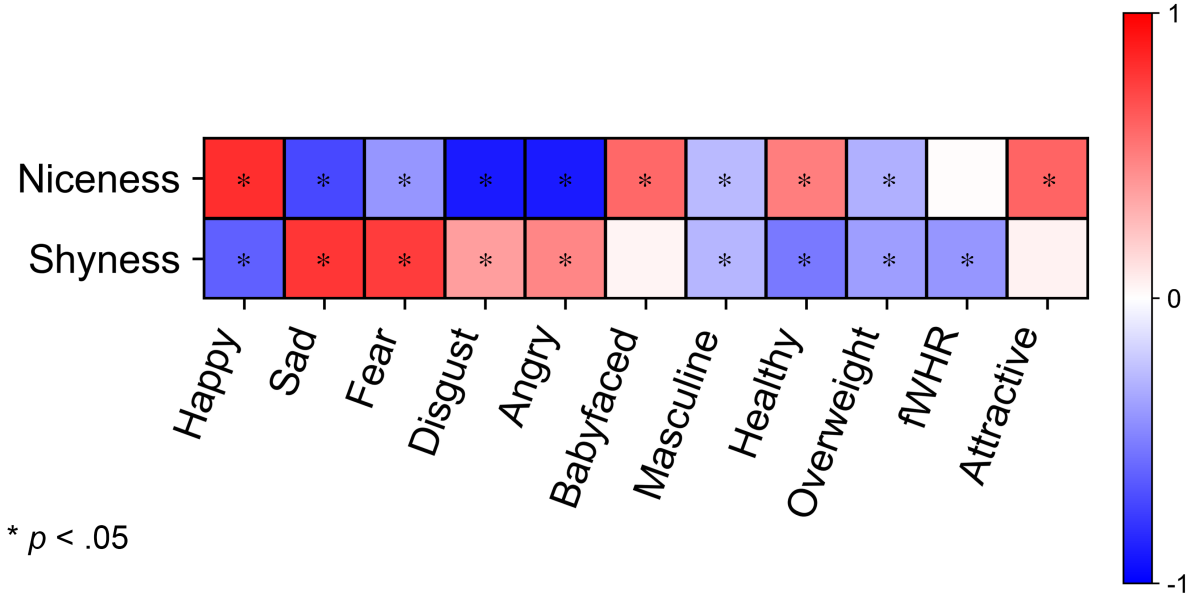

Figure 3. Heat maps representing the correlations between the factor scores for the dimensions (niceness and shyness) and potential facial cues. Pearson's $r = 0$ (white), 1 (red), or -1 (blue). *$p < .05$ (**To be printed in color**).

In summary, converging evidence from Study 3 suggests that the niceness dimension for children's faces is fairly comparable, but not identical, with the trustworthiness dimension for adult faces. Niceness correlated highly with judgments of trustworthiness, warmth and valance, just as has been found for the adult face dimension of trustworthiness (Sutherland et al., 2016; Sutherland et al., 2015). Furthermore, adults draw on similar facial cues in child and adult faces to form these impressions. For example, niceness correlated positively with the emotional expression of happiness, and negatively with disgust and anger, just as has

been found for adult faces (e.g. Oosterhof & Todorov, 2008; Said et al., 2009). However, important differences do exist between the niceness and trustworthiness dimensions. For example, here we found a strong negative correlation between dominance ratings and the niceness dimension for children's faces, but dominance ratings do not substantially correlate with the trustworthiness dimension for adults' faces (Oosterhof & Todorov, 2008). Nevertheless, niceness and trustworthiness are similar to the extent they can both be conceptualized as a valence dimension.

The shyness dimension for children's faces was not directly comparable to the dominance dimension for adult faces. Judgments of dominance for children's faces did not correlate with shyness judgments. Moreover, dominance ratings did not load onto the shyness dimension, even when variables closely related to the dominance dimensions were included in a factor analysis. In addition, the relationship between facial cues and the shyness dimension found here, was not the same as the relationship between cues and the dominance dimension found previously for adult faces (Said et al., 2009; Oosterhof & Todorov, 2008). For example, for adult faces dominance correlates negatively with babyfaceness (Montepare & Zebrowitz, 1998; Oosterhof & Todorov, 2008), but here the shyness dimension did not correlate with babyfaceness. Furthermore, for adult faces dominance correlates positively with angry (Oosterhof & Todorov, 2008; Said et al., 2009; Sutherland et al., 2016) and negatively with happy (Said et al., 2009; Zebrowitz, Kikuchi & Fellous, 2007) expressions. If shyness is simply the opposite of dominance, then shyness should correlate negatively with angry and positively with happy expressions. However, shyness correlated positively with angry and negatively with happy expressions.

Nevertheless, although there are clear differences between the shyness and dominance dimensions in terms of emotional expression cues, they do both correlate with fWHR. Shyness correlated negatively with fWHR whereas dominance correlates positively with

fWHR, as one might expect if shyness was simply the opposite of dominance or a similar dimension. Speculatively, this similarity raises the interesting possibility that the shyness dimension for children's faces is the antecedent of the dominance dimension for adult faces.

Finally, it was interesting that shyness correlated negatively with impressions of weight, somewhat akin to Sheldon's ectomorph classification (Sheldon, Stevens & Tucker, 1940), i.e. the perception that individuals with little body fat are self-conscious, socially anxious and introverted, which are all traits closely related to the shyness dimension. It is interesting that our novel data-driven study produces the same classifications, suggesting that these are commonly held beliefs.

Table 4. *Pearson product-moment correlations of the niceness and shyness factor scores with a) variables important in social psychology literature, and b) variables likely to be used as facial cues for impressions. Large correlations (>.5; Cohen, 1988) are in bold. Significance tests (Fisher Z transformation; Lee & Preacher, 2013; Steiger, 1980) examine whether each variable is significantly correlated more with one dimension over the other. Note that z tests compared positive correlations (i.e. so we could test the absolute difference in correlation strength).*

a)	Trustworthiness	Dominance	Sociability	Warmth	Competence	Valence
Niceness Dimension	*r* = **.947** *p* < .001	*r* = **-.868** *p* <.001	*r* = **.763** *p* < .001	*r* = **.946** *p* < .001	*r* = **.712** *p* <.001	*r* = **.884** *p* < .001
Shyness Dimension	*r* = -.154 *p* =.257	*r* =-.239 *p* = .076	*r* = **-.696** *p* <.001	*r* = -.378 *p* = .004	*r* = -.371 *p* = .005	*r* = -.411 *p* = .002
Sig. Test	*z* = 8.94 *p* < .001	*z* = 5.77 *p* < .001	*z* = .77 *p* = .444	*z* = 7.41 *p* < .001	*z* = 2.64 *p* = .008	*z* = 5.00 *p* < .001

b)	Happy	Sad	Fear	Disgust	Angry	Babyfaced	Masculinity	Healthy	Over-weight	fWHR	Attractive
Niceness Dimension	*r* = **.813** *p* < .001	*r* = **-.717** *p* < .001	*r* = -.411 *p* = .002	*r* = **-.894** *p* < .001	*r* = **-.897** *p* < .001	*r* = **.589** *p* < .001	*r* = *-.273* *p* = *.042*	*r* = **.502** *p* < .001	*r* = -.306 *p* = .022	*r* = .014 *p* = .916	*r* = **.608** *p* < .001
Shyness Dimension	*r* = **-.621** *p* <.001	*r* = **.782** *p* <.001	*r* = **.760** *p* < .001	*r* = .381 *p* =.004	*r* = .475 *p* < .001	*r* = .045 *p* = .741	*r* = *-.287* *p* = *.032*	*r* = **-.522** *p* < .001	*r* = -.381 *p* = .004	*r* = *-.409* *p* = *.002*	*r* = .047 *p* = .733
Sig. Test	*z* = 2.17 *p* = .030	*z* = 0.82 *p* = .414	*z* = 2.92 *p* = .004	z = 5.44 *p* < .001	*z* = 4.97 *p* < .001	*z* = 3.55 *p* < .001	*z = 0.08* *p = .467*	*z* = 0.14 *p* = .887	*z* = 0.47 *p* = .641	*z = 2.38* *p = .018*	*z* = 4.62 *p* < .001

STUDY 4: Validating the Two Dimensions

In Study 1-3, we found that young children's faces are evaluated on niceness (broadly comparable to trustworthiness) and shyness (dissociable from dominance) dimensions. Here, we aimed to validate these dimensions by showing that they could be visualized in children's faces. We also wanted to show that adults are sensitive to these impressions when they are visually manipulated in children's faces and tested with new raters (following Oosterhof & Todorov, 2008; Sutherland et al., 2013; Walker & Vetter, 2009; Walker & Vetter, 2016). This test was important, because it let us check that the labels we assigned to the dimensions were appropriate, using an independent sample of participants. First, we used morphing techniques to visualize the dimensions of niceness and shyness (comparable to Sutherland et al., 2013). We then used these face morphs (or composites) to check whether participants were sensitive to subtle facial manipulations of the dimensions. If cues that are indicative of niceness and shyness are captured by the face composites then participants should be able to infer judgments of niceness and shyness from these faces. Furthermore, manipulating the saliency of these cues should impact judgments of respective traits.

We used morphing and transforming techniques to enhance/reduce the salience of these cues to test whether trait judgments could be manipulated along a continuum. Enhancing/reducing cues that are indicative of niceness and shyness should lead to respective increases/decreases in judgments of niceness and shyness. To this end, we asked participants to rate a set of children's faces that were transformed along the dimensions, for how nice or shy they look. We also wanted to check that the two dimensions generalize to a new sample of Caucasian adults, from another culture. In studies 1-3, participants were American adults recruited online. Here, they were Australian adults who were tested individually in the lab.

As well as validating that adults are sensitive to the niceness and shyness dimensions, Study 4 generated a new set of children's faces that vary systematically along the important

dimensions of niceness and shyness (whilst remaining photorealistic). We are working with the authors of the CAFE database (LoBue & Thrasher, 2014) to make these images publicly available for scientific research. These images provide a valuable tool for future research into the effects of impressions of children, for example, in social, educational, and legal settings.

Method

Participants

Sixty-five (29 female, $M = 23.2$, $SD = 7.9$, range = 18 - 53 years) Caucasian adult volunteers were recruited from the University of Western Australia. One participant was excluded because they repeated the same numeric rating across an entire block of faces. Sample size was pre-determined by a power calculation using the effect size observed from previous, similar studies (Study 2 and 3; Walker and Vetter, 2016: $r = .35$), with a significance level of .05 and power of 0.8. Based on this calculation we aimed for 62 participants and recruited 65.

Stimuli

First, to visualize the two dimensions we created high and low composite faces by averaging together faces with the five highest or five lowest factor scores on each dimension (calculated from Study 2), as shown in Figure 4. Averaging faces generates prototypical representations of the facial cues driving specific judgments (for a review see Sutherland, Rhodes, et al., 2017), and has been used to validate the adult face dimensions (Sutherland, Liu, et al., 2018; Sutherland et al., 2013).

To ensure that our averages most accurately captured the cues indicative of our dimensions, we allowed faces to contribute to more than one face average if they were high/low on both dimensions (one face; high nice and low shy). We included all images

across face sex to help understand whether sex contributed as a cue (following Oosterhof & Todorov, 2008; Sutherland et al., 2013; for face-sex specific composites see Supplementary Materials Figure S1) and to keep the procedure consistent with that used by Oosterhof & Todorov (2008).

We constructed the face averages using Psychomorph (Tiddeman et al. 2011). To create the averages, 179 fiducial points were positioned to landmark specific features of the face. Psychomorph uses this information to average both the color/texture and shape of the images. We used these resulting face composites to create the stimuli for this study.

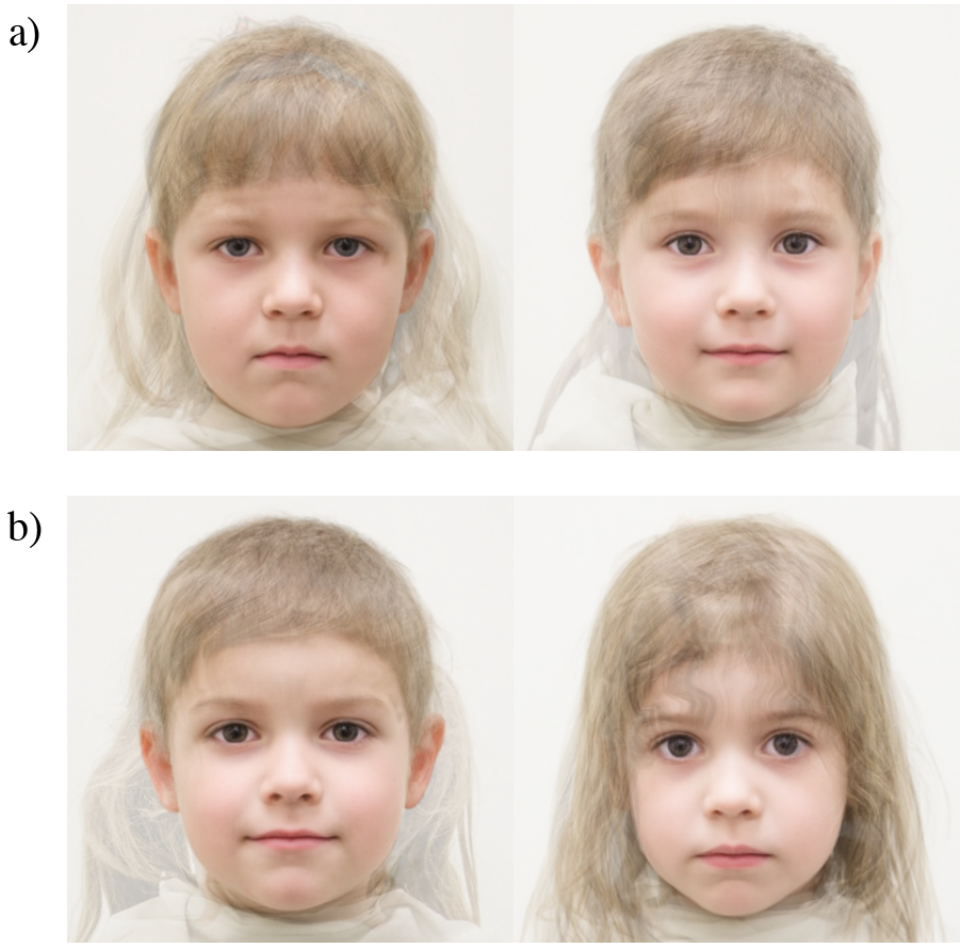

Figure 4. Composite faces created for the a) niceness dimension (left: low-nice, right: high-nice) and b) shyness dimension (left: low-shy, right: high-shy). Each composite was created

by averaging together the five faces (roughly representing 10% of the total faces) with the highest or lowest factor scores for each dimension.

Composite-face continuum: We created two face continua using the composite face images. One face continuum was created by morphing from the low-nice (0%) to high-nice (100%) composite, in incremental steps of 25%. The other face continuum was created by morphing from the low-shy (0%) to high-shy (100%) composite, in incremental steps of 25%. Therefore, in total there were 10 face images created by morphing along each of the two dimensions (5 niceness, 5 shyness; See Figure 5a, 5b).

Individual-face continuum: We used the four face composites (high-nice, low-nice, high-shy, low-shy: Figure 4), to transform new, individual face photographs along a continuum. Transforming a face differs from morphing because it transforms a single face along a continuum by applying the difference between two other face images (see Sutherland, et al. 2017). This method is particularly advantageous because it allows us to manipulate the saliency of the dimensions while controlling other variables, such as face identity. Moreover, facial transforming generates photorealistic images, making them an appealing alternative to face composites. Here, we transformed 20 new face photographs (10 female; from the CAFE database) along the niceness and shyness dimensions. We chose the 10 female and 10 male faces with the lowest combined factor score, i.e., they were the least distinctive of either dimension and thus easiest to transform in either direction.

To transform along the niceness dimension, we morphed a new identity towards/away from the high-nice face average, and vice versa for the low-nice face average. To transform along the shyness dimension, faces were morphed towards/away from the high-shy face average, and vice versa for the low-shy face average. Transformations were made at 25% and 50%, in both directions (e.g. 50% low-shy, 25% low-shy, 25% high-shy and 50% high-shy), so that each face identity had 4 values along each dimension, plus its original image, resulting

in 100 face stimuli per dimension (see Figure 5c, 5d for an example). We chose ±50% as a limit because face transformations started to look distorted beyond this point. Two additional faces from the CAFE database (1 female, 1 male) were also transformed to be used as practice stimuli. As our priority was to create photorealistic images, we transformed the face identities on shape (following Perrett et al., 1999; face transformations with color made very little difference other than introducing a ghosting shadow).

In total, there were 210 face stimuli (100 individual faces transformed along niceness dimension, 100 individual faces transformed along shyness dimension, 5 face composites morphed along the niceness dimension, and 5 face composites morphed along the shyness dimension).

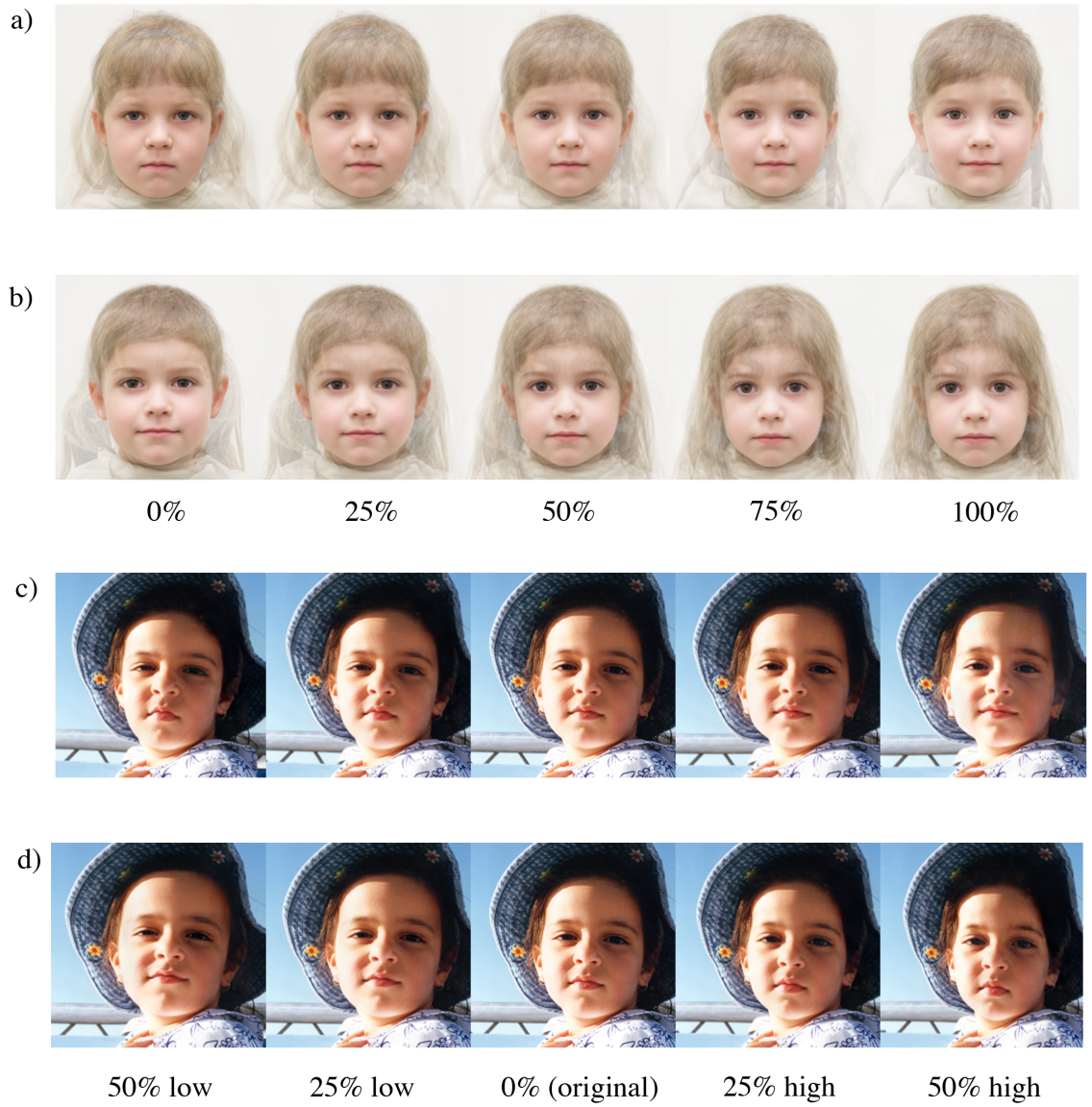

Figure 5. Face-composite continua morphed from a) low-nice (0%) to high-nice (100%) and b) low-shy (0%) to high-shy (100%) averaged composites. Individual-face continua transformed along the c) niceness dimension (50% low-nice to 50% high-nice), and d) the shyness dimension (50% low-shy to 50% high-shy), in steps of 25%. Note, this face identity did not come from the CAFE database, but is an example of how the faces were visually manipulated.

Procedure

Participants completed the experiment individually on a computer (display resolution 21.5-inch, 1920x1920) at the University of Western Australia. Participants saw the 105 faces

manipulated along the niceness dimension (5 face composites, 100 individual faces), and the 105 faces manipulated along the shyness dimension (5 face composites, 100 individual faces), in separate blocks (order randomized). Participants were asked to rate each set of faces for their corresponding attribute, on a scale of 1 (Not at all nice/shy) to 9 (Extremely nice/shy). As in Study 2, participants were encouraged to go with their spontaneous first impressions, and not to spend too long on the responses. Each face remained on screen until participants responded. Face order was randomized within the blocks and each block started with 10 practice stimuli (corresponding face transformations for one female and male face; data not analysed). Following the rating task, participants were asked basic demographic questions about themselves (e.g. age, sex, ethnicity) and were debriefed.

Data Analysis

To test whether participants were sensitive to the manipulations of niceness and shyness we ran a linear trend analysis on the mean trait ratings. As a first step, we examined inter-rater reliability of the ratings. Reliability was high for judgments of niceness (Cronbach's alpha = .96) and shyness (Cronbach's alpha = .96).

Results and Discussion

We ran linear trend analyses on the mean ratings to test whether participants' ratings tracked the manipulations of niceness and shyness. There was a significant linear trend for ratings of niceness, $F_{linear}(1, 64) = 179.40$, $p < .001$ $\eta_p^2 = .737$. Interestingly, participants were more sensitive to variations in the lower levels of niceness, $F_{quadratic}(1, 64) = 55.59$, $p < .001$, $\eta_p^2 = .465$, just as has been found for the adult-face dimension of trustworthiness (Oosterhof & Todorov, 2008). There was also a significant linear trend for ratings of shyness, F_{linear} (1,

64) = 311.45, p <.001, η_p^2 = .830 (See Figure 6). There were no higher order trends, all Fs < 3.51, ps > .066, η_p^2 < .052.

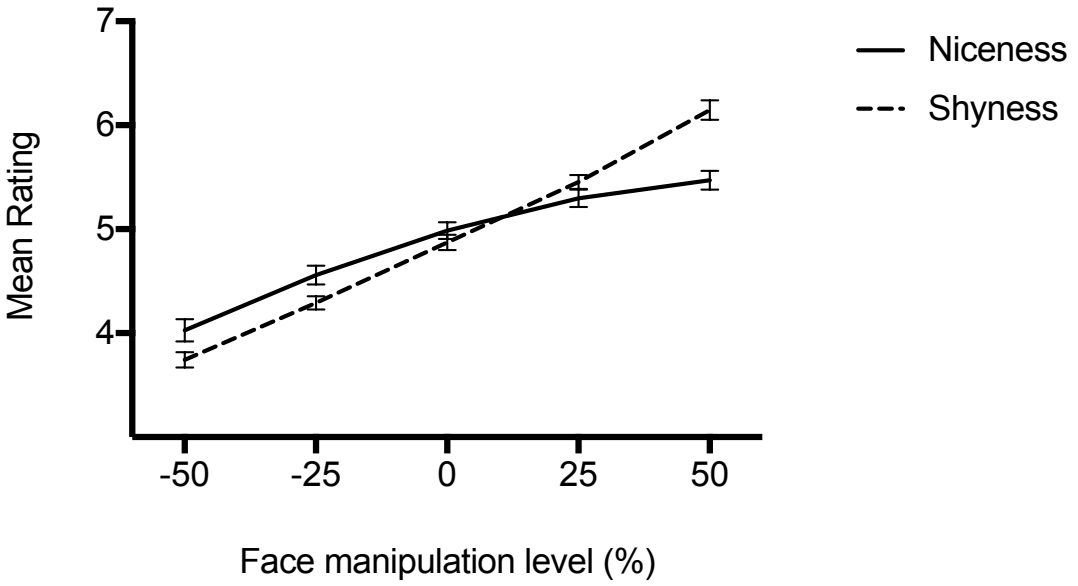

Figure 6. Mean face ratings for faces manipulated in incremental steps for niceness and shyness. Standard error bars are shown.

These results confirmed that niceness and shyness can be visualized in children's faces. A new, independent sample of adults was sensitive to subtle deviations of niceness and shyness in children's faces, thus validating our two-dimensional model. We have also produced an ecologically valid set of children's face stimuli that vary systematically along the niceness and shyness dimensions. This set is particularly valuable given that child face stimuli are scarce.

STUDY 5: Validating the Two Dimensions and Testing their Specificity

In Study 5, we further validated our dimensions in two ways. First, we tested whether participants could distinguish *between* our dimensions (divergent validity). If niceness and

shyness capture independent concepts, then adults should be able to distinguish between them. It is important that participants can discriminate between the dimensions because this would confirm that participants are not just sensitive to the valence of each dimension, given that the two dimensions are correlated (although this correlation is weak; r = -.17, Study 2) and that positive valence cues contributed to both dimensions (Study 3). Therefore, we employed the approach of past researchers (Walker & Vetter, 2016) to test whether participants could tell apart faces that were manipulated on niceness from faces manipulated on shyness (i.e. between-dimension pairs). This method offers a strict test of the discriminability of the two dimensions.

Second, we tested whether participants could distinguish along our dimensions (criterion validity). That is, participants should be able to tell apart faces that were manipulated to enhance niceness/shyness, from faces manipulated to reduce niceness/shyness (i.e. same-dimension pairs). These results would replicate the results from Study 4, confirming that adults are sensitive to variations of niceness and shyness in children's faces.

Method

Participants

Twenty-seven Caucasian adult volunteers (12 females, M = 20.7, SD = 2.2, range = 18 - 27 years) participated at the University of Western Australia. We chose this sample size based on a power calculation using the effect size from a similar study (i.e. Walker & Vetter, 2016; Study 4: d = 1.44), with a significance level of .05 and power of 0.8. The analysis revealed we would only need 6 participants. However, our study was not directly comparable to Walker and Vetter's (2016; e.g. we used children's faces instead of adult faces and manipulated our faces on two dimensions instead of five). Therefore, we re-calculated this

based on a more conservative effect size (i.e. 0.5), and aimed to recruit a sample of 33 participants.

Stimuli

We used a subset of the 20 children's faces from Study 4 with strongly enhanced/reduced features (i.e. transformed at 50%). Therefore, each of the 20 individual child faces had four variations (50% high-nice, 50% low-nice, 50% high-shy, 50% low-shy). We also included the four face composites (high-nice, low-nice, high-shy, low-shy), originally used to create the face transformations. In total, there were 84 face stimuli. We used two additional child face images (one female and one male; the same as in Study 4) with four variations as practice stimuli.

Procedure

Participants were shown two face photographs side-by-side, and were required to make a two-alternative forced choice between them regarding which child looked shyer or nicer. Each pair of images consisted the same child's face, either manipulated on niceness or shyness. Face pairs could differ *along* a dimension continuum (i.e. same-dimension pairs), e.g., a face identity transformed to enhance niceness and the same face identity transformed to reduce niceness (e.g. Figure 7a). Alternatively, face pairs could differ *between* the two dimension continua (i.e. between-dimension pairs). For example, a face identity transformed to enhance niceness paired with the same face identity transformed to reduce shyness (e.g. Figure 7b), or vice versa. As the dimensions were negatively correlated (e.g. faces rated higher for niceness were rated lower for shyness), we paired together manipulations in the opposite direction. This procedure offered the most conservative test of discriminability, as

these pairs should share greater perceptual similarity than pairs in the same direction (as can be seen in Figure 4).

Pairs were presented in four blocks separated by judgments; two same-dimension blocks and two between-dimension blocks. In the same-dimension blocks, participants saw face pairs manipulated along one dimension, niceness (reduced vs. enhanced) and shyness (reduced vs. enhanced). In the between-dimension blocks participants saw face pairs manipulated between the two dimensions (reduced shyness vs. enhanced niceness; and reduced niceness vs. enhanced shyness).

For each pair, participants were asked to pick the face that was either enhanced for niceness or shyness. The correct face choice (i.e. the face which was enhanced for each respective trait) appeared on the left for half the trials. Face pairs were shown twice in separate blocks, once with the correct choice on the left and once with the correct choice on the right. Therefore, participants saw a total of 168 face pairs (4 judgment blocks x 21 face pairs x 2 repetition). Participants saw the four judgment blocks in a randomized order. Trials within each block were also randomized. Each block started with two practice stimuli pairs (1 female, 1 male; order randomized). Afterwards participants provided demographic information (e.g. age, sex, ethnicity) and were debriefed.

Figure 7. Examples of face pairs that were either a) same-dimension pairs (left: reduced niceness; right: enhanced niceness), or b) between-dimension pairs (left: enhanced niceness; right: reduced shyness). Note, this face did not come from the CAFE database to protect the identity of the child participants.

Results and Discussion

As an initial step, we calculated the proportion of trials on which participants chose the "correct" face for each of the four conditions (enhanced vs. reduced niceness, enhanced vs. reduced shyness, enhanced niceness vs. reduced shyness, enhanced shyness vs. reduced niceness).

Importantly, participants were significantly more accurate than chance (0.5) at discriminating pairs in all four conditions (enhanced vs. reduced niceness, $M = .92$, $SD = .10$, $t(26) = 21.80$, $p < .001$, $d = 4.19$; enhanced vs. reduced shyness, $M = .98$, $SD = .04$, $t(26) = 61.92$, $p < .001$, $d = 11.88$; enhanced niceness vs. reduced shyness, $M = .65$, $SD = .22$, $t(26) =$

3.51, $p = .002$, $d = 0.68$; enhanced shyness vs. reduced niceness, $M = .88$, $SD = .11$, $t(26) = 17.27$, $p < .001$, $d = 3.31$). Therefore, participants were sensitive to faces that varied in opposite directions along the same dimension (criterion validity), and could also discriminate between the two dimensions (divergent validity).

As a final test of our model, if the dimensions vary as we expect them to then we would predict a greater difference between face pairs that vary on the same dimension, than faces that vary between dimensions. To test this assumption, we conducted a paired samples *t*-test on accuracy for same-dimension pairs versus between-dimension pairs. As expected, participants were more accurate at discriminating same-dimension than between-dimension pairs, $t(26) = 7.27$, $p < .001$, $d = 1.40$, See Figure 8.

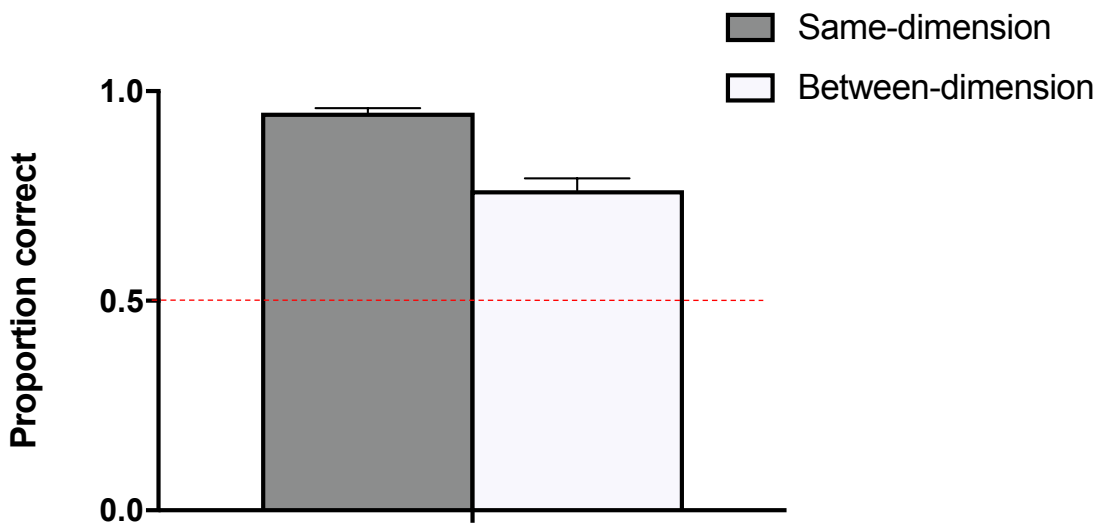

Figure 8. Proportion correct when discriminating between same-dimension and between-dimension face pairs. Standard error bars are shown. Chance level (0.5) is indicated with a dashed line.

STUDY 6: Examining the Social Consequences of the Child Face Dimensions

Across the first five studies, we have provided strong evidence that adults evaluate young children's faces on the dimensions of niceness and shyness. In Study 6, we ask whether these judgments have behavioral consequences for children. In other words, do adults behave differently towards children who vary along the dimensions of niceness and shyness? This is important to consider because such biases are likely to have serious social consequences for children. For example, adults may behave differently towards children who look particularly nice/shy, and in turn this may reinforce nice/shy behaviors in such children (i.e. self-fulfilling prophecy; Zebrowitz et al., 1996).

We examined the potential behavioral consequences of first impressions by investigating their influence on adults' behavioral expectations of children. Past research has demonstrated that children's facial appearance influences behavioral expectations of them (e.g. Barocas & Black, 1974; Berkowitz & Frodi, 1979; Clifford & Walster, 1973; Kenealy et al., 1988; Langlois et al., 2000; Ritts et al., 1992; Salvia et al., 1977; Zebrowitz et al., 1992; Zebrowitz et al., 1991). For example, babyfaced children are assumed to be more honest than their counterparts (Zebrowitz & Montepare, 1992; Zebrowitz et al., 1996).

Our approach was influenced by this previous research, but also went further in two distinct ways. First, we focused on the variables that we found to be particularly important for children's faces given our theoretical model (niceness and shyness), rather than on variables chosen based on *a priori* hypotheses about specific cues (e.g. babyfaceness and attractiveness; Zebrowitz & Montepare, 2008). Secondly, the child face stimuli we use varied only on the dimensions we have manipulated (i.e. we could control for confounding variables such as identity or sex), which allowed for a stronger test of causality.

We used the validated face sets from Studies 4 and 5 to experimentally investigate whether children's facial appearance alters adults' behavioral expectations of them. Using the

same procedure as Study 5, we asked participants to choose between child face pairs for specific trait-related tasks. For example, participants were asked to decide which child they would award for good behavior (for the niceness dimension). If facial appearance does influence adults' expectations of children, then participants should systematically choose the faces that are enhanced/reduced for niceness and shyness, in line with our predictions.

Method

Participants

We recruited 16 Caucasian adult participants (7 female; age $M = 24.6$, $SD = 10.2$, range = 20 – 33 years) who completed the task at the University of Western Australia. To be conservative, sample size was estimated using the smallest effect size observed from Study 5 (from the comparable condition; i.e. same-dimension pairs), which had a very similar methodology (power = .8, significance level = .05, $d = 4.19$). This power analysis revealed we would need three participants. Considering this sample was so small, we re-calculated sample size based on the comparable, yet more conservative results from Walker and Vetter ($d = 1.44$). This analysis revealed we would need 6 participants (and we recruited 16).

Stimuli and Procedure

We used the same 84 face stimuli as Study 5. We also followed the same procedure as Study 5, but with two differences. First, participants only saw same-dimension pairs because we were interested in investigating behavioral expectations of children who varied along each dimension. Second, participants were asked hypothetical behavioral questions about the face pairs.

Participants were asked two questions (one phrased positively and one negatively), which aimed to discriminate between an enhanced-nice and reduced-nice looking child

(which child would you be most likely to give an award to for good behavior; which child would you watch more closely for rough play, respectively), and two questions aimed to discriminate between an enhanced and reduced-shy child (which child would be less likely to answer a question in class; which child would you choose to lead the class discussion, respectively). For questions phrased positively, we predicted participants would choose the face enhanced for niceness/shyness, and for questions phrased negatively, we predicted participants would choose the face reduced for niceness/shyness. For example, for the question, 'which child would you be most likely to give an award to for good behavior?', we predicted that participants would choose the face enhanced for niceness. For the question, 'which child would you watch more closely for rough play?', we predicted that participants would choose the face reduced for niceness. Face pairs were shown twice, once with the correct choice on the left and once with the correct choice on the right. In total, participants saw 168 face pairs; 21 identities (20 transformed identities and 1 composite face) x 4 dimension questions (positive-nice, negative-nice, positive-shy, negative-shy) x 2 repetitions (left / right presentation). Trials were blocked by question. Pair order was randomized within each block. Block order was also randomized and each block started with two practice trials (one female and one male; as Study 5).

Results and Discussion

We first calculated the proportion of trials on which participants chose the "correct" face for each question. We then collapsed across the two questions for each dimension in order to control for positive and negative wording (i.e. we were not interested in, nor expected, any such difference). We compared these proportions to chance level (0.5). Importantly, we found that children's facial niceness and shyness did influence adults' behavioral expectations of them. Participants chose the expected face, which aligned with the

behavioral expectations, for faces manipulated on niceness and shyness, above chance level, niceness; $t(15) = 13.94$, $p < .001$, $d = 3.48$, shyness; $t(15) = 5.45$, $p < .001$, $d = 1.36$ (see Figure 9). Participants chose behaviors that reinforced the appearance of the children. Participants' hypothetical behavior was more influenced by faces manipulated for niceness than shyness, $t(15) = 3.27$, $p = .005$, $d = 0.82$, but nevertheless was well above chance across both dimension conditions. Indeed, these results suggest that adults' impressions of children's faces are likely to impact the behavioral expectations and outcomes of children.

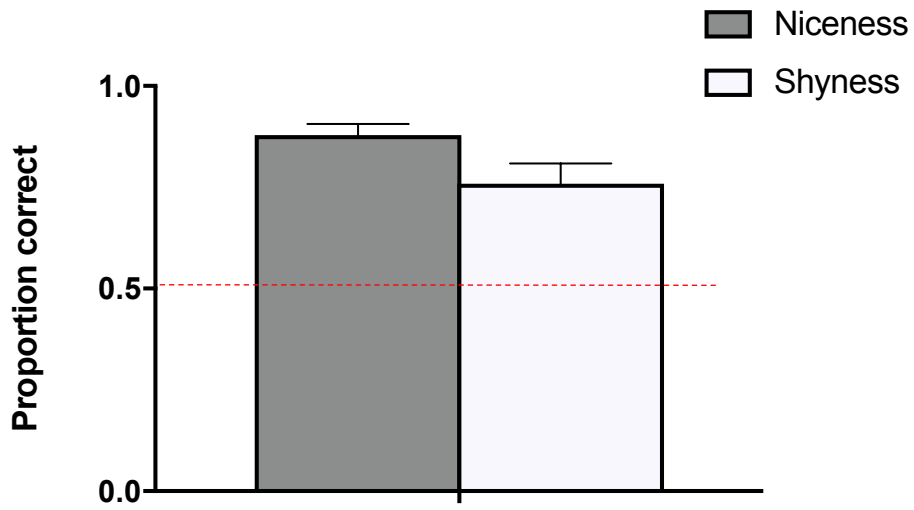

Figure 9. Average proportion that participants chose the face that aligned with the trait-related expectation, when faces where manipulated for niceness and shyness. Chance level (0.5) is indicated with a dashed line.

Study 7: Testing the Robustness of the Dimensions

We have found strong evidence that niceness and shyness underlie adults' impressions of children's faces (Studies 1-3), that adults are sensitive to visual cues of niceness and shyness in children's faces (Studies 4-5) and that impressions of niceness and shyness influence adults' expectations of children (Study 6). Here, we test the robustness of these dimensions, by examining whether they generalize to an independent, larger and more heterogeneous database of children's faces. Importantly, not only is this database three times the size of that in Study 2, but it consists of naturalistic, everyday ambient images instead of the highly controlled stimuli used in Study 2. This characteristic of the database is important, because ambient face images are more similar to those that we encounter in everyday life (Jenkins, et al., 2011, Sutherland, et al., 2017) and therefore allow for a strong test of the robustness of the niceness and shyness dimensions. Moreover, the substantial variability introduced by the ambient images could alter the structure of the dimensions, or introduce entirely new dimensions as found for adult faces (Sutherland, et al., 2013). Therefore, if the same dimensions of niceness and shyness emerge, we would have strong evidence for their robustness.

We collected ambient images of 154 children's faces from the internet. Following the method introduced by Sutherland et al. (2013), we deliberately sourced face images that naturally varied in expressions, poses, hair style, and so forth, as these are the conditions we experience faces in everyday life. We tested our two-dimensional model by collecting trait ratings for this large, heterogeneous face database and carried out a factor analysis on these ratings, following the method of Study 2.

Method

Participants

We recruited a new sample of 256 adult Caucasian participants from M-Turk (Female = 128, *M* = 39, *SD* = 12, range = 19 – 70 years[1]). We invited participants to complete the study who had already completed a short qualifier on TurkPrime (i.e. participants were already pre-screened to be Caucasian and had passed a short attention check). As in previous studies, we excluded an additional 18 participants (13 participants did not pass our attention check, 1 reported technical issues, 1 entered the same response for an entire block, and 3 participants were not in America during the time of testing). Participants were recruited until a minimum of 10 participants (i.e. five of each sex) had rated each trait (See Supplementary Materials Table S19), as in Study 3. This sample size has been found sufficient for good reliability at the group level (e.g. Sutherland et al., 2013; Sutherland et al., 2016).

Stimuli

Ambient images of 154 Caucasian children's faces (77 female) were taken from the internet (these images were labelled as free for any use). We used eight search terms: child, kid, 4-year old child, 5-year old child, 6-year old child, 4-year old kid, 5-year old kid, and 6-year old kid, to source the images. Equal numbers of female and male images were selected for each search term. Images were required to meet a predetermined set of criteria. Specifically, images were required to be in color, have no occlusions to the face, have details of the face clearly visible, be of adequate image quality (>80 KB), and with eyes open. Only images of Caucasian children were included, as in Study 2. We tried to focus our sample on faces of children 4-6 years old (as reflected in our search terms). However, as the actual age of the children could not be determined we decided to restrict our sample to "pre-pubescent"

[1] Note, one participant's age was unknown and was therefore excluded from this analysis.

children. Three independent raters checked the images to ensure all children looked pre-pubescent, but older than 4 years. This process left us with a total of 154 children's faces. A sample of 2.5 times the size of the original is required for a replication study (Simonsohn, 2015), and here our sample size exceeds this recommendation.

Images were cropped in a rectangle around the face and standardized in height to 250pixels, but otherwise were left unmodified. As images were purposefully ambient, they were free to vary on expression, pose, lighting, and so forth (see Figure 10 for examples of these ambient images).

Figure 10. Examples of ambient images (four female and four male). These face images reflect the variable conditions under which we see faces in everyday life and contrast the highly controlled images from Study 2. For copyright reasons, these are not the actual images used in Study 7, although they are representative of our stimuli.

Procedure

Participants were randomly assigned to rate the 154 children's faces on one of 19 attributes, as in Study 2. For each face, participants were asked to rate, "How [trait] is this child?" on a scale of 1 (Not at all [trait]) to 9 (Extremely [trait]). To minimize fatigue, participants rated each face in this large sample twice, instead of three times as in Study 2. Otherwise, the procedure was identical to that of Study 2. The task started with two additional faces (one female and one male) as practice trials, which were also sourced online. As a final step, participants answered demographic questions about themselves and completed the attention check, as in Study 2. The experiment took approximately 25min, for which participants were reimbursed (US $3).

Analysis

Intra-rater reliability was adequate (See Supplementary Materials Table S19), as in Study 2. Therefore, we averaged across trait ratings for each participant and calculated inter-rater reliability. Cronbach's alpha also revealed good inter-rater reliability (See Supplementary Materials Table S19).

To determine the dimension structure, we performed a factor analysis on the average ratings of the 19 traits with oblique rotation, as in Study 2. Results from previous studies revealed no participant sex or face sex differences, and therefore we ran the factor analysis on the data as a whole (but also see Supplementary Materials Table S21 and S22 for separate analyses). Data were suitable for factor analysis (most correlation coefficients > .3, see Supplementary Material S20; Bartlett's test of sphericity < .001; Kaiser-Meyer-Olkin > .6: Kaiser, 1970; Kaiser & Rice, 1974). Kaiser criterion, the scree plot and parallel analysis unanimously supported the extraction of two factors and therefore we extracted a two-dimensional model.

Results and Discussion

To evaluate our two dimensions, we interpreted the structure matrix (See Table 5). Reassuringly, two dimensions emerged which were almost identical to the two dimensions from Study 2 (niceness and shyness).

The first dimension extracted had highest loadings from energetic, shy and serious ratings, and closely resembled the shyness dimension from Study 2 (See Table 5). All of the traits that loaded most onto the shyness dimension in Study 2 (i.e. shy, confident, timid, quiet, lonely and energetic) loaded most onto the first dimension here. For the second dimension, positive judgments loaded positively (e.g. sweet, innocent, nice) and negative judgments loaded negatively (e.g. troublemaker and tough) onto the dimension. This pattern is identical to that of the niceness dimension from Study 2, and clearly resembles a valence dimension. As in Study 2, the two dimensions were only weakly correlated (niceness vs. shyness correlation Study 2 r = -.17, Study 7 r = -.15) The same two dimensions also emerged when we performed separate analyses for female and male faces, for female and male participants, and when we performed a PCA with Varimax rotation as performed by Oosterhof and Todorov (2008; See Supplementary Materials Table S21, S22, S23 and S24).

Table 5

Principal axis factor analysis on the 19 traits for children's faces with oblique rotation, as derived from the structure matrix. Numbers are factor loadings (and can be interpreted as correlations between the dimensions and original trait variables). The dotted line divides the traits that load most onto each dimension. Factor loadings > .4 are in bold.

Trait	Dimension 1: Shyness	Dimension 2: Niceness
Energetic	**.98**	.05
Shy	**-.96**	-.10
Serious	**-.95**	-.32
Lonely	**-.92**	**-.43**
Timid	**-.91**	.14
Funny	**.91**	.02
Cheerful	**.88**	**.50**
Friendly	**.79**	**.68**
Confident	**.77**	**.52**
Quiet	**-.75**	**.45**
Sweet	.26	**.94**
Innocent	.25	**.89**
Trouble-Maker	.10	**-.87**
Kind	**.55**	**.86**
Nice	**.60**	**.83**
Tough	-.25	**-.78**
Intelligent	-.08	**.75**
Attractive	-.19	**.74**
Curious	.19	**.68**
Variance explained: equivalent varimax PCA	**53.5%**	**30.9%**

Although the dimensions extracted here and in Study 2 share considerable overlap, one difference between studies is the relative amount of variance each dimension explains; that is,

in Study 7, the dimensions switch in order of variance explained. However, replication of a factor analysis does not require the order of the dimensions to be identical, so long as the same common dimensions emerge with similar items loading on them (Fabrigar & Wegener, 2011). Our factor analysis meets these replication criteria.

In summary, we found that the dimensions extracted in Study 2 (from a small sample of controlled stimuli) generalized to a large, independent sample of ambient images. This generalizability provides strong evidence for the robustness of niceness and shyness as important dimensions underlying impressions of children's faces.

General Discussion

Facial first impressions have received intense research interest over the past decade. To date, trustworthiness and dominance have been conceptualized as universal, functional evaluations because together they signal the potential threat of others (Oosterhof & Todorov, 2008). Here, we proposed a test of the underlying functional assumption of this model and a test of the flexibility of these dimensions. Specifically, we considered whether the same dimensions also apply to a population of faces from whom it would not be functional to infer threat, namely children's faces. Adult-child interactions are focused on establishing a nurturing and caretaking relationship, and if the dimensions that underlie adult's impressions of children's faces are functional, then they should reflect these social goals.

A functional basis of first impressions

For the first time, we found a population of faces which is not evaluated on identical dimensions to those used for adult faces. Specifically, we found robust evidence that adults' impressions of young children's faces are based on the dimensions of niceness and shyness (Study 1 and 2). We validated these dimensions in three further studies and showed that they

transfer to a new population of participants (from a different culture; Study 4 and 5) and to an independent sample of heterogeneous, ambient images (Study 7). Study 3 suggested that the first dimension from the child (niceness) and adult (trustworthiness; Oosterhof & Todorov, 2008) face models are more similar than different. From a theoretical perspective, it is likely that niceness and trustworthiness tap into the same general underlying construct. This construct can be conceptualized as a valence dimension. For children's faces, adults spontaneously use words such as *nice* and *sweet* to describe this valence dimension, but do not spontaneously infer trustworthiness. Here, we were motivated by our data-driven approach to label the dimension as niceness.

Not only are the niceness and trustworthiness dimension empirically similar, but they seem to be influenced by comparable cues (e.g. overgeneralization of emotion cues) and both were strongly based on valence, suggesting that they are based on similar perceptual mechanisms. However, it is interesting to note that some differences exist between the dimensions. For example, dominance strongly negatively correlated with the niceness dimension for children's faces, but does not correlate strongly with the trustworthiness dimension for adult faces (Oosterhof & Todorov, 2008), suggesting that these apparently similar dimensions are not identical. Therefore, there are also important difference between the dimensions which are captured in our choice of label as niceness, instead of trustworthiness. Nevertheless, overall, it is likely that adults use trustworthiness (or something similar) to signal the intentions of others and that this inference is functional for faces across the lifespan.

In contrast, we found that the second dimension underlying impressions of young children's faces, that we have labelled as shyness, was distinct from the dominance dimension for adult faces (Oosterhof & Todorov, 2008). Although adults were able to make consistent judgments of dominance from children's faces, these judgments did not relate to

the shyness dimension. Critically, judgments of dominance (and traits related to dominance) loaded more onto the first (niceness) than second (shyness) dimension. This result is striking, because it reveals that shyness is an independent dimension for children's faces, that is dissociable from the adult face dimension of dominance.

Adults also used mostly different cues to make inferences of shyness from children's faces, compared to those used to make inferences of dominance from adult faces (Oosterhof & Todorov, 2008), suggesting there is more than just a semantic difference between the shyness and dominance dimensions. We also found that shyness was dissociable from sociability, consistent with their behavioral dissociation (Asendorpf & Meier, 1993; Coplan & Armer, 2007; Coplan, Prakash, O'neil, & Armer, 2004). Finally, the shyness dimension was distinguishable from competence, which has also been suggested as an alternative dimension to dominance for adult faces (Sutherland, Liu, et al., 2018; Sutherland et al., 2013).

The observed differences between the shyness and dominance dimensions highlights the flexibility of the dimensions underlying first impressions. We argue that children's faces are not evaluated on dominance/competence, because such judgments are not relevant to the social goals associated with adult-child interactions. Instead, adults are sensitive to a dissociable dimension for children's faces: shyness. The fact that we found such a different dimension for children's faces is theoretically remarkable because it provides important support for the underlying assumption that impressions are functional (Oosterhof and Todorov, 2008), and acknowledges the flexibility of these dimensions (for a similar argument when examining dimensions across culture, see Sutherland, Liu et al., 2018).

These results raise an interesting question: why do adults evaluate young children's faces on shyness? If the dimensions underlying first impressions are functional (Oosterhof & Todorov, 2008), then we should expect children's faces to be evaluated on dimensions that

are relevant for such faces. Adult-child interactions are predominantly established on providing a caregiving and protecting relationship (Bowlby, 1969). Given this context, it might be important for adults to be sensitive to shyness in order to gauge which children might need more sensitive nurturing. For example, adults may infer shyness from children's faces to guide how to behave towards children (e.g. a shy-looking child might need a more a nurturing and dependent relationship). Indeed, the shyness dimension was strongly related to the facial cue of fear, which might give rise to inferences about vulnerability rather than potential threat or capability. Alternatively, perhaps adults are sensitive to cues in children's faces that signal how *threatened* the child is. Nevertheless, our results align with the broader context of the importance of nurturing and caretaking behaviors towards children (e.g. Aradhye, Vonk, & Arida, 2015; Franklin & Volk, 2017; Kringelbach, Stark, Alexander, Bornstein, & Stein, 2016; Luo, Kendrick, Li, & Lee, 2015; Picci & Scherf, 2016). The fact that we find different dimensions for children's faces is also interesting from an evolutionary perspective, as it suggests that responses to children's faces are not just a by-product of mechanisms that have evolved to detect threat from adult faces (as similarly proposed by DeBruine, Hahn, & Jones, 2016).

It is also possible that bottom-up stimulus attributes contribute to the differences between the child- and adult-face dimensions. For example, adults may be perceptually unable to evaluate children's faces on dominance. Dominance is heavily influenced by sexually dimorphic cues (e.g. facial masculinity: Oosterhof & Todorov, 2008; Sutherland et al., 2013) that are enhanced after puberty (Kesterke et al., 2016) and thus visual cues to dominance may simply be lacking in children's faces. Against this idea, we found that ratings of dominance for children's faces did show similar variance to ratings of trustworthiness (dominance: $M = 4.6$, $SD = 0.8$, trustworthiness: $M = 5.2$, $SD = 0.7$), suggesting that adults could discriminate faces based on this trait. More generally, it remains possible that the facial

structure of children's faces contributes to the different dimensions found for children's faces. It is important to note that this idea does not weaken our top-down theory. Instead, it is likely that bottom-up and top-down processes are complementary, and work together to shape the underlying structure of facial impressions (as also suggested by Stolier, Hehman, & Freeman, 2017; Sutherland et al., 2017).

In Study 4, we found evidence for the role of emotion, masculinity, babyfaceness, fWHR, attractiveness and health cues in signaling niceness and shyness. Future research could further explore the physical cues underlying these dimensions, and particularly for shyness, which was investigated for the first time here. For example, research could systematically model the specific physical attributes related to shyness using a neural network approach (comparable to Vernon, Sutherland, Young, & Hartley, 2014). Future research could also investigate which regions of the face are involved when drawing these social perceptions. For example, research could employ reverse correlation (Dotsch & Todorov, 2012), or the bubbles technique (Gosselin & Schyns, 2001). Our data-driven approach here opens up these new lines of inquiry, which can be investigated using complementary hypothesis driven approaches.

Flexibility of facial impression dimensions

Our results help answer an important theoretical question: how flexible are the facial dimensions which underlie first impressions? They suggest considerable flexibility, in that different social goals associated with young children's faces may drive different impressions of such faces. Future research could seek to further test the flexibility of these dimensions. For example, do different dimensions underlie evaluations of faces across the lifespan? A functional perspective would suggest that the dimensions should change during adolescence, a developmental stage where social goals transition from reflecting the needs of childhood, to

the needs important during adulthood. Perhaps during adolescence, entirely new dimensions emerge to align with unique social goals of adolescents. Indeed, social goals associated with adolescence (specifically, pubertal development) have been found to influence other face-processing abilities (Motta-Mena & Scherf, 2016; Picci & Scherf, 2016; Scherf, Behrmann, & Dahl, 2012). Alternatively, the child face dimensions may transition to more closely resemble the adult face dimensions, to align with the transition of social goals at this age. Speculatively, it is possible that the shyness dimension for children is the antecedent of dominance (or competence) for adults. In support of this idea, both shyness and dominance correlated with fWHR and masculinity. Ideally, a longitudinal approach (similar to the approach of Schwartz, Snidman & Kagan, 1996, who track shy behavior over time) would allow researchers to track the changes in dimensions across the lifespan while controlling for face identity.

It would also be useful for future research to test the functional basis of first impressions by actively manipulating social goals. For example, if children's faces were presented in a threatening context, would the change of context be enough to drive a dimension that more closely resembles dominance? Furthermore, research could consider whether this flexibility also applies to other broad social cognitive dimensions, such as warmth and competence (Fiske et al., 2007).

Interestingly, recent research has found other instances where the structure of the adult face dimensions diverges from the original two-dimensional model (Oosterhof & Todorov, 2008). In some cases, the correlation or valence of the trait judgments differs (e.g. Oh, Dotsch, Porter, & Todorov, 2018; Sutherland et al., 2015) and in other cases, entirely new dimensions emerge (Sutherland et al., 2013). For example, Sutherland et al. (2013) found a new (attractiveness-age) dimension emerged for adult faces when they used ambient image face stimuli with a wide age range (see also South Palomares et al., 2017; Sutherland

et al 2018, Wolffhechel et al., 2014). This result is particularly interesting because it also reveals that the dimensions can be somewhat flexible. Moreover, it highlights the sensitivity of impressions to differences between face populations.

Although research has found new dimensions can emerge across different populations of adult faces (Sutherland et al. 2013), here we find considerable consistency in the dimensions extracted from two independent populations of children's face. In study 2 we used highly controlled face images of children between 4-6 years old. In Study 7, we used heterogeneous ambient images that varied naturally as faces do in everyday life. Study 7 also likely contained a wider age range of faces than Study 2. Regardless of these substantial stimuli differences, we found more similarities than differences between the dimensions extracted. In Study 7, two dimensions emerged which resembled the dimensions of niceness and shyness, although the dimensions switched in order. Nevertheless, these results provide strong evidence for the robustness of the child face dimensions.

It is remarkable that at a broader level, relatively consistent dimensions emerge across numerous psychological stimuli (see Fiske et al., 2007). Indeed, there are striking conceptual similarities between our child face dimensions and other social psychology dimensions. The primary dimension that emerges within social psychology consistently resembles a valance (good/bad) dimension (e.g. Asch, 1946; Fiske et al., 2007; Oosterhof & Todorov, 2008; Rosenberg et al., 1968), similar to our niceness dimension. In contrast, the second dimension varies depending on both context and stimuli. For example, a second dimension has emerged as dominance (Oosterhof & Todorov, 2008), competence (Fiske et al., 2007), intellectual ability (Rosenberg et al., 1968), social status (South Palomares et al 2017), and even potency (Osgood, 1952). Here, we find a second dimension that emerges as shyness. Our results show that shyness is not identical to the adult facial impression dimension of dominance (or

competence), although it may share some overall conceptual similarity to a dimension such as potency.

Finally, it is important to consider the extent to which our results generalize to other face/perceiver populations, particularly across cultures. Our results here capture Caucasian adults' impressions of Caucasian, young children's faces. For adult faces, there is considerable cross-cultural agreement for the dimensional structure of impressions (Sutherland, Liu, et al., 2018; Walker, Jiang, Vetter, & Sczesny, 2011). This stability across cultures suggests that the dimensions for children's faces are also likely to generalize across cultures. However, it is possible that cultural factors might impact the dimensional structure for children's faces. In particular, shyness is not seen as maladaptive across all cultures. In China, self-restraint and dependency are encouraged in children (Ho, 1986; Ho & Kang, 1984). Quiet and shy children are praised, and called "guai," meaning "good". In turn, shyness is positively related to perceptions of sociability and peer acceptance in Eastern but not Western societies (Chen, Rubin, & Sun, 1992). Considering this cross-cultural difference, it would be interesting for future research to investigate whether the two dimensions we find here transfer to a Chinese sample (comparable to Sutherland, Liu, et al., 2018) or whether Chinese adults are less sensitive to shyness manipulations in children's faces (comparable to Walker et al., 2011). It is possible that just one dimension will emerge for Chinese adults' impressions of children's faces, if shyness judgments are strongly valenced.

On a similar note, it would be interesting to examine the generalizability of these results to impressions of other-race faces (e.g. Caucasian participants' ratings of other-race faces). Sutherland et al., (2017) found substantial resemblance in the dimensions from British and Chinese populations, suggesting that impressions of own-race and other-race faces are more similar than they are different. However, it would be interesting to see what the dimensions look like for groups of other races, particularly Caucasian's impressions of

African American faces. Caucasian people tend to associate more threat with African-American than Asian faces (Dixon, 2006), and this difference in social goals might drive different dimensions for such faces.

Future research could also consider how these dimensions generalize across face/perceiver sex. Here, our results highlight more similarities than differences between both female/male perceivers and female/male faces, consistent with the work on adult faces to date (South Palomares et al., 2017). As these analyses were only exploratory, it would be interesting for future research to more thoroughly investigate potential sex differences.

Finally, future research could also consider how the dimensions we find for young children (4 – 6 years old) generalize to children of other ages (e.g. post-pubescent children). Critically, the primary aim of our study was not to explain impressions of all children, but to test the functional basis of impressions on a population of faces for whom it would not be relevant to detect threat from. Therefore, we do not necessarily expect the dimensions for young children's faces to generalize to older children, particularly if different social goals are associated with such children (as we discuss above). In fact, finding that these dimensions do not generalize to older, post-pubescent children would actually support our theoretical stance. One might also expect these dimensions to change again, when considering children's impressions of children's faces. Nevertheless, our data-driven approach here establishes a foundation for these exciting new lines of enquiry.

Social implications of facial impressions

We also sought to investigate the potential practical consequences of first impressions in light of our new theoretical model. In Study 6, we found that children's facial niceness and shyness influenced adults' behavioral expectations of them. It is possible that these biases may contribute to self-fulfilling (or self-defeating) prophecies in the real world (see

Zebrowitz et al., 1996). For example, if teachers are more likely to give an award for good behavior to a nice-looking than not-nice-looking child, this may reinforce nice behaviors in such children. Interestingly, recent research has found that a relationship between facial trustworthiness and real-world trustworthiness is already present at 8 years (Q. Li et al., 2017), potentially due to phenomenon such as the self-fulfilling prophecy. It is therefore important that we educate adults (especially those within educational settings) about the biases that may influence their behavior towards children.

Here, we found strong evidence that facial impressions influence adults' behavioral expectations of children. Expectations are a major driver of behavior (Armitage & Conner, 2001), and therefore these impressions will likely influence real behavior towards children, just as they do for adults (see Todorov et al., 2015). Nevertheless, future research could examine whether similar expectations transfer to adults' *actual* behavior towards children. For example, are teachers actually less likely to choose a shy-looking child to answer a question than their peers? We see our current study as providing a foundation that will structure future endeavors into the behavioral consequences of first impressions. Future research could also consider the role of other sources of information in interpersonal interactions (e.g. knowledge about past behaviour; Chang, Doll, van't Wout, Frank, & Sanfey, 2010; T. Li, Liu, Pan, & Zhou, 2017; Rezlescu, Duchaine, Olivola, & Chater, 2012), and whether this information can override the impressions we form from a face.

Although our results from Study 6 suggest that adults are likely to behave differently to children who vary along these dimensions, it is not obvious whether there is any accuracy behind these impressions. This possibility raises an interesting question; are children who look more nice/shy, actually more nice/shy? Recent research suggests that there may be some accuracy behind impressions of trustworthiness (cf. "niceness") for children's faces (Q. Li et al., 2017). To date, no study has investigated whether impressions of shyness in children's

faces are a valid cue for actual shyness. Interestingly, research has found a link between facial appearance and the *development* of shy behavior in children (Arcus & Kagan, 1995; Zebrowitz et al., 2015) and aggressive behavior in later life (Schwartz, et al., 1996). Arcus and Kagan (1995) found that high-reactive infants who were developmentally more likely to become shy children, had wider faces than infants with low-reactive temperaments and who were more likely to develop into outgoing children. Here, we also found that fWHR was a cue for the shyness dimension (Study 3) and previous research has found that fWHR is a cue to dominance (Alrajih & Ward, 2014; Mileva, Cowan, Cobey, Knowles, & Little 2014) and aggressive behavior (Haselhuhn, Ormiston & Wong, 2015) in adult faces. Together, this evidence is suggestive of potential accuracy in impressions of shyness.

A new database of children's faces

To help with future investigations, we also created a validated set of children's faces that vary along the psychologically relevant dimensions of niceness and shyness (Study 4 and 5). This approach will be particularly useful for future research into the behavioral consequences of first impressions of young children's faces. In the past, research has used children's faces that naturally vary along specified attributes. For example, research comparing behavior towards attractive versus unattractive children has used real face stimuli which have been previously rated for facial attractiveness (e.g. Berkowitz & Frodi, 1979; Salvia et al., 1977). However, this approach cannot control for other stimuli differences (e.g. face identity, sex, age) that might also drive behavioral responses. In contrast, our approach is systematically controlled, making these kinds of images appealing to use in such research. Importantly, using this approach these faces can be made to look photorealistic, despite these manipulations, which also make them a useful alternative to computer-generated images. We

are currently working to make these images publicly available and linked to the CAFE database.

Conclusions

To date, trustworthiness and dominance have been conceptualized as universal evaluations of faces that function to signal the potential threat of others. Here, we test the functional assumption by considering whether these same dimensions apply to a population of faces not associated with threat. We found that young children's faces are evaluated on dimensions that are not identical to the adult face dimensions (niceness and shyness vs. trustworthiness and dominance; Oosterhof & Todorov, 2008) and importantly, not related to threat. The first child face dimension, niceness, was empirically and theoretically similar (although not identical) to the adult face valence dimension of trustworthiness. In contrast, the second child face dimension, shyness, was clearly dissociable from the adult face dimension of dominance. These results represent a significant step forward for models of first impressions because they provide support for a key assumption of current models, namely that impressions have a functional basis. Furthermore, these results suggest that the dimensions underlying facial impressions (and perhaps social psychology, more broadly) are more flexible than previously conceptualized. Finally, our findings also indicate that impressions of children's faces strongly affect adults' expectations of children and therefore are likely to also impact adults' behavior towards them. Childhood is a crucial time of development (see Phillips & Shonkoff, 2000), so it is particularly important that we consider the social consequences that these impressions are likely to have.

References

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of personality and social psychology, 93*(5), 751.

Abir, Y., Sklar, A. Y., Dotsch, R., Todorov, A., & Hassin, R. R. (2017). The determinants of consciousness of human faces. *Nature Human Behaviour*, 111(27), 9792-9797.

Abraham, E., Hendler, T., Shapira-Lichter, I., Kanat-Maymon, Y., Zagoory-Sharon, O., & Feldman, R. (2014). Father's brain is sensitive to childcare experiences. Proceedings of the National Academy of Sciences, 111(27), 9792-9797.

Adolphs, R., Nummenmaa, L., Todorov, A., & Haxby, J. V. (2016). Data-driven approaches in the investigation of social perception. *Phil. Trans. R. Soc. B, 371*(1693), 20150367.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological monographs, 47*(1), i.

Alrajih, S., & Ward, J. (2014). Increased facial width-to-height ratio and perceived dominance in the faces of the UK's leading business leaders. British Journal of Psychology, 105(2), 153-161.

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.

Aradhye, C., Vonk, J., & Arida, D. (2015). Adults' responsiveness to children's facial expressions. *Journal of Experimental Child Psychology, 135*, 56-71.

Arcus, D., & Kagan, J. (1995). Temperament and craniofacial variation in the first two years. *Child development, 66*(5), 1529-1540.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British journal of social psychology, 40*(4), 471-499.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*(3), 258.

Asendorpf, J. B., & Meier, G. H. (1993). Personality effects on children's speech in everyday life: Sociability-mediated exposure and shyness-mediated reactivity to social situations. *Journal of personality and social psychology, 64*(6), 1072.

Barocas, R., & Black, H. K. (1974). Referral rate and physical attractiveness in third-grade children. *Perceptual and Motor Skills, 39*(2), 731-734.

Bauman, C. W., & Skitka, L. J. (2012). Corporate social responsibility as a source of employee satisfaction. *Research in Organizational Behavior, 32*, 63-86.

Berkowitz, L., & Frodi, A. (1979). Reactions to a child's mistakes as affected by her/his looks and speech. *Social Psychology Quarterly, 42*(4), 420-425.

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: the impact of age-related craniofacial changes on social perception. *Psychological bulletin, 100*(1), 3.

Boothroyd, L. G., Jones, B. C., Burt, D. M., & Perrett, D. I. (2007). Partner characteristics associated with masculinity, health and maturity in male faces. *Personality and Individual Differences, 43*(5), 1161-1173.

Bowlby, J. (1969). *Attachment and Loss: Attachment. V*: Basic Books.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*(2), 135-143.

Carré, J. M., & McCormick, C. M. (2008). In your face: facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society of London B: Biological Sciences, 275*(1651), 2651-2656.

Caulfield, F., Ewing, L., Burton, N., Avard, E., & Rhodes, G. (2014). Facial trustworthiness judgments in children with ASD are modulated by happy and angry emotional cues. *PloS one, 9*(5), e97644.

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive psychology, 61*(2), 87-105.

Chen, X., Rubin, K. H., & Sun, Y. (1992). Social reputation and peer relationships in Chinese and Canadian children: A cross-cultural study. *Child development, 63*(6), 1336-1343.

Clifford, M. M., & Walster, E. (1973). The effect of physical attractiveness on teacher expectations. *Sociology of education, 46*(2), 248-258.

Coetzee, V., Perrett, D. I., & Stephen, I. D. (2009). Facial adiposity: a cue to health? *Perception, 38*(11), 1700-1711.

Cogsdill, E. J., & Banaji, M. R. (2015). Face-trait inferences show robust child–adult agreement: Evidence from three types of faces. *Journal of Experimental Social Psychology, 60*, 150-156.

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces a developmental study. *Psychological science, 25*(5), 1132-1139.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences 2nd edn: Erlbaum Associates, Hillsdale.

Collova, J. R., Kloth, N., Crookes, K., Burton, N., Chan, C. Y., Hsiao, J. H., & Rhodes, G. (2017). A new other-race effect for gaze perception. *Journal of Experimental Psychology: Human Perception and Performance, 43*(11), 1857.

Coplan, R. J., & Armer, M. (2007). A "multitude" of solitude: A closer look at social withdrawal and nonsocial play in early childhood. *Child Development Perspectives, 1*(1), 26-32.

Coplan, R. J., Prakash, K., O'neil, K., & Armer, M. (2004). Do you" want" to play? Distinguishing between conflicted shyness and social disinterest in early childhood. *Developmental Psychology, 40*(2), 244.

DeBruine, L. M., Hahn, A. C., & Jones, B. C. (2016). Perceiving infant faces. *Current Opinion in Psychology, 7*, 87-91.

Dixon, J. C. (2006). The ties that bind and those that don't: Toward reconciling group threat and contact theories of prejudice. *Social Forces, 84*(4), 2179-2204.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*(5), 562-571.

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.

Elfenbein, H. A., & Ambady, N. (2002). Is there an in-group advantage in emotion recognition?

Enlow, D. H., & Moyers, R. E. (1982). *Handbook of facial growth*: WB Saunders Company.

Ert, E., Fleischer, A., & Magen, N. (2015). Trust and Reputation in the Sharing Economy: The Role of Personal Photos on Airbnb. *Tourism Management, 55*(1), 62-73.

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*: Oxford University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods, 4*(3), 272.

Field, A. (2009). *Discovering statistics using SPSS*: Sage publications.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences, 11*(2), 77-83.

Franklin, P., & Volk, A. A. (2017). A Review of Infants' and Children's Facial Cues' Influence on Adults' Perceptions and Behaviors.

Geary, D. C., & Bjorklund, D. F. (2000). Evolutionary developmental psychology. *Child development, 71*(1), 57-65.

Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PloS one, 10*(7), e0132726.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of personality and social psychology, 106*(1), 148.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision research, 41*(17), 2261-2271.

Haselhuhn, M. P., Ormiston, M. E., & Wong, E. M. (2015). Men's facial width-to-height ratio predicts aggression: A meta-analysis. PLoS One, 10(4), e0122637.

Havighurst, R. J. (1948). *Developmental tasks and education*. Chicago, US.: Chicago Press Developmental tasks and education.

Havighurst, R. J. (1972). Developmental tasks and education . New York: David McKay Company. *Inc. I l979*.

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. British Journal of Psychology, 99(1), 45-56.

Hehman, E., Sutherland, C. A., Flake, J. K., & Slepian, M. L. (2017). The Unique Contributions of Perceiver and Target Characteristics in Person Perception.

Ho, D. Y. F. (1986). Chinese patterns of socialization: A critical review.

Ho, D. Y. F., & Kang, T. K. (1984). Intergenerational comparisons of child-rearing attitudes and practices in Hong Kong. *Developmental Psychology, 20*(6), 1004.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.

Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual review of psychology, 68*, 269-297.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313-323.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement, 20*(1), 141-151.

Kenealy, P., Frude, N., & Shaw, W. (1988). Influence of children's physical attractiveness on teacher expectations. *The Journal of Social Psychology, 128*(3), 373-383.

Kesterke, M. J., Raffensperger, Z. D., Heike, C. L., Cunningham, M. L., Hecht, J. T., Kau, C. H., . . . Marazita, M. L. (2016). Using the 3D Facial Norms Database to investigate craniofacial sexual dimorphism in healthy children, adolescents, and adults. *Biology of sex differences, 7*(1), 23.

Kringelbach, M. L., Stark, E. A., Alexander, C., Bornstein, M. H., & Stein, A. (2016). On Cuteness: Unlocking the Parental Brain and Beyond. *Trends in cognitive sciences*.

Landy, J. F., Piazza, J., & Goodwin, G. P. (2016). When it's bad to be friendly and smart: The desirability of sociability and competence depends on morality. *Personality and social psychology bulletin, 42*(9), 1272-1290.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological bulletin, 126*(3), 390.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: the importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of personality and social psychology, 93*(2), 234.

Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

Lewis, G. J., Lefevre, C. E., & Bates, T. C. (2012). Facial width-to-height ratio predicts achievement drive in US presidents. *Personality and Individual Differences, 52*(7), 855-857.

Li, Q., Heyman, G. D., Mei, J., & Lee, K. (2017). Judging a Book by Its Cover: Children's Facial Trustworthiness as Judged by Strangers Predicts Their Real-World Trustworthiness and Peer Relationships. *Child development*.

Li, T., Liu, X., Pan, J., & Zhou, G. (2017). The interactive effect of facial appearance and behavior statement on trust belief and trust behavior. *Personality and Individual Differences, 117*, 60-65.

LoBue, V., & Thrasher, C. (2014). The child affective facial expression (CAFÉ) set. Databrary.

LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults. *Frontiers in Psychology, 5*, 1532.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). Karolinska Directed Emotional Faces (Psychology Section, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm.

Luo, L., Kendrick, K. M., Li, H., & Lee, K. (2015). Adults with siblings like children's faces more than those without. *Journal of Experimental Child Psychology, 129*, 148-156.

Maack, J. K., Bohne, A., Nordahl, D., Livsdatter, L., Lindahl, Å. A., Øvervoll, M., . . . Pfuhl, G. (2017). The Tromso Infant Faces database (TIF): development, validation and application to assess parenting experience on clarity and intensity ratings. Frontiers in Psychology, 8.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3.

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2012). The social evaluation of faces: a meta-analysis of functional neuroimaging studies. *Social cognitive and affective neuroscience, 8*(3), 285-299.

Mileva, V. R., Cowan, M. L., Cobey, K. D., Knowles, K. K., & Little, A. C. (2014). In the face of dominance: Self-perceived and other-perceived dominance are positively associated with facial-width-to-height ratio in men. Personality and Individual Differences, 69, 115-118.

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27*(4), 237-254.

Montepare, J. M., & Zebrowitz, L. A. (1998). Person perception comes of age: The salience and significance of age in social judgments. *Advances in experimental social psychology, 30*, 93-161.

Motta-Mena, N. V., & Scherf, K. S. (2016). Pubertal development shapes perception of complex facial expressions. *Developmental science*.

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5*(2), 159-168.

Nunnally, J. C. (1978). *Psychometric theory* (2 ed.). New York: McGraw Hill.

O'Toole, A. J., Peterson, J., & Deffenbacher, K. A. (1996). An ‘other-race effect’for categorizing faces by sex. *Perception, 25*(6), 669-676.

Oh, D., Dotsch, R., Porter, J., & Todorov, A. (2018). Gender Biases in Impressions from Faces: Empirical Studies and Computational Models. *Retrieved from psyarxiv. com/fxvcu*.

Olivola, C. Y., Sussman, A. B., Tsetsos, K., Kang, O. E., & Todorov, A. (2012). Republicans Prefer Republican-Looking Leaders Political Facial Stereotypes Predict Candidate Electoral Success Among Right-Leaning Voters. *Social Psychological and Personality Science, 3*(5), 605-613.

Olivola, C. Y., & Todorov, A. (2010a). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*(2), 83-110.

Olivola, C. Y., & Todorov, A. (2010b). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology, 46*(2), 315-324.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *PNAS, 105*(32), 11087-11092.

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological bulletin, 49*(3), 197.

Paquette, D. (2004). Theorizing the father-child relationship: Mechanisms and developmental outcomes. *Human development, 47*(4), 193-219.

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition, 24*(5), 607-640.

Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior, 20*(5), 295-307.

Phillips, D. A., & Shonkoff, J. P. (2000). *From neurons to neighborhoods: The science of early childhood development*: National Academies Press.

Picci, G., & Scherf, K. S. (2016). From caregivers to peers: Puberty shapes human face perception. *Psychological science, 27*(11), 1461-1473.

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law, 16*(6), 477-491.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PloS one, 7*(3), e34293.

Ritts, V., Patterson, M. L., & Tubbs, M. E. (1992). Expectations, impressions, and judgments of physically attractive students: A review. *Review of Educational Research, 62*(4), 413-426.

Rohner, R. P., & Rohner, E. C. (1981). Parental acceptance-rejection and parental control: Cross-cultural codes. *Ethnology, 20*(3), 245-260.

Rosenberg, S., Nelson, C., & Vivekananthan, P. (1968). A multidimensional approach to the structure of personality impressions. *Journal of personality and social psychology, 9*(4), 283.

Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological science, 19*(2), 109-111.

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of personality and social psychology, 104*(3), 409.

Said, C. P., Haxby, J. V., & Todorov, A. (2011). Brain systems for assessing the affective value of faces. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 366*(1571), 1660-1670.

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*(2), 260.

Salvia, J., Algozzine, R., & Sheare, J. (1977). Attractiveness and school achievement. *Journal of School Psychology, 15*(1), 60-67.

Scherf, K. S., Behrmann, M., & Dahl, R. E. (2012). Facing changes and changing faces in adolescence: A new model for investigating adolescent-specific interactions between pubertal, brain and behavioral development. *Developmental Cognitive Neuroscience, 2*(2), 199-219.

Scherf, K. S., & Scott, L. S. (2012). Connecting developmental trajectories: Biases in face processing from infancy to adulthood. *Developmental psychobiology, 54*(6), 643-663.

Schwartz, C.E.. Snidman, N., & Kagan, J. (1996). Early childhood temperament as a determinant of externalizing behavior in adolescence. *Development and Psychopathology, 8*(3), 527-537.

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), Person Perception and Interpersonal Behavior (pp. 300 –315). Stanford, CA: Stanford University Press.

Sheldon, W. H., Stevens, S. S., & Tucker, W. B. (1940). The varieties of human physique.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological science, 26*(5), 559-569.

South Palomares, J. K., Sutherland, C. A., & Young, A. W. (2017). Facial first impressions and partner preference models: Comparable or distinct underlying structures? *British Journal of Psychology*.

South Palomares, J. K., & Young, A. W. (2017). Facial first impressions of partner preference traits: trustworthiness, status, and attractiveness. *Social Psychological and Personality Science*, 1948550617732388.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological bulletin, 87*(2), 245.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust male facial width and trustworthiness. *Psychological science, 21*(3), 349-354.

Stolier, R. M., Hehman, E., & Freeman, J. B. (2017). A Dynamic Structure of Social Trait Space. *Trends in cognitive sciences*.

Sutherland, C. A., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: data-driven modelling of Chinese and British perceivers' unconstrained facial impressions. *Personality and social psychology bulletin*.

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition, 127*(1), 105-118.

Sutherland, C. A., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition, 157*, 257-267.

Sutherland, C. A., Rhodes, G., & Young, A. W. (2017). Facial image manipulation: A tool for investigating social perception. *Social Psychological and Personality Science, 8*(5), 538-551.

Sutherland, C. A., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology, 106*(2), 186-208.

Sutherland, C.A.M., Young, A.W., & Rhodes, G. (2017). Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties. *British Journal of Psychology, 108*, 397-415.

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion, 13*(4), 724.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Psychology, 66*(1), 519-545.

Vernon, R. J., Sutherland, C. A., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences, 111*(32), E3353-E3361.

Walker, M., Jiang, F., Vetter, T., & Sczesny, S. (2011). Universals and cultural differences in forming personality trait judgments from faces. *Social Psychological and Personality Science, 2*(6), 609-617.

Walker, M., & Vetter, T. (2009). Portraits made to measure: Manipulating social judgments about individuals with a statistical face model. *Journal of Vision, 9*(11), 12-12.

Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived Big Two and Big Five personality factors modeled in real photographs. *Journal of personality and social psychology, 110*(4), 609.

Walker, M., & Wänke, M. (2017). Caring or daring? Exploring the impact of facial masculinity/femininity and gender category information on first impressions. *PloS one, 12*(10), e0181306.

Wang, H., Han, C., Hahn, A. C., Fasolt, V., Morrison, D. K., Holzleitner, I. J., ... & Jones, B. C. (2019). A data-driven study of Chinese participants' social judgments of Chinese faces. *PloS one, 14*(1), e0210315.

Wolffhechel, K., Fagertun, J., Jacobsen, U. P., Majewski, W., Hemmingsen, A. S., Larsen, C. L., . . . Jarmer, H. (2014). Interpretation of appearance: the effect of facial features on first impressions and personality. *PloS one, 9*(9), e107721.

Zebrowitz, L. A. (2017). First Impressions From Faces. *Current Directions in Psychological Science, 26*(3), 237-242.

Zebrowitz, L. A., Brownlow, S., & Olson, K. (1992). Baby talk to the babyfaced. *Journal of Nonverbal Behavior, 16*(3), 143-158.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and social psychology review, 1*(3), 204-223.

Zebrowitz, L. A., Collins, M. A., & Dutta, R. (1998). The relationship between appearance and personality across the life span. *Personality and social psychology bulletin, 24*(7), 736-749.

Zebrowitz, L. A., Franklin, R. G., & Boshyan, J. (2015). Face shape and behavior: Implications of similarities in infants and adults. *Personality and Individual Differences, 86*, 312-317.

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin, 28*(2), 238-249.

Zebrowitz, L. A., Kendall-Tackett, K., & Fafel, J. (1991). The influence of children's facial maturity on parental expectations and punishments. *Journal of Experimental Child Psychology, 52*(2), 221-238.

Zebrowitz, L. A., Kikuchi, M., & Fellous, J. M. (2007). Are effects of emotion expression on trait impressions mediated by babyfaceness? Evidence from connectionist modeling. *Personality and Social Psychology Bulletin, 33*(5), 648-662.

Zebrowitz, L. A., & Lee, S. Y. (1999). Appearance, stereotype-incongruent behavior, and social relationships. *Personality and social psychology bulletin, 25*(5), 570-585.

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced individuals across the life span. *Developmental Psychology, 28*(6), 1143.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497-1517.

Zebrowitz, L. A., Olson, K., & Hoffman, K. (1993). Stability of babyfaceness and attractiveness across the life span. *Journal of personality and social psychology, 64*(3), 453.

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). " Wide-Eyed" and" Crooked-Faced": Determinants of Perceived and Real Honesty Across the Life Span. *Personality and social psychology bulletin, 22*(12), 1258-1269.