

Evidence for a Kernel of Truth in Children's Facial Impressions of Children's Niceness, but not Shyness

*Jemma R. Collova¹, Linda Jeffery¹, Gillian Rhodes¹, Ellen Bothe¹, Clare A.M. Sutherland^{1,2}

¹School of Psychological Science, The University of Western Australia, Crawley, Australia

²School of Psychology, University of Aberdeen, Aberdeen, Scotland

*Corresponding author: Dr Jemma Collova
jemma.collova@uwa.edu.au
The University of Western Australia, 35 Stirling Hwy,
Crawley, Western Australia 6009

Acknowledgements: We are grateful to the parents and children who helped make this research possible. We would like to thank Romina Palermo for providing us the opportunity to contact her sample of participants and to use some existing data. We also thank Lou Ewing for sharing the Zeb the Alien Scientist testing materials, and Saba Siddique for comments regarding a manuscript draft. Finally, we would like to thank Kaitlyn Turbett, Dielle Horne, Saba Siddique, Chloe Giffard, and Maira Vicente Braga, for help testing participants.

JC, LJ, GR, and CS conceived the study and helped to draft and edit the manuscript. JC programmed the experiment, collected most participant data, performed the statistical analyses and drafted the first manuscript draft. EB coordinated image collection and testing schedules. All authors participated in the study design, and read, provided critical revisions and approved the final manuscript. The study methods, hypotheses and analyses were pre-registered (<https://osf.io/kjtva/registrations>).

Funding: This research was supported by an APR Internship Academic Mentor Grant to CS, an Australian Research Council (ARC) Centre of Excellence Grant award to GR [CE110001021], ARC Discovery Early Career Research Award to CS [DE190101043], ARC Discovery Grant to GR and CS [DP170104602], ARC Discovery Grant to LJ [140101743], and a Research Training Program stipend to JC.

Abstract

Adults teach children not to ‘judge a book by its cover’. However, adults make rapid judgments of character from a glance at a child’s face. These impressions can be modestly accurate, suggesting that adults may be sensitive to valid signals of character in children’s faces. However, it is not clear whether such sensitivity requires decades of social experience, in line with the development of other face-processing abilities (e.g. facial emotion recognition), or whether this sensitivity emerges relatively early, in childhood. An important theoretical question therefore, is whether or not children’s impressions are at all accurate. Here, we examined the accuracy in children’s impressions of niceness and shyness from children’s faces. Children (aged 7–12 years, ~90% Caucasian) and adults rated 84 unfamiliar children’s faces (aged 4-11 years, 48 female, ~80% Caucasian) for niceness (Study 1) or shyness (Study 2). To measure accuracy, we correlated facial impressions with parental responses to well-established questionnaires about the actual niceness/shyness of those children in the images. Overall, children and adults formed highly similar niceness ($r = .94$) and shyness ($r = .84$) impressions. Children also showed mature impression accuracy: Children and adults formed modestly accurate niceness impressions, across different images of the same child’s face. Neither children nor adults showed evidence for accurate shyness impressions. Together, these results suggest that children’s impressions are relatively mature by middle childhood. Furthermore, these results demonstrate that any mechanisms driving accurate niceness impressions are in place by 7 years, and potentially before.

Keywords: accuracy, facial impressions, children, development, face perception

Evidence for a Kernel of Truth in Children's Facial Impressions of Children's Niceness, but not Shyness

People spontaneously infer personality characteristics from a glimpse of a stranger's face (for a review; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). For example, people form facial first impressions of trustworthiness, friendliness and competence (Oosterhof & Todorov, 2008; Rule, Krendl, Ivcevic, & Ambady, 2013). These facial impressions are pervasive: they form rapidly (Willis & Todorov, 2006), and influence peer interactions and social outcomes for both adults and children. Impressions of an adult's face influence their likelihood of winning political elections (Olivola & Todorov, 2010) and receiving harsh criminal sentences (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Zebrowitz & McDonald, 1991). Impressions of a child's face influence the intensity of discipline that they receive from adults (Berkowitz & Frodi, 1979), as well as their likelihood of being perceived as a leader (Zarbatany & Marshall, 2015) and being trusted by other children (Ewing, Sutherland, & Willis, 2019). Given the widespread influence of impressions on social outcomes, it is important to understand the origin and development of these crucial social judgements. Indeed, understanding the development of impression formation is a current priority within the field of person perception, and a topic of theoretical deliberation (see Mondloch et al., 2019; Over & Cook, 2018; Sutherland et al., 2020; Terrizzi, 2020).

Children's Trait Impressions

The tendency to form facial impressions emerges early in life. Evidence from preferential looking paradigms suggests that infants as young as seven months old are sensitive to facial signals of trustworthiness in adult faces (Jessen & Grossmann, 2016, 2019; Sakuta, Kanazawa, & Yamaguchi, 2018). Young children also show high agreement in character judgments from faces (Charlesworth, Hudson, Cogsdill, Spelke, & Banaji, 2019;; Palmquist, Cheries, & DeAngelis, 2019; Tang, Harris, Zou, & Xu, 2019). For example, three-

year old children agree with each other when assigning “nice” and “mean” judgments to pairs of computer-generated adult faces (Cogsdill, Todorov, Spelke, & Banaji, 2014).

Despite compelling evidence that young children form impressions from faces, there are mixed results regarding the age at which these impressions are mature. To measure impression maturity, children and adults are typically asked to discriminate between pairs of carefully controlled face images that have been pre-selected, or computer modified, to vary on a certain characteristic (e.g. Baccolo & Cassia, 2019; Charlesworth et al., 2019; Cogsdill & Banaji, 2015; Cogsdill et al., 2014; Mondloch, Gerada, Proietti, & Nelson, 2019; Nelson, Kennedy-Costantini, Lee, & Dixon, 2019; Palmquist et al., 2019; Terrizzi et al., 2018). In such studies, five-year-old children can show adult-like impressions, suggesting that the tendency to form facial impressions reflects a fundamental social cognitive capacity that emerges relatively early in life (Cogsdill et al., 2014; Ewing, Caulfield, Read, & Rhodes, 2015). However, other evidence from implicit impression judgments (e.g. through examining participant behavior) reveals that impressions continue to mature until at least 12 years of age (Mondloch et al., 2019; but see Ewing et al., 2019). This result suggests that the social experiences accumulated by late childhood are not adequate for adult-like impressions, but rather, that extended cultural learning of face-trait associations (Eggleston, et al., 2020; Over & Cook, 2018; Zebrowitz, et al., 2012) may be necessary for mature impression formation.

Impression Accuracy

An interesting feature of adults’ facial impressions is that they are sometimes (very) modestly accurate (but see Rule et al., 2013). For example, adults’ impressions of honesty (Bond, Berry, & Omar, 1994; Verplaetse, Vanneste, & Braeckman, 2007), health (Kramer & Ward, 2006), and sexual unfaithfulness (Leivers, Simmons, & Rhodes, 2015; Rhodes, Morley, & Simmons, 2013) from other adults’ faces show above-chance accuracy. This evidence suggests that there may be some valid signals of character in faces, and that adults

are sensitive to these cues. From an evolutionary perspective, accurately detecting these traits would be adaptive in the context of adult peer relationships. For example, accurately detecting valid signals of sexual unfaithfulness would help adults to avoid the costs of investing in an unfaithful partner (Rhodes et al., 2013). However, accurate impressions can also be explained by a self-fulfilling prophecy effect, whereby an individual who looks particularly honest is more likely to be trusted, thus encouraging that person to act more honestly (Zebrowitz et al., 1996).

Despite various theoretical accounts for accuracy, the question of whether or not first impressions are at all accurate remains a contentious topic of debate (Bonnefon et al., 2015). For instance, some impressions clearly stem from inaccurate group stereotypes (e.g. McCrae et al., 2013) and some trait impressions from faces show no evidence of accuracy (Efferson & Vogt, 2013; Rule et al., 2013). We also know relatively little about the etiology of accuracy in adults' impressions. For instance, it is not clear whether any such ability requires decades of accumulated social experiences. An important question, therefore, is whether or not children's impressions are at all accurate. Observing any accuracy in children's impressions would be theoretically important, as this result would suggest that social experiences and perceptual expertise prior to adolescence are adequate for impression accuracy.

Might Children's Impressions be Accurate?

To our knowledge, only one study has directly examined the accuracy in children's trait impressions. Li, Heyman, Mei, and Lee (2017) investigated the accuracy in Chinese children's impressions of trustworthiness from other Chinese children's faces, motivated by the predicted importance of trust in child peer relationships (e.g. Bernath & Feshbach, 1995). To examine impression accuracy, the authors took standardized images of children's faces (e.g. controlled for expression, background, lighting, and so on) and also measured trait trustworthiness as reported by the peers of the photographed children. The authors then used

child (aged 8.4 – 11.5 years) and adult trustworthiness impressions of those faces to predict measures of reported trait trustworthiness. Both children's and adults' impressions of trustworthiness were modestly accurate (children: $r = .27$; adults: $r = .26$). This study is important, as it reveals that children's impressions have the potential to be accurate, at least for standardized face images and in a Chinese population context.

Other evidence suggestive of accuracy comes from investigating the overall maturity of children's impressions. Children's and adults' impressions from standardized face images can show high agreement (Cogsdill & Banaji, 2015; Cogsdill et al., 2014). Assuming that there are valid signals of character in faces, and that children and adults use these cues in similar ways, then children's impressions might also show similar accuracy to those of adults. For example, both 8-year old children and adults use facial width-to-height-ratio to cue impressions of aggression from adult faces (Short, et al., 2012). However, there may be subtle differences between children's and adults' impressions (Baccolo & Cassia, 2019; Caulfield, Ewing, Bank, & Rhodes, 2016; Ma, Xu, & Luo, 2015; Mondloch et al., 2019; Palmquist et al., 2019). For example, compared with adults, children from 8-10 years old are less likely to use attractiveness to cue impressions of trustworthiness (Ma et al., 2015), and children from 4-10 years old do not always rely on the same emotional cues when forming implicit trait judgments of adult faces (Mondloch et al., 2019). Children may therefore, form less accurate impressions if they are less sensitive to any potentially valid signals, as compared to adults.

There are also other reasons why children may form less accurate impressions from faces than adults. Generally speaking, children show less mature face processing skills (e.g. Baker, Laurence, & Mondloch, 2017; Mondloch, Geldart, Maurer, & Le Grand, 2003; Mondloch, Le Grand, & Maurer, 2002; Pellicano & Rhodes, 2003) and social cognitive abilities (e.g. Flavell, 1999; Liu, Gelman, & Wellman, 2007; Miller & Aloise, 1989),

compared to adults. For example, young children are relatively poor at predicting social behaviors from previous trait-relevant behaviors (Liu et al., 2007). These immature visual and social-cognitive abilities may limit children's ability to form accurate facial impressions. Indeed, De Neys, Hopfensitz, and Bonnefon (2015) found that adolescents (aged 13 years) formed less accurate impressions of trustworthiness than adults (aged 18), suggesting that impression accuracy may be a skill which continues to develop over time. Accuracy might improve as social cognitive abilities mature and people learn about valid face-trait associations with cultural and social experience (consistent with FeldmanHall, et al., 2018; Over & Cook, 2018; Sutherland, et al., 2020; Zebrowitz, et al., 2012).

In summary, it remains an open question as to whether or not children's impression accuracy is adult-like, and research directly addressing this question is limited. It is not clear whether previous evidence for accuracy in children's impressions from standardized images of Chinese children's faces (Li, et al., 2017) would generalize to more naturalistic images, and to images of children from other cultures. Furthermore, it is not clear whether impression accuracy is dependent on mature facial impressions in general. Finally, to date, most studies have examined impression accuracy for adult faces (e.g. De Neys et al., 2015). If there is any accuracy in children's impressions, it is likely to be for impressions of other children's faces (consistent with other own-age advantages in face perception: see Rhodes & Anastasi, 2012).

Present Studies

Here, we investigated the accuracy in primary-school aged children's impressions of other primary-school aged children's faces. Children may be especially sensitive to any valid signals in other children's faces. Moreover, children frequently form impressions of their peers, particularly across the primary-school-aged years as their social world rapidly expands. From a practical perspective, it is therefore especially important to understand the nature of

children's impressions of their peers, although almost nothing is known about these impressions (for exceptions, see Cogsdill & Banaji, 2015; Ewing et al., 2019; Li et al., 2017).

To date, the studies which have investigated children's impressions of children's faces have focused on specific traits predicted to be important in child-peer relationships (e.g. trustworthiness; Li et al., 2017). Here, we examined the accuracy of two traits, niceness and shyness, that have been identified as important for children's faces through a previous data-driven approach (Collova, Sutherland, & Rhodes, 2019). Niceness and shyness impressions capture the most variance in adults' impressions of children's faces, and may reflect the by-product of mechanisms which have evolved to guide nurturing and caregiving behaviors towards children (Collova et al., 2019). Considering the importance of these traits, it is possible that there are valid signals of niceness/shyness in children's faces. Interestingly, one recent study found that adults' impressions of niceness, but not shyness, from children's faces were modestly accurate (Collova et al., 2020). The theoretical importance of these traits has only recently been understood, and so no study has asked whether children's impressions of these traits are accurate. It might be important for children in social interactions to accurately detect whether another child will play nicely (i.e. niceness), or be willing to socially engage (i.e. shyness). This idea aligns with an ecological perspective of person perception (McArthur & Baron, 1983; Zebrowitz & Montepare, 2006), whereby the qualities that we perceive in other people serve an adaptive function in guiding our behavior. Alternatively, impressions of these traits might be accurate due to self-fulfilling prophecy effects.

In two studies, we investigated children's impressions of niceness (Study 1) and shyness (Study 2) from the faces of their peers, using images of children from a previous study (Collova et al., 2020). So that we could benchmark the accuracy in children's impressions, we also collected adults' (the caregivers of the child participants) impressions of those same faces. Participants who came into the lab were randomly assigned to rate either

niceness or shyness. We pre-registered the study methods, hypotheses and analyses (<https://osf.io/kjtva/registrations>).

We specifically recruited children 7-12 years of age because general face processing abilities are still developing during this period (e.g. emotional facial recognition and facial identity recognition; Kadosh & Johnson, 2007; Meinhardt-Injac et al., 2020; Thomas et al., 2007). There is also mixed evidence as to whether or not impression judgments by 12 years of age are mature. Although some evidence reveals mature impressions by 12 years of age (Ewing et al., 2019), other evidence reveals impressions are still developing (Mondloch et al., 2019). Furthermore, research investigating the maturity of children's impressions has so far mainly used faces that have been pre-selected to vary on a specific trait based on adults' ratings (Caulfield et al., 2016), examined children's impression of computer-modified (Ewing et al., 2019; Mondloch et al., 2019) or computer-generated (Charlesworth et al., 2019; Cogsdill et al., 2014; Palmquist et al., 2020) faces, and has measured impression maturity by comparing proportion responses in two-alternative forced choice tasks (Cogsdill & Banaji, 2015; Palmquist, et al., 2019). Although this research has been pivotal in establishing the early emergence of impression formation, these approaches have not left children and adults much room to vary. We propose that a strict measure of impressions maturity would be to examine the overlap in children's and adults' trait ratings from naturalistic face images, that vary as faces do in everyday life. It remains an open question as to whether or not children aged 7-12 years will show more subtle differences in their impressions as compared to adults, when forming impressions under more naturalistic conditions, with naturalistic face images.

We had two main aims. First, we wanted to determine whether children's impressions of children's faces were at all accurate, and if so, whether this accuracy was similar to any accuracy in adults' impressions. If children show an adult-like pattern of accuracy, this result would suggest that children and adults rely on the same valid signals when forming these

impressions. Furthermore, this result would suggest that any social-cognitive abilities, or social experiences accumulated by childhood are adequate for impression accuracy.

To measure accuracy, we correlated children's and adults' niceness and shyness impressions of unfamiliar children's faces, with actual measures of niceness and shyness for the children shown in the images. To index actual niceness/shyness, we analysed parents' responses on well-established questionnaires regarding nice and shy behaviors, following Collova et al. (2020). To the extent that both children and adults are sensitive to valid signals in faces, we might expect modest accuracy for impressions of niceness, but not shyness (consistent with Collova et al., 2020), although we were open to other results, as discussed above. For example, children's impressions might be less accurate than adults' given children's less mature face processing (e.g. Mondloch et al., 2003) and social cognitive (e.g. Liu et al., 2007) abilities. Alternatively, children may be more accurate than adults, in the context of own-age advantages (e.g. Kuefner, Macchi Cassia, Picozzi, & Bricolo, 2008; for a review see Rhodes & Anastasi, 2012) as we used children's faces. As an exploratory analysis, we examined whether the amount of contact with children acted as a potential mediator for accuracy, as an own-age advantage account would predict.

Our second aim was to determine the similarity of children's and adults' impressions, irrespective of accuracy. Investigating the maturity of children's facial impressions is an area of current theoretical interest (Mondloch et al., 2019; Nelson et al., 2019; Over & Cook, 2018; Terrizzi et al., 2020). To date, the face images used to investigate children's impressions have not been very natural. For example, face images have been taken under highly controlled lab conditions (Caulfield et al., 2016; Cogsdill & Banaji, 2015; Ewing et al., 2019; Li et al., 2017; Mondloch et al., 2019; Short et al., 2012), computer-modified (e.g. morphed Caulfield et al., 2016; Ewing et al., 2019; Mondloch et al., 2019), or entirely computer-generated (Charlesworth et al., 2019; Cogsdill et al., 2014; Ma et al., 2015;

Palmquist et al., 2019; Tang et al., 2019). Furthermore, in most of these studies children were asked to choose between two faces that had been carefully pre-selected to vary on a specific characteristic. These stimuli allow precise control over variables of interest, and therefore have proven very useful in determining whether or not children are at all sensitive to particular facial information, as a starting point. However, it is not clear that children will show a similar level of impression maturity when dealing with more natural faces that vary on many attributes, such as, expression, viewpoint, lighting and pose. It is important to investigate children's impressions of more naturalistic face images, as these conditions better represent people's experience with faces in everyday life (also see Laurence & Mondloch, 2016). If children do not show mature impression formation for more naturalistic face images, then the impressions they form in everyday life might be less mature than initially predicted.

We examined children's impressions using naturalistic ambient images that varied in everyday cues (for importance of ambient images, see Jenkins, White, Van Montfort, & Burton, 2011; Laurence & Mondloch, 2016; Sutherland et al., 2013). We randomly assigned children to rate the images of children's faces for either niceness/shyness. To measure impression maturity, we then directly compared children's trait impressions with adults' impressions (collected here), and also with an independent sample of young adults' impressions (collected previously; Collova, et al., 2020). By directly comparing these explicit trait ratings, we were able to examine the overlap of children's and adults' impressions at the face level. We expected that asking children to provide trait ratings would elicit meaningful variance in their impressions, perhaps not captured in prior research using two-alternative forced choice tasks. This method may therefore provide a strict test of impression maturity.

Study 1: Are Children's Impressions of Niceness Accurate?

We examined the accuracy in children's and adults' impressions of niceness, a trait which represents the primary dimension for impressions of children's faces. We recruited child (aged 7-12 years) and adult (the caregivers of those children) participants. Participants judged the niceness of unfamiliar children's faces. To determine impression accuracy, we correlated niceness impressions with parental responses on well-established questionnaires assessing nice behaviors for those children (from here on referred to as actual niceness).

We were primarily interested in children's and adults' impression accuracy at the group level. As an exploratory analysis, we also considered individual-rater accuracy (following Carré et al., 2009; Collova, et al., 2020; Foo, Loncarevic, Simmons, Sutherland, & Rhodes, 2019; Sutherland et al., 2018) which provides insight into the extent to which individual raters should rely on their impressions, and offers a strict test of accuracy as it does not average across individual-rater error. Our sample size for individual-rater analyses was relatively small, and so we report these analyses in our Supplementary Materials.

Method

Participants and Power

Our final participant sample, after exclusions (see below), consisted of 21 children (12 female, $M = 10$ years 4 months, $SD = 1$ year 7 months, range = 7 years -12 years 7 months, 17 Caucasian: See Supplementary Materials Figure S5 for age distribution), and 20 adults (all female, $M = 42$, $SD = 3$, range = 34 - 49 years¹, 17 Caucasian). Adult participants were the caregivers of the children (all mothers). The child participants had previously volunteered for experiments in our lab, and attended a wide range of different primary schools.

Our primary analyses were at the face level. That is, for each face, we averaged across participants' ratings and analyses were performed on these averaged ratings. Critically,

¹ One adult participant's age was not known.

therefore, power was determined by the number of face identities in our experiment. A power analysis with a power of .80 and an effect size from a comparable previous study (Collova et al., 2020: Pearson's $r = .368$), revealed we would need 55 face identities to observe a similar level of accuracy for impressions of niceness. Our face stimuli sample size exceeded this number ($N = 84$ face identities, 420 total images: see Materials for detail).

To ensure reliable face ratings at the participant-group level, we recruited as many child and adult participants as possible during the two-week school holiday period. We set a minimum sample size of 15 participants per group, which has been found sufficient for good reliability at the group level (Little, Roberts, Jones, & DeBruine, 2012) including for comparable ambient images (Collova et al., 2019; Sutherland et al., 2013). Our participant sample size exceeded this limit, although our critical analyses were at the face level.

We excluded one additional adult participant who gave the same rating for an entire block. We excluded an additional eight child participants whose sibling had also completed the task (chosen at random, unless there was reason to exclude one sibling over the other, e.g. poor attention). Our rationale for these exclusions was that siblings may form similar impressions, consistent with heritability for face identity recognition (Wilmer et al., 2010). Because the child and adult participants were not independent (i.e. adults were the caregivers of children), we also compared children's impressions to impressions from an independent sample of young adults' impressions, collected previously (Collova, et al., 2020). Sample size and exclusion criteria were pre-registered. All exclusions were made before any analysis.

Materials

Child Face Images

Stimuli consisted of five ambient images for each of the 84 children (420 total images; 48 female, $M_{\text{age}} = 8$, range = 4-11 years, 68 Caucasian, 9 Asian, 4 Maori, 2 African, 1 Anglo Indian). Different images of the same person can lead to different impressions

(Jenkins et al., 2011; Sutherland, Young, & Rhodes, 2017; Todorov & Porter, 2014). For example, it is possible for the same person to look highly trustworthy in one image, and highly untrustworthy in another image, because of subtle changes in expression. It is therefore not certain that accurate impressions of one image would generalize to other images of that same person's face. Here, we tested whether accuracy would hold across different images of the same child's face despite within-person variability in impressions, providing a stringent test of accuracy (following Collova et al., 2020).

The ambient child-face images were previously collected for another study (Collova, et al., 2020). We obtained permission to use 84 of the original 86 identities used in that previous study. Images were sent in by the parents of the children, and were free to vary on expression, hairstyle, viewpoint, lighting, and so forth (see Figure 1 for example images). There were no restrictions regarding variability, as long as the child's face was clearly visible. Parents were asked to send the five most recent photographs they had on their phones, although we cannot be certain this is how parents selected the images. If parents sent in more than five photographs we selected the highest quality images (determined by image resolution). Stimuli were cropped with an oval mask around the face, were re-sized to a standardized size of 180pixels wide (resolution = 72KB; approx. 4cm wide on screen), but were otherwise unmodified.

Figure 1

An example of five ambient images of one female (top) and one male (bottom) face identity.

These two identities were the practice stimuli in our experiment. The authors received consent for these images to be published in this article.



Trait Measure: Niceness

We used parents' responses on the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) to index actual niceness for those children we had images of. The SDQ is a well-established, short questionnaire comprised of five scales. To measure niceness, we averaged scores from the two scales that were most relevant to the niceness dimension; the prosocial behavior and conduct problem scales (following our pre-registration; we never analyzed results from the other scales). The prosocial scale includes five statements about positive social behavioral tendencies, such as, "my child is considerate of other people's feelings". The conduct problem scale includes five items related to behavioral problems, such as, "my child often fights with other children or bullies them". Parents responded to each statement on a three-point scale (0 = not true, 1 = somewhat true, 2 = certainly true).

To create an overall actual niceness measure, we reverse scored the conduct problems scale, and then averaged the sum of the two scale scores together (following Collova et al., 2020). Therefore, scores could range from 0-10, with higher scores signalling greater

niceness. The SDQ has good validity and reliability (Hawes & Dadds, 2004; Seward, Bayliss, Stallman, & Ohan, 2018; Stone, Otten, Engels, Vermulst, & Janssens, 2010). Reliability was also good for our current sample (prosocial scale: $\alpha = .84$; conduct problems scale: $\alpha = .65$; combined scales: $\alpha = .78$).

Procedure

Participants completed the study on a computer in our lab during the two-week school holidays. Participants rated individual images of each child's face for how nice they looked on a scale of 1 (*not at all nice*) to 9 (*extremely nice*). The five images of each child's face were distributed across five different image sets (i.e. the distributed-image task: Collova et al., 2020). Image set and face order were randomized for each participant.

We embedded the task within a story about 'Zeb the Alien Scientist' for both the child and adult participants (adapted from Caulfield, Ewing, Burton, Avard, & Rhodes, 2014). Participants helped Zeb the Alien complete a mission to learn more about humans. We asked participants to help teach Zeb about how nice people on Earth look. First, to make sure children understood this term, we asked them to explain what it "means to be nice". Children provided verbal responses to the experimenter. Second, to familiarise participants with the niceness scale, and to make sure they understood the scale ends, participants completed practice sentences which described a nice behavior ("how nice is someone who shares their toys with other children?") and a not-nice behavior ("how nice is someone who laughed at their best friend when they had to get glasses?"). To help quantify the scale intervals, under each scale number (1-9) we included a visualisation of a cup with increasing volumes of water (following Caulfield et al., 2014: Figure 2). Responses to these questions confirmed that all participants understood the term nice and the scale endpoints. Adult participants were not asked to provide verbal responses, and self-progressed through the instructions on a computer. Importantly, during these instructions, no face images were shown to participants.

so we also asked participants to press ‘space’ for any face they knew. On average, each child participant indicated they knew two other children, and these trials were treated as missing.

Participants came into the lab, and completed the task on a computer. To minimize any potential influence of an adults’ presence on their child’s impressions (or vice versa) each child and adult completed the study in separate rooms. An experimenter stayed with each child for the duration of the task, and the adult completed the task at their own pace. The task lasted approximately 40 min, with breaks. At the end of the testing session, adult participants answered some basic demographic questions about themselves and their child (e.g. age, sex, amount of contact with children: see Supplementary Materials Section 1). Participants also completed another task (after this task), as part of a different study and were reimbursed \$20. Adults and children provided written and verbal consent to participate. Ethical approval was obtained from the Human Research Ethics Office at the University of Western Australia (RA/4/1/6805: Why does face identification ability improve during childhood?; and RA/4/8710: First impressions of children’s faces).

Results

Actual Niceness

We used scores from the SDQ to index actual niceness (see Table 1 for descriptive statistics). The SDQ scores from our sample were similar to norms reported elsewhere (Mellor, 2005). Scores were not normally distributed (most scores were in the higher end of the distribution; Shapiro-Wilk (84) = .862, $p < .001$), and therefore we report both parametric (Pearson’s r) and non-parametric (Spearman’s ρ) measures of association throughout, although they produce essentially identical patterns of results.

Accuracy

Our primary analyses were at the face level. That is, within each image set, we calculated a mean niceness impression for each face by averaging niceness impressions

across participants, separately for the child (Figure 3a) and adult (Figure 3d) participants (Table 1 for descriptive statistics). For each face, we excluded the rating of any participant who indicated they knew any of the face images for that same child.

Table 1

Descriptive statistics for actual niceness scores (as measured on the Strengths and Difficulties Questionnaire) and impressions of niceness, averaged across participants' ratings, separately for children and adults

		<i>M</i>	<i>SD</i>	Range	Skew	Kurtosis
<i>Criterion:</i>	<i>Actual Niceness</i> ^a	8.5	1.4	3.5 – 10	-1.42	2.15
Children	Image Set 1	6.2	1.0	2.6 – 7.8	-1.1	1.3
	Image Set 2	5.9	1.1	3.0 – 7.7	-0.6	-0.4
	Image Set 3	6.0	1.0	3.5 – 7.8	-0.9	-0.1
	Image Set 4	5.8	1.2	2.3 – 7.9	-0.8	0.2
	Image Set 5	6.0	1.0	2.9 – 7.6	-0.8	0.2
	Average impression	6.0	0.7	3.8 – 7.3	-0.5	-0.1
Adults	Image Set 1	6.3	0.8	3.4 – 7.7	-0.7	0.7
	Image Set 2	6.2	0.9	3.8 - 7.6	-0.6	-0.1
	Image Set 3	6.3	0.9	3.9 – 7.7	-0.8	-0.2
	Image Set 4	6.1	1.0	3.4 – 8.0	-0.8	0.5
	Image Set 5	6.3	0.84	3.9 – 7.9	-0.6	0.1
	Average impression	6.3	0.63	4.7 – 7.7	-0.2	-0.1

^a Scores could range from 0 – 10.

N = 84 faces.

We were interested in whether any accuracy would hold across the five image sets. This question is theoretically important because recent evidence reveals that different photos of the same person can convey very different impressions (Jenkins et al., 2011; Todorov & Porter, 2014). To measure accuracy, we correlated the niceness impressions with actual niceness scores, separately for each image set. That is, we ran five separate correlation analyses, so that each correlation only contained the average trait rating for one image of each child (and ensuring there was no nested structure in our analyses). We observed small-medium significant correlations in almost all of the image sets, for both the child and adult participant groups (Child participants: r range = .145 to .247; Adult participants: r range = .179 to .291; Table 2), revealing modest accuracy for impressions of niceness across different images of the same child's face.

We were also interested in comparing the variability in niceness impressions from different images of the same child, with the variability in impressions from images of different children. Evidence from impressions of adults' faces (Todorov & Porter 2014) has found that different photos of the same person (within-person variability) can convey impressions that vary as much as impressions from images of entirely different people (between-person variability). To measure within-person variability, we calculated the average variability in impressions across the five different images of each child's face. We calculated variance on the judgments averaged across the five images of each child, consistent with previous approaches (Todorov & Porter 2014; Collova et al., 2020). Indeed, we also found that impressions of niceness showed as much variability across different images of the same child, as across images of different children. Importantly, our correlational analyses revealed that impressions were modestly accurate in each image set, and therefore, our results demonstrate that impression accuracy can withstand at least some degree of variation across images of the same person.

Returning to the question of accuracy, we also examined the accuracy of impressions averaged across the five image sets. Consistent with the results for the individual image sets, there was a medium, positive, and significant correlation between niceness impressions and actual niceness for both child and adult participant groups, and no evidence for a difference in accuracy between the two groups (Table 2, Figure 3c, 3f). Meta-analytical statistics (weighted by sample size) revealed almost identical results to the correlational analyses, that is, child and adult niceness impressions were accurate across the five image sets (Figure 3b, 3e). There was no compelling evidence that impressions became more accurate when averaged across the five image sets (Supplementary Materials section 2, Table S1), suggesting that any noise in the individual images did not reduce accuracy. Results also remained significant after controlling for face sex (see Supplementary Materials section 3.1 and Table S3). Finally, results at the individual-rater level (see Supplementary Materials Section 4.1 for individual-rater level analyses) revealed a consistent pattern of modest accuracy for the child and adult participants (see Table S5 for significance tests against chance). These results demonstrated that impression accuracy was not driven by a few extreme participants, and that not every child or adult formed significantly accurate impressions.

Table 2

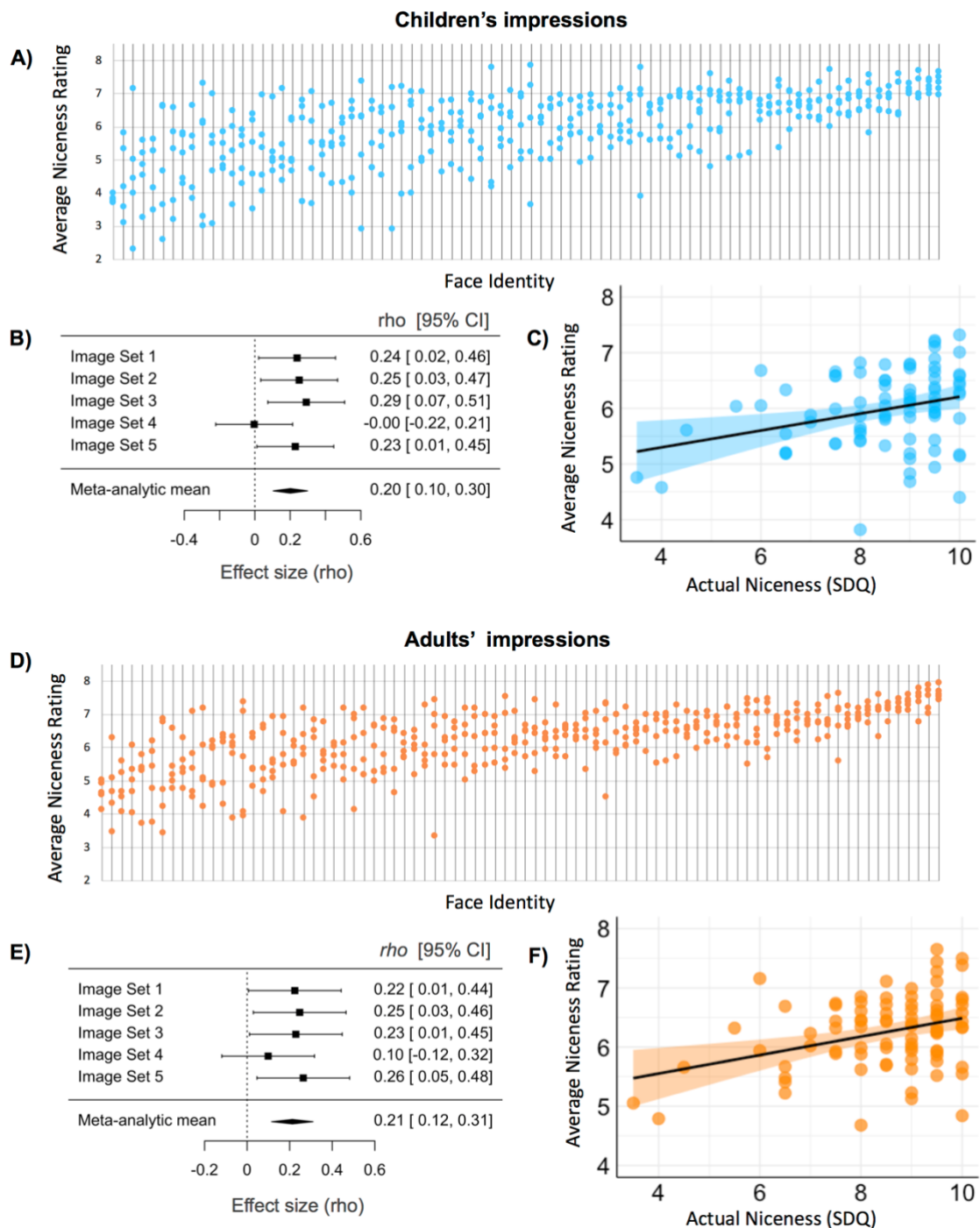
Spearman's rho (rho) and Pearson's r (r) correlations between actual niceness and niceness impressions averaged across participants, separately for children and adults. Significance tests (Fisher Z transformation; Steiger, 1980) comparing accuracy (rho and r) between the children's and adults' average impressions.

Group-level accuracy		Spearman's rho		Pearson's r	
		rho	p	r	p
Children	Image set 1	.239	.029	.135	.220
	Image set 2	.251	.021	.246	.024
	Image set 3	.291	.007	.247	.023
	Image set 4	-.004	.974	.145	.188
	Image set 5	.229	.036	.223	.042
	Average Rating	.265	.015	.304	.005
Adults	Image set 1	.224	.040	.179	.103
	Image set 2	.247	.023	.262	.016
	Image set 3	.229	.036	.269	.013
	Image set 4	.099	.371	.228	.037
	Image set 5	.264	.015	.291	.007
	Average Rating	.259	.017	.348	.001
Fisher's z	Average Rating	z = 0.15	.883	z = 1.18	.236

N = 84 faces.

Figure 3

Child (A-C) and adult (D-F) participants' impressions of niceness at the face level: A/D) Niceness impression for each of the five images of each face identity, averaged across participants. Each vertical line represents one face identity ($N = 84$). Each circle represents the average rating for one of the five images of each identity and demonstrates the large variability in niceness impressions across the different images. Identities are ordered on the x-axes according to their mean niceness impression. B/E) Parameter estimates (effect size and confidence intervals calculated by weighted sample size) for impression accuracy within each individual image set, and their overall meta-analytic mean effect size (Spearman's rho). C/F) Scatterplot of the association between actual niceness (from the Strength and Difficulties Questionnaire; SDQ) and niceness impressions averaged across the five image sets, revealing modest accuracy. Each point represents one face identity, with darker transparency representing more points superimposed. We plot the line of best fit and 95% confidence intervals.

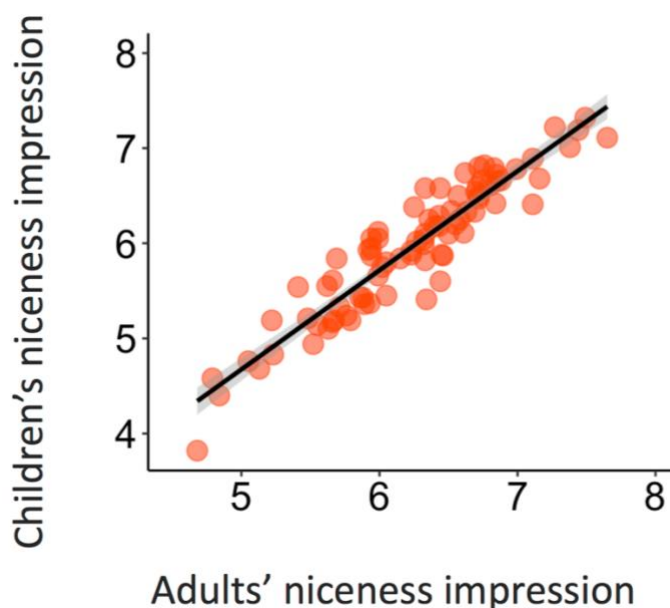


How Similar are Children’s and Adults’ Niceness Impressions?

Setting the question of accuracy aside, we also directly examined the similarity between children’s and adults’ impressions of niceness, averaged across the image sets. There was a large, positive correlation between children’s and adults’ impressions collected here, Pearson’s $r = .94$, $p < .001$, $N = 84$ (Figure 4). We also compared children’s impressions with impressions from an independent sample of young adult participants, collected previously (Collova et al., 2020), and found a large correlation, Pearson’s $r = .84$, $p < .001$, $N = 84$ (see Supplementary Materials Section 5.1., and Table S7 for more detail). Thus, impressions of niceness made by children and adults were highly similar, irrespective of their accuracy.

Figure 4

Scatterplot of the association between children’s and adults’ impressions of niceness ($r = .94$, $p < .001$). Each point on the figure represents ratings averaged across the five images for one child-face identity ($N = 84$). We plot the line of best fit and its confidence intervals ($\pm 95\%$).



Summary

Children showed modest accuracy for impressions of niceness. Niceness impressions were accurate across different images of the same child's face, at both the participant-group and individual-rater level, and despite considerable variability in impressions of niceness across those different images. Across all analyses, children showed an adult-like pattern of both impressions, and impression accuracy.

Study 2: Are Children's Impressions of Shyness Accurate?

In Study 2, we examined the accuracy of shyness impressions, a trait which represents the second dimension for adults' impressions of children's faces (Collova et al., 2019). A new sample of child and adult participants completed the same procedure as in Study 1, but here judged the shyness of the children's faces. To measure accuracy, we correlated shyness impressions with parental reports of shy behavior (referred to as actual shyness).

Method

Participants

Our final participant sample (after exclusions) consisted of 17 children (11 female, $M = 10$ years 3 months, $SD = 1$ year 5 months, range = 8 years 8 months - 12 years 8 months, Caucasian $N = 16$: See Supplementary Materials Figure S5 for age distributions), and 17 adults (13 female, $M = 45$ years, $SD = 10.6$, range = 30-75 years, Caucasian $N = 16$). Adult participants were the caregivers of the children (12 mothers, 4 fathers, 1 grandmother). Participants were recruited from the same sample of participants in Study 1, and also came into the lab during a two-week school holiday period.

We followed the same sample size and exclusion criteria as in Study 1 (as also specified in our OSF pre-registration). We excluded two child participants (siblings) who did not understand the task, and one adult participant who did not complete the entire task. We

also excluded data from two child participants whose siblings had completed the task (chosen at random, data never analysed) so that children's observations remained independent.

Materials

Trait Measure: Shyness

Data regarding actual shyness were collected from the same sample of children as in Study 1. To measure shyness, we used parents' responses on the Colorado Childhood Temperament Inventory (CCTI; Buss & Plomin, 1984; Rowe & Plomin, 1977). We analysed scores from the five items that comprised the shyness subscale of the CCTI, for example, my "child tends to be shy" (following Collova et al., 2020). Parents responded to these items on a 5-point scale, ranging from 1 (*not at all like this / strongly disagree*) to 5 (*a lot like this / strongly agree*). Scores could range from 5-25, with higher scores indicative of shyer behaviors. The CCTI has good validity (Webster-Stratton & Eyberg, 1982) and reliability (shyness scale $\alpha = .88$; Buss & Plomin, 1984). Reliability was also good in our sample (shyness scale $\alpha = .85$).

Stimuli and Procedure

We used the same stimuli and procedure as in Study 1, except that participants rated the children's faces for how shy they looked on a scale of 1 (*not at all shy*) to 9 (*extremely shy*). In the context of our Zeb the Alien Scientist story, participants helped teach Zeb about how shy people on Earth look. We asked children to explain to Zeb what it means to be shy. Participants also answered two practice questions, describing a shy behavior ("how shy is someone who never wants to answer a question in front of the whole class?") and a non-shy behavior ("how shy is someone who does a performance in front of the whole school?"). Responses to these questions confirmed that all children understood the term shy and the scale endpoints.

Results

Actual Shyness

To index actual shyness, we analysed parents' responses to questions from the shyness scale of the CCTI (Table 3 for descriptive statistics). Scores from this scale were not normally distributed; Shapiro-Wilk (84) = .955, $p = .005$. Most scores were in the lower end of the distribution, and were slightly lower than norms reported elsewhere (Rowe & Plomin, 1997), revealing our sample of children was not especially shy. Therefore, we report both parametric (Pearson's r) and non-parametric (Spearman's ρ) measures of association here and throughout, although the pattern of results is consistent between the two.

Table 3

Descriptive statistics for actual shyness scores (shyness scale of the Colorado Childhood Temperament Inventory) and impressions of shyness, averaged across participants' ratings, separately for the children and adults.

		<i>M</i>	<i>SD</i>	Range	Skew	Kurtosis
<i>Criterion:</i>	<i>Actual Shyness^a</i>	10.9	4.0	5 - 21	0.4	-0.7
Children	Image Set 1	4.3	0.9	2.5 - 6.9	0.5	0.0
	Image Set 2	4.5	1.2	1.9 - 7.2	0.2	-0.9
	Image Set 3	4.5	1.2	2.4 - 7.1	0.5	-0.6
	Image Set 4	4.3	1.2	2.0 - 7.5	0.6	-0.1
	Image Set 5	4.3	1.1	2.0 - 7.1	0.4	-0.3
	Average Rating	4.4	0.7	2.8 – 6.7	0.3	0.6
Adults	Image Set 1	4.0	1.1	1.8 - 6.7	0.3	-0.6
	Image Set 2	4.1	1.2	1.8 – 7.0	0.1	-0.9
	Image Set 3	4.2	1.2	1.7 - 7.1	0.3	-0.5
	Image Set 4	4.2	1.2	1.7 - 6.8	0.1	-0.8
	Image Set 5	4.1	1.1	2.0 - 6.8	0.2	-0.5
	Average Rating	4.1	0.7	2.3 – 6.4	0.2	0.5

^a Scores could range from 5-25.

N = 84 faces.

Accuracy

Within each image set, we averaged shyness impressions across participants separately for the child (Figure 5a) and adult (Figure 5d) participant groups, and correlated these impressions with actual shyness (see Table 4). The pattern of accuracy was identical for

both children and adults: impressions were not significantly accurate in any image set (Child participants: r range = $-.033$ to $.106$; Adult participants: r range = $-.079$ to $.201$; Table 4). It was possible that impressions were not accurate in each of the different image sets because impressions varied substantially across the different images of the same child's face. Indeed, the within-person variability in shyness judgments was particularly high compared to the between-person variability (Child participants: within-person variance = 1.05 , between-person variance = 0.47 ; Adult participants: within-person variance = 1.13 , between-person variance = 0.49).

When impressions were averaged across the five image sets, the correlation was not significant (Figure 5c, 5f). There was no significant difference in accuracy between impressions from the child and adult participant groups (Table 4). Meta-analytic statistics (weighted by sample size) also showed the same result as the correlational analyses, that is, no compelling evidence for accuracy for neither children nor adults (Figure 5b, 5e). There was no evidence that the averaged impressions were more accurate than impressions from the individual image sets, suggesting that any random noise introduced by single images did not confound accuracy (Supplementary Materials Table S2). Results were consistent for female and male faces, and remained non-significant after controlling for face sex (See Supplementary Materials Section 3.2. and Table S4). Individual-rater level analyses revealed no compelling evidence for accuracy for individual child or adult participants (Supplementary Materials Section 4.2 and Table S6).

Table 4

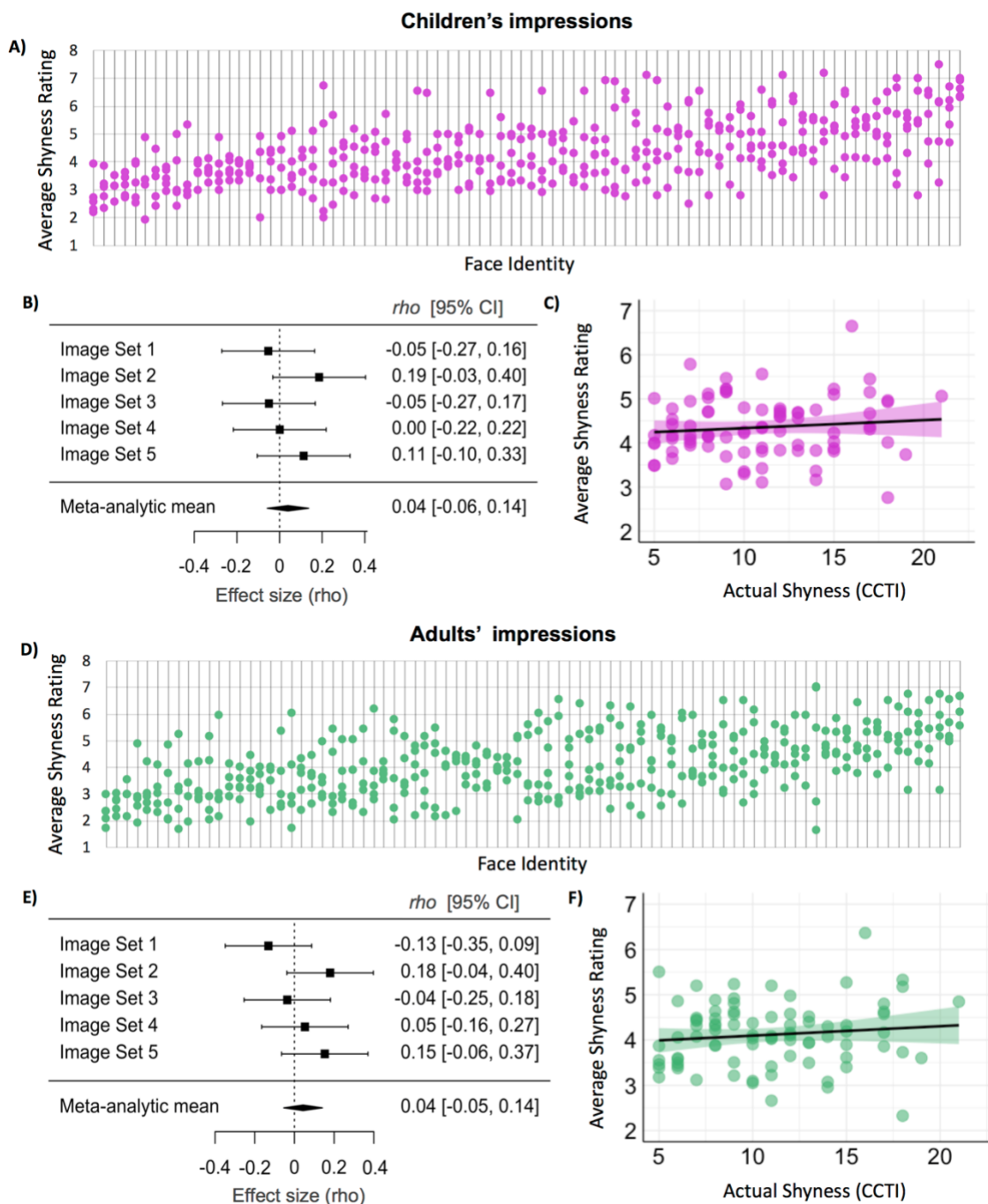
Spearman's rho (rho) and Pearson's r (r) correlations between actual shyness and impressions of shyness averaged across participants, separately for children and adults. Significance tests (Fisher Z transformation; Steiger, 1980) comparing accuracy (rho and r) between the children and adults.

Group level accuracy		<i>Spearman's rho</i>		<i>Pearson's r</i>	
		<i>rho</i>	<i>p</i>	<i>r</i>	<i>p</i>
Children	Image set 1	-.053	.632	.037	.736
	Image set 2	.186	.090	.207	.059
	Image set 3	-.050	.651	-.033	.763
	Image set 4	.001	.998	-.001	.998
	Image set 5	.113	.306	.099	.369
	Average Rating	.094	.394	.106	.338
Adults	Image set 1	-.131	.235	-.079	.476
	Image set 2	.180	.102	.201	.066
	Image set 3	-.036	.749	.024	.829
	Image set 4	.053	.629	.052	.637
	Image set 5	.153	.164	.147	.183
	Av Rating	.118	.284	.119	.281
Fisher's Z	Average Rating	$z = 0.34$.734	$z = 0.21$.835

All N = 84.

Figure 5

Child (A-C) and adult (D-F) participants' impressions of shyness at the face level: A/D) Shyness impression for each of the five images of each face identity, averaged across participants. Each vertical line represents one face identity ($N = 84$). Each circle represents the average rating for one of the five images of each identity, and demonstrates the large variability in shyness impressions across the different images. Identities are ordered on the x-axes according to their mean shyness impression. B/E) Parameter estimates (effect size and confidence intervals calculated by weighted sample size) for impression accuracy within each individual image set, and their overall meta-analytic mean effect size (Spearman's ρ). C/F) Scatterplot of the association between actual shyness (from the Colorado Childhood Temperament Inventory: CCTI), and shyness impressions averaged across the five image sets, revealing no compelling accuracy. Each point represents one face identity, with darker transparency representing more points superimposed. We plot the line of best fit and 95% confidence intervals.

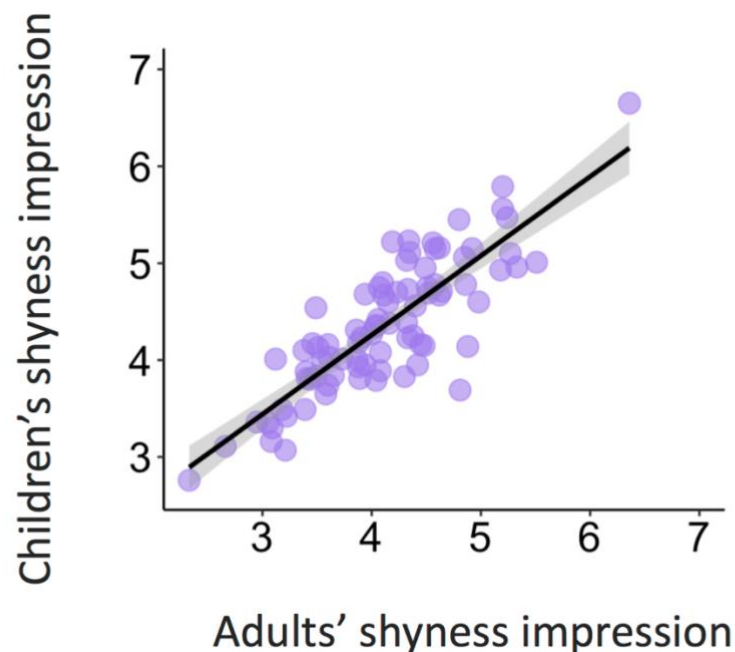


How Similar are Children's and Adults' Shyness Impressions?

There was a large correlation between children's and adults' impressions of shyness, suggesting these impressions are highly similar: Pearson's $r = .84$, $p < .001$, $N = 84$ (Figure 6). Children's shyness impressions were also highly similar to an independent sample of young adult participant's impressions, collected previously (Collova et al., 2020), $r = .82$, $p < .001$, $N = 84$ (for more details see Supplementary Materials Section 5.2 and Table S8).

Figure 6

Scatterplot of the association between children's and adults' impressions of shyness. Each point on the figure represents ratings averaged across the five images for each child-face identity ($N = 84$). We plot the line of best fit and its confidence intervals (+/- 95%).



Summary

Overall, there was no compelling evidence that children or adults formed accurate impressions of shyness. It is possible that accuracy was lacking because of the high within-person variability in shyness impressions, indicating that these impressions varied

substantially across the five different images of each child's face. Irrespective of accuracy, children and adults did form highly similar impressions of shyness.

General Discussion

Children in both experiments showed an adult-like pattern of impressions and accuracy. Children and adults formed highly similar impressions of niceness and shyness. Furthermore, in Study 1, children's and adults' facial impressions of niceness modestly correlated with parental reports of nice behavior for those children we had face images of. In Study 2, neither children nor adults showed evidence for accurate impressions of shyness. Together, these results contribute to the current theoretical debate surrounding impression accuracy and provide support for modest accuracy in niceness (but not shyness) impressions for children aged 7-12 years.

A Developmental Perspective

Our results speak to the maturity of both impressions and impression accuracy in middle childhood. To our knowledge, we present the first study investigating children's trait impressions from naturalistic, ambient face images. We show that children's and adults' impressions of niceness ($r = .94$) and shyness ($r = .84$) are remarkably similar in highly variable set of face stimuli. Our stimuli contrast previous research which has mostly used highly-controlled or computer-generated face images to examine children's impressions (e.g. Cogsdill, et al., 2014; Ewing et al., 2019; Palmquist et al., 2019, 2020). Although these studies have provided very important insight into how children form impressions, it is crucial that we understand children's impressions from naturalistic face images because these images represent the nature in which children experience faces in everyday life. Indeed, there are important differences in adults' processing of real and computer-generated faces. Computer-generated faces do not fully tap adults' face expertise (Crookes et al., 2015) and disrupt the

social perception of trait evaluations (Balas & Pacella, 2017). Here, we confirm sophisticated impression formation in children, using real and naturalistic face stimuli. We also show high impression similarity at the level of face trait ratings, providing stronger evidence for impression maturity than previous research using two-alternative forced choice tasks (e.g. Charlesworth et al., 2019; Cogsdill & Banaji, 2015; Cogsdill et al., 2014).

We found no evidence for a difference in the degree of accuracy between children's and adults' impressions, probably because these impressions were so highly correlated. These results indicate that sensitivity to whatever cues are being used to make these accurate niceness judgments are mature by middle childhood, and possibly before (we consider what these potential cues might be later in the discussion). Taken together, these results demonstrate an adult-like pattern of impressions and their accuracy in middle childhood, consistent with previous research (Li et al., 2017). Interestingly, this adult-like pattern of impression formation stands in sharp contrast with other face-processing abilities, which continue to mature well into late adolescence and adulthood (e.g. facial emotional recognition: Thomas, De Bellis, Graham, & LaBar, 2007; identity recognition: Germine, Duchaine, & Nakayama, 2011; Mondloch, Le Grand, & Maurer, 2010).

We also found considerable within-person variability in both children's and adults' impressions of niceness and shyness, suggesting that these impressions are highly dependent on specific image characteristics. This result is consistent with person perception research investigating adults' facial impressions (e.g. Collova et al., 2020; Jenkins, et al., 2011; Sutherland et al., 2017; Todorov & Porter, 2014), and also aligns with recent evidence that children's facial recognition is influenced by image-specific characteristics (Laurence & Mondloch, 2016). To better understand children's impressions, it is critical that researchers consider the variability in face stimuli, and test the generalizability of their results to other face identity samples. Here, we provide converging evidence for accuracy in children's

impressions of niceness, consistent with accuracy in children's impressions of trustworthiness given standardized child-face images (Li et al., 2017). Our results also demonstrate that the accuracy observed previously for impressions of Chinese children's faces (Li et al., 2017) can generalise across culture, to a majority Caucasian sample of faces.

Why are Impressions of Niceness Accurate?

The question of whether or not impressions are at all accurate remains widely debated amongst face perception researchers (Bonafon, Hopfensitz, & De Neys, 2015; Todorov, Funk, & Olivola, 2015). Here, our results reveal some accuracy for impressions of niceness from children's faces, albeit very modest. We also show that accuracy is not image-specific, but generalises across different images of the same child's face. These results replicate the accuracy for niceness found previously (Collova et al., 2020) in two new participant samples, demonstrating that this accuracy is reproducible in this image set. There are several possible explanations that could account for this accuracy (also see Zebrowitz & Collins, 1997).

First, it is possible that impressions of niceness were accurate because of the self-fulfilling prophecy effect (see Rosenthal, 1994; Zebrowitz & Collins, 1997). A self-fulfilling prophecy effect would predict that children who look nice are treated positively, and therefore end up behaving in ways that increase the accuracy of this positive impression. Here, we showed that both children and adults formed highly similar impressions of niceness, indicating that treatment by both adults and children could contribute to a self-fulfilling prophecy effect. Indeed, our results align with evidence that is suggestive of self-fulfilling prophecies for similar traits such as honesty (e.g. Zebrowitz et al., 1996) and trustworthiness (Li et al., 2017) from standardized images of children's faces. If the accuracy we observed was due to a self-fulfilling prophecy effect, this result would suggest that treatment by others before middle childhood is sufficient to affect behaviour (but see Alley et al., 2019). Any

attempts to reduce such effects would therefore need to take place early in development, or at least before middle childhood.

A second possible explanation for this accuracy, is that people have adaptive mechanisms for identifying niceness/trustworthiness traits from faces. Impressions of niceness and trustworthiness from children's faces are highly similar (Collova et al., 2019). From an evolutionary perspective, impressions of these traits may signal important information about those who have good versus harmful intentions towards us (Oosterhof & Todorov, 2008; Zebrowitz, 2004; Zebrowitz & Montepare, 2006). For example, it might be adaptive for both children and adults to infer whether a stranger is likely to steal their resources, or betray their trust. In support of this adaptive argument, we show that experience during childhood (at least by 7 years of age) is enough to extract valid signals of niceness, although it would be important for future research to test exactly how early this accuracy emerges. Nevertheless, the finding that younger children show less mature impression accuracy would not necessarily rule out an adaptive account. The ability to detect valid signals of character in faces may be a skill which is adaptive, but which also develops over time, akin to other face processing skills (e.g. face identity recognition; see Young & Burton, 2018). Clearly, a skill can develop whilst also being important for survival and so these two accounts are not mutually exclusive (Sutherland, Collova, et al., 2020).

Alternatively, it is possible that that parents may have been biased when selecting the photographs, or when reporting the actual niceness (or shyness) for their child. For instance, parents who might be biased to choose nice looking photos of their child, might also report their child as behaving nicely. Nevertheless, our study was designed to minimise these potential biases. For instance, parents were naïve to the exact predictions of the study, and filled out the questionnaires months prior to our request for face images. Parents also completed the questionnaires as part of a lengthy questionnaire battery, and so it is unlikely

that they remembered the specific questions related to niceness/shyness when selecting the images. Furthermore, the questionnaire that we used to measure actual niceness is well established and reliable (Stone et al., 2010). Finally, our results agree with other research which has also found modest accuracy using images of children taken by a stranger, and peer judgments of actual niceness/trustworthiness of those children as the criterion (Li et al., 2017). Therefore, it seems unlikely that our results could be entirely attributed to any parental social desirability bias.

Why Did we not Find Compelling Evidence for Accurate Shyness Impressions?

Shyness impressions might not have shown evidence for accuracy because we observed a relatively limited range of actual shyness scores (mostly in the lower end of the distribution). It would be important to test whether any accuracy exists for children with higher scores. Nevertheless, we did observe accurate niceness impressions despite a similarly limited range of niceness scores. Alternatively, shyness impressions may not be accurate if these impressions are less likely to lead to self-fulfilling prophecy effects, as compared to niceness impressions. Indeed, impressions of shyness are less likely to influence how adults report they would behave towards children as compared to impressions of niceness (Collova et al., 2019). In this context, shy behaviors may not be reinforced in children who look shy.

Finally, it is possible that there are valid cues of shyness in children's faces, but that these cues were not captured in our ambient stimuli. Shyness impressions were particularly variable across different images of the same child's face, and this may have limited the potential for accuracy. This variability suggests that shyness impressions are highly dependent on image specific information. If there are valid cues of shyness in faces, then these cues might be especially disrupted by changes in image characteristics (e.g. dynamic changes in expression, or pose). In contrast, valid cues of niceness may be more invariant across different images of the same person's face, for example, stable facial cues of an

upturned mouth or high eyebrows, that might resemble emotional expressions (see Said, Haxby, & Todorov, 2011).

Future Directions and Limitations

Here, we found mature impressions in children aged 7-12 years old. Evidence from two-alternative forced choice tasks show that children as young as 5-6 years old may already form adult-like impressions of faces (Cogsdill et al., 2014), although these impressions might become more fine-grained with development (Baccolo & Cassia, 2019). It would be interesting to test whether impressions of niceness and shyness are also adult-like at this younger age. Niceness is a general valence (positive-negative) dimension. Indeed, valence evaluations underlie a range of different social cognitive stimuli (e.g. Fiske, Cuddy, & Glick, 2007; Todorov, 2008), and sensitivity to valence-related traits may reflect a fundamental ability that emerges early in life (see Jessen & Grossmann, 2016, 2019). In contrast, it is possible that other more specific traits, such as shyness, require sensitivity to more subtle face attributes, and therefore might be slower to develop (also see Palmquist & DeAngelis, 2020). Previous evidence reveals that individual differences in impressions are shaped more by personal experience than genes or shared environments (Sutherland et al., 2020), and it would be interesting for developmental researchers to investigate the age at which these experiences are most influential in shaping impressions. Indeed, culture learning and social experiences during childhood are likely to shape impressions (Over, Eggleston & Cook, 2020). It would also be interesting to test whether impression *accuracy* is adult-like at a younger age. If young children do form accurate impressions, this finding would provide further support for early sensitivity to valid signals of character.

Another important future direction will be to consider which facial cues accurately signal niceness in children's faces. Potential candidates include emotional expressions (e.g. happiness, sadness, anger), facial attractiveness and babyfaceness, which holistically cue

impressions of niceness (Collova et al., 2019). For example, children who were reported by their parents as behaving more nicely, might have a face structure that resembles a happy expression (e.g. a high set brow, or an upward turn in their lips). Indeed, children between the ages of 5–8 years show an adult-like use of emotional cues when forming impressions of children's faces (Ewing et al., 2019; but also see Mondloch et al., 2019), and so it is possible that resemblance to emotional expression acted as a valid signal to niceness.

It is also important to consider whether our results here from ambient images, would generalize to standardized images of children's faces (previous evidence suggests that they will; Li et al., 2017). The ambient images we used were sent in by the parents of the children, and were usually images taken recently on those parents' phones. These ambient images may have especially captured valid signals of niceness. For example, particularly nice children might have been more likely to smile for a photograph taken by their parent. Participants may have been sensitive to the valid cue of a smile when forming these accurate impressions. It is important to note that although smiling may cue impressions of niceness in children (Collova et al., 2020), facial impressions are influenced by a holistic combination of multiple facial cues (Vernon, et al., 2014), and so it is unlikely that smiling (or any single cue) could account for this accuracy. Instead, it is likely that participants are integrating multiple different cues in their judgements of niceness. Our results are also consistent with previous evidence for accuracy in children's impressions of trustworthiness (a trait closely related to niceness; Collova et al., 2019) from standardized images of children's faces ($r = .27$; Li et al., 2017). Importantly, for these stimuli, children held neutral expressions and were not smiling. To the extent that valid cues of niceness/trustworthiness are overlapping, it is unlikely that smiling would fully account for accuracy in impressions of niceness. Nevertheless, ambient images may better capture behavioral tendencies than standardized face images. Future research could directly test this possibility (see Bjornsdottir & Rule, 2017).

Conversely, the nature of these ambient images might account for why shyness impressions were not accurate. The concept of shyness relates to a person's behavioral tendencies around strangers (e.g. CCTI; Rowe & Plomin, 1977). However, the images were likely taken by the child's parent and in an environment familiar to the child (e.g. at their home, or school), and therefore, these photographs might not capture the cues relevant to a child's behavior around strangers. From this perspective, the context in which images are taken may be critical in conveying relevant trait-relevant behavioral information. Indeed, Verplaetse et al. (2007) showed that accuracy for impressions of trustworthiness from adult faces depended on the context in which the photograph was taken. There was modest impression accuracy for photographs taken of subjects while they were engaging in a non/cooperative behavior, but not photographs that were taken prior to that subject engaging in the behavior (also see Slepian & Ames, 2016). Thus, it is possible that impressions of shyness would be accurate for images taken in a context where shy children displayed actual shy tendencies. As noted by an anonymous reviewer, it would be especially interesting for future research to consider whether school class photographs (usually taken by a stranger, in an unfamiliar setting) capture shy tendencies.

One potential limitation of the present study is that we used explicit trait ratings to measure children's trait impressions. These explicit judgments might fail to fully capture the tendency for children to form impressions in everyday life (see Mondloch et al., 2019; but also Ewing et al., 2019), or to spontaneously use valid cues in interactions with others. For instance, children might not consistently assign explicit niceness trait ratings to faces because of an immature understanding of this term, but still implicitly rely on these impressions when choosing which child to approach in the playground. Explicit trait ratings might therefore underestimate children's impression maturity. Nevertheless, in our study, we carefully ensured that children understood how to use the scale endpoints and that they conceptually

understood the terms niceness and shyness. Moreover, we found that children's impressions correlated highly with adults' impressions, suggesting that these explicit ratings did provide a valid measure of impressions in our sample of children.

As a final note, it is important to acknowledge the very modest effect sizes of our accuracy results, and that not every individual rater was accurate. Thus, people should certainly not be relying on these impressions in everyday life, particularly when there is other valid information available (such as actual behaviour; Hooper et al., 2018). Nevertheless, our effect sizes are comparable to those observed by Collova and colleagues (2020), and more generally, to other effect sizes in the field of psychology (see Schäfer & Schwarz, 2019). Researchers should not overlook evidence for accuracy because of the modest size of these effects. Even a small degree of accuracy experienced during childhood could lead to the perception that impressions contain some predictive power, and over time, these accumulative effects might reinforce the tendency to form these snap judgments.

Conclusions

The development of facial impressions is a topic of current theoretical interest (Sutherland, et al., 2020). Here, for the first time, we show that children and adults form highly similar impressions of niceness and shyness from ambient images of children's faces. Our results also provide novel insight into the development of impression accuracy. Like adults, children formed accurate impressions of niceness from children's faces, with no evidence of accuracy in shyness impressions. These results reveal that perceptual expertise and social experiences accumulated by 7 years of age (and potentially before), are adequate for detecting any valid signals of niceness in children's faces.

References

- Alley, Z. M., Kerr, D. C., Wilson, J. P., & Rule, N. O. (2019). Prospective Associations Between Boys' Substance Use and Problem Behavior Histories and Their Facial Trustworthiness in Adulthood. *Journal of Social and Clinical Psychology, 38*(8), 647-670. <https://doi.org/10.1521/jscp.2019.38.7.647>
- Baccolo, E., & Macchi Cassia, V. (2019). Age-Related Differences in Sensitivity to Facial Trustworthiness: Perceptual Representation and the Role of Emotional Development. *Child Development. https://doi.org/10.1111/cdev.13340*
- Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition, 161*, 19-30. <https://doi.org/10.1016/j.cognition.2016.12.012>
- Balas, B., & Pacella, J. (2017). Trustworthiness perception is disrupted in artificial faces. *Computers in Human Behavior, 77*, 240-248. <https://doi.org/10.1016/j.chb.2017.08.045>
- 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. doi.org/10.2307/3033814
- Bernath, M. S., & Feshbach, N. D. (1995). Children's trust: Theory, assessment, development, and research directions. *Applied and Preventive Psychology, 4*(1), 1-19. [https://doi.org/10.1016/S0962-1849\(05\)80048-4](https://doi.org/10.1016/S0962-1849(05)80048-4)
- Bjornsdottir, R. T., & Rule, N. O. (2017). The visibility of social class from facial cues. *Journal of personality and social psychology, 113*(4), 530. <https://doi.org/10.1037/pspa0000091>

- Bond, J., Charles F, Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology*, *15*(4), 523-534.
https://doi.org/10.1207/s15324834basp1504_8
- Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2015). Face-ism and kernels of truth in facial inferences. *Trends in cognitive sciences*, *19*(8), 421-422.
<https://doi.org/10.1016/j.tics.2015.05.002>
- Caulfield, F., Ewing, L., Bank, S., & Rhodes, G. (2016). Judging trustworthiness from faces: Emotion cues modulate trustworthiness judgments in young children. *British Journal of Psychology*, *107*(3), 503-518. <https://doi.org/10.1111/bjop.12156>
- 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. <https://doi.org/10.1371/journal.pone.0097644>
- Charlesworth, T. E., Hudson, S.-k. T., Cogsdill, E. J., Spelke, E. S., & Banaji, M. R. (2019). Children use targets' facial appearance to guide and predict social behavior. *Developmental Psychology*. <http://dx.doi.org/10.1037/dev0000734>
- 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. <https://doi.org/10.1016/j.jesp.2015.05.007>
- 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. doi: 10.1177/0956797614523297
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences 2nd edn. In: Erlbaum Associates, Hillsdale.
- Collova, J. R., Sutherland, C. A., & Rhodes, G. (2019). 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, 117(5), 900–924.

<https://doi.org/10.1037/pspa0000167>

Collova, J. R., Sutherland, C. A., Jeffery, L., Bothe, E., & Rhodes, G. (2020). Adults' facial impressions of children's niceness, but not shyness, show modest accuracy. *Quarterly Journal of Experimental Psychology*, 73(12), 2328-2347.

<https://doi.org/10.1177/1747021820957575>

Crookes, K., Ewing, L., Gildenhuis, J.-d., Kloth, N., Hayward, W. G., Oxner, M., . . .

Rhodes, G. (2015). How well do computer-generated faces tap face expertise? *PloS one*, 10(11). <https://doi.org/10.1371/journal.pone.0141353>

Degner, J., & Dalege, J. (2013). The apple does not fall far from the tree, or does it? A meta-analysis of parent-child similarity in intergroup attitudes, *Psychological Bulletin*, 139(6), 1270-1304. <https://doi.org/10.1037/a0031436>

De Neys, W., Hopfensitz, A., & Bonnefon, J.-F. (2015). Adolescents gradually improve at detecting trustworthiness from the facial features of unknown adults. *Journal of Economic Psychology*, 47, 17-22. <https://doi.org/10.1016/j.joep.2015.01.002>

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

Efferson, C., & Vogt, S. (2013). Viewing men's faces does not lead to accurate predictions of trustworthiness. *Scientific Reports*, 3, 1047. <https://doi.org/10.1038/srep01047>

Eggleston, A., Flavell, J. C., Tipper, S. P., Cook, R., & Over, H. (2020). Culturally learned first impressions occur rapidly and automatically and emerge early in development. *Developmental Science*, e13021. <https://doi.org/10.1111/desc.13021>

- Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental science*, *18*(2), 327-334.
<https://doi.org/10.1111/desc.12218>
- Ewing, L., Sutherland, C. A., & Willis, M. L. (2019). Children show adult-like facial appearance biases when trusting others. *Developmental Psychology*.
10.1037/dev0000747
- FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences*, *115*(7), E1690-E1697.
<https://doi.org/10.1073/pnas.1715227115>
- 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.
<https://doi.org/10.1016/j.tics.2006.11.005>
- Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual review of psychology*, *50*(1), 21-45. <https://doi.org/10.1146/annurev.psych.50.1.21>
- Foo, Y. Z., Loncarevic, A., Simmons, L. W., Sutherland, C. A., & Rhodes, G. (2019). Sexual unfaithfulness can be judged with some accuracy from men's but not women's faces. *Royal Society Open Science*, *6*(4), 181552. <https://doi.org/10.1098/rsos.181552>
- Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, *118*(2), 201-210.
<https://doi.org/10.1016/j.cognition.2010.11.002>
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*, *38*(5), 581-586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>

- Hawes, D. J., & Dadds, M. R. (2004). Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry*, 38(8), 644-651. 10.1080/j.1440-1614.2004.01427.x
- 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.
<https://doi.org/10.1371/journal.pone.0122637>
- Hooper, J. J., Sutherland, C. A., Ewing, L., Langdon, R., Caruana, N., Connaughton, E., . . . Rhodes, G. (2018). Should I trust you? Autistic traits predict reduced appearance-based trust decisions. *British Journal of Psychology*.
<https://doi.org/10.1111/bjop.12357>
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313-323. doi.org/10.1016/j.cognition.2011.08.001
- Jessen, S., & Grossmann, T. (2016). Neural and behavioral evidence for infants' sensitivity to the trustworthiness of faces. *Journal of Cognitive Neuroscience*.
https://doi.org/10.1162/jocn_a_00999
- Jessen, S., & Grossmann, T. (2019). Neural Evidence for the Impact of Facial Trustworthiness on Object Processing in a Gaze-cueing Task in 7-month-old Infants. *Social neuroscience*. <https://doi.org/10.1080/17470919.2019.1651764>
- Kadosh, K. C., & Johnson, M. H. (2007). Developing a cortex specialized for face perception. *Trends in cognitive sciences*, 11(9), 367-369. DOI: 10.1016/j.tics.2007.06.007
- Kuefner, D., Macchi Cassia, V., Picozzi, M., & Bricolo, E. (2008). Do all kids look alike? Evidence for an other-age effect in adults. *Journal of Experimental Psychology: Human Perception and Performance*, 34(4), 811. doi.org/10.1037/0096-1523.34.4.811

- Laurence, S., & Mondloch, C. J. (2016). That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology*, *143*, 123-138. <https://doi.org/10.1016/j.jecp.2015.09.030>
- Leivers, S., Simmons, L. W., & Rhodes, G. (2015). Men's sexual faithfulness judgments may contain a kernel of truth. *PloS one*, *10*(8), e0134007. <https://doi.org/10.1371/journal.pone.0134007>
- 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*. <https://doi.org/10.1111/cdev.12907>
- Little, A. C., Roberts, S. C., Jones, B. C., & DeBruine, L. M. (2012). The perception of attractiveness and trustworthiness in male faces affects hypothetical voting decisions differently in wartime and peacetime scenarios. *Quarterly Journal of Experimental Psychology*, *65*(10), 2018-2032. <https://doi.org/10.1080/17470218.2012.677048>
- Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child development*, *78*(5), 1543-1558. <https://doi.org/10.1111/j.1467-8624.2007.01082.x>
- Ma, F., Xu, F., & Luo, X. (2015). Children's and adults' judgments of facial trustworthiness: the relationship to facial attractiveness. *Perceptual & Motor Skills*, *121*(1), 179-198. <https://doi.org/10.2466/27.22.pms.121c10x1>
- McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological review*, *90*(3), 215. <https://doi.org/10.1037/0033-295X.90.3.215>
- McCrae, R. R., Chan, W., Jussim, L., De Fruyt, F., Löckenhoff, C. E., De Bolle, M., ... & Allik, J. (2013). The inaccuracy of national character stereotypes. *Journal of research in personality*, *47*(6), 831-842. <https://doi.org/10.1016/j.jrp.2013.08.006>

- Mellor, D. (2005). Normative data for the Strengths and Difficulties Questionnaire in Australia. *Australian Psychologist*, 40(3), 215-222. <https://doi.org/10.1037/0033-295X.90.3.215>
- Meinhardt-Injac, B., Kurbel, D., & Meinhardt, G. (2020). The coupling between face and emotion recognition from early adolescence to young adulthood. *Cognitive Development*, 53, 100851. <https://doi.org/10.1016/j.cogdev.2020.100851>
- Miller, P. H., & Aloise, P. A. (1989). Young children's understanding of the psychological causes of behavior: A review. *Child development*, 257-285. doi.org/10.2307/1130975
- Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology*, 86(1), 67-84. [https://doi.org/10.1016/S0022-0965\(03\)00102-4](https://doi.org/10.1016/S0022-0965(03)00102-4)
- Mondloch, C. J., Gerada, A., Proietti, V., & Nelson, N. L. (2019). The influence of subtle facial expressions on children's first impressions of trustworthiness and dominance is not adult-like. *Journal of Experimental Child Psychology*, 180, 19-38. <https://doi.org/10.1016/j.jecp.2018.12.002>
- Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, 31(5), 553-566. doi.org/10.1068/p3339
- Mondloch, C. J., Le Grand, R., & Maurer, D. (2010). Development of expertise in face recognition.
- Nelson, N. L., Kennedy-Costantini, S., Lee, A. J., & Dixson, B. J. (2019). Children's judgements of facial hair are influenced by biological development and experience. *Evolution and Human Behavior*. <https://doi.org/10.1016/j.evolhumbehav.2019.06.002>

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

<https://doi.org/10.1007/s10919-009-0082-1>

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *PNAS*, *105*(32), 11087-11092. <https://doi.org/10.1073/pnas.0805664105>

Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition*, *170*, 190-200. Doi: 10.1016/j.cognition.2017.10.002

Over, H., Eggleston, A., & Cook, R. (2020). Ritual and the origins of first impressions. *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Palmquist, C. M., & DeAngelis, E. R. (2020). Valence or traits? Developmental change in children's use of facial features to make inferences about others. *Cognitive Development*, *56*. 10.1016/j.cogdev. 2020.100948

Palmquist, C. M., Cheries, E. W., & DeAngelis, E. R. (2019). Looking smart: Preschoolers' judgements about knowledge based on facial appearance. *British Journal of Developmental Psychology*. <https://doi.org/10.1111/bjdp.12303>

Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological science*, *14*(6), 618-622. https://doi.org/10.1046/j.0956-7976.2003.psci_1474.x

Rhodes, G., Morley, G., & Simmons, L. W. (2013). Women can judge sexual unfaithfulness from unfamiliar men's faces. *Biology letters*, *9*(1), 20120908. <https://doi.org/10.1098/rsbl.2012.0908>

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological bulletin*, *138*(1), 146. Doi: <https://doi.org/10.1037/a0025750>

- Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science*, 3(6), 176-179.
<https://psycnet.apa.org/doi/10.1111/1467-8721.ep10770698>
- Rowe, D. C., & Plomin, R. (1977). Temperament in early childhood. *Journal of personality assessment*, 41(2), 150-156. https://doi.org/10.1207/s15327752jpa4102_5
- 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. <https://doi.org/10.1037/a0031050>
- 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. <https://doi.org/10.1098/rstb.2010.0351>
- Satchell, L. P., Davis, J. P., Julle-Danière, E., Tupper, N., & Marshman, P. (2019). Recognising faces but not traits: Accurate personality judgment from faces is unrelated to superior face memory. *Journal of Research in Personality*, 79, 49-58.
<https://doi.org/10.1016/j.jrp.2019.02.002>
- Schäfer, T., & Schwarz, M. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, 10, 813. doi: 10.3389/fpsyg.2019.00813
- Seward, R. J., Bayliss, D. M., Stallman, H. M., & Ohan, J. L. (2018). Psychometric Properties and Norms for the Strengths and Difficulties Questionnaire Administered Online in an Australian Sample. *Australian Psychologist*, 53(2), 116-124.
<https://doi.org/10.1111/ap.12325>
- Short, L. A., Mondloch, C. J., McCormick, C. M., Carré, J. M., Ma, R., Fu, G., & Lee, K. (2012). Detection of propensity for aggression based on facial structure irrespective of

face race. *Evolution and Human Behavior*, 33(2), 121-129.

<https://doi.org/10.1016/j.evolhumbehav.2011.07.002>

Slepian, M. L., & Ames, D. R. (2016). Internalized impressions: The link between apparent facial trustworthiness and deceptive behavior is mediated by targets' expectations of how they will be judged. *Psychological science*, 27(2), 282-288. doi: 10.1177/0956797615594897.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245-251. <https://doi.org/10.1037/0033-2909.87.2.245>

Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clinical child and family psychology review*, 13(3), 254-274. doi: 10.1007/s10567-010-0071-2

Sutherland, C. A., Burton, N. S., Wilmer, J. B., Blokland, G. A., Germine, L., Palermo, R., ... & Rhodes, G. (2020). Individual differences in trust evaluations are shaped mostly by environments, not genes. *Proceedings of the National Academy of Sciences*, 117(19), 10218-10224. <https://doi.org/10.1073/pnas.1920131117>

Sutherland, C. A., Collova, J. R., Palermo, R., Germine, L., Rhodes, G., Blokland, G. A., ... & Wilmer, J. B. (2020). Reply to Cook and Over: Social learning and evolutionary mechanisms are not mutually exclusive. *Proceedings of the National Academy of Sciences*, 117(28), 16114-16115. <https://doi.org/10.1073/pnas.2009587117>

Sutherland, C. A., Martin, L. M., Kloth, N., Simmons, L. W., Foo, Y. Z., & Rhodes, G. (2018). Impressions of sexual unfaithfulness and their accuracy show a degree of universality. *PloS one*, 13(10). <https://doi.org/10.1371/journal.pone.0205716>

- 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. DOI: 10.1080/02699931.2018.1449735
- Sutherland, C. A., 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*(2), 397-415. Doi: 10.1111/bjop.12206
- Tang, Y., Harris, P. L., Zou, H., & Xu, Q. (2019). The impact of emotional expressions on children's trust judgments. *Cognition and Emotion*, *33*(2), 318-331.
<https://doi.org/10.1080/02699931.2018.1449735>
- Terrizzi, B. F., Brey, E., Shutts, K., & Beier, J. S. (2019). Children's developing judgments about the physical manifestations of power. *Developmental psychology*, *55*(4), 793. DOI: 10.1037/dev0000657
- Terrizzi, B.F. Conceptual Development and Change Precede Adults' Judgments About Powerful Appearance. *Adaptive Human Behavior and Physiology* **6**, 194–211 (2020).
<https://doi.org/10.1007/s40750-020-00135-3>
- Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental science*, *10*(5), 547-558. <https://doi.org/10.1111/j.1467-7687.2007.00614.x>
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*(1), 208-224. DOI: 10.1196/annals.1440.012
- Todorov, A., Funk, F., & Olivola, C. Y. (2015). Response to Bonnefon et al.: Limited 'kernels of truth' in facial inferences. <https://doi.org/10.1016/j.tics.2015.05.013>

- 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. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological science*, 25(7), 1404-1417. <https://doi.org/10.1177/0956797614532474>
- Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: the sequel.: A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, 28(4), 260-271. <https://doi.org/10.1016/j.evolhumbehav.2007.04.006>
- Webster-Stratton, C., & Eyberg, S. M. (1982). Child temperament: Relationship with child behavior problems and parent-child interactions. *Journal of Clinical Child & Adolescent Psychology*, 11(2), 123-129. <https://doi.org/10.1080/15374418209533075>
- Willis, J., & Todorov, A. (2006). First impressions making up your mind after a 100-ms exposure to a face. *Psychological science*, 17(7), 592-598. <https://doi.org/10.1111/j.1467-9280.2006.01750.x>
- Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., . . . Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, 107(11), 5238-5241. <https://doi.org/10.1073/pnas.0913053107>
- Zarbatany, L., & Marshall, K. G. (2015). Are First Impressions of Unknown Children and Early Adolescents Affected by the Facial Attractiveness of Their Best Friend? *Merrill-Palmer Quarterly*, 61(4), 468-485. <https://doi.org/10.13110/merrpalmquar1982.61.4.0468>
- Zebrowitz, L. A. (2004). The origin of first impressions. *Journal of Cultural and Evolutionary Psychology*, 2(1-2), 93-108. <https://doi.org/10.1556/jcep.2.2004.1-2.6>

- 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. https://doi.org/10.1207/s15327957pspr0103_2
- Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and human behavior*, 15(6), 603-623. <https://doi.org/10.1007/BF01065855>
- Zebrowitz, L. A., & Montepare, J. M. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. *Evolution and social psychology*, 81-113.
- Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., & Godoy, R. (2012). First impressions from faces among US and culturally isolated Tsimane' people in the Bolivian rainforest. *Journal of Cross-Cultural Psychology*, 43(1), 119-134. DOI: 10.1177/0022022111411386